

# 4 The Conversation Analytic Approach to Transcription\*

---

ALEXA HEPBURN

*Loughborough University*

GALINA B. BOLDEN

*Rutgers University*

## 1 Introduction

In 1965, as an undergraduate student enrolled in Harvey Sacks' lecture course and later in her position of 'clerk/typist', Gail Jefferson undertook the task of typing out everything that was said in the tape-recorded conversations Sacks had collected (Lerner, 2004). By the late 1960s this apparently simple task had generated most of the comprehensive system<sup>1</sup> for transcribing talk and other conduct in talk-in-interaction that conversation analysts now rely on.

A key insight of conversation analytic research is that various features of the delivery of talk and other bodily conduct are basic to how interlocutors build specific actions and respond to the actions of others (see Drew, this volume, on turn design; Levinson, this volume, on action). It is for this reason that Jefferson developed, and other conversation analysts continue to develop, ways of representing talk and other conduct that capture the rich subtlety of their delivery. Jefferson's system of conventions evolved side by side with, and was informed by the results of, interaction analysis, which continues to show that there are many significant things going on in talk that parties to the interaction treat as relevant, and that simple orthographic representation misses. However, transcripts are necessarily selective in the details that are represented and thus are never treated by conversation analysts as a replacement for the data.

Conversation analysts' insistence on capturing not only what is said but also details of how something is said, including interactants' visible behaviors, is based on the assumption that "no order of detail in interaction can be dismissed *a priori* as disorderly, accidental, or irrelevant" (Heritage, 1984: 241).<sup>2</sup> Conversation analytic transcripts need to be detailed enough to facilitate the analyst's quest to discover and describe orderly practices of social action in interaction.

In this chapter, we first discuss the key conventions used in transcribing, with illustrations of the importance of using them. We then explore new frontiers and emerging challenges for transcription. To conclude, we engage with some of the criticisms that have been made about the conversation analytic approach to transcription.

## 2 Transcription conventions

In the following we discuss the most commonly used conventions for transcribing vocal conduct in talk-in-interaction (see Jefferson, 2004a). Jefferson transcription conventions are intended to build intuitively on familiar forms of literary notation (underlining for emphasis, capital letters for volume, arrows for pitch movement, and so on), which makes learning transcription conventions relatively straightforward. The conventions are organized into the following five categories:

- i. Transcript layout (section 2.1)
- ii. Temporal and sequential relationships, which concern how different parts of talk are related in time (section 2.2)
- iii. Aspects of speech delivery, including changes in pitch, loudness, tempo, degrees of emphasis, and voice quality (section 2.3)
- iv. Metacommentary and uncertain hearings (section 2.4)
- v. Ways of representing various other activities in interaction that do not necessarily involve speech, such as sighing, laughing and crying (section 2.5)

We will take each of these in turn and briefly illustrate their interactional significance.

### 2.1 *Transcript layout*

As seen in this chapter and elsewhere in this volume, conversation analysts use a standard layout for setting out transcripts, which illustrates three key features (illustrated in Excerpt (1)): speakers are identified at each point where speaker transition is relevant (Sacks, Schegloff & Jefferson, 1974); talk is represented as it is produced, not as it might have been intended or as it 'should' have been produced; and a fixed width font is used to align overlapping talk and/or visible behavior.

(1) NSPCC DG Daughter and Fleas. 1.59

```
1  CPO:      It's a bit acrimonious [is it,]
2  Caller:      [Ye:s ] iddis.
3  CPO:      Ri:ght. Okay, .hh ER:m (0.3) tk (.) Ri:ght.
4              (0.5)
5  CPO:      .hh I MEAN THERE MAY BE (.) VARIous rea:sons,
```

Note also the use of line numbers: each line of the transcript is numbered to facilitate reference to specific points in the interaction during analysis.

### 2.2 *Temporal and sequential relationships*

One of the most consequential contributions of CA to the study of talk and social action was the introduction of timing and sequential position in interaction. CA research has amply demonstrated the incredible precision with which interlocutors coordinate their talk (e.g. Jefferson, 1973, 1984a, 1986, 1989; Sacks, Schegloff & Jefferson, 1974; Schegloff, 2000). Timing is therefore taken very seriously in CA transcription conventions.

#### 2.2.1 **Overlapping talk**      Simultaneous talk by two or more interlocutors is

represented by lining the overlapping talk up and marking it with *square brackets*.<sup>3</sup> Overlap onset is marked with the left square bracket ([); overlap offset, when marked, is indicated with the right square bracket (]), as in Excerpt (2):

(2) Hyla & Nancy

5 Nan: Oh ril[ly I'm'nna go loo:k,]  
6 Hyl: [In the view section.]

There are a number of reasons why marking overlap is important (relatedly, see Clayman, this volume, on the transition relevance place). Careful transcription of overlaps and gaps facilitated the discovery that turn taking is not a result of speakers simply waiting for one another to finish talking. Rather speakers project possible completions of turn constructional units, aiming to start their talk at a transition-relevance place (Sacks, Schegloff & Jefferson, 1974):

(3) Ibid.: 702

1 Desk: What is your last name [Lorraine.  
2 Call: [Dinnis

When they get this wrong, speakers tend to drop out of overlap very quickly (Schegloff, 2000). Overlap therefore provides evidence for the rules of turn taking and also exposes a range of practices that speakers employ to subvert those rules (cf. Sacks, Schegloff & Jefferson, 1974). A careful transcription of the onset of overlap is necessary for a grounded analysis of the interactional consequences of simultaneous talk. This illustrates the symbiotic relationship between careful transcription and new analytic insight.

**2.2.2 Latching** Latching (represented by equals signs) marks the absence of any discernible silence between two turns or between parts of one turn. Latching indicates that an ordinary ‘beat of silence,’ which represents a normal transition space (or an unmarked next position) between one turn and another, is not there (Jefferson, 1984a; Schegloff, 2000). Latching can therefore occur between different speakers’ turns:

(4) Wilson 02 4:34

1 Mum: Ah don't think they'd like each other,=  
2 Lisa: =The'd try an kill each other.=

Or within the same speaker’s turn, as when a speaker ‘rushes through’<sup>4</sup> to extend his/her turn after a possibly complete TCU has been produced (see Walker, this volume):

(5) Wilson 02 3:43

1 Mum: I think 'e might li:ke it.=Might'n'e.

Latched production can be critical for understanding what is being accomplished interactionally. For instance, Local (2004) shows how the latched together unit *and=uh(m)* can mark any subsequent talk as not related to immediately prior talk, but rather resuming an earlier topic.

Equals signs can also show that a speaker's talk, broken up into separate lines on the transcript to accommodate the placement of overlapping talk, is nevertheless 'through produced,' as in Excerpt (6) (lines 3/6 and 4/7):

(6) Location, Location, Location

3 Phil: [I don't [think you're gonna give 'im]=  
 4 Hazel: [a h H A H H A ↑HA ↑HAH ]=  
 5 Kir: =[hih ha hah hah ha .h] [i.h] h h]=  
 6 Phil: =[a MO:MENT'S CHOI:CE: ] [Hh ]  
 7 Haz: =[ahah hah ha hah .hh] [h h h ]=

**2.2.3 Gaps and pauses**

Conversation analysts distinguish between *gaps*, which are silences that occur between TCUs, and *pauses*, silences within a TCU (Sacks, Schegloff & Jefferson, 1974). Silences are measured to the nearest tenth of a second and placed in parenthesis: between lines of talk for gaps and on the same line as the preceding talk for pauses. A period in parentheses indicates a *micropause*, a hearable silence that is less than two-tenths of a second.

The transcription conventions were developed measuring silence relative to speech rhythm so that silence is understood relative to the tempo of the talk. To time silences following this method, one acclimates to the pacing of the talk and then, when the silence is reached, begins counting by saying a counting phrase at the pace of the preceding talk (Auer, Couper-Kuhlen & Müller, 1999; T. P. Wilson & Zimmerman, 1986). For example, some people use 'none one thousand, one one thousand...' etc. to time the silence as following: if the silence is broken immediately following 'none,' a 0.2 second silence is indicated; if the silence is broken at 'none one,' a 0.5 second silence is marked; if the silence is broken at 'none one thou-,' a 0.7 second silence is indicated; if the silence is broken at 'none one thousand,' the silence is a full second in length, and so on. If there is a noticeable silence but it is shorter than 'none,' it is a micropause. If it occurs between TCUs, then it is most likely a 'beat of silence'—the normal amount of time left between TCUs—and therefore will remain an unmarked transition.

Some CA studies (Couper-Kuhlen, in press; Stivers, et al., 2009) rely on absolute measures of silences, often obtained through computer-aided methods, and here it is important to listen for the beat of silence between TCUs and leave it unmarked.

There are many reasons why marking silences is important. For example, whether a turn transition space is compressed (e.g. when the next turn is latched as shown above), extended (gap), or unmarked, has interactional consequences for interlocutors. For example, in Excerpt (7) below, the gap at line 2 alerts A to an upcoming dispreferred response (Pomerantz, 1984; Schegloff, 2007) to her suggestion to have partners at a bridge game they are organizing, which is evident by A backing off at line 3:

(7) SBL3-3

1 A: Why wd'n it be nice tih play pa::rtners.  
 2 (0.7)  
 3 A: Or w'tchu li:ke tha:t.  
 4 (2.8)  
 5 L: tch We:ll, (0.2) I don't know how wud we get partners.

Speakers also show an orientation to the length of silence. For example, Jefferson (1989)

noted that a silence of approximately one second might be a ‘standard maximum’ allowance for silence, at which point interlocutors begin some activity designed to resolve the problem (e.g. lines 2-3 of Excerpt (12)). Moreover, precise measures of silence have been used as evidence for a cyclical application of the turn-allocation rules (M. Wilson & T.P Wilson, 2005; T. P. Wilson & Zimmerman, 1986), as well as to investigate possible cross-cultural variability in turn-taking (Stivers, et al., 2009).

## 2.3 *Speech delivery and intonation*

In this section we will discuss how to represent unit-final intonation, stretches of words and phrases, cut-offs and jump-starts, emphasis, volume, and pitch change, and will focus on their interactional relevance.

**2.3.1 Unit-final intonation** CA transcripts use punctuation marks (period, comma, question mark, inverted question mark) to represent TCU-final intonation, not grammatical punctuation. Specifically:

- A period (.) indicates a falling intonation contour, not necessarily an assertion (see lines 1, 3 and 5 of (7)).
- A *question mark* (?) indicates strongly rising intonation, but not necessarily an interrogative (see line 4 of (16)).<sup>5</sup>
- A *comma* (,) indicates slightly rising intonation, not necessarily a clause boundary and not necessarily marking that the speaker is continuing (e.g. line 1 of (4)).
- Some transcriptionists have used the *inverted question mark* (¿) or a *question mark followed by a comma* (?,) to indicate a pitch rise that is stronger than a comma but weaker than a question mark.
- An *underscore* (e.g. “it\_”) at a turn ending represents level intonation.<sup>6</sup>

Note that many researchers discuss TCU-final intonation by making reference to this punctuation notation. Thus, one may describe the end of line 5 in Excerpt (7) above as displaying *falling* or *period* intonation. Earlier in the same line, “well,” may be identified as having *slightly rising* or *comma* intonation. These terms (e.g. *falling* and *period*) are used interchangeably.

Unit-final intonation conveys important information about actions accomplished by the turn-constructive unit, for example, in terms of the speaker’s epistemic stance relative to that of the recipient (see Drew, this volume, on turn design; Levinson, this volume, on action formation). Similarly to interrogative morpho-syntax, unit-final rising intonation may indicate the speaker’s low epistemic access or entitlement to what is being said (as in an intonational question), while falling intonation may convey a higher degree of access/entitlement (cf. Bolden, 2010; Raymond, 2009). Unit-final intonation also plays an important role in turn-taking and specifically in projecting and indicating a TCU’s possible completion point (cf. Couper-Kuhlen & Ford, 2004; Fox, 2001). In fact, due to the importance of prosody in talk-in-interaction, and the difficulty some transcriptionists have in transcribing it accurately, some (interactional linguists) advocate more sophisticated methods than the symbols outlined above (e.g. Couper-Kuhlen & Selting, 1996). For example, some projects may benefit from speech analysis software that can create graphical representations of prosodic contours.<sup>7</sup> Note, however, that a contour that appears to rise or fall on a graph may, in conjunction with other phonetic cues, be perceived by the human ear differently (on phonetics more generally, see Walker, this volume).

**2.3.2 Volume** *Underlining* (e.g. “Why”) is used to indicate some form of stress or emphasis, either by increased amplitude or by higher pitch (often both). The more underlining within a word, the greater the emphasis.<sup>8</sup> For example, in (7), line 1, “pa:rtners” carries greater emphasis than “Why”.

When marking emphasis, it is important to think about the natural stress that words carry with them. For example, words like *information* or *interruption* are typically pronounced with a mild stress (on the *-may-* and *-rup-* syllables, respectively), so no marking is required unless the speaker does further emphasis.

Especially loud talk (e.g. shouting) is indicated by *upper-case letters* (e.g. “a MOMENT”). Competing in overlap is a common environment for this (Schegloff, 2000).

*Degree signs* (e.g. “°yes°”) are placed around talk that is markedly quiet or soft. Double degree signs (e.g. “°°yes°°”) indicate a particularly quiet voice, such as *sotto voce*, or whispering, or mouthing (i.e. forming words with one’s mouth, with very little sound emerging), which may accompany extreme upset (Hepburn, 2004).

**2.3.3 Pitch variations** Intonation contours that lack the more marked shift in pitch represented by arrows are marked by underlining, sometimes in combination with colon signs (which alone indicate sound stretching; see the following section). An underlined element followed by a colon indicates an up-down contour through the word: for example, in line 1 of Extract (8) below, pitch moves up then down through the production of the word “pa:ssing”. A colon underlined indicates pitch movement sliding up through the work—e.g. “les’n:s” (line 5 of (8)):

(8) Holt:SO88:2:11:2.

1 Gle: Yes I will[do. when I'm: (.) when I'm next pa:ssing.  
 2 Les: [ .hhh  
 3 Les: .hhh Right. .h Now the other thing is, as part of his  
 4 Christmas presen:' I would like t'give Gordon a few  
 5 les'n:s=

These markings can be especially useful where speakers’ intonation slides up and down throughout a word. For example, starting up, sliding down and then back up (as in “No::;”) allows speakers to inflect their talk with ‘warning’ intonation (Hepburn & Potter, 2011).<sup>9</sup>

Even relatively slight variations in pitch (and volume) may be interactionally significant. For example, Schegloff (1998: 238, emphasis in original) shows that “a speaker can regularly project by a pitch-peak...the *next* possible completion at which the turn-unit or the turn has been designed to end”.

It is useful to try to distinguish sharper rises or falls in pitch from variations in intonation contour marked by underlines and colon/underlines. Such sharp changes upward in pitch are indicated with an *up arrow* (↑) (or, alternatively, by a *caret* (^)), while downward shifts are marked with a *down arrow* (↓) (a *breve* (˘) or a *pipe* (|)). Arrows may also mark resetting of the pitch register at which the talk is being produced. Pitch variations can be marked within a word, as with “mat:↓ress:” below, or across a string of words (surround the string with the appropriate symbols) as with “↑we pl’s e bring↑” below:

(9) Wilson 02 15:25

1 Ellie: Can ↑we pl'se bring↑ the mat:↓ress:.

Doubled symbols (e.g. (↑↑ and ↓↓ or ^^ or ||) can also be used for particularly sharp pitch resets (line 10 of Excerpt (18)).

Pitch shifts play an important role in accomplishing a wide range of actions in conversation. Elevated pitch is often associated with heightened emotion as discussed below (see Ruusuvuori, this volume, on emotion). It can also accompany laughter (as in (6) above) and contribute to an enactment of 'surprise' (Selting, 1996; Wilkinson & Kitzinger, 2006). Pitch, together with other aspects of prosody, has been found to play a role in delivering and receiving news as 'good' or 'bad' (Freese & Maynard, 1998).

**2.3.4 Speed/tempo of speech** The combination of *greater-than* and *less-than* symbols (> <) indicates that a stretch of talk is compressed or rushed (line 3 of (10) below).

(10) Wilson 10 3:09

1 Lisa: Remember (1.0) my pyjamas.  
2 (0.4)  
3 Mum: Oh pyjamas. >We mustn't forget °that.<=  
4 =Which ones a'ye gunna wear.

Speeding up talk can display that it may be slightly superfluous—at line 3 Mum's second TCU is hearably doing a 'note to self'. Used in the reverse order (< >), the signs indicate that the talk between the symbols is slower or more drawn out than the surrounding talk (lines 1/3 of (11) below). This can add emphasis to talk; here it surrounds reported speech.

(11) Wilson 13 11:59

1 Ell: Angela s[ai:d:, .hhh <follow> ]=  
2 Mum: [(°°ooh:) ( )]=  
3 Ell: =<your nose.>=[°so I followed my nose.°]

The less-than symbol by itself (<) indicates that the immediately following talk is *jump-started*: "a practice by which speakers bring off a start to the following talk that sounds earlier than it is, and seems to be produced by an over-loud first syllable" (Schegloff, 2005: 473; for a phonetic account, see Local & Walker, 2004). This is also known as a *left push*. The left push is one way of holding on to the turn that has come to a possible completion (a rush through, discussed above, is another practice):

(12) Wilson 03 7:47

1 Mum: Ah think John enjoyed it didn'e.<He's ↑ni:ce  
2 isn'e.=John:,

Here Mum is disagreeing with a prior turn in her first TCU, and she quickly jump-starts into a change of topic.

*Colons* (:) are used to indicate the prolongation or stretching of the sound just preceding

them (e.g. recall line 3 of (7)). The more colons, the longer the stretching.<sup>10</sup> Stretching talk can be implicated in overlap, e.g. when a speaker's stretch on the last syllable and the next speaker's start on the current TCU's projected completion lead to terminal overlap (Jefferson, 2004c).

A *hyphen* (-) after a word or part of a word indicates a *cut-off* (phonetically a glottal stop):

(13) Wilson 03 8:07

1 Mum: ...I al:ways thought that people from Jamaica  
2 [were always bla- very dark skinned.]

Cut-offs are used, for example, to initiate self-repair (Schegloff, Jefferson & Sacks, 1977; see also Kitzinger, this volume ) as in (13), where Mum starts to produce “bla[ck]” but cuts this word off and replaces it with “very dark skinned” (line 2).

**2.3.5 Voice quality** Aside from various forms of pitch, stretch, volume and emphasis, there are other features of the delivery of talk that are transcribed. For example:

- *Smiley voice*, or suppressed laughter, marked or enclosed by the British pound sign (£). Jefferson (2004b) notes its role in acknowledging, but not joining in with, laughter;
- *Creaky voice* (#) can appear during upset or turn endings (Local, Wells & Sebba, 1985);
- *Tremulous voice* (~) can signal upset, even where no other signs are present (Hepburn, 2004).

## 2.4 *Transcriptionist's comments and uncertain hearings*

*Double parentheses* are used to mark a transcriptionist's description of events rather than representations of them: e.g. ((cough)) or ((telephone rings)). When the transcriptionist is uncertain about a hearing, single parentheses are used to represent a possible hearing, as in line 1 below:

(14) Wilson 02

1 Dad: [ (That wz) ]  
2 Lisa: [They wouldn't] remember each other.=

If there are two possible hearings, they are separated with a forward slash within the parenthesis:

(15) Holt:88U:1:5:5

1 Dan: Yhhou (can't Oh/get on-) w'that's why you [get on  
2 with im.=.hh

## 2.5 *Features accompanying talk*

This section will focus on the transcription of aspiration, laughter and crying.

**2.5.1 Aspiration** Hearable aspiration (breathing) is shown by the letter *h*—the greater the number, the longer the aspiration. Hearable inhalation (in-breath) is shown with a period before the letter(s) (e.g. .hh).<sup>11</sup>

Aspiration may be involved in conveying a range of emotions, such as extreme upset or hysteria. Whalen and Zimmerman (1998) showed that call takers in 911 calls (CT) use the term ‘hysteria’ to account for failing to collect required information. Here is a selection from their transcript of a caller subsequently classified as ‘hysterical’:

(16) Whalen & Zimmerman (1998: 148)

```

1 CT: Nine=one=o[ne, what is your emergency?
2 C: [HUHHHHH .HHHHHH HHHHHHH .HHHHH
3 HUHHHH .HHH ((loudly gasping/out of breath))
4 CT: Hello?
5 C: we(HHH)- .HHHH we've(HHH)- HUHHH .HHHH HEEHHH-
6 the address- HHHHH .HHH we need (ambu-ln-) .HHH

```

Kidwell (2006) shows that common features of ‘hysteria’ include crying, panting and sobbing.

**2.5.2 Laughing** Gail Jefferson showed that laughter is organized in fine detail to coordinate with and sometimes sustain ongoing actions (Jefferson, 1984b, 1985, 2004b; Jefferson, Sacks & Schegloff, 1987). For example, in her paper on laughter in troubles telling, she notes that laughter may display the teller’s “good spirits” and bravery, so is “troubles-resistant” (Jefferson, 1984b: 367). Other researchers have focused on the appearance and coordination of laughter in different settings (see Glenn, 2003, for a review; see also Haakana, 2002; Lampert & Ervin-Tripp, 2006).

Laughter is typically made up of tokens such as *huh/hah/heh/hih*, meaning that there can be different voiced vowels *within* the aspiration, and, as line 2 below shows, the same interlocutor may use a variety of forms.

(17) Location, Location, Location (extended (6))

```

1 Bill: Th[e shed-]
2 Kirsty: [H H h i] h [h h u h h huh hhah hah ]=
3 Phil: [I don't [think you're gonna give 'im]=
4 Hazel: [a h H A H H A ↑HA ↑HAH ]=
5 Kir: =[hih ha hah hah ha .h] [i.h] h h]=
6 Phil: =[a MO:MENT'S CHOI:CE: ] [Hh ]
7 Haz: =[ahah hah ha hah .hh] [h h h ]=
8 Bil: [hihyeahh]=
9 Phi: =[.Hhugh I thi(h)nk you'll gedover it W:ON:'T you:!=

```

These elements may have elevated volume and pitch (line 4), and may contain differing degrees of aspiration. They may also contain consonants such as “g” in “.hugh” on line 9 (see Jefferson, 2010).

Laughter may interlace speech so that *laughter particles* are produced simultaneously with talk. Laughter particles may be one or more elements of plosive aspiration (in parenthesis, such as in “thi(h)nk” in line 9). They may also be simply *breathy* (as in “yeahh” in line 8 of (17) above).<sup>12</sup>

**2.5.3 Crying** Hepburn (2004) described a range of features of crying that can alert interlocutors to upset. Excerpt (18) below, taken from a reality television show, illustrates some typical features:

- Reduced volume or ‘mouthing’ (line 1) indicating trouble with speaking;
- Sniffs (line 5);
- Other embodied features include touching face/wiping eyes (line 4), and hiding face/avoiding eye contact (line 12);
- Silence where other talk might be due (lines 9/11);
- Increased pitch (line 10); and
- Increased aspiration (line 13).

(18) Pop idol 1

1 Cont: [((mouthing, shakes head, [smiling))  
 2 Dec: [((puts hand  
 3 Dec: [on Cont.’s shoulder) ]  
 4 Cont: [ ((thumb wiping eye)) ]  
 5 Cont: [°.snih°]  
 6 ??: [ ( ) ]  
 7 Dec: [Y’er[’t. ]  
 8 Ant: [ [ (°↓ar]ight.°)  
 9 (0.3)  
 10 Cont: ↑↑Yeah,  
 11 [ (1.2) ]  
 12 [((thumb stays on eye, looking down))]  
 13 Cont: °.hh° (.) .SHHIIH

Preliminary studies (e.g. Hepburn, 2004; Hepburn & Potter, 2007) suggest that interlocutors can respond empathetically even when only one or two of these signs of upset are present.

In this section, the focus has been on vocal conduct in English language talk-in-interaction. More recently, conversation analysts have started to tackle different kinds of data—e.g. recordings in other languages and video-recordings—which has raised new transcription challenges. In the following section, we discuss some of these new developments.

### 3 Transcribing talk in languages other than English

While English transcription conventions are well established, conversation analysts working in other languages face unique challenges. The following is a very brief discussion of some of the issues involved in transcribing talk in other languages and presenting it to audiences unfamiliar with the language.

#### 3.1 *Orthographic representation*

If the language under study is not written with the Roman alphabet, the transcriptionist needs to decide how to represent the language orthographically. The options include using: (a) the writing system of the target language (for languages that have a standard written form), (b) a standard phonetic system, such as a simplified version of the International Phonetic Alphabet (e.g.

Enfield, 2007; Moerman, 1988), or (c) a Roman transliteration system (e.g. for Russian, Bolden, 2008; for Korean, Lee, 2006). The latter, commonly adopted option has a number of advantages, including relative accessibility of the transcripts to English-speaking audiences and the ease of publication in English-language venues.

### 3.2 *Transcribing tonal languages*

Approximately 42 percent of word languages are tonal (Maddieson, 2008). In tonal languages (such as Mandarin Chinese and Thai), pitch variations (*tones*) are phonemic (i.e. pitch contours change word meanings) rather than pragmatic (as they are in English). Because conversation analytic transcription was developed for a non-tonal language (English), there are no standard ways of representing tones. Lexical tones may be represented using a number system (see, for example, Enfield, 2007, for Lao), or—especially for languages with simpler tonal systems—diacritics (e.g. Hanks, 2007, for Yucatec Maya; Moerman, 1988, for Thai).

Furthermore, a transcriptionist working on a tonal language must decide whether (and how) unit-final intonation is to be marked. In English, unit final prosody is an important aspect of the transcription system (as discussed above) because final intonation contours carry important pragmatic information about the action(s) accomplished by the turn-constructive unit. However, in tonal languages, final pitch contours are greatly affected by the lexical tone of the final word (e.g. Tao, 1996), which raises the question of how unit-final intonation is to be transcribed.

### 3.3 *Presenting data to English-speaking audiences*

When presenting transcripts of talk in languages other than English to English-speaking audiences, typically a multi-linear transcription is used (e.g. Sidnell, 2009). The most common practice is to provide a three-line transcription (as in (19) below), where the first line represents the original talk (in the adopted orthography), the second line is a morpheme-by-morpheme English gloss of the original (a combination of word translations and grammatical information in an abbreviated way<sup>13</sup>), and the third line is an idiomatic English gloss that attempts to capture the local, interactional meaning of the original.

(19) PF41 (Russian)

```
1 GRE: <A rushskix eh mnoo?ga krugom/ All/  
PRT Russians many around NAME  
Are there many Russians around, Alla?
```

The advantage of a three-line transcription (especially for languages that do not follow the English word order) is that it allows an English speaker some understanding of the talk as it unfolds, such as providing information about when (relative to what has been said) overlaps occur.<sup>14</sup> A two-line transcription—that is, one containing only an orthographic representation of the original talk and an idiomatic English gloss—is typically used for languages that closely follow the English word order.

Presenting data to an audience unfamiliar with the language under investigation involves translation of the original talk, which is difficult and wrought with problems (see, for example, Bilmes, 1996; Moerman, 1988; Schegloff, 2002). Schegloff (*ibid.*: 263) warns that translation

inevitably suggests analysis and points out an additional burden that conversation analysts working on languages other than English have when presenting their data to English audiences:

The translation needs to be rendered in a fashion sensitive not only to the detail and nuance of the material being studied in its language-of-occurrence, but also sensitive to the detail and nuance in comparable English language interactions as revealed in the already extant literature.

In other words, translation is not just a perfunctory step in the transcription process, but is part of the analysis of the original talk that has to be informed by what is known about English.

#### 4 Transcribing visible conduct

Conversation analytic transcription conventions were developed primarily on audio-recorded data (such as, telephone conversations), and thus capture interlocutors' vocal conduct only. The use of video in conversation analytic research, pioneered by Charles Goodwin in the 1970s (Goodwin, 1981), is widespread today (see Heath & Luff, this volume). In face-to-face interactions, participants' visible conduct is instrumental to how social actions are accomplished and coordinated, which means that it has to be represented on a transcript. Although the transcription of both talk and visible behavior is necessarily selective, the transcription of visible behavior may be even more so due to the substantial number of parameters. Moreover, visible behavior involving facial expressions, body posture, gestures and gaze can occur in overlap with each other and with talk.

Conversation analytic studies have relied on various methods (sometimes in combination) for transcribing visible conduct:

- transcriptionist comments,
- specialized notational systems (see chapters in this volume by Heath & Luff, Mondada, and Rossano), and
- visual representations, such as drawings and video frame grabs.

Using transcriptionist comments—in double parenthesis, such as ((A looks at B))—is, perhaps, the simplest way of transcribing visible conduct (see Excerpt (20) below). However, limited information about how visible behaviors unfold incrementally over time can be presented in this way, as comments can provide only a simplified gloss of the observable behavior.<sup>15</sup>

(20) Crouch 02 0.15

```
1 Mum: Kath'rine >c'you move<
2 [ a[long] a littw bit ple[ase. ]=
3 Mum: [((pu[shes] Anna's chair next [to Katherine))]=
4 Anna: [.huh] [ .hhu:] =
5 Kath: =[((swings legs round to block chair))]
6 Kath: =[I'wn- I wanna sit [on my oh°::: ]
```

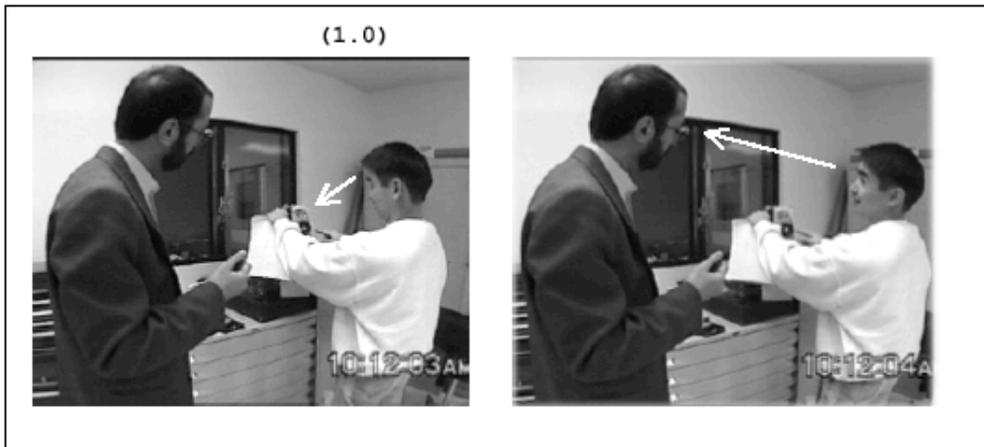
Specialized notational systems—such as those developed for eye gaze (Goodwin, 1981; Rossano, Brown & Levinson, 2009) and gestures (Heath, 1986; Streeck, 1993)—are designed to capture such details. For example, Goodwin's conventions for representing eye-gaze facilitate the analysis of the coordination of gaze and talk, and the coordination of gaze between the participants. (This has led to a number of fascinating findings; see Rossano, this volume).

Visual representations (e.g. drawings and digital video frames) accompanying a transcript of vocal conduct have the advantages of being easily interpretable and more holistic in representation. Moreover, digital frame grabs can be selected to capture, with a high degree of precision, the temporal unfolding of visible conduct in coordination with speech. With the advent of digital video, frame grabs can be quite easily produced and edited for clarity of presentation (and, if desired, to disguise participants' identifiable features).

For one example of the use of video frame grabs, see Excerpt (21): Alex is on the left; Bob on the right.

(21) WB17 (Bolden, 2003:204)

1     Bob:        The problem with it like- (0.4)  
 2                we had it (0.8) no:w     (0.6) i:s (0.2)  
 3                we only ha:ve



4 → Alex:     conjunction.



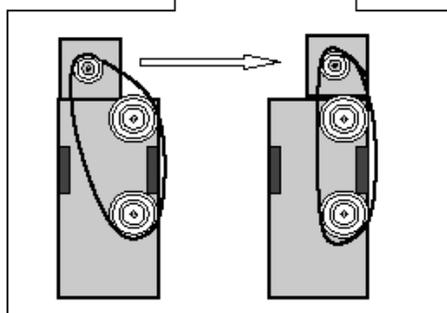
5                (0.5)  
 6     Alex:        (punch) =

Here, eye gaze is represented visually and in relation to the vocal conduct. Additionally, other

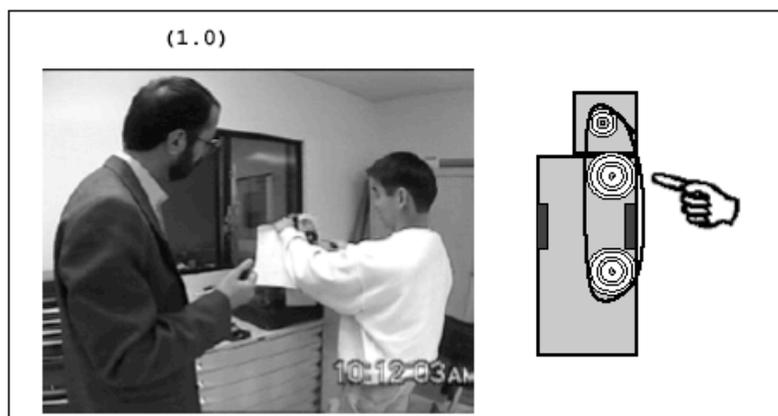
visible aspects, such as body orientations, gestures, and the surrounding environment, can be seen and taken into an account. When material objects become relevant to the interaction, the transcriptionist may need to represent them as well. The following Excerpt (22) attempts to capture Bob's manipulation of the machine he is talking about.

(22) WB17 (Ibid.: 198)

1 Bob: The problem with it like- (0.4)  
2 we had it (0.8) no:w (0.6) i:s (0.2)



3 we only ha:ve



4 → Alex: conjunction.  
5 (0.5)  
6 Alex: (punch) =

Visual representations are sometimes the most accessible way of representing video data in public presentations and in print. However, as seen from the two excerpts above, this way of representing data is very space consuming, so only very short transcripts can be included in publications.

Note that both (21) and (22) represent the same stretch of interaction but selectively highlight different aspects of the participants' visible conduct. This illustrates the fact that

decisions about how to present visual data (at conferences and in publications) have to be guided by the researchers' analytic goals. However, the video-recording itself (and not its inevitably selective transcript) has to remain the object of the analysis.

We have focused up to now on some of basic principles and practices of transcription, and on the different ways that transcription continually evolves. Part of that evolution must involve opening ourselves up to scrutiny by other researchers outside CA, and we outline below some critical concerns that have been raised.

## **5 Concerns with the CA approach to transcription**

We can differentiate two types of (often overlapping) concerns about transcription which we call *epistemological* and *practical*. The following subsections outline and assess these concerns.

### **5.1 *Epistemological concerns***

Epistemological objections to transcription have to do with whether transcription is a valid research tool and to what degree transcription embodies researchers' theoretical orientations.

**5.1.1 Criticisms** Some critics (e.g. Bogen, 1992; Mishler, 1991) object to detailed transcription as a research tool. They claim that most researchers would be better off doing different styles of research, as all the detail in the transcript just gets in the way of what is analytically important. Other critics accept the importance of transcribing the specifics of interaction, but focus on the need for transcriptionists to be more reflexive about the product. For example, Ochs (1979) suggests that there are assumptions about the world wired into transcription, and that transcriptionists should consider the way they work and be cautious about these assumptions (see Ashmore & Reed, 2005, for a more critical version of this). Similarly, Bucholtz (2000) advocates a reflexive transcription practice that highlights the assumptions made in the process.

**5.1.2 Conversation analysts' response** The idea that the transcript is theorized is unproblematic if it is taken as a reminder that there are different things that can be encoded in a transcript and that what the researcher chooses to encode will depend on their research questions and analytic perspective. However, wherever researchers are concerned in one way or another with talk as action, arguments would need to be developed, grounded in materials, which showed that the lack of attention to such features of talk would not be analytically consequential (see Potter & Hepburn, 2005, in debate with Mischler, 2005).

### **5.2 *Practical concerns: inadequacy of CA transcription***

One of the most striking features of conversation analytic transcription is its representation of talk in non-standard orthography. This approach has been criticized from two sides: those who prefer the ease of using standard English language orthography (typically qualitative social scientists) and those who prefer using a standard phonetic transcription system, such as the International Phonetic Alphabet (typically linguists).

**5.2.1 Criticisms** On the one hand, advocates of the orthographic transcription claim that CA transcription is too difficult to read, transcriptionist-dependent, and presents speech in a caricatured fashion of ‘comic book’ orthography, which makes speakers look stupid (see Jefferson, 1983, for a response to these critics). Furthermore, some critics argue that the complexity of CA transcription is not necessary if conversation is not the primary object of study (e.g. Du Bois, 1991; Ochs, 1979 develop this argument). On the other hand, linguists often criticize conversation analytic transcription for being unsystematic in its representation of phonetic details and advocate the adoption of a standard phonetic transcription system, such as the International Phonetic Alphabet (IPA).

**5.2.2 Conversation analysts’ response** Orthographic transcription imposes the conventions of written language designed to be broadly independent of specific readers. Such a transformation systematically wipes out evidence of intricate coordination, talk as action and recipient design. It encourages the analyst to interpret talk by reference to an individual speaker (e.g. perceptions) or focus on abstract relations between word and world (e.g. discourses). If talk is a medium for action, then forms of representation that try to capture elements of action rather than ‘just the words’ are needed.

On the other hand, using the International Phonetic Alphabet or a similar system in Conversation Analysis would highlight a range of features of speech production at the expense of information about overlap, delay, gesture and so on. The point is that IPA is tuned to the theoretical concerns of certain areas of linguistics and speech pathology but not designed to capture talk as action (cf. Kelly & Local, 1989; Ochs, 1979). The basic principle of conversation analytic transcription is “to get as much of the actual sound as possible into [the] transcripts, while still making them accessible to linguistically unsophisticated readers” (Sacks, Schegloff & Jefferson, 1974: 734).

The bottom line here is that everyone involved in CA has the opportunity to highlight the limitations of the transcription system and contribute to a better one.

## 6 Conclusions

Jeffersonian transcription provides a shared, standard system for rendering talk-in-interaction in a way that can be textually reproduced. It is compact, transportable and reproducible, and provides for easy random access unlike audio or video records. CA transcription is a fundamental resource for data sessions, presentations and journal articles, and, as such, it is often the medium through which analysts encounter and evaluate each others’ work. It is therefore at the center of the epistemic culture of Conversation Analysis. CA transcription has evolved, and will continue to evolve, with the gradual progression of conversation analytic studies of interaction.

## NOTES

---

\* We would like to thank Tanya Stivers for excellent editorial input, Jonathan Potter for some useful discussion of critical perspectives, and Kathryn Howard, Roi Estlein, Makoto Hayashi and Estefania Guimaraes for assistance with the non-English transcription section.

- 1 One of the first full descriptions of the conversation analytic transcription system can be found in Sacks, Schegloff and Jefferson (1974); its modern incarnation in, for example, Jefferson (2004a).
- 2 Sacks' (1984: 22) refers to this notion as "order at all points".
- 3 Older transcripts represent the beginning of overlapping talk with two forward slashes (//) on the line being overlapped by an incoming speaker.
- 4 "In a 'rush-through' a speaker who is approaching upcoming possible turn completion increases the pace of the talk, does not decelerate, talks through the momentary silence which regularly intervenes between the end of a turn and the start of a next, and launches a next turn-constructive unit" (Schegloff, 1998: 241). See the following section on how speech tempo is marked.
- 5 In English, questions (especially, *wh*-interrogatives) may be produced with falling or only slightly rising intonation (Couper-Kuhlen, in press; Geluykens, 1988). For example, the two interrogatives in Excerpt (7) (lines 1 and 3) are produced with falling intonation.
- 6 Some transcriptionists leave this intonation contour unmarked; however, the use of the underline makes it clearer that the intonation was transcribed as flat rather than simply left out.
- 7 Praat is one commonly used program, available for free download at: <http://www.fon.hum.uva.nl/praat/>. ToBI (Tones and Break Indices) conventions may also be useful (see <http://www.ling.ohio-state.edu/~tobi/>).
- 8 There are different practices about placement of underlines: the first letter of the word adds some emphasis to the whole word; underlining only the vowel sound means emphasis is not retained in the rest of the word.
- 9 Currently not all CA transcriptionists use a combination of colon and underline for representing pitch movement; some rely on the caret and pipe, or arrows, for all pitch changes (both pitch resets and within-word pitch changes). However it is notable that Jefferson found this an important element in her transcripts up to her most recent output in the Nixon tapes (Jefferson, 2004a); and Schegloff (2007) also employs the colon/underline notation. As indicated, it has been indispensable when adequately transcribing interaction between parents and young children.
- 10 Graphically stretching a word on the page by inserting blank spaces between letters does not indicate how it was pronounced; it is used to allow alignment with overlapping talk (cf. line 4 of Excerpt (16)).
- 11 In older transcript, the period is superscripted.
- 12 Potter and Hepburn (2010) suggest that the category *laughter* may be over-interpretative and propose replacing the laughter terminology with *interpolated particles of aspiration*.
- 13 For the interlinear gloss, researchers adopt glossing practices that accord with the linguistic properties of the language under study. The glosses may follow disciplinary conventions (such as the Leipzig Glossing Rules for linguistics: <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>) and/or reflect the study's objectives (e.g. be more or less specific with regard to the identified linguistic features).
- 14 Though, as Bilmes (1996) observes, this might be of limited use since the reader unfamiliar with the language under study still will be unable to accurately assess the TCU's projectability, for example.
- 15 Some researchers (e.g. Kidwell, 2005; Kidwell & Zimmerman, 2006) combine transcriptionist comments with one or several of the more sophisticated representation systems discussed below.

## References

- Ashmore, M., & Reed, D. (2005). Innocence and nostalgia in Conversation Analysis: The dynamic relations of tape and transcript. *Historical Social Research, 30*, 73-94.
- Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time: The rhythm and tempo of spoken interaction*. New York: Oxford University Press.
- Bilmes, J. (1996). Problems and resources in analyzing Northern Thai conversation for English language readers. *Journal of Pragmatics, 26*, 171-188.
- Bogen, D. (1992). The organization of talk. *Qualitative Sociology, 15*, 273-296.
- Bolden, G. B. (2003). Multiple modalities in collaborative turn sequences. *Gesture, 3*, 187-212.
- Bolden, G. B. (2008). Reopening Russian conversations: The discourse particle *-to* and the negotiation of interpersonal accountability in closings. *Human Communication Research, 34*, 99-136.
- Bolden, G. B. (2010). 'Articulating the unsaid' via *and*-prefaced formulations of others' talk. *Discourse Studies, 12*, 5-32.
- Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics, 32*, 1439-1465.
- Couper-Kuhlen, E. (in press). Some truths and untruths about prosody in English question and answer sequences. In J. P. De Ruiter (Ed.), *Questions: Formal, functional and interactional perspectives*. Cambridge: Cambridge University Press.
- Couper-Kuhlen, E., & Ford, C. E. (Eds.). (2004). *Sound patterns in interaction: Cross-linguistic studies from conversation*. Amsterdam: John Benjamins.
- Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in conversation: Interactional studies* (Vol. 12). New York: Cambridge University Press.
- Du Bois, J. W. (1991). Transcription design principles for spoken discourse research. *Pragmatics, 1*, 71-315.
- Enfield, N. J. (2007). Meaning of the unmarked: how 'default' person reference does more than just refer. In N. J. Enfield & T. Stivers (Eds.), *Person Reference in Interaction* (pp. 97-120). Cambridge: Cambridge University Press.
- Fox, B. A. (2001). An exploration of prosody and turn projection in English conversation. In M. Selting & E. Couper-Kuhlen (Eds.), *Studies in interactional linguistics* (pp. 286-315). Philadelphia: John Benjamins.
- Freese, J., & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society, 27*, 195-219.
- Geluykens, R. (1988). The myth of rising intonation in polar questions. *Journal of Pragmatics, 12*, 467-485.
- Glenn, P. J. (2003). *Laughter in interaction*. New York: Cambridge University Press.
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
- Haakana, M. (2002). Laughter in medical interaction: From quantification to analysis, and back. *Journal of Sociolinguistics, 6*.
- Hanks, W. F. (2007). Person reference in Yucatec Maya conversation. In N. J. Enfield & T. Stivers (Eds.), *Person Reference in Interaction* (pp. 149-171). Cambridge: Cambridge

- University Press.
- Heath, C. (1986). *Body movement and speech in medical interaction*. New York: Cambridge University Press.
- Hepburn, A. (2004). Crying: Notes on description, transcription, and interaction. *Research on Language and Social Interaction*, 37, 251-290.
- Hepburn, A., & Potter, J. (2007). Crying receipts: Time, empathy and institutional practice. *Research on Language and Social Interaction*, 40, 89-116.
- Hepburn, A., & Potter, J. (2011). Threats: Power, family mealtimes and social influence. *British Journal of Social Psychology*, 50, 99-150.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. New York: Polity Press.
- Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, 9, 47-96.
- Jefferson, G. (1983). Issues in the transcription of naturally occurring talk: caricature versus capturing pronunciation particulars. *Tilburg Papers in Language and Literature*, 34.
- Jefferson, G. (1984a). Notes on some orderlinesses of overlap onset. In V. D'Urso & P. Leonardi (Eds.), *Discourse Analysis and Natural Rhetorics*. Padua: cleup editore.
- Jefferson, G. (1984b). On the organization of laughter in talk about troubles. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 346-369). Cambridge: Cambridge University Press.
- Jefferson, G. (1985). An exercise in the transcription and analysis of laughter. In T. A. v. Dijk (Ed.), *Handbook of Discourse Analysis Volume 3* (pp. 25-34). New York: Academic Press.
- Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Human Studies*, 9, 153-183.
- Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silences of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166-196). Clevedon: Multilingual Matters.
- Jefferson, G. (2004a). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13-31). Philadelphia: John Benjamins.
- Jefferson, G. (2004b). A note on laughter in 'male-female' interaction. *Discourse Studies*, 6, 117-133.
- Jefferson, G. (2004c). A sketch of some orderly aspects of overlap in natural conversation. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 43-59). Philadelphia: John Benjamins.
- Jefferson, G. (2010). Sometimes a frog in your throat is just a frog in your throat: gutturals as (sometimes) laughter implicative. *Journal of Pragmatics*, 42, 1476-1484.
- Jefferson, G., Sacks, H., & Schegloff, E. A. (1987). Notes on laughter in the pursuit of intimacy. In G. Button & J. R. E. Lee (Eds.), *Talk and social organization* (pp. 152-205). Philadelphia: Multilingual Matters.
- Kelly, J., & Local, J. K. (1989). On the use of general phonetic techniques in handling conversational material. In P. Bull & D. Roger (Eds.), *Conversation: An Interdisciplinary Perspective* (pp. 197-212). Philadelphia: Multilingual Matters.
- Kidwell, M. (2005). Gaze as social control: How very young children differentiate "the look" from a "mere look" by their adult caregivers. *Research on Language and Social Interaction*, 38, 417-449.

- Kidwell, M., & Zimmerman, D. (2006). "Observability" in the interactions of very young children. *Communication Monographs*, 73, 1-28.
- Lampert, M., & Ervin-Tripp, S. M. (2006). Risky laughter: Teasing and self-directed joking among male and female friends. *Journal of Pragmatics*, 41, 1248-1271.
- Lee, S.-H. (2006). Second summonings in Korean telephone conversation openings. *Language in Society*, 35, 261-283.
- Lerner, G. H. (2004). Introductory remarks. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 1-11). Philadelphia: John Benjamins.
- Local, J. K. (2004). Getting back to prior talk: and-uh(m) as a back-connecting device in British and American English. In E. Couper-Kuhlen (Ed.), *Sound patterns in interaction: Cross-linguistic studies of phonetics and prosody for conversation* (pp. 377-400). Amsterdam: John Benjamins.
- Local, J. K., & Walker, G. (2004). Abrupt-joins as a resource for the production of multi-unit, multi-action turns. *Journal of Pragmatics*, 36, 1375-1403.
- Local, J. K., Wells, W. H. G., & Sebba, M. (1985). Phonology for Conversation: Phonological aspects of turn delimitation in London Jamaican. *Journal of Pragmatics*, 9, 309-330.
- Maddieson, I. (2008). Tone (chapter 13). In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (Eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library (Available online at <http://wals.info/feature/13>).
- Mishler, E. G. (1991). Representing discourse: The rhetoric of transcription. *Journal of Narrative & Life History*, 1, 255-280.
- Mishler, E.G. (2005). Commentary. *Qualitative Research in Psychology*, 2, 314-317.
- Moerman, M. (1988). *Talking culture: Ethnography and conversation analysis*. Philadelphia: University of Pennsylvania Press.
- Ochs, E. (1979). Transcription as theory. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental Pragmatics* (pp. 43-72). New York: Academic Press.
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 57-101). New York: Cambridge University Press.
- Potter, J. & Hepburn, A. (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology*, 2, 38-55.
- Potter, J., & Hepburn, A. (2010). Putting aspiration into words: 'Laugh particles', managing descriptive trouble and modulating action. *Journal of Pragmatics*, 42, 1543-1555.
- Raymond, G. (2009). Grammar and social relations: Alternative forms of yes/no type initiating actions in health visitor interactions. In A. F. Freed & S. Ehrlich (Eds.), *Why do you ask?: The function of questions in institutional discourse* (pp. 69-107). Oxford: Oxford University Press.
- Rossano, F., Brown, P., & Levinson, S. C. (2009). Gaze, questioning, and culture. In J. Sidnell (Ed.), *Conversation Analysis: Comparative Perspectives* (pp. 357-406). Cambridge: Cambridge University Press.
- Sacks, H. (1984). Notes on Methodology. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 21-27). Cambridge: Cambridge University Press.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696-735.
- Schegloff, E. A. (1998). Reflections on studying prosody in talk-in-interaction. *Language and*

- Speech*, 41, 235-263.
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29, 1-63.
- Schegloff, E. A. (2002). Reflections on research on telephone conversation: Issues of cross-cultural scope and scholarly exchange, interactional import and consequences. In K. K. Luke & T.-S. Pavlidou (Eds.), *Telephone calls: Unity and diversity in conversational structure across languages and cultures* (pp. 249-281). Amsterdam: John Benjamins.
- Schegloff, E. A. (2005). On integrity in inquiry ... of the investigated, not the investigator. *Discourse Studies*, 7, 455-480.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361-382.
- Selting, M. (1996). Prosody as an activity-type distinctive cue in conversation : the case of so-called "astonished" questions in repair initiation. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (Vol. 12, pp. 231-270). Cambridge, UK: Cambridge University Press.
- Sidnell, J. (Ed.). (2009). *Conversation Analysis: Comparative Perspectives*. Cambridge: Cambridge University Press.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *PNAS*, 106, 10587-10592.
- Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs*, 60, 275-299.
- Tao, H. (1996). *Units in Mandarin conversation: Prosody, discourse, and grammar*. Amsterdam: John Benjamins.
- Whalen, J., & Zimmerman, D. H. (1998). Observations on the expression and management of emotion in naturally occurring activities: The case of "hysteria" in calls to 911. *Social Psychology Quarterly*, 61, 141-159.
- Wilkinson, S., & Kitzinger, C. (2006). Surprise as an interactional achievement: Reaction tokens in conversation. *Social Psychology Quarterly*, 69, 150-182.
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review*, 12, 957-968.
- Wilson, T. P., & Zimmerman, D. (1986). The structure of silence between turns in two-party conversation. *Discourse Processes*, 9, 375-390.