12

RECURRENT PROBLEMS AND RECENT EXPERIMENTS IN TRANSCRIBING VIDEO

Live transcribing in data sessions and depicting perspective

Eric Laurier and Tobias Boelt Back

Introduction – tracing back to Jefferson

In Ethnomethodology and Conversation Analysis (EMCA), Gail Jefferson's conventions for transcribing talk are widely considered the default transcription system (Ayaß, 2015), the one that most trained conversation analysts routinely turn to in transcribing talk-in-interaction. Getting acquainted with how to write and read the Jeffersonian format has long been considered an important first step on the road to becoming competent in "the ways of the CA tribe" (ten Have, 2007, p. 11). Over the years, these conventions have been continuously reworked and developed to represent newfound and increasingly complex phenomena. In this chapter, we try to recapture the spirit of Jefferson's transcribing as a way of attending to features that are overlooked by other forms of transcription. EMCA has its own peculiarities in transcribing because its analysts aren't sure what to transcribe at the outset, given its ethos of unmotivated attentiveness. The unmotivated ethos allows our attention to be caught by other things than those which the "literature" has found to be significant and would guide us to scrutinise. The task of EMCA transcribing is itself inescapably inventive in offering solutions to the problem of rendering newly relevant phenomena found in flight.

Although Jefferson's formats for marking details of talk are well known to scholars in EMCA, that she was experimenting and also attempting to capture the specifics of embodied action has been largely overlooked (though, see Albert et al., 2019; Hepburn & Bolden, 2017). As Charles Goodwin recalled, during their data sessions in the 1970s Jefferson would put a piece of transparency over a TV screen and outline prominent bodily features and experiment with adapting existing dance notation systems (Goodwin & Solomon, 2019). These attempts offer us an inkling of Jefferson's open-ness to alternatives to the line-by-line, ASCII

DOI: 10.4324/9781003424888-16

standard characters, text-based format that she became better known for. In the digital realm, the current generation of EMCA researchers have used screen grabs as their way of tracing embodied action from video screens.

In this chapter, we will return to Jefferson's warrants for changing existing transcription formats and reflect on the abiding problems of transcribing in EMCA studies. We will compare the Jefferson format with the graphic transcript and present our experiments with using the graphic transcript for transcribing "live" during data sessions, as well as for documenting social phenomena in final publications. The graphic transcript foregrounds other phenomena and in different ways than the Jeffersonian system, particularly visually and spatially available phenomena of order: bodily actions, objects, movements, environmental features, etc. As Jefferson did, we will raise the problems that we have faced in transcribing while also working through cases of phenomena that show how and why we shape and reshape the forms of our transcripts.

A short history of transcription conventions and their abiding problems

> While those of us who spend a lot of time making transcripts may be doing our best to get it right, what that might mean is utterly obscure and unstable. It depends a great deal on what we are paying attention to. It seems to me, then, that the issue is not transcription per se, but what it is we might want to transcribe, that is, attend to.
>
> *(Jefferson, 1985, p. 25)*

The Jeffersonian breakthrough was in transcribing as a means of becoming more attentive to orderliness in otherwise heard-but-un-noticed elements of talk, both for researchers and for their readers. Indeed, she asked us not to become overly fixated on the transcript itself because its rightness remains "utterly obscure and unstable". In practice this meant that when she noticed an un-noticed thing in talk such as a speaker saying "nyem", she did not follow the convention to correct it as a defectively spoken "yes" or "no" or "ehm" (Jefferson, 1978). Thereby she created a new translatable entity from the hearable details of speech into the textual details of a transcript. She compared and evaluated what she was creating with the renderings of speech in standard and dialect orthographies in newspapers, novels, and theatre scripts, showing that these formats missed variation in the use of "dialect signatures" (Jefferson, 1983, p. 9) and pronunciation: variation which was relevant, accountable, and reportable yet unnoticed by linguists, sociologists, anthropologists, etc. The point, then, was not to conform to existing standards for representing speech nor to iron out its wrinkles but to listen first to speakers' voices and then work out how to register what was hearable in order to reveal the orderliness of those wrinkles.

Mondada's amendment and supplementation of the Jeffersonian transcription system was in itself quietly radical. It not only revealed but also departed from

several assumptions of transcripts of audible speech. In Jeffersonian transcripts, speakers and their speech appear as a next line or sentence (in fact, this is the case for almost all transcription formats). The line-by-line format itself neatly fits with the rule used by speakers: one person speaks at a time. The transcriber and analyst need only attend to one person's actions at a time. Phenomena that break that rule, overlapping speech being the most common, are notated and easily visible in Jeffersonian transcripts. It was never the case, of course, that all action from other parties halts while one person is speaking. What Mondada's (2018) system showed, from the outset, were the gestalts of doing things together, in which speaking was but one part. While a person speaks, other parties to the event would be, for example, frowning in time to an announcing of bad news, halting their walking at the place name just mentioned, picking up a pencil to establish themselves as a potential next speaker, and so on.

Relatedly, by registering non-verbal actions Mondada's system displaced speech itself, the feature which the Jeffersonian format assumed to be "the main course of action" (Mondada, 2018, p. 98). Her augmentation shifted analytic attention toward embodied actions of all sorts. For example, Mondada transcribed a video recording of a team of three people carrying around heavy paintings during the preparations of an art exhibition at a museum. As shown by her transcribing, one staff member lifts a large object. In response, another member of the staff assists him to lift the object via the haptic initiation, sensed through the object, without a word (a word being more than un-necessary). The transcribing reveals a course of haptic action which is being analysed by its recipient and their relevant response – help – is given. Moreover, Mondada's transcribing shows the party assisting the lifter in a timely way and the very absence of a verbal request for help produces a distinctive accountability to their help. To have asked "do you need a hand?" would have had its distinct accountabilities in being understood, for example, to have avoided immediate assistance or implying that a hand was not really needed.

Mondada is clear that a central quality of the transcript that she retains from the Jeffersonian format is time and timing and, in the example above of lifting an object, we can understand why. Her format maintains the use of bracketed numbers for marking the length of actions in tenths of seconds. Without each line of action being registered in parallel timelines, as lines of text, we could not see timing as one of the measures for producing and recognising the promptness or lateness of actions. Our first example is of two occupants of a car during a journey and it is an example we will look at first in Mondada format and then later as a graphic transcript.

In Transcript 1, line 01 the driver's inspection of his instruments lasts six tenths of a second, while the passenger brushing his trousers continues for 2.6 seconds plus the 0.6 seconds of the driver's inspection. Mondada's orthography adds complexity to the reading of the text when the reader seeks to assemble the timings shown via the brackets of clock-time measures, inserted symbols, and parallel formulations of non-verbal actions. The transcribers' continuing achievement is

TRANSCRIPT 1 Example of Mondada transcript created by Eric (with help from Mondada & Deppermann) from Deppermann et al. (2018).

```
01          ^(0.6)^                    (2.6)%
      dri:  ^inspects instruments^
      pas:  >> brushing trousers -----------%

02          * (1.6) +(1.2)+ (0.2)*
      dri:          +gz rvm+
      pas:  *gz PAS window-------*

03          (0.3)*(0.7)+ (0.6)*(0.5)+
      dri:             +gz rsm------+
      pas:       *gz F window-* *gz PAS window----→> (5.8)

04          (0.9)+    (1.8)     +
      dri:       +gz instruments+
```

to render the relative timings of actions so that we can recover just when actions start, their duration, and their temporal trajectory.

However, to only focus on Mondada's supplement to the existing Jeffersonian format would lead us to miss that, like Jefferson, she has led EMCA's experiments and creativity in transcribing outside of that format. For example, she hosted a workshop in Rome in 2017 bringing together artists, graphic designers, and EMCA researchers, including Eric, to experiment with distinct diagrammatic, graphic, and other ways of transcribing action that extended beyond ASCII text and figures. In her work, she continues to experiment with alternatives that would help us register phenomena in other ways than our conventional text-based transcripts.

Alongside bodily actions, the Mondada orthography deals with what we might broadly call the material environment or, as Mondada sometimes calls it, the *local ecology* of objects, technologies, architecture, vegetation, etc. Accompanying the addition of the local ecology is a further choice about what features to render, which "are potentially infinite because they exceed conventional forms and include situated, ad hoc resources that depend on the type of activity and its specific ecology" (Mondada, 2018, p. 95). There is always a tension between the infinite elements, the *plenum* as Garfinkel might put it, the absences of the recording, and the desire to transcribe what the recording has preserved in a "similar, coherent and robust way that is essential for systematic analyses" (p. 95).

Reflections on hiding, estranging, and formulating things when transcribing

There are four central problems of transcription in EMCA that we want to consider ahead of reflecting on our own graphic transcription practices. The first: detail, is its seeming strength, and one of the earliest sustained reflections on the

problem of excessive detail in CA transcripts, was David Bogen's (1992, 1999). Bogen argued that Jeffersonian transcripts feature in a complementary literary pairing with the main text where, at one level, they serve a mimetic effect to achieve the sense that events recorded in this way, really happened in this way. In this sense, the transcript's purpose is no longer to narrate the earlier events but, instead, to provide an orthographic emplotment of the relevant acoustic (and, by extension, videographic) details. The literary pairing is manifest in the task of the text, accompanying the transcript, to help the reader recognise in the "hyperabundance of transcriptural detail what is significant about the original event" (Bogen, 1999, p. 203). The problem that Bogen identified was that as the level of detail increases it begins to obscure rather than show the phenomenon, after which, the trick of the analyst becomes to reveal the relevant feature for the reader which has become hidden in amongst the sea of details.

Defending Jeffersonian detailing, Hepburn and Bolden (2017) argued that Bogen, in his 1992 paper, did not prove that his criticisms of transcription diminish the findings from its use. They stressed the need for continuing to include hearable intricacies of recorded phenomena in transcripts, in part, because "a good, precise transcript is a display of quality, of taking the words and actions seriously" (Hepburn & Bolden, 2017, p. 192). Disagreeing with Bogen's characterisation of the Jeffersonian transcript, they argued that the features transcribed are warranted as being those which co-participants demonstrably find consequential in the course of interaction. Hepburn's (2004) work on crying and Glenn's (2010) and others' work on laughter, demonstrating the Jeffersonian spirit of transcribing as a way of attending to features that members are attending to. Crying and laughter's identifying details are absent from, or are caricatured in, their conventional renderings in literature, newspapers, and, of course, most work in psychology, linguistics, sociology, etc. In transcribing significant details EMCA researchers become all the more intimate with their logics and uses, even as the familiar features of laughter or tears are rendered in ways that seem strange, as texts, to their original producers. Yet Bogen does not exclude attention to details of speech or gestures per se but leaves us instead wondering about how to transcribe in ways that do not obscure the phenomena, and perhaps might even preserve its familiarity without the requirement of the analyst to recover it. Though, as we will discuss later, graphic transcripts become just as entangled in this thicket around detail, strangeness, and literary pairings.

Our second reflection on transcribing emerges around the problems of how we are inevitably categorising and formulating members and their actions when we are transcribing. With no god's eye perspective for us, as transcribers, we are partial and we are always, also, finding and losing the phenomenon. Rod Watson (1997) reminded us that the convention of presenting a speaker by the same identification (e.g., "dri", an abbreviated "driver", for Transcript 1) at every transition relevance place is at odds with the shifting identities emerging from

the very activities transcribed. For example, for events happening in a vehicle we label its occupants in the transcript unchangingly even though, as the actions unfold, various pairings will become relevant: adult + child, aunt + nephew, driver + passenger, driver + navigator, etc. The salience of category identification is particularly marked in workplace settings, yet, of course, applies in other settings as well. As Watson puts it, these speaker formulations are "categorical-incumbency-as-transcribed" and "[t]he very transcription procedures for such occasions of speech exchange indicate a background reliance on the provision of membership categories" (1997, pp. 51–52). As such, the speaker's categorisation as e.g., "TV producer" precedes the transcribed action and thereby the transcript steers its readers to "hear" each turn-of-talk as tied to the category in the identification.

Our third reflection builds on Watson's categorisation problems but in relation to transcribing non-verbal action, given each non-verbal activity requires formulation for transcribing it. Again, it steers the reader toward what kind of action they should understand the person, or other agent, as doing. For example, in Transcript 1, Eric transcribed the driver's action as "inspecting instruments" rather than "glances at radio" or "looks down", etc. Equally, in line 2, the use of "gz" (an abbreviated "gaze") selects amongst the possible practices of looking. It occurs to us that we have come to use "gz" commonly in transcribing because it serves as a placeholder term for practices of looking, which leaves what kind of looking it is to be specified in the paired text. Trying to use more neutral or generic identifications of non-verbal actions does not solve the problem. Inevitably they lose the local recognisability and availability of the participants' ongoing production of their actions with their eyes, heads, hands etc.

Our fourth reflection on non-verbal action is around the granularity we use in describing embodied actions in transcripts where "gz" is itself a case in point. In their study of instructional demonstrations, Lindwall and Lymer (in press) build on Schegloff's (2000) analysis of the granularity of descriptions as a member's concern and resource. According to Schegloff, "[k]nowing how granularity works matters then not just substantively, but methodologically" (p. 719) because EMCA descriptions of practices stand in contrast to other social sciences' representations, in their greater attentiveness to identifying locally relevant details. Schegloff's point is that what details to include and when to be more or less detailed, should be tied to the level of detail being used by participants. Lindwall and Lymer contrast videos of instructions by medical professionals for students with Youtube instructional videos for a general audience. Via that comparison, they show that to standardise the level of detail in the transcriber's description then misses the divergent levels of detail the instructed actions provide and that characterises them. In short, transcripts that have built on Jefferson's pioneering work are not without problems that are internal and of abiding interest to EMCA. They should not be seen as the endpoint for our efforts to depict practices, document action, and do justice to the voices of others.

Transcribing with the comic strip form does not dissolve the problems of transcribing

It might appear that by reminding us of these four problems facing transcribers we are going to offer the comic strip form as a solution to those problems. We are not. Transcribing actions building on comic strip conventions provides us with distinct responses to those problems rather than solutions. Perplexingly for those hoping to escape the betrayals of transcription, our experience is that they add further problems for the analyst. A first place to begin is that, in Western popular culture, the comic strip form is associated with the trivial, the informal, and the fun. It stands in sharp contrast to the technical aesthetic of Jeffersonian and Mondada transcripts.

Or does it? In the early days of Jefferson's experiments, she contrasted what she was trying out with the existing transcription systems in linguistics:

> Phonetic transcripts are not accessible to most readers. And the sort of "comic book" orthography I use (e.g., for "What are you doing?", "Wutche doin?") is considered objectionable in that it makes the speakers look "stupid"; it seems to caricature them rather than illuminate features of their talk.
>
> *(Jefferson, 1983, p. 3)*

In a first look at a graphic transcript (Transcript 2) of the recording transcribed in Transcript 1, comic strip orthography, compared to a traditional CA transcript, caricatures the participants and their activities simply by resembling a comic. It seems also to have lost detail, particularly the details that were registered in the text in Transcript 1.

Jefferson (1983) balanced the risk of caricaturing speakers and their speech against the gain in accessibility of her "comic book" orthography. Her complaint was similar to Bogen's, but about phonetic transcription, which Jefferson argued was obscuring rather than showing identifying details. More than that, she showed that what might at first seem to be a trivial feature of speech, such as a "nyem" or

TRANSCRIPT 2 Graphic transcript of recording transcribed in Transcript 1.

the details of laughter particles, are deeply consequential for what is meant and done by participants in interaction. In publications, the seriousness of what the graphic transcripts is showing, is similarly vouched for in the accompanying analysis that reshapes first impressions of them as trivialising human affairs. Though, as we noted earlier, the association between comics and the inconsequential, and readers' expectations and conventions for reading them are variable across cultures of reading. At this point, it might still seem as if comic strips will be more fun to read, but the graphic transcript is by no means as captivating and vivid as classic comics. As with any type of transcribing, there is a meeting between the need for under-appreciated detail and the potential lack of imagination in depicting that detail. Part of our criteria in producing graphic transcripts, keeping in mind Bogen's complaint, is to show rather than obscure what is happening ahead of returning to it, as a members' accomplishment, in our analysis. There is a familiar grammar of comic strips for the EMCA scholar to draw on and adapt (see McCloud (1993) for an exemplary and fun guide). It is more than a visual grammar, it is what Grennan (2017) calls a *lexicogrammar*, which has distinct ways of representing speech, timing, time, actions, motion, etc., yet one that is familiar as part of, not just its use in comic strips, but also in the many advertisings, instructions manuals, etc. that share its grammar.

Perhaps the most central problem of graphic transcripts is the loss of a singular sense of, and measure of, the relative timing of actions. As mentioned above, a unitary measure of timing is implicit and explicit in the Jeffersonian format. The length of the line of ASCII text provides a measure of its timing, with variations on that measure marked in slower or faster delivery (e.g., '<slower>' and '>faster<'). There are ways to try and establish a graphic transcript with a comparable measure of timing, for example, you can add time markers explicitly in captions (see further experiments in showing time measures in Back, 2020; Laurier, 2019). In reflecting on the problem of showing a measure of timing, it is not that a sense of timing is missing from the comic strip form. The temporality of, and across, panels has always been multiple because they have the timing associated with the actions depicted in the image, the duration added by speech bubbles or other sounds, and the caption box's use in both timing of the panel as well as placing it in the past or future (see Laurier, 2014a, 2019). The lack of a single measure is also a possibility. Multiple temporalities are produced, drawn upon, recognised, etc. by members in and as part of multiple courses of action. Those multiple timings are a challenge in transcribing that Mondada (2018) also responds to.

The problem of timing takes a further twist in the graphic transcript. To provide for their intelligibility to readers as singular images in a panel, actions are routinely captured from video recordings at a moment in their course that makes them most recognisable as what they are seen to be by the analyst (Jayyusi, 1993). Actions' trajectories are broken into stages, which is more often in relation to their iconic recognisability than the shape of their trajectory, or they may be selected to represent other features of analytic interest visible in their course. However,

in trying to show trajectories, they are captured at the points when they start (for example, at their "home position"), at a mid-point, then again at their completion (for example, a return to home position). In selecting for their representativeness in any panel, they run into Jefferson's original complaint over stereotyping in transcription. Yet, one development of the comic strip EMCA has pursued is using distinctive images, the equivalent of Jefferson's *nyem*, to help show a thing found in the recording that a comic strip writer would discard as ambiguous or unrecognisable. The way those ambiguities of the screen capture are routinely resolved is by captioning the image (e.g., "inspects instruments" in Transcript 2) in order to instruct the reader on how to see specific features of the image (Lindwall & Lymer in press). Such a caption, of course, leads us back to the categorisation and formulation problem of embodied actions that besets transcripts, so reflection is required on the warrants for categorisations formulated by the analyst.

To balance out what will likely sound like an overly cautious introduction we will now offer succour by considering our recent experiments in transcribing. Firstly, we will introduce a data session transcribing technique that displaces the typical pairing of a Jeffersonian or Mondada transcript with playing video recordings. Secondly, we will describe how a graphic transcript has served us well, if not better than a textual transcript in showing phenomena related to visual perspectives.

Live transcribing in data sessions via rapid triptychs

One of the great strengths of the line-by-line Jefferson transcript is its flexibility to serve as a worksheet in data sessions. It roughly follows the timeline of the video and is correctable, augmentable, and annotatable (see for example: Antaki et al., 2008; Bolden & Hepburn, 2017; Laurier, 2014b). By contrast, over the past decade we, Eric and Tobias, had come to accept that crafted graphic transcripts were almost always considered useless for the annotation and note-taking of data session participants. It appeared to us that the kind of video annotation tools being developed were, in fact, the more promising avenue for supporting video data sessions (Albert et al., 2019; McIlvenny, 2021). Our understanding was that having graphic transcripts at hand distracted from paying unmotivated attention to the video materials, and that these transcripts best arrived late in the day, long after the first noticings of things.

As part of the background work to writing this chapter, we began experimenting with graphical transcribing "live" during data sessions. To share their noticings, participants were asked to make three-panel strips, or a *triptych*, though they were left with the freedom to have more panels if they wished. Using this many panels was a suggestion rather than a prescription. A pair of panels could work, or they could have experimented with many more panels than three. Data sessions routinely vary in how many noticings are suggested by each person, according to local rules of thumb, numbers of analysts present, time available, etc. While we

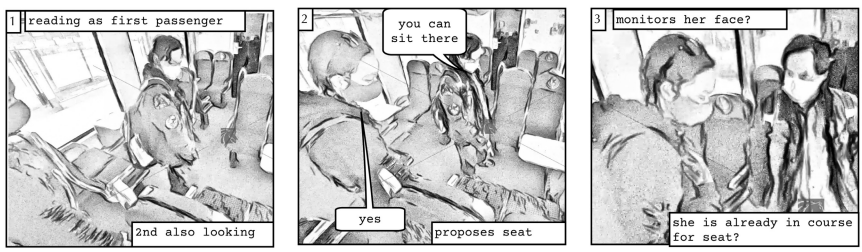
did not exclude transcribing by drawing from the video, or, in a nod to Jefferson, tracing paper over video screens, the fastest process was to frame-grab from the video recording (which is also the easiest way to re-grab and replace an image). For making the images, providing individual copies of the video was the only practical solution we could find[1]. We shared templates in Comic Life, PowerPoint, and Keynote formats for making comic strips. Participants, once they had made a preliminary noticing, would depict the thing which they noticed using three or more panels.

In one data session, we used video data of public transport during the coronavirus pandemic. Transcript 3 was made by a participant in that session. The session was used to explore how passengers select seats on a train as part of the joint accomplishment of following pandemic restrictions. Every other seat on the train had a *non-seating* sticker on it, serving as a resource and reminder for passengers of distancing guidelines. The sketch filter was applied to the original video for anonymisation purposes rather than during the transcribing in the data session. The transcript was used by the participant to support their describing of how seat-selecting and seat-proposing is divided between the passengers.

In the data session, each analyst usually only had time to produce one graphic transcript. These were sketches toward analyses. Even in that short time, the sketching routinely involved several grabs at the video frame to represent not just for each individual panel, but also, participants re-grabbed images for their fit in building sequences across the panels, thereby creating the narrative of the practice. Participants cropped to focus (e.g., panel 3 in Transcript 3), they broke up actions into shorter or longer durations across and within the panels, as ways of attending to different kinds of details and temporalities.

The participants reported that the depictional work of grabbing frames and captioning them shifted their attention toward the embodied, visual, and spatial aspects of the video recorded event. From the outset, in roughing out the transcripts, participants were inspired to pursue more adequate renderings of the embodied practices that they had become interested in. There was a stimulating dis-satisfaction with what they could render. At the same time, and in a limited

TRANSCRIPT 3 Three-panel noticing from data session in a graphic transcript workshop in June 2021.



time, the very process of producing the three-panel renderings became an occasion for the participants' initial noticings to foster more noticings at the point of documenting, as is so often the case in transcribing. Indeed, transcribing was fruitful rather than mechanical. For example, when Transcript 3 was produced, one researcher noted how, in cropping the video-grabs to focus on gaze and hand gesture, his attention was drawn to the literal *foot work* of the three passengers as part of proposing where to sit. To shift perspective to the feet was going to require cropping the image to foreground the feet of the passengers while also grabbing a differently timed set of frames from the video. Thus, the data-session graphic transcript was not an endpoint and instead pushed the participants through revealing recurrently what was missing. Live graphical transcribing altered other practices of the data session as well. Participants shifted away from replaying the source video, using their transcript to show the phenomena that they were talking about. This sharing of the document rather than replaying the video, lessened the time taken for each person to present in our online data sessions, though this will vary dramatically depending on local ecologies of room hardware and software for sharing screens, projecting video, etc. However, the change in presenting material was at the expense of staying with the video, with the attendant risk that transcripts supplanted the finding of details in the video recording.

One of the overlooked qualities of transcripts in data sessions is that they are usually passed back by members of the data session to the presenter as part of the sharing of collective analyses. The passing back shows the organisation and accountability of the workplace in terms of whose data it is and who should one day make something from the session. The multiple graphic transcripts continued to support this practice of passing back, yet as a different sort of resource. In a simple sense they were digital documents rather than paper, which fitted well to the fact that the sessions were online during the COVID pandemic. More significantly, they provided parallel, sometimes convergent, sometimes divergent, iterations of the very transcription process, rather than the one stabilised transcript with each member's annotation. The destabilisation was falling on a different side of the tensions between standardisation and inventiveness in transcribing. The presenter, at the end of a data session, had not only community noticings to draw on but also a portfolio of ways of representing the thing being analysed.

Depicting perspectives

When narratives of comic strips are drawn, rather than created with screen grabs, shifts of perspective are used to both show a perspective on things, people, and places and, relatedly, to show perspective as the perspectives of particular characters in the narrative. When we transcribe from the given perspectives of cameras, our ability to show different perspective is limited. In what follows we will, in that Jeffersonian spirit, use an analysis of perspectives to consider how it is more than simply the angle and location of the camera. The recognising, sharing, and

diverging of perspectives are at the heart of the spatialising of members' sense-making and organisation of their action and so, in itself, showing see-able perspectives requires Jeffersonian experimentation with the comic strip form.

Perspective is an optical matter and an analytic one. It encompasses the use of different angles of lenses from 45 degree to 360 degree, the location of single or multiple cameras, what view into a setting is being presented, the use of close-ups, mid-shots, and wides. All of which are compositional and narrative matters, not just in cinematography but also in video recordings of activities for analysis in EMCA (Heath et al., 2010; Mondada, 2014). Cinematography has developed visual (and audio) perspective as a resource and grammar for indexing experience, character, opinion, and more. If we take this meaning of perspective further, it brings us to more familiar EMCA concepts of stance, orientation, alignment, and so on. Moreover, as we ourselves produce transcripts that are derived from single or multiple cameras' perspectives, we quickly come to guide the reader's recognition through instructed seeing of the preceding recordings (Goodwin, 2000; Lindwall & Lymer, in press).

As we have said, graphic transcripts are both able to draw upon and find themselves limited by the field of view and location of the video camera or multiple cameras. Luff and Heath (2012) remind us that rather than being simply a practical and technical concern, the task of placing a camera, and thus selecting its point-of-view as well as its point-viewed-from, "uncovers methodological concerns that reveal the distinctive demands that video places on researchers concerned with the detailed analysis of naturally occurring social interaction" (p. 257). Shooting with a camera is, as Mondada (2014) argues, a "proto-analysis" of the phenomenon. Working from what is shot may or may not be convergent with that proto-analysis. The question here is something of a practical one. How do we then use given perspectives either way? As transcribers we do not share the illustrator's luxury of being able to freely select amongst possible perspectives, yet the production of the graphic transcript need not be simply subservient to the details captured by a video recording. As we have hinted already, whenever we are working with screen grabs, we can crop the still image; with multiple cameras we can select between their perspectives and locations. The 360-camera provides further possibilities in showing perspective because we can select from a complete circle of view around the camera, even if the viewpoint's point of origin remains the same. The 360-camera effectively displaces the selection of what angle to look at, from the event of recording, to the event of viewing the recording and, of course, widens the perspective selection possibilities for the transcriber.

Selecting and sequencing images from video recordings

Transcript 4 draws upon the camera-perspectives available from a dashcam in order to show members' perspectives as a matter of concern for these members and how perspectives are produced via their actions on the road. The dashcam

TRANSCRIPT 4 Transcript from Laurier et al. (2020) dashcam with conventional perspective pairings.

has two lenses, one showing the occupants of the car (panel 1) and one showing part of the view out of the front of the car (panels 2–6). In a sense it prefigures the 360-camera given that many 360-cameras end up being edited or rendered to produce dashcam grammars of forward and backward perspectives (see CAVA360VR in McIlvenny, 2021). The dashcam has a fixed position within the vehicle on the dashboard, which means that the forward camera perspective is moving when the vehicle itself is moving. The graphic transcript was produced to show an event that helps us understand how the beeps from car horns are produced to be hearable by other road users.

The transcript uses a single image of the occupants from the rear-facing camera in panel 1 to provide the reader with a sense of *whose* perspective they will come upon in the next frame. As we have noted earlier, perspective provides more than locating where we are looking from, it is to whom this view belongs. The pairing of perspectives of a scene with its members looking, is routinely used in comic strips and film editing to show the subject and afterwards show what they are looking at. The graphic transcript builds on that adjacently-paired perspective relationship. Panel 1 provides more, an excess, yet one that is a relevant resource. It is the driver's perspective but also visible in the image is a front seat passenger and a child in the rear, all looking ahead. In the following sequence of video frames, it shows the visual perspective as a member's concern and practical accomplishment.

The analytic phenomena that Eric sought to show in creating Transcript 4, for the publication on driving-in-traffic, was what he had from the recording: the use of the car itself as a perspective producing device. The auto-rickshaw in front of

the car is positioned, for the driver, as an obstacle to looking ahead. When the driver alters the car position as shown in panel 3, we can see, as the driver can see, a lack of vehicles ahead on the pavement side. In panel 5 we can see, through the graphic transcript, how returning to the middle of the dual lanes shows the view ahead for the slot to be overtaken into but that the view of the pavement side has been obscured again. The transcript is seeking to show the driver's work of finding perspectives into the traffic ahead. Eric selected the video-stills in panel 2 and 3 to show that gap, and for a certain perspective to be visible in its pairing with the image before and after it.

The sequencing in Transcript 5 is similar to that in Transcript 4. We see a passenger leave the train station platform, walk through a staircase to an underground tunnel while being interviewed about his daily commuting during COVID-19. The frames are grabbed from a 360-degree camera recording, which allowed Tobias to crop from a full sphere. The aim of the transcript is to show, from the passenger's perspective, the distribution of co-passengers in a shared space. Showing the

TRANSCRIPT 5 Building a passenger's perspective (Back, 2021).

passenger-interviewee looking and pointing towards the tunnel in panel 1, accompanied by his continuous voice-over in the connected speech bubbles throughout the subsequent panels, establishes him as the one to whom the perspectives in panels 2 and 3 belong.

An earlier version of Transcript 5 included the passenger from panel 1 in all three panels walking down the stairs on the right side of each panel (similar to the composition of panel 1). However, the analysis shifted from a concern with the multiactivity of walking and talking as intersubjective practice to the formations of the group of co-passengers walking in front of the camera and the interviewee. Consequently, Tobias re-cropped the panels to exclude the talking passenger and guide the reader's attention towards the group of passengers.

As with the auto-rickshaw, the moving passengers are recognisably the same across the three panels. The changing environments: the railway platform in panel 1, the stairs in panel 2, and the underground level in panel 3 are seen as sequentially related. They show the reader, from the interviewee's perspective, how the passengers ahead move together across the three sites, while maintaining some degree of social distancing due to COVID-19 regulations. The reader's attention is redirected from timing as a member's resource, as emphasised in the line-by-line transcript, to spacing. That is, the transcript focuses its reader's attention on the relative location of bodies from one site to the next and within those sites. The recognisability of each scene as a next, but not incrementally progressing, does however produce a sense of the longer stretches of time between each panel. The interviewee's words are read in terms of how and when he sees the spatial distribution of co-passengers from panel to panel and site to site, progressing from the reader seeing the crowd as dense as "herrings in a barrel" in panel 2 to revealing the later assessment of the interviewee on the number of passengers in the tunnel as "not so bad" in panel 3. Given the different ways of assessing the busy-ness of public spaces during the COVID pandemic, it is possible that our interviewee would have seen even a small group like the one shown as too dense for him to feel comfortable.

From panel to panel there are undoubtedly other elements to the participants' work of seeing what is happening: glancing to the sides, drivers using their mirrors, drivers and walkers looking over their shoulder, noticing co-passenger formations, etc. Being tied to the footage, graphic transcripts sometimes produce strange and ambiguous visual sequences. Their strangeness requires the use of caption boxes and further instruction through the accompanying text to help the reader make sense of unconventional perspective pairings which again returns us to the problems of analysing the category organisation of the scene.

In re-examining the work of the graphic transcript, we can begin to understand not just that it allows for complex assemblies of members' practices but also that multiple perspectives and temporalities are at work where those perspectives and temporalities are relevant to members' practices. Our concern has been to show how we might try and address those via the renderings of the graphic transcript.

What we want to remind you of, before concluding, is that the transcripts in this section were, of course, the result of a number of versions where different grabbed images were tried, various layouts, speech bubbles shifted around, etc.

Conclusion

In opening this chapter, we examined, at some length, the reasons for and some of the abiding problems of transcription in EMCA studies. We considered the introduction of "live" graphic transcribing in data sessions as a novel way of showing what each analyst had noticed. We underlined how rendering a video recording as a comic strip shifted the analysts' attention toward the embodied, visual, and spatial aspects of members' practices and how it changes the sharing of noticings with the presenter. In a further reflection on our recent experiments in transcribing, we demonstrated how, in finalising our transcripts for publication, we responded to the phenomena of visual perspective and the intimately related problem of reproducing members' perspectives from the perspective(s) of our cameras.

As we argued, in the comic strip form, we have distinct possibilities to respond to the action formulation and actor categorisation problems: For example, in relation to *gaze*, by showing direction of view, and to whom a specific view belongs, rather than merely describing who is looking/gazing at what when. We have shown how transcribing using comic strip conventions hybridised with transcription criteria was shaped around particular analytic foci. We used the case of perspective as one such spur to us, as part of participants' sense-making and action production, finding perspectives into traffic, guiding a co-passenger through a crowded tunnel, distributing seats on a train, etc. As Jefferson herself showed, this required both drawing upon and disrupting existing conventions for exhibiting phenomena.

Transcribing, in the comic strip form, shares EMCA's commitment to transcription as a way of highlighting and pointing up practices. As John Heritage recalls the 1970s, when Jefferson first circulated her recording and transcripts:

> [T]he transcripts highlighted features of the recordings that we might have otherwise overlooked, and pointed up practices in the talk which turned out to be highly relevant to our analyses. Because the tapes and transcripts were quite widely circulated, they contributed to a 'culture of transcribing', to the development of a set of common standards that we learned to share, and tried to live up to. The standards allowed us to recognize failure – our own and other people's – as well as success.
>
> *(Heritage quoted in Hepburn & Bolden, 2017, p. 181)*

While Heritage's comments emphasise the common standards, transcribing has also always been a site for supplementing or altering those standards when we reveal further orders of social practices. What we would urge caution around is reproduced standardised forms being the criteria that we use to recognise

Jefferson's rightness ahead of inquiries that are "utterly obscure and unstable". Indeed, we do not wish to propose a fixed set of standards for the many elements of the graphic transcript (e.g., panels, captions, textualised speech, layout of bubbles). For us, the Jeffersonian breakthrough emerged from questioning the existing and dominant traditions of representation of hearable and consequential speech. Even as Jefferson was building a new system because existing representations of talk were ignoring the hearable and accountable features of speech, she was wary around how speakers might then be caricatured in her transcribing. There is no existing standard for graphic transcript orthography in EMCA that we wish to depart from, it is already a departure from Jeffersonian and Mondada formats. Building graphic transcripts for EMCA inquiries remains instead, as it shouldn't surprise EMCA scholars to learn, an ad hoc endeavour.

Our aim here has not been to pull you into a puzzling transcription procedure, a tutorial of a primary problem in EMCA studies, that any recording and any transcript misses the haecceities of the thing. Garfinkel (2002) taught us that lesson in his summoning phones tutorial, where transcribing was instructively impossible and the tutorial in transcribing was at best a secondary one for his students and, more likely, a useful ruse. Our aim instead has been to help EMCA researchers develop their craft in good-enough representations of events, weighing the balance between their identifying details and gratuitous granularity. Elsewhere, Garfinkel did not abandon attempts to show phenomena via vivid ethnographic descriptions, diagrams, photographs, and re-use of members' own representations of a phenomenon. Our warrant for experimenting with graphic transcripts is to maintain the Jeffersonian spirit of showing what is missed or otherwise caricatured. Using the comic strip's forms of panels, speech bubbles, etc. is a different way of picking up Jefferson's pencil and tracing paper to settle upon what we are attending to.

Note

1 Depending on the sensitivity of the materials, this may not be possible and/or may require trusting members to destroy their copies of the video at the end of the session.

References

Albert, S., Heath, C., Skach, S., Harris, M. T., Miller, M., & Healey, P. G. T. (2019). Drawing as transcription: How do graphical techniques inform interaction analysis. *Social Interaction: Video-Based Studies of Human Sociality*, *2*(1). https://doi.org/10.7146/si.v2i1.113145

Antaki, C., Biazzi, M., Nissen, A., & Wagner, J. (2008). Accounting for moral judgments in academic talk: The case of a conversation analysis data session. *Text and Talk*, *28*(1), 1–30. https://doi.org/10.1515/TEXT.2008.001

Ayaß, R. (2015). Doing data: The status of transcripts in conversation analysis. *Discourse Studies*, *17*(5), 505–528. https://doi.org/10.1177/1461445615590717

Back, T. B. (2020). *One more time with feeling: Resemiotising boundary affects for doing 'emotional talk show' interaction for another next first time* (ISBN 978-87-7210-627-4) [PhD Thesis, Aalborg University]. Aalborg University Press.

Back, T. B. (2021). *Building a passenger's perspective.* [Unpublished transcript from the project 'Travelling Together']. Aalborg University. Copies available from the author.

Bogen, D. (1992). The organization of talk. *Qualitative Sociology*, *15*(3), 273–295. https://doi.org/10.1007/BF00990329

Bogen, D. (1999). *Order without rules*. Suny Press.

Deppermann, A., Laurier, E., Mondada, L., Broth, M., Cromdal, J., De Stefani, E., Haddington, P., Levin, L., Nevile, M., Rauniomaa, M. (2018). Overtaking as an interactional accomplishment. *Gesprachsforschung - Online-Zeitschrift zur verbalen Interaktion*, *19*, 1–131.

Garfinkel, H. (2002). *Ethnomethodology's program*. Rowman & Littlefield.

Glenn, P. J. (2010). *Laughter in interaction*. Cambridge University Press.

Goodwin, C. (2000). Practices of seeing: Visual analysis - An ethnomethodological approach. In T. v. Leeuwen & C. Jewitt (Eds.), *Handbook of visual* (1st ed., pp. 157–182). Sage Publications.

Goodwin, C., & Salomon, R. (2019). Not being bound by what you can see now. Charles Goodwin in conversation with René Salomon. *Forum: Qualitative Social Research*, *20*(2). https://doi.org/10.17169/fqs-20.2.3271

Grennan, S. (2017). *A theory of narrative drawing*. Springer.

Have, P. t. (2007). *Doing conversation analysis, A practical guide*. Sage.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. Sage Publications.

Hepburn, A. (2004). Crying: Notes on description, transcription, and interaction. *Research on Language and Social Interaction*, *37*(3), 251–290. https://doi.org/10.1207/s15327973rlsi3703_1

Hepburn, A., & Bolden, G. B. (2017). *Transcribing for social research*. Sage Publications.

Jayyusi, L. (1993). The reflexive nexus: Photo-practice and natural history. *Continuum: The Australian Journal of Media & Culture*, *6*(2), 25–52. https://doi.org/10.1080/10304319309359397

Jefferson, G. (1978). What's in a `Nyem. *Sociology*, *12*(1), 135–139. https://doi.org/10.1177/003803857801200109

Jefferson, G. (1983). Issues in the Transcription of Naturally-Occurring Talk: Caricature versus capturing pronunciational particulars. Tilburg Papers in Language and Literature, *34*.

Jefferson, G. (1985). An exercise in the transcription and analysis of laughter. In T. A. Van Dijk (Ed.), *Handbook of discourse analysis* (pp. 25–34). Academic Press.

Laurier, E. (2014a). The graphic transcript: Poaching comic book grammar for inscribing the visual, spatial and temporal aspects of action. *Geography Compass*, *8*(4), 235–248. https://doi.org/10.1111/gec3.12123

Laurier, E. (2014b). The lives hidden by the transcript and the hidden lives of the transcript [Unpublished manuscript].

Laurier, E. (2019). The panel show: Further experiments with graphic transcripts and vignettes. *Social Interaction: Video-Based Studies of Human Sociality*, *2*(1). https://doi.org/10.7146/si.v2i1.113968

Laurier, E., Muñoz, D., Miller, R., & Brown, B. (2020). A bip, a Beeeep, and a beep beep: How horns are sounded in Chennai traffic. *Research on Language and Social Interaction*, *53*(3), 341–356. https://doi.org/10.1080/08351813.2020.1785775

Lindwall, O., & Lymer, G. (in press). Detail, granularity, and laic analysis in instructional demonstrations. In M. Lynch & O. Lindwall (Eds.), *Instructed and instructive actions*. Routledge.

Luff, P., & Heath, C. (2012). Some "technical challenges" of video analysis: Social actions, objects, material realities and the problems of perspective. *Qualitative Research*, *12*(3), 255–279. https://doi.org/10.1177/1468794112436655

McCloud, S. (1993). *Understanding comics: The invisible art*. Kitchen Sink Press.

McIlvenny, P. (2021). New Technology and Tools to Enhance Collaborative Video Analysis in Live 'Data Sessions'. QuiViRR: Qualitative Video Research Reports, 1. https://doi.org/10.5278/ojs.quivirr.v1.2020.a0001

Mondada, L. (2014). Shooting as a research activity studies of video practices. In M. Broth, E. Laurier, & L. Mondada (Eds.), *Studies of video practices: Video at work* (pp. 33–63). Routledge.

Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, *51*(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Schegloff, E. (2000). On granularity. *Annual Review of Sociology*, *26*(1), 715–720. https://doi.org/10.1146/annurev.soc.26.1.715

Watson, R. (1997). Some general reflections on 'categorization' and 'sequence' in the analysis of conversation. In S. Hester & P. Eglin (Eds.), *Culture in action: Studies in membership categorization analysis* (pp. 49–76). University Press of America.