How to Transcribe Multimodal Interaction?

Jeff Bezemer
Institute of Education, University of London
j.bezemer@ioe.ac.uk

Draft chapter for C.D. Maier & S. Norris (eds). Texts, Images and Interaction: A Reader in Multimodality. Mouton de Gruyter.

May 2012

PLEASE DO NOT CITE WITHOUT PERMISSION FROM THE AUTHOR

Abstract

This working paper looks at some of the key issues to consider when transcribing multimodal interaction. Making a transcript is an invaluable analytical exercise: by forcing yourself to attend to the details of a strip of interaction you gain a wealth of insights into the situated construction of social reality, including insights in the collaborative achievements of people, their formation of identities and power relations, and the socially and culturally shaped categories through which they see the world. In this paper I reflect on the process of making a multimodal transcript. I discuss the following steps: (1) Choosing a methodological framework (Reviewing multimodal frameworks and Considering rhetorical status of the transcript); (2) Defining purpose and focus of transcript (Selecting episode and features to transcribe and Defining questions to address; (3) Designing the transcript (Creating a template, Defining transcription conventions and Filling in the template); (4) Reading the transcript (Annotating and Recounting the transcript); (5) Drawing conclusions (Addressing research questions and Making connections with other studies and theoretical constructs).

Introduction

A recurring challenge in my work and that of many others who are interested in multimodal interaction is transcription. Whenever I analyse a video recording of an interaction the question I am having to address is, How can I represent gesture, for instance, or gaze, or speech, and the alignment between all these different means of communication in writing and perhaps in video stills or drawings? There are at least two reasons why so many researchers take on that challenge. First, making a transcript is an invaluable analytical exercise: by forcing yourself to attend to the details of a strip of interaction you gain a wealth of insights into the situated construction of social reality, including insights in the collaborative achievements of people, their formation of identities and power relations, and the socially and culturally shaped categories through which they see the world. That is the epistemological function of transcription. Second, transcripts can be included in academic publications, which by and large are still paper-based. That way the

transcript becomes verifiable 'evidence' of the argument that is developed in the publication. That is the rhetorical function of transcripts.

There are many different approaches to multimodal transcription. In previous, collaborative work we tried to make sense of some of the differences and similarities. By comparing a number of different published transcripts we reconstructed some of the epistemological and rhetorical choices that the transcribers made (See Bezemer & Mavers, 2011). Note that in acknowledgement of the significance of all these choices and its analytical potential multimodal transcription is usually done by the researchers themselves, and not 'outsourced' to external transcribers (who commonly transcribe interviews, for instance). Indeed, like any text, transcripts reveal as much about what is represented as they do about the text maker and the context in which they were produced. For instance, transcribers make choices about which clips to transcribe, which of the modes captured in the clip to transcribe, and how to represent these. All of these choices reflect the interests of the transcribers, their professional vision. And they have epistemological implications: the *re*-making of video-recorded interaction as a multimodal transcript leads to fresh insights.

In this chapter I provide a reflexive account of how I made a detailed multimodal transcript for one particular study. The study was on communication in the operating theatres of an inner-city hospital in London (Bezemer et al. 2011). The question that the transcript was designed to help me answer was, how do the members of a surgical team communicate to accomplish an operation? The transcript is a representation of an exchange between two surgeons and a scrub nurse. It was selected from a set of ten video-recorded operations. Instead of using wide-angle video cameras to capture what happens around the operating table, we used a camera fitted to a light handle, allowing us to capture the features that the participants in the interaction typically orient to, that is, their hands and those of their colleagues, their instruments, and the parts of the patient's body that they operated on. We used a wireless microphone worn by one of the surgeons to record the audio. In addition to these recordings we took photographs and kept detailed field notes of all the operations that we observed, particularly noting changes in the spatial configuration of participants around the operating table.

The transcript I will discuss might look quite different from many other multimodal transcripts, but I take it that researchers go through a number of similar steps to make them. These steps are outlined in Fig. 1. In what follows I discuss each of these steps, illustrating them using the research described above as a case study.

Fig. 1: How to transcribe multimodal interaction?

Choose a methodological framework
- Review multimodal frameworks
- Consider rhetorical status of the transcript

Define purpose and focus of transcript
- Select episode and features to transcribe
- Define questions to address

Design the transcript
- Create a template
- Define transcription conventions

DRAFT

Fill in the template

Read the transcript
Annotate the transcript
Recount the transcript

Draw conclusions
Address questions
Make connections with other studies and theoretical constructs

Choosing a methodological framework

Making a multimodal transcript starts with choosing a methodological framework that is apt for doing multimodal analysis. These frameworks have distinctly different takes on multimodality and work with different notions of validity (see Chapter x of this volume), and that is reflected in the multimodal transcripts. The transcript that I am discussing here was made for a paper (Bezemer et al, 2011) in which my co-authors and I adopted a conversation analytic approach. Originally focused on the study of talk, a growing body of work in Conversation Analysis (CA) now deals with a range of modes of communication. Much of this work is focused on medical work. Some of these studies are focused on surgical activity (Mondada, 2003, Koschmann, 2011, Svensson et al. 2009). Adopting the methodology underlying this body of work had a number of important implications for transcription.

First, it meant that I would select a small timescale, say, snippets of no more than a minute or so, so that the selected video clip can be reviewed frame-by-frame (there are 30 frames in one second). It is probably fair to say that the more detailed the multimodal transcript, the smaller the timescale that the researcher can afford to select. Had I adopted a different approach then I might have chosen to analyze clips second-by-second, or minute-by-minute, allowing me to cover more material. That might then have enabled me to made different claims, for instance, about the frequency of occurrence of a certain analytical category.

Second, adopting the methodology underlying the 'multimodal' studies in conversation analysis on medical work meant that I was going to make the transcript not just for myself but also for an audience, i.e. for the readers our our paper. Not just that, it was going to feature as 'evidence' for the arguments I was going to develop in the paper, giving the transcript a particular status within the manuscript. In CA, transcripts are not presented as illustration of a main body of text, they are presented as the object of analysis, as the main text that its surrounding text comments on, 'contextualizes'; a bit like the semiotic relation between a painting in a museum and its caption. All claims made in the surrounding text need to be grounded in the transcript (and not exclusively in, say, an interview with one of the people featuring in the transcript or in ethnographic insights of the researcher). CA transcripts tend to follow transcription conventions originally defined by Gail Jefferson, one of the 'founders' of CA. However she only suggested conventions for the transcription of speech. A variety of forms are now used to transcribe communication in other modes in a way that is 'acceptable' and convincing to

DRAFT

conversation analysts. That means, for instance, that it has to represent in detail the temporal unfolding of the interaction in all modes included for transcription.

Third, it meant that I was going to use some of the categories and concepts that are consistent with the approach I adopted. In the penultimate section of the chapter I will connect the insights I gained from making the transcript with the observations of other researchers who studied multimodal interaction. For instance, I draw on Kendon (1990) and Norris (2004) to explore how people participate in more than one activity at the same time; and I draw on Goffman (1971), Hindmarsh & Pilnick (2007) and Scollon & Scollon (2004) to understand how people read the bodies of others. Paradoxically, in this study, while clearly 'multimodal' in its outlook, I didn't use the notion of 'mode' to separate out different *sets* of socially and culturally organised meaning making resources (Kress, 2009). Instead I separated out the different parts of the body that were available to and used by the participants as resources for making meaning (and visible on the recording) alongside the use of speech: head, upper body and arms and hands. The reason for not identifying modes, at least not *a priori*, was my being a complete outsider of the professional community of operating theatre staff. Hence I did not know how its means of communication were organized.

Define purpose and focus

Researchers select episodes for transcription for any number of reasons. In previous research I was often drawn to occasions where the interaction order is disturbed, making visible some of the ideologies operating in that context. For instance, I transcribed an excerpt from a video recording of a secondary school classroom in which a student threw a pack of chewing gum to his classmate while the teacher, who noticed this, was giving instruction to the class (Bezemer, 2008). The transcript allowed me to investigate the regulation of displaying orientation through the body. The episode described in this chapter was selected for the opposite reason. Now my attention was drawn to the apparent seamlessness of the interaction between a junior surgeon on the one hand, and a consultant surgeon and a scrub nurse on the other hand.

Having replayed the clip a number of times I noticed that two activities occurred at the same time, involving different sets of participants. The consultant is tying knots, which is facilitated by the first assistant, a senior surgeon, who provides the necessary traction by holding a retractor in place. The consultant and the first assistant are talking about the management of beds. The SHO, a junior surgeon or 'senior house officer' (SHO), is within earshot of this conversation but is not a "ratified participant" (Goffman 1981) in it. The episode starts at the point where the SHO makes a request for and subsequently receives scissors from the scrub nurse. It ends after the SHO has applied the scissors to cut a suture that the consultant has just finished tying. All this happens in less than 15 seconds. I noticed that the SHO was ready to apply the scissors exactly at the point where the consultant needed someone to cut the suture he had just tied. That raised questions such as, How did the SHO know that she was expected to cut the suture at that point in time and at that point of the suture without having received any spoken instructions from the consultant or first assistant? And where did these scissors come from? It took me (and all others I showed this clip to) several rounds of playing the episode in slow-motion to discover how the consultant used his body to signal to the SHO where and

DRAFT

when to cut the suture; and to discover how the SHO had used her body (and speech) to make a request to the scrub nurse for scissors.

Having roughly defined what the boundaries were of the episode I then had to chose exactly what the beginning and the end points of my transcript were going to be. As starting point I chose what I came to see as the onset of the SHO's request, that is the point that she begins to turn her body away from the operative field and towards the scrub nurse. As end point I chose the point immediately after the SHO had cut the suture. I also need to consider what features I wanted to transcribe. That was shaped, in part, by the partiality of the frame of the video recording. Since this camera was inside one of the operating lights, it moved along as the surgeons adjusted the positioning of the light. In the clip selected for transcription the SHO can be seen from her back and from a high angle. Her face is invisible, and so are her left arm and hand and her legs. What I could transcribe were the movements she makes with her head (suggesting the direction of her gaze), her trunk/upper body, and her right arm and hand, and her use of speech. So that's what I chose to transcribe. More specifically, I identified three dimensions of the use of head, trunk and arm and hand that I wanted to detail: their temporal and their spatial organization (up/down, left/right). I excluded the talk between the consultant and the first assistant as they play no role in the interactions that the SHO is engaged in. I also excluded the body movements of the first assistant (whose task is to hold a big retractor) and I only selectively transcribed the movements of the consultant.

As I was selecting the episode and the interactional features I was formulating questions at the same time. I went back and forth between defining and redefining questions and honing in on a particular episode. Being both 'commissioner' and 'designer' of the transcript, I had to brief myself before doing any transcription at all. I was particularly interested in two questions. First, how does the SHO communicate with the scrub nurse and with the consultant, respectively. For instance, how does she signal to the scrub nurse that she requires scissors? And how does the consultant signal to the SHO when and where to cut the suture he's holding? Second, how does the SHO manage to remain involved in two activities at the same time, namely the requesting and passing of an instrument, and the knot tying and cutting? Drawing on the CA studies cited above and other multimodal research I assumed that the SHO and the other participants would use their bodies to achieve all this. So I had to find a way to represent how the various movements of body parts map onto each other. More specifically, the transcript had to show a) how the body movements of one person are (dis)aligned, for instance to make a request; b) how the body movements of different people are (dis)aligned, for instance, when a request is acknowledged.

Design the transcript

Having established the focal episode and the purpose of the multimodal transcript I then had to design it, that is, creating a template and defining the conventions for transcribing the features I had included. I considered a number of different designs. Multimodal transcripts are not only 'multimodal' in that they represent multimodal interaction, they are also multimodal in that multiple modes operate in the transcript, usually a combination of writing, typography, image, and/or layout. Each of these reshape the focal interaction in particular ways. Since I had chosen to represent the selected episode frame-by-frame the use of photographic stills seemed inapt – it

DRAFT

would take 300 stills to represent one second. However by drawing lines of various kinds on grid paper I felt I could represent the timing and direction of the movements that the SHO made with the body parts I was focusing on. This approach was inspired by the work of Christian Heath and colleagues (2010).

As I was interested in the temporal unfolding and the synchrony between body movements I designed a template in which temporality is arranged horizontally, with the different body parts to be detailed separated out on the vertical axis. I drew a horizontal time line on a piece of grid paper and worked out how many frames or seconds each millimetre would stand for. For instance, you could take 1mm per frame, that's 30mm per second. That means that in landscape orientation you could fit 9 seconds across the full width of the sheet. If you took 1cm per second, that's 1mm per 10msecs, you could fit 27 seconds on one line. I had just under 15secs to transcribe, so I did 1mm per 3 frames. I could now draw lines below the time line, each line representation a different body (part).

I then defined what I wanted to transcribe on each line. The first line was going to represent the consultant's actions; they were broadly described in terms of "tying knots" and "holding thread tight." The following four lines were to represent the SHO; one line for her use of upper body, one for her use of right arm, one for her use of head, and one for use of speech. I also defined conventions for expressing movement and fixation of these body parts. I used a dotted line for movement, and a continuous line for fixation. Discontinuations of lines indicate "invisibility" of a feature on the video record, for instance, when the SHO's head temporarily blocks the view of the consultant's hand movements. I described these movements as 'up/down' and 'left/right'. Speech is used at only one point in this episode and was transcribed using conventional orthography and placed on a separate line. The time lapse of speech is detailed as a dotted line.

With the template designed, I started to fill it in using a media player that allowed me to forward frame-by-frame and to vary the playing speed. I focused on one body part at the time as that is the easiest way to keep track of the minute changes I was interested in. As I replayed the clip I decided where a movement started and where it ended, and in what direction it went, and translated that into pen strokes and annotations on my template. That way I filled in the empty fields on the template. Some time later, as I was preparing the manuscript in which the transcript was to appear, I 'digitized' the transcript by remaking it in Windows Paint. That's the version of the transcript that is reprinted here as Fig. 2.

[insert Fig. 2 about here]

Read the transcript

Transcripts don't speak for themselves. You need to read it, mark points of interest, annotate it, and refine it. For instance I found it useful to draw horizontal lines from the time line downwards at what seemed 'critical moments' to see if the reconfiguration of bodies does indeed mark a shift of some kind (I subsequently deleted these lines again from the transcript). I also add stills to depict bodily configurations at certain points in the epsiode. A good way to then organize your thoughts is to begin writing your interpretation of the transcript. As author you are expected to take readers through your transcript. Now the challenge is not to translate a video clip into a multimodal transcript, but to translate the multimodal

DRAFT

transcript into a (written) ‘recount’. As in the transcription stage you make selections and highlight interactional features for your audience, and you gain new insights by systematically describing what you see in the visual representation you made. Fig. 3 reflects what I wrote about my transcript.

Fig. 3: A recount of the multimodal transcript (Source: Bezemer et al., 2011:406-408. Time is indicated in minutes:seconds:frames).

Upper body

The SHO stands between the patient’s legs, which are bent and which rest on frames (see Fig. 2). The scrub nurse and the consultant are on her left, and the registrar is on her right. The SHO has little room to maneuver, yet she can and does move her *upper body* sideways and upward and downward to get the view and make the hand movements that she needs. Moving her trunk downward and upward allows her to respectively get physically closer to and to move away from the operative field. Moving slightly to the left allows her to get physically closer to the scrub nurse. Thus the movements of her upper body are suggestive of her orientation toward the various actions of others unfolding around her. Tilting toward the operative field suggests an increased engagement with the consultant’s manual actions. Erecting her upper body and bending slightly to the left but without actually pivoting her trunk suggests a decreased engagement with the consultant’s manual actions, but without a complete withdrawal, while suggesting the onset of a temporary engagement with the scrub nurse. Thus she can display her engagement with the actions of two people who at that point are themselves not oriented toward each other.

The SHO begins to tilt toward the operative field as the consultant begins to tie knots. At 9:46:00 she begins to move back up and tilts slightly to the left, suggesting engagement with the scrub nurse, without entirely withdrawing from the consultant’s manual actions. Three seconds later she moves back to the middle of the space between the patient’s legs, where her trunk remains in a more or less stable position for just over a second. Thus she has disengaged with the actions of the scrub nurse and displays an increased orientation to the operative field. She then turns slightly to the left again, stays in that position for half a second and then moves farther upward for a second and a half. In this way she displays engagement with the scrub nurse again while sustaining the consultant’s actions. She maintains that position for just over one and a half seconds and then begins to tilt toward the operative field again, until 9:57:21. She has disengaged with the scrub nurse again and is now displaying orientation to the operative field only. She maintains that position for the rest of the clip. So in this fifteen-second clip the SHO engages momentarily with the scrub nurse twice, without ever completely moving away from the operative field and the consultant’s manual actions.

Arms and hands

The SHO’s *left arm* is invisible in the clip, but it looks like she is resting that arm on the patient. Her *right arm* and *hand* play a crucial function in her coordination. At the start of the clip, when the consultant is tying knots, her right hand is still resting on the patient. At 9:46:15 she moves her hand up, only to let it rest again. Moments later she moves her hand up again while stretching her arm and moving it to the left. This happens as she is moving her upper body up and tilts slightly to the left. This accentuates her engagement with the scrub nurse. Her right hand is now in front of her trunk, and while she soon starts shifting away from the scrub nurse, again she

maintains that position of arm and hand until she has received the scissors from the scrub nurse at 9:53:08. That way she continues to display orientation to the scrub nurse, signaling to the scrub nurse that she is available for receiving the scissors she requested while moving closer to the operative field again. The scrub nurse is likely to have anticipated the request. Only moments ago she has passed a stitch to the consultant, and she is well aware that the attached thread will have to be cut when he has completed the knot tying.

When the SHO has received the scissors she moves her arm back to the right, accentuating her withdrawal from the scrub nurse's actions. As she moves her arm closer to her trunk she twists and grasps the scissors, moving her fingers into its rings (see Fig. 2, Picture 3). From 9:55:03 she holds the scissors in a fairly fixed position, apparently ready to apply them (see Fig. 2, Picture 4). At 9:56:18 the consultant stops moving his *hands and fingers*, holding the thread in a fixed position using both hands, in such a way that it can be cut (only) by someone else (see Fig. 2, Picture 5). His hand positioning "broadcasts" to the SHO the request that the thread needs to be cut. The SHO then moves closer to the thread, and at 9:57:21, just over a second after the consultant started holding the thread fixed and tight between his fingers, she cuts it (see Fig. 2, Picture 5). She then moves her arm and hand back to her trunk while rotating the scissors on her fingers and grasping it such that it points upward.

Head gaze

The SHO's shifting head position and gaze direction is a further indication of her orientation toward the actions of others around her. Moving her trunk without also moving her head allows her to display dual orientation toward the operative field and the scrub nurse, and to gradually engage more with the one and less with the other. When she first turns her head to the left, at 9:47:12, she has already begun to move her trunk up and to tilt it slightly to the left; she also has already begun to move her arm and hand toward the scrub nurse. Thus the turn of her head, which allows her to direct her gaze to the scrub nurse, is the third indicator of her increased engagement with the scrub nurse, alerting her to an upcoming request. Then, at 9:48:00, she verbalizes the request: "scissors please" (see Fig. 2, Picture 2). Now that the request has been completed, she turns her head back to the right, allowing her to look into the operative field again. She knows that soon the scrub nurse will offer her the scissors, and she still holds her open hand within the scrub nurse's reach. At 9:51:00 she turns her head to the scrub nurse again. The positioning of her head and direction of her gaze is not visible on the video record until she receives the scissors just over two seconds later and turns her head back to the operative field.

Drawing conclusions from the transcript

As I had completed the recount I returned to the questions I had asked myself when I selected a strip of interaction and began to connect the insights I had gained to work in which similar actions are described. My question was, first, how does the SHO manage to remain involved in two activities at the same time, namely the requesting and passing of an instrument, and the knot tying and cutting? Second, how does the SHO communicate with the scrub nurse and with the consultant to achieve those two activities?

Engaging in simultaneously unfolding activities
The SHO managed to help sustain and complete one activity (knot tying) through the timely initiation and completion of another (instrument exchange), each with its own set of participants. The SHO managed to simultaneously engage in those two activities by using her body in particular ways. This observation relates to findings on multimodal interaction in other settings. For instance, Norris transcribed a video recording of a school crossing guard helping children cross a street. She shows that

"while the crossing guard shifts her focal attention from directing the cars to directing the children and back again, she simultaneously engages in two-higher level actions on different levels of attention/awareness. In other words, when the traffic guard is directing traffic, she is engaged in focused interaction with the drivers, while she is simulatenously engaged in interaction with the children at the corner: making sure that they are safe. Her engagement in the interaction with the children at this time is not as focused as her interaction with the drivers, but is clearly ongoing." (p. x).

Like the crossing guard, the SHO is juggling between two activities, each of which require her attention: she needs to keep her eyes on the consultant, so as to know when and where to cut. She also needs to keep an eye on the scrub nurse, first to make a request for scissors and then to receive the scissors. The transcript shows how the SHO uses her body to manage this multiple engagement, namely through what Kendon (1990) calls the "f-formation." Looking at "informal" gatherings, he shows how participants standing in a social circle with two others can temporarily turn their head away from the center point of this "f-formation" while sustaining their involvement in the talk. They keep their lower body in line with the center of the f-formation to express engagement with the talk and use the upper body to engage, temporarily, with someone situated outside the formation. The multimodal transcript suggests that the SHO positions herself in a similar way, with her lower body still aligned with the operative field and the consultant, and her head turned away momentarily at various points to engage with the scrub nurse.

If I had to summarize this in one sentence, I might say, the body is a vital resource for managing one's participation in simultaneously unfolding activities.

Reading bodies
By mapping the bodily movements of the SHO I made visible what Goffman (1971) calls a 'body gloss', i.e., a gloss to broadcast one's interactional positioning. The SHO made a *body gloss* designed to signal, first, to the scrub nurse that she is about to ask her to pass some instrument and, second, to the consultant, that she is ready to apply the scissors whenever he is ready. She can see that the consultant is tying knots, she knows that soon the thread will need to be cut, and she knows that the consultant will be expecting her to do that. She has to calculate the time it will take her to request and receive the scissors, and she has to consider the availability of the scrub nurse for receiving the request, as she may be occupied by, for instance, talking to one of the circulating nurses. We also saw the consultant displaying a body gloss designed to signal to the SHO where and when she was expected to cut: in this case, a discontinuation of knot tying and stretching of the thread.

In other words, the transcript makes visible how the participants in the interaction "gloss" with and "read" bodies, and how that enabled them to "seamlessly coordinate emerging activities" (Hindmarsh & Pilnick, 2007: 1413). Similar

DRAFT

observations have been made in studies of interaction in classrooms. For instance, Scollon & Scollon (2004) describe the actions involved in a teacher handing a paper to a student. They note that before the handing can occur two conditions have to be met: “First, the two participants in this action have to come to agreement that it is going to occur, then, secondly, they must position themselves appropriately.” (p. 64). In their example, the first condition is met by the teacher calling the student’s name and the teacher and student establishing eye contact; the second condition is met by the teacher and students positioning themselves at the right distance (e.g. the teacher walks to the student, the student reaches out his arm). The actual handing is achieved through “micro-movements that are adjusted to the weight of the object and the timing of the movements of their hands toward each other”. To some extent, these actions correspond with the actions involved in the request for and passing of the scissors. For instance, I noted the SHO’s body movements that signal ‘what is going to occur’ and enable her to establish eye contact and ultimately receive the scissors from the scrub nurse. The name calling is not required at the operating table; by default, all requests for instruments are directed to the scrub nurse. By making comparisons like these we can begin to recognize, at a micro level, how recurring actions (such as handing objects) are differently and similarly achieved across different social (institutional) contexts.

Concluding remarks

As with any multimodal transcript, the one discussed here is only a partial representation of the interaction it is based on. This partiality is the result of both spatial and temporal limitations. The temporal partiality is rather strong (although perhaps backgrounded in the transcript and its recount): I transcribed less than 15secs from an operation that lasted more than 3hrs. An example of the spatial limitation is that what was framed by the camera excludes a great number of other, simultaneously unfolding activities. For instance, at the patient’s head end an anaesthetic team was engaged in activities which had a direct effect on the work of the surgeons. It is useful to reflect on these limitations, to check that the transcript does indeed match your original questions, your interpretation and your conclusions, and, last but not least, to anticipate what effect the transcript might have on its readers. As I pointed out at the beginning of this chapter, different research traditions have different ways of transcribing multimodal interaction, so it’s important to pitch your transcript to the right audience. What may seem an entirely ‘valid’ transcript in one (academic) context may seem to be lacking in validity in others.

The multimodal transcript discussed in this chapter was made without any transcription software package, of which there are now quite a few available. When choosing transcription software it is important to still consider all the choices discussed in this chapter. A lot of them are being made for you by the software developers, and it’s worth checking that they match yours. For the purposes of the transcript discussed here ELAN, a free package developed by the Max Plank Institute of Psycholinguistics, would have been a good option. In ELAN (Wittenburg et al. 2006) you can create the type of template I made on grid paper, with a horizontal time line as a base, and ‘tiers’ below it, each of which describe a particular interactional dimension (in my case, ‘upper body’, ‘right arm’ and ‘head gaze’.) ELAN is particularly helpful if you aim to analyze small strips of interaction second-by-second or in even greater detail.

DRAFT

I like to conclude this chapter by going back to the comparison between the passing of the scissors in my multimodal transcript and the passing of the paper in the account of Scollon & Scollon. The Scollons use the notion of the 'historical body' to highlight the embodied resources that people develop through practical experience. The teacher and student, and the SHO and the scrub nurse had been involved in 'handing' certain objects in their respective settings many time before, which allowed them to perform these actions relatively effortlessly in the instances described. It is unlikely that student or SHO were ever 'taught' how to hand these objects, or that they found instructions on this in the textbooks or syllabi they studied. The notion of the 'historical body' connects with a range of theories about knowledge, learning and the body, in philosophy, anthropology, sociology and psychology (think about terms like, 'tacit knowledge', 'practical knowledge', 'procedural knowledge', 'habitus', et cetera). While these theories acknowledge the significance of the body in social interaction they are rarely accompanied by multimodal transcripts that visually represent that significance. That is an important function of multimodal transcripts: to show aspects of social interaction which often remain unarticulated in the narratives provided by researchers and the people they study. It is these 'hidden' dimensions that can provide inroads into understanding substantive issues (in my case, the safety and quality of surgical care).

Suggestions for Further Readings

Flewitt, Rosie, Mirjam Hauck, Regina Hampel and Lesley Lancaster (2009). What are multimodal data and transcription? In Carey Jewittt (ed) The Routledge Handbook of Multimodal Analysis. London: Routledge.

Heath, Christian, Jon Hindmarsh, and Paul Luff. 2010. Video in Qualitative Research: Analysing Social Interaction in Everyday Life. Los Angeles: Sage.

Norris, S. (2004). *Analyzing Multimodal Interaction*. London: RoutledgeFalmer.

Project idea

Choose a focus for a small study on multimodal interaction. Use a video on Youtube that speaks to that question and discuss how you would go about making a multimodal transcript of that video using the steps outlined in this chapter.

Acknowledgements

This chapter was written with support from MODE: Multimodal Methodologies for Researching Digital Data and Environments, a node of the National Centre for Research Methods which is funded by the Economic and Social Research Council.

References

Bezemer, J. & D. Mavers (2011). Multimodal Transcription as Academic Practice: A Social Semiotic Perspective. *International Journal of Social Research Methodology* 14, 3, 191-207. (Special Issue on Video based Social Research: Theory and Practice).

Bezemer, J., Murtagh, G., Cope., A., Kress, G. and Kneebone, R. (2011). "Scissors, Please" The Practical Accomplishment of Surgical Work in the Operating Theatre. *Symbolic Interaction* 34, 3, 398-414.

Goffman, Erving. 1971. Relations in Public: Micro Studies of Public Order. New York: Basic Books.

DRAFT

Goffman, Erving. 1981. Forms of Talk. Philadelphia: University of Pennsylvania Press.
Heath, Christian, Jon Hindmarsh, and Paul Luff. 2010. Video in Qualitative Research: Analysing Social Interaction in Everyday Life. Los Angeles: Sage.
Hindmarsh, Jon and Alison Pilnick. 2007. “Knowing Bodies at Work: Embodiment and Ephemeral Teamwork in Anaesthesia.” Organization Studies 28:1395–416.
Kendon, Adam. 1990. Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge: Cambridge University Press.
Kress, G. (2009). “What is mode?” In Jewitt, C. (ed). The Routledge Handbook of Multimodal Analysis. London: Routledge.
Koschmann, Timothy, Curtis LeBaron, Charles Goodwin, and Paul Feltovich. 2011. “‘Can You See the Cystic Artery Yet?’: A Simple Matter of Trust.” Journal of Pragmatics 43:521–41.
Mondada, Lorenza. 2003. “Working with Video: How Surgeons Produce Video Records of Their Actions.” Visual Studies 18:58–73.
Svensson, Marcus, Christian Heath, and Paul Luff. 2007. “Instrumental Action: The Timely Exchange of Implements during Surgical Operations.” Pp. 41–60 in *Proceedings of European Conference on Computer-Supported Cooperative Work 2007*, edited by L. Bannon, I. Wagner, C. Gutwin, R. Harper, and K. Schmidt. Limerick: Springer.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.

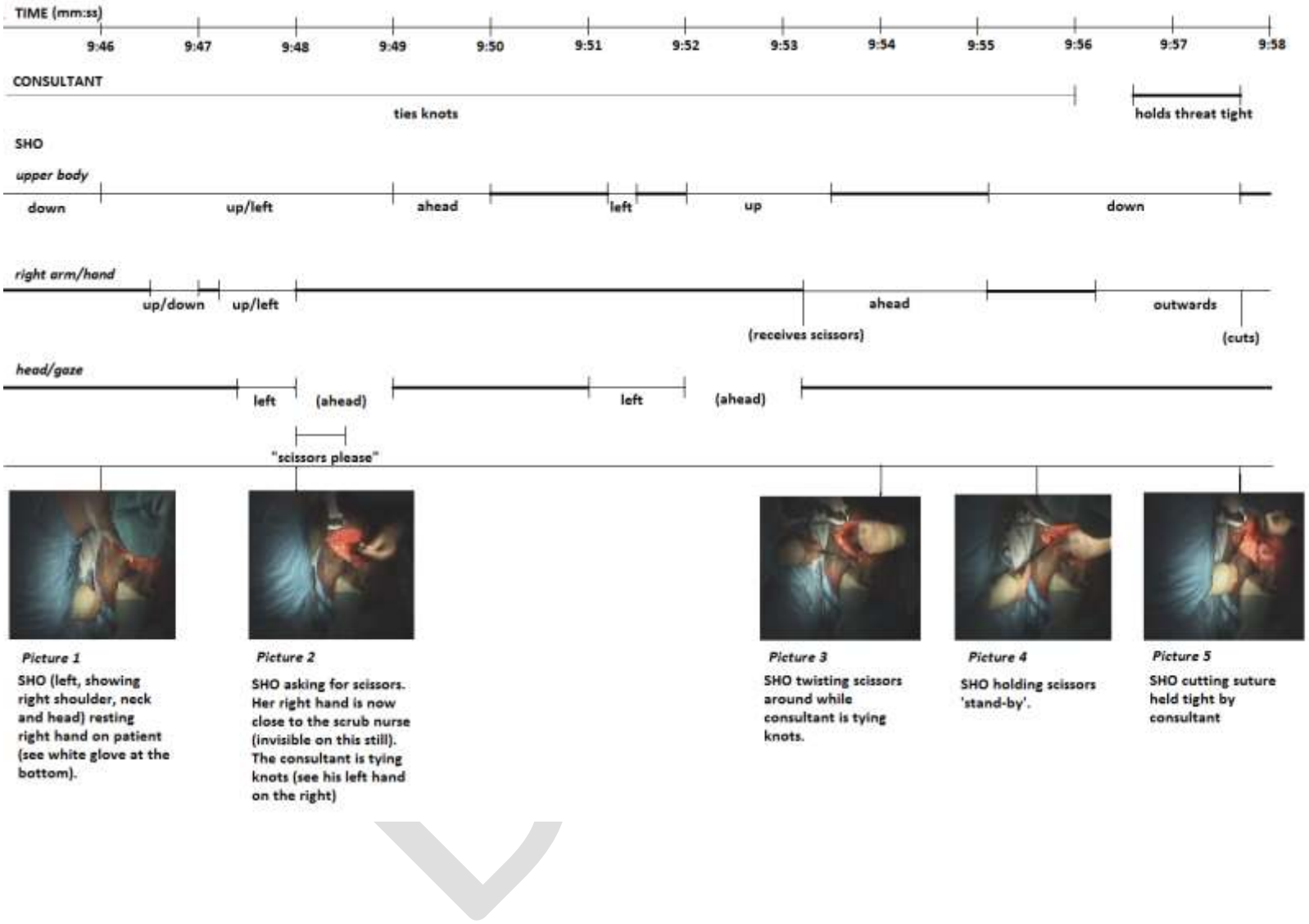
short bio note

Dr Jeff Bezemer is Senior Research Fellow at the Institute of Education, University of London and Deputy Director of MODE, a node of the National Centre for Research Methods that is focused on developing multimodal methodologies for researching digital data and environments (mode.ioe.ac.uk). He has published in the areas of workplace learning and professional communication; multimodality, literacy and pedagogy; multilingualism and education; and multimodal research methods.

contact details

London Knowledge Lab, Institute of Education, 23-29 Emerald Street, London WC1N 3QS
j.bezemer@ioe.ac.uk

DRAFT



Picture 1

SHO (left, showing right shoulder, neck and head) resting right hand on patient (see white glove at the bottom).

Picture 2

SHO asking for scissors. Her right hand is now close to the scrub nurse (invisible on this still). The consultant is tying knots (see his left hand on the right)

Picture 3

SHO twisting scissors around while consultant is tying knots.

Picture 4

SHO holding scissors 'stand-by'.

Picture 5

SHO cutting suture held tight by consultant

DRAFT