National Centre for Research Methods Working Paper

05/12

Transcribing video

Diane Mavers

Institute of Education, University of London

NCRM
National Centre for
Research Methods

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Transcribing video

MODE Working Paper 05/12
August 2012
Diane Mavers
Institute of Education, University of London

Contents

Introduction

With increasing availability and use of video recording in contemporary empirical research in the social sciences, questions around how footage is to be transcribed have become urgent. What are the various ways in which this can be done? What issues arise? What are the benefits and limitations of different approaches? This paper investigates how video materials can be remade on the page or page-like screen; how the different modes of embodied communication (e.g. speech, gaze, gesture and posture) can be re-presented as writing or image. Through examining published transcripts, it suggests some of the features that are sustained, lost and added when embodied expression and interaction are reconfigured as graphic transcripts. Overarching aims include to suggest factors for critical reflection in transcribing video materials and to open up issues for debate.

Issues in transcribing video footage

Transcribing video materials crosses disciplinary boundaries; it is undertaken in Cultural Studies, Film and Media, Sociology, Communication Studies, Psychology, and so on. Framed by the research focus, how video footage is transcribed may entail attendance to various modes and hence a multimodal perspective. Multimodal transcription is not methodologically exclusive and might be adopted in, for example, ethnography or action research, and may be helpful in discourse analysis or as a complement to conversation analysis. As a relatively new approach that is being taken up across different disciplines, theoretical frames and methodologies, whilst descriptive and analytical vocabulary associated with multimodality is becoming increasingly prevalent, different approaches are resulting in diversity of use. Understandings of commonly shared terms are not entirely settled (e.g. 'mode', 'multimodality'), some notions are subject to contestation (e.g. 'affordance'), and fresh terminology continues to emerge. Some concepts are closely aligned to the social semiotic theory from which they derived (e.g. 'sign', 'arbitrariness', 'metafunctions'), whilst others, even if shaped by this framework, can be readily integrated into other theoretical approaches (e.g. 'design', 'ensembles', 'resource'). For definitions of terms relating to multimodality, please see the MODE glossary at: http://multimodalityglossary.wordpress.com/

Transcribing the multimodality of video footage is predicated on prior methodological resolution with regard to what is needed in order to respond to the research question. Take, for example, investigating how an initiative was implemented. If background facts are sought as a basis for mapping the area, the lexis of what was said in interviewing may be sufficient. Adopting a multimodal approach presupposes that 'modes' beyond speech are worthy of analysis and relevant for interpretation.

> 'In contrast to words, nonverbal signs have often been excluded from study on the grounds that they are problematic for data collection and analysis, ancillary to learning through spoken or written modes and are idiosyncratic or arbitrary, characterized by personal and cultural variations with limited functional potential that render them unsuitable for systematic forms of analysis' (Flewitt, 2006: 27).

If the study seeks to identify people's views, the emphases and hesitations of what they say could be enlightening, and perhaps other embodied expressions such as facial expression and gesture. In observing what actually went on, access to what people did as well as what they said may be pivotal. Decisions reached in advance of empirical work as well as those made in the moment shape what is gathered as video footage, for example how many camcorders are used, when to begin and end recording, who is chosen for inclusion, closeness and angle of shot, and the quantity of materials. An advantage of video as against note-taking is that episodes can be multiply revisited in refining a transcript – indeed, the significance of what is said or done may not become clear until after the moment. Transcription is complex. The process of repeated viewings of extracts in real time, at speed and frame-by-frame, with and without sound, can be time consuming and challenging, and this should be factored into the research design.

With fast-moving technological developments, multimodal data can be presented and disseminated in ways not possible even ten years ago. It is possible to create digital resources where the 'raw' data is hyperlinked to transcripts and analyses (Andrews et al, 2012). There are well-established conventions for transcribing speech, including the fine-grained detail of colons to suggest lengthening, underlining to indicate emphasis, brackets to mark simultaneity and dashes to signify cut-offs (e.g. Sacks et al, 1974; Jefferson, 1984). There is no such standardization in multimodal transcription. With Even so, print currently remains prevalent for publication. Building on earlier work (e.g. Ochs, 1979; Erikson, 1986), contemporary researchers continue to experiment with transcribing video data in a range of ways: in writing and various forms of image, as well as diverse layout. This trend not to be restrictive, even within individual pieces (e.g. Goodwin, 2000; Mavers, 2011) can be seen as a strength in that it enables transcribers to select the method most apt to the particular need.

How can the multimodality of expression and interaction be transcribed on the page or the page-like screen? Remaking activity and exchange is not at all straightforward because what is represented bodily may no longer be available graphically. As resources may not be shared between modes, what was originally communicated must be reconfigured. The lexis of speech can be re-presented as writing, but sounds produced orally, or the motion of gesture created actionally through space and time, must somehow be suggested as marks on the fixity of the page or the page-like screen. The process of remaking across modes has been variously named – 'transduction' (Kress, 1997), the 'transmodal moment' (Newfield, 2009) and 'transmodal redesign' (Mavers, 2011) – each with particular theoretical nuances. The question is how to retain constancy of meaning between the video footage and the transcript, and indeed whether this can be done.

Modes of transcription

My first experience of video was recording six-year-olds learning how to make a stop-frame animation and then teaching their peers (Mavers, 2011). Keen to capture the richness of the activities, I began by transcribing all that was said over approximately two hours of footage, amounting to almost 18,000 words. In the extract below, the language of what was said and who said it are documented, as well as indications of speech-like lexis in ''cos' and 'one's', certain articulations in 'oooh' and 'laughs', and one instance of pausing in '(.)'.

Ebony: oooh no (.) just leave that one there 'cos that one's
Zak: that one's [inaudible]
Ebony: and that one
Farida: shoot
Olivia: action
Farida and Ebony: shoot
Olivia: action (laughs)
Farida: shoot
Olivia and Zak: action
Farida and Ebony: shoot
Zak: action
Olivia: action

Page after page of the children repeating 'action' (getting ready for the shot) and 'shoot' (capturing the image) suggests a humdrum activity, and perhaps an associated assumption that there was little going on. Actually, much of the interactional exchange was constituted in modes beyond speech: in actions on and gesturing around the small-world figures, gaze, facial expression, bodily orientation, and so on.

I subsequently developed ‘thick descriptions’ (Geertz, 1973) in a narrative vignette style in order to provide fuller information about what went on.

> ‘‘Now,’ says Ebony. ‘What was next? We need the paper.’ She disappears to find the storyboard. On her return, Ebony rights and repositions the monkey that Ambareen and Jeselle have knocked over. She then prises the two pirate figures from Jeselle, protesting, ‘Oh don’t put it there, no, no, no, no, no, no, no.’ Ebony looks intently at the storyboard, then, grasping the two pirates in her right hand, rehearses their movement towards the monkey as she reads aloud, ‘They find the monkey.’ Dangling the figures between her fingers, she continues to read the next caption: ‘The pirates are cross.’ Jeselle points to this frame on the storyboard and then to the succeeding one, which Ebony attends to with concentration, and they engage in some discussion. Turning back to the ‘stage’ as she examines the storyboard, Ebony concludes ‘Okay, we’ve done that bit’, and continues to study the page intently, presumably with a view to what comes next’ (Mavers, 2011: 119).

This written account is not restricted to speech. What I find interesting in retrospect is my choice of words. These include presence and absence (e.g. ‘disappears’, ‘return’), actions on objects (e.g. ‘grasping’, ‘rehearses’, ‘dangling’), speech (e.g. ‘protesting’, ‘continues to read’), gaze (e.g. ‘looks intently’, ‘attends to with concentration’) and gesture (e.g. ‘points’). Some verbs are fairly ‘neutral’ (e.g. ‘rights’, ‘knocked over’, ‘repositions’). Other vocabulary is more loaded. For example, I wrote that Ebony ‘prises’ – not ‘takes’ or ‘snatches’ or ‘yanks’ – the two pirate figures from Jeselle. All vocabulary is a selection from a reservoir of possibilities. It is not that this choice is good or mistaken, but that it is a redesign of Ebony’s action. In framing how the interaction is understood, it is already an interpretation. Furthermore, the linearity of writing means that what was done and said concurrently, and viewed and listened to simultaneously in the video extract, is ordered sequentially. In responding to my research question, Ebony became the main focus, and certain things done by the other children were bypassed. What is important is that this ‘vignette’ is not the original interaction but a redesign of it.

> ‘Even the most richly detailed vignette is a reduced account, clearer than life. Some features are selected in from the tremendous complexity of the original event […] and other features are selected out of the narrative report. Thus the vignette does not represent the original *event itself*, for this is impossible. The vignette is an abstraction; an analytic caricature (of a friendly sort) in which some details are sketched in and others left out; some features are sharpened and heightened in their portrayal […] and other features are softened or left to merge with the background’ (Erickson, 1986:150).

The term ‘multimodal transcription’ begs the question as to whether it is the materials being transcribed or the transcript to which the description ‘multimodal’ belongs. (Indeed, deriving from the Latin *trans* (‘over’) *scribere* (‘write’), the etymology of the term is writing.) Embodied expression and interaction are always multimodal. A transcript is multimodal when it contains more than one mode. Including image as well as writing as a means of transcription forces the transcriber to decide which meanings will reside where and how these modes relate. Image varies. A series of photographic stills provides certain information at a glance, such as features of the setting, objects and what people look like, which may or may not be included in a written version. When moving image is re-presented as still image, the dynamism of action is lost. It might be presumed that the 'reader' will imagine movement, or this might be implied in consecutive and cumulative transcripts frames. Tracing photographic stills offers choice in what to include or exclude, and with what detail and emphasis. For example, social relations in the body positions, gaze and gesture of four girls eating together are re-presented in a tracing, whilst precisely what they were eating or wearing is indistinct, as this is not relevant to the research focus (Goodwin, 2009: 55). Other features are realized in writing, such as specification of moment-by-moment gestures (e.g. 'raises hands to head', 'pointing to', 'slaps hand'), gaze (e.g. 'closes eyes', 'lowers head with shut eyes') and actions (e.g. 'turn away', 'walk away'), which are integrated in brackets alongside lexis (e.g. 'disgusting') and articulations (e.g. 'OU:::') (ibid). Through this combination of 'linguistic, intonational and corporeal resources' (Goodwin, 2009: 43), Goodwin, with an interest in social exclusion, argues that an African-American, working class girl is marginalized and treated with contempt by upper middle class peers. Drawing rather than tracing offers the option of conflating what happened over time. Different moments might be represented in order to convey key aspects of what went on.

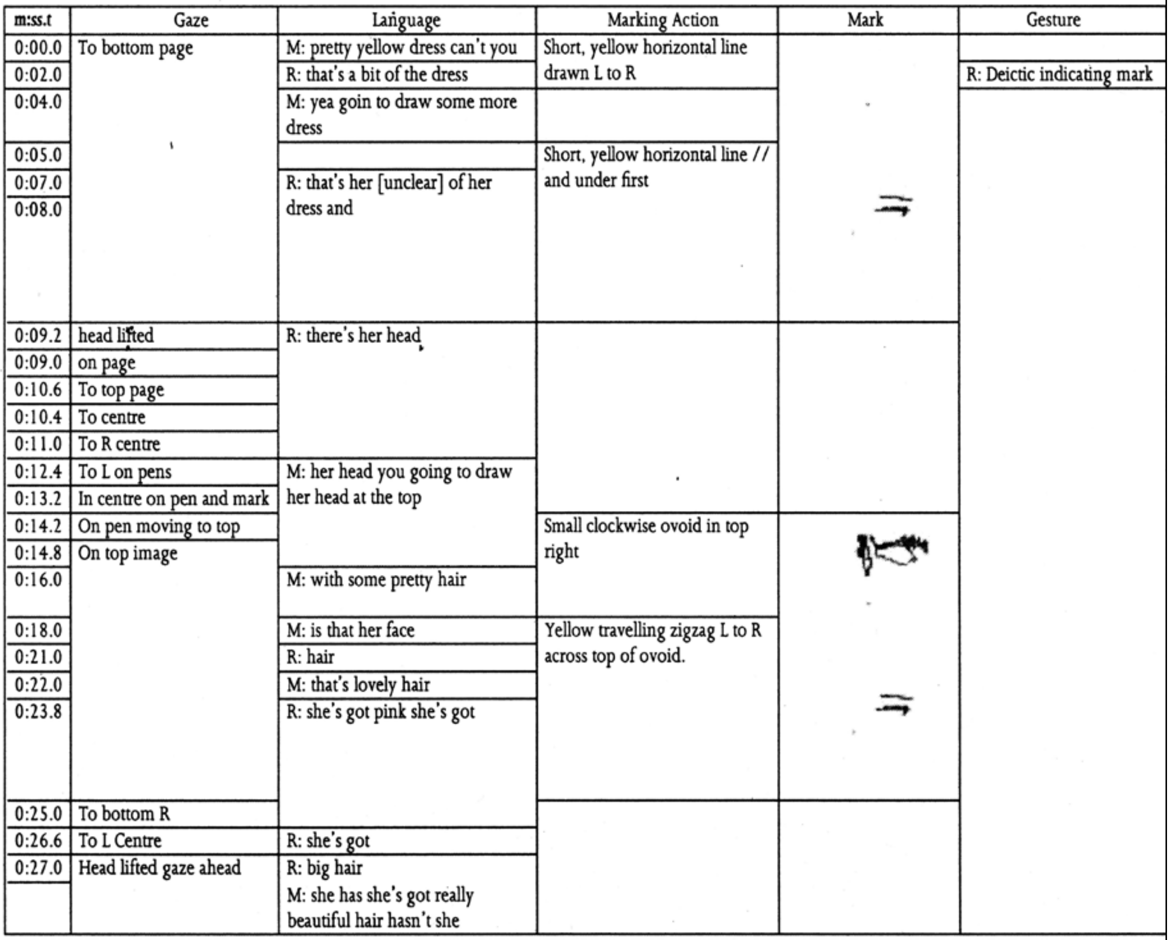
How can speech, articulation, gaze, gesture, action and body position be re-presented as writing or image? With the aim of investigating a variety of ways in which authors have transcribed video materials on the page, four published transcripts are described and analysed below. Using a template approach in order to attend to the same issues across the different transcripts, the examination is divided into five main sections, each split into sub-divisions: the publication (bibliographic details, a concise introduction to the author, the paper's focus and the data gathered in the research); the transcript(the actual transcript, its subject matter, its location in the publication and its relation to other transcripts in the paper); the multimodal configuration of the transcript (a brief description of the modes of transcription and examination of each mode of transcription); how the transcript is used (in relation to the body text and the associated argument); and finally, authors were invited to correct, suggest modifications and comment on the description/analysis and to provide further detail such as reflections on the processes of creating the transcript and more general views on transcribing video materials. Each author gave consent to use the transcript, and permission was also granted by the journal publishers.

Example 1: Writing with cumulative scans of the graphic product (Lancaster)

The publication	
Publication details	Lancaster, L. (2007) 'Representing the ways of the world: how children under three start to use syntax in graphic signs', *Journal of Early Childhood Literacy, 7*(2): 123-154. DOI: 10.1177/1468798407079284. This paper can be accessed online at: http://ecl.sagepub.com/content/7/2/123
The author	Lesley Lancaster has a background in Applied Linguistics and Multimodal Communication. Her particular research interest is the sign making of children below the age of three years
Focus of the publication	This paper investigates the principles brought by the under-threes to their graphic text making, and challenges the view that their mark making lacks intentionality and is devoid of 'representational significance' (Lancaster, 2007:

	123-127). Lancaster explores how the same and different inscriptions are used for particular purposes and examines the organizational relationships or 'grammaticization' of the marks young children make. She argues that the current policy emphasis on phonics in early literacy frames writing as the transcription of speech rather than an 'independent graphic mode in its own right' (Lancaster, 2007: 142), and that this assumption fails to build on what preschoolers already know and can do before they enter formal education (Lancaster, 2007: 149).
Methods of data collection	Materials gathered across the paper comprise video footage of children, with a parent alongside, making graphic texts, and children's graphic productions.

The transcript



m:ss.t	Gaze	Language	Marking Action	Mark	Gesture
0:00.0	To bottom page	M: pretty yellow dress can't you	Short, yellow horizontal line drawn L to R		
0:02.0		R: that's a bit of the dress			R: Deictic indicating mark
0:04.0		M: yea goin to draw some more dress			
0:05.0			Short, yellow horizontal line // and under first		
0:07.0		R: that's her [unclear] of her dress and			
0:08.0					
0:09.2	head lifted	R: there's her head			
0:09.0	on page				
0:10.6	To top page				
0:10.4	To centre				
0:11.0	To R centre				
0:12.4	To L on pens	M: her head you going to draw her head at the top			
0:13.2	In centre on pen and mark				
0:14.2	On pen moving to top		Small clockwise ovoid in top right		
0:14.8	On top image				
0:16.0		M: with some pretty hair			
0:18.0		M: is that her face	Yellow travelling zigzag L to R across top of ovoid.		
0:21.0		R: hair			
0:22.0		M: that's lovely hair			
0:23.8		R: she's got pink she's got			
0:25.0	To bottom R				
0:26.6	To L Centre	R: she's got			
0:27.0	Head lifted gaze ahead	R: big hair M: she has she's got really beautiful hair hasn't she			

Transcript 1 – Belle (the first of three pages of the transcript, page150)
Re-printed from Lancaster, L. (2007) 'Representing the ways of the world: how children under three start to use syntax in graphic signs', *Journal of Early Childhood Literacy, 7*(2): 123-154, with permission from SAGE Publications and the consent of the author.

Subject matter	Ruby (age 32 months) is making a representation of the Disney character, Belle, with her mother alongside.
Location	The transcript appears as an appendix (Lancaster, 2007: 150-152).
Relation to other transcripts	This is the only transcript in the paper. Successive inscriptions made by Ruby, as well as descriptions of what went on, are integrated into analyses in the body text.

Multimodal configuration of the transcript

Description	This is a written transcript that documents gaze, 'language', marking action and gesture, along with a column that includes successive and cumulative scans of what Ruby inscribed. Other modes such as facial expression and

	posture are excluded.
Writing (alphabetical)	*Gaze:* Written documentation of gaze consists of four components: where Ruby looked on the page (e.g. 'bottom', 'top', 'centre', 'R' and 'L'); what she looked at (the available pens and the pen she was using, and inscriptions she had just and previously made); looking away from the page ('gaze ahead'); and adjustment of gaze (e.g. 'moving to'). This specifies in a way image may not. *Language:* Under the heading 'language', Lancaster implies speech through spelling (e.g. 'yeah', 'goin') and elision ('can't', 'that's', 'there's', 'she's', 'hasn't'). There is no attempt to represent the sounds and rhythms of orality. That speech is not sentenced is implicit in her exclusion of capital letters and punctuation marks (there are no full stops or question marks). Whilst the remainder of the transcript provides details about what Ruby looked at and did, this column also includes what was said by her mother. *Marking action:* Writing provides information about sequentiality and directionality in the process of production (e.g. 'L to R', 'clockwise', 'travelling zigzag'), as well as describing the product in terms of mark (e.g. 'short, yellow horizontal line', 'ovoid', 'zigzag',), spatial orientation ('horizontal') and position (e.g. 'under', 'top right', 'across top'). Hence, writing provides actional information and specifies visual detail. *Gesture:* Lancaster writes that gesture was transcribed when it arose (Lancaster, 2007: 128). The single entry in this first page of the transcript classifies the gesture ('deitic') and what was pointed out ('mark').
Writing (numerical)	Numerical information provides the timing of each modal instantiation. The extract examined here represents 27 seconds. The duration of the full transcript is given as 1 minute 13 seconds. It is partitioned temporally into 23 parts, at an average of 1.17 seconds per division.
Scanned inscriptions	Still shots of Ruby's text-in-the-making, located in the column entitled 'mark', shows what she inscribed. This visuality provides the 'reader' with the actual lines and shapes of her representation, which would be difficult in written description alone.
Layout	This transcript is set out as a table. The five modes selected for inclusion – 'gaze', 'language', 'marking action', 'mark' and 'gesture' – are represented in vertical columns. This layout allows the reader to follow any one mode in sequence, and thereby to track, for example, shifts in gaze or the sequential and cumulative marks of graphic representation. For example, the reader can follow how Ruby and her mother described representations, who initiated ideas and how they were or were not taken up in inscription. Multimodal simultaneity is established horizontally. Where Ruby was looking and what she said as she made marks can be tracked. Unevenness of temporal divisions is a consequence of modal expressions moving beyond one-second boundaries, indicating the complexity of multimodal configurations.
How the transcript is used	
Relation to the body text	The episode being investigated is introduced in the main body text with a brief description of where the episode took place, who was involved and which inscriptional resources were available, followed by a summary of what the finished product consisted of (Lancaster, 2007: 131-132). This leads to a detailed account of the marks Ruby made, how she went about making them, her associated gaze, gesture and speech, what her mother said and how she responded (Lancaster, 2007: 132-133). What Ruby inscribed is reproduced in a

	series of five figures within the body text which progressively show each of the five 'signs' of the graphic text in the sequence in which they were produced. This is followed by analysis and discussion of approximately 1,000 words. As an appendix, the transcript contextualizes features that are picked out for attention in the body text, and also provides fuller detail.
Argument	Lancaster contends that Ruby drew the cartoon character pictographically rather than pictorially, and that the parts of the representation (clothing, head, hair, arms, legs, mouth, eyes) are positioned in ways that are 'relationally appropriate' if 'spatially separate', thereby resembling 'appropriate relative location rather than physical connection' (Lancaster, 2007: 133-134). She goes on to argue that, even though Ruby did not draw or write in a conventional way, her mark making was principled and her representational strategies entailed features typical of both writing and drawing.

Author's response

Traditionally, children's textual intentions have been extrapolated from their material productions after the event, sometimes with the addition of evidence from language, asking them for clarification about aspects of meaning. In this respect, interpretation of children's texts is largely an appraisal of outcomes and reports, separated from their process of production, rather than an evaluation of a sequence of activities taking place in real time. This creates particular problems when the children are too young to be producing conventionally understood marks and signs, and are still developing as language users. Multimodal transcription of video footage was used for this project with the purpose of opening up the communicative and semiotic processes in which the children were engaging in real time, and associating these with the evolution of their graphic signs as they were being produced, providing detailed evidence to support interpretations of their meaning. A second-by-second, micro-analytic time frame was used in order to open up the process and examine it. A linear, tabulated format was chosen, as the researchers wanted a 'working' transcription that was easy to use. Gaze was used as an 'anchor' mode as shift of gaze tended to coincide with transferral of attention from one segment of activity within an episode to the next, and starting the transcription of episodes by recording these shifts made the transcribing process easier by providing a sequence of frames to work with within each episode. The decision was made not to use any of the available CAQDAS packages, as the actual process of transcribing proved an excellent means of starting to generate soundly based interpretations of the children's semiotic activity.

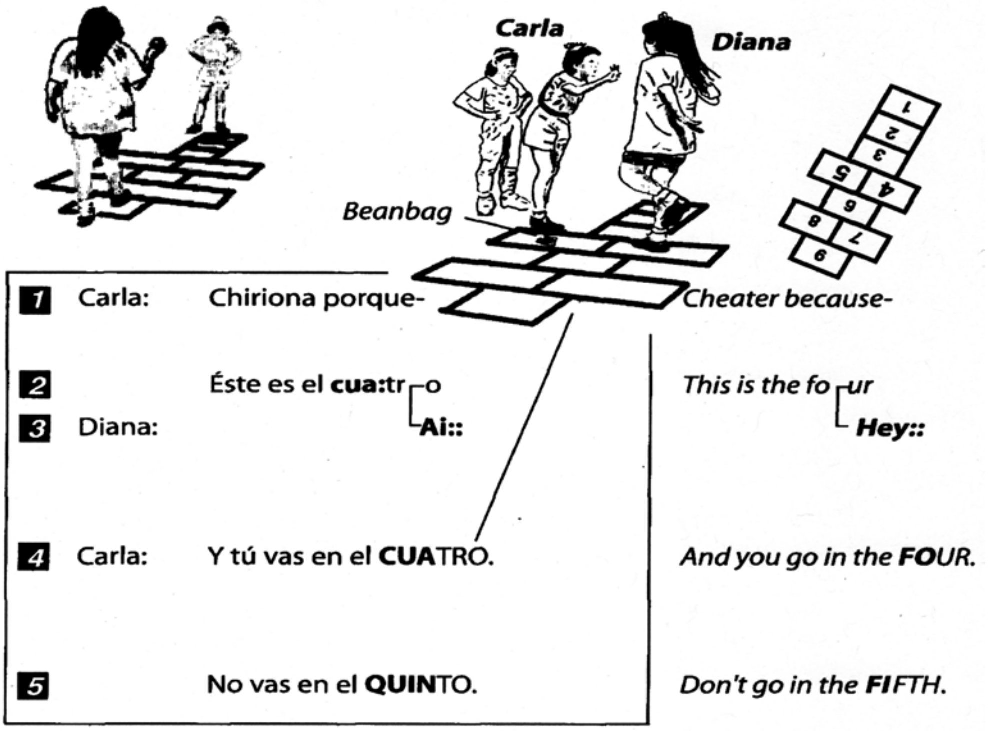
However, the transcription process was very time consuming, and this placed limitations on what could be done during a time-limited study. The selection of modes that should be transcribed gave rise to much discussion, informed by the need to balance the representation of all bodily modes that contributed to the generation of signs, including those involved in the interpersonal communication that was a central part of that process, against these practical constraints. Gaze was a very important mode, since tracking its direction indicated focus of intention, including the constant movement between graphic activity and interpersonal communication. Facial expression, where significant, was noted in this column in the interests of economy. Posture, on the other hand, was fairly invariable, since on the whole the children chose to be seated whilst they were marking, with the adult sitting nearby, and so was not included within the transcription – though that is not to say that it might not have provided productive insights if it had. Constraints were certainly placed on the extent to which the differential functions of the modes, particularly in the case of language, could be separately transcribed – prosodic and syntactic features, for example. This might well have provided a useful layer of evidence, but the time needed to do this would have been considerable.

The decision to concentrate primarily on the children during the filming process was dictated largely by the objectives of the project, though practical and financial considerations also played a part. The degree to which the interactions between adults and children were an integral part of not just the sign making process, but of the signs themselves was one of the most important findings of the project. In terms of future research in this area, the use of several cameras, and the development and extension of techniques of multimodal transcription that can both extend the description of the differential functions of modes, and can be used to map the distribution of interactive networks involving participants, and tools, objects, and physical space will need to be prioritised.

Example 2: Drawing and writing (Goodwin)

The publication	
Publication details	Goodwin, C. (2000) 'Action and embodiment within situated human interaction', *Journal of Pragmatics, 32*(10): 1489-1522. This paper can be accessed online at: http://dx.doi.org/10.1016/S0378-2166(99)00096-X
The author	Charles Goodwin works in the area of Linguistic Anthropology. His research interests include talk and other embodied means of communication as interactively organized social practices.
Focus of the publication	This paper investigates the simultaneous deployment of different semiotic resources in everyday communication. Challenging a common assumption underlying analysis of social interaction – that language is 'both primary and autonomous' whilst everything else is 'context' – Goodwin investigates 'the public visibility of the body as a dynamically unfolding, interactively organized locus for the production and display of meaning and action' (Goodwin, 2000: 1490).
Methods of data collection	Video recording is the method used for recording three young girls playing hopscotch.

The transcript



1	Carla:	Chiriona porque-	*Cheater because-*
2		Éste es el **cua:tr**[o	*This is the fo*[*ur*
3	Diana:	[**Ai::**	[***Hey::***
4	Carla:	Y tú vas en el **CUATRO**.	*And you go in the **FOUR**.*
5		No vas en el **QUINTO**.	*Don't go in the **FIFTH**.*

The third of six transcripts relating to the analysis (page1497)

Re-printed from Goodwin, C. (2000) 'Action and embodiment within situated human interaction',

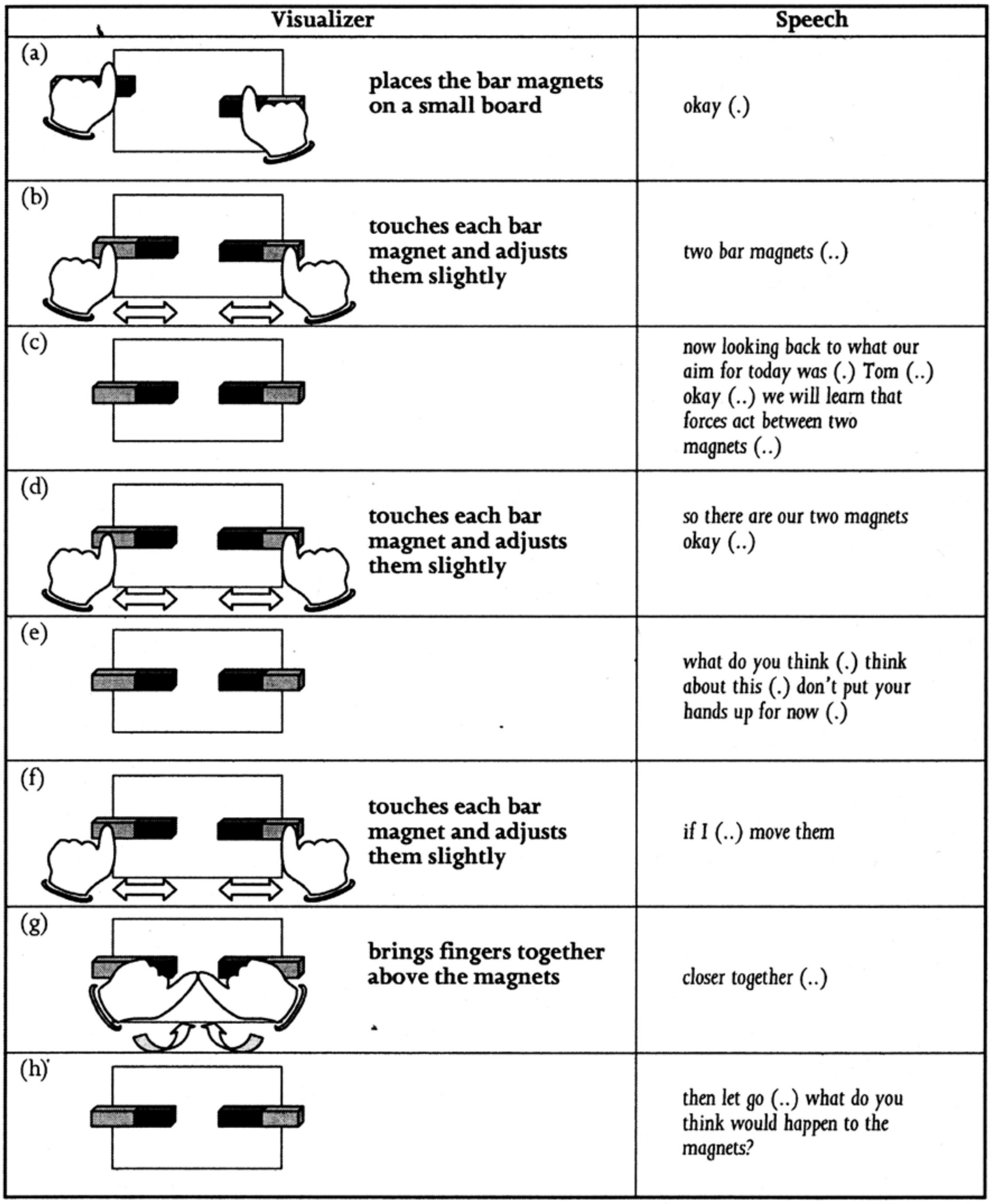
Journal of Pragmatics, 32(10): 1489-1522, with permission from ELSEVIER Publications and the consent of the author.	
Subject matter	The transcript is a re-presentation of an altercation between two girls during a game of hopscotch.
Location	The transcript is located in the analysis of the paper's body text.
Relation to other transcripts	Six interrelated transcripts are interspersed across the analysis of this social interaction. In a layered argument, Goodwin variously repeats, excludes, modifies and extends features of drawing, writing and symbols as he builds his argument. For example, the drawing of the altercation appears in three transcripts and is also extracted for separate attention. The transcript selected for examination here is the third in this section of the paper.
Multimodal configuration of the transcript	
Description	• At the top, a drawing of the altercation includes three girls and a hopscotch grid, with the two focal participants and the beanbag labelled. • Underneath, to the left and inside a frame is a transcript of what was said and by whom, with each chunk numbered, along with drawings of a footprint linked to a grid and three hands, one with a circling arrow. • To the right of the above is a translation into English, with the drawings of the footprint and hands repeated.
Drawing from video still	*Activity:* For those familiar with hopscotch, at a glance the drawing specifies the game through the grid and the hopping action of the nearest child (Diana). Omission of the details of the setting excludes clues as to where the game took place (i.e. the street, the park). *Participants:* The written account in the body text confirms that the participants are young girls. Their clothing implies the informality of play in a not-too-cold climate, features that are otherwise absent from the paper. An advantage of drawing and tracing is that anonymity can be readily preserved; they remove ethical concerns regarding identification of the children, which would be more difficult in a screen shot. *Embodied modes:* The posture of the girl to the left implies that she is a bystander. That she is drawn implies that she was in some sense involved in the game (other people present, if any, are excluded). Carla's represented orientation, gaze and gesture towards Diana suggest interaction, the detail of which is elucidated in the written transcript below. Diana's bent knee, backwardly angled arm and flying hair imply action. Her twisting posture becomes pivotal in the analysis.
Diagrammatic image	The footprint indicates foot position. A shift from four fingers to five digits suggests that a numerical point is being made.
Symbols	*Gesture:* The bold semi-circular arrow located at the wrist represents twisting of the hand. *Simultaneity:* Square brackets denote synchronous speech and vocalized exclamation. *Labelling lines:* Lines connect labels with drawn items.
Writing (alphabetical)	*Labelling:* Writing labels two of the participants. That the child to the left is not named diminishes her relevance to the analysis, whilst specifying the beanbag marks representation of an object that might otherwise pass by unnoticed. The speakers are clearly labelled. *Speech:* The lexis of what was said is re-presented in writing. Emboldening and

	capitalization indicate emphasis, with intonational pitch contours specified in a previous transcript. Following the conventions of Conversation Analysis, double colons represent pausing.
Writing (numerical)	As in Conversational Analysis, what was said is chunked and numbered. This assigns prominence to speech as the 'base' mode consistently transcribed throughout, whereas selected action, position and gesture are integrated when they contribute to the argument. Goodwin subsequently commented that, crucially, citing line numbers enabled him to refer the 'reader' to 'specific places in a large, complex transcript' (email communication in response to the description/analysis).
Layout	In the transcript as a whole, each part links with and is dependent on the other. Almost titular, the labelled drawing at the top provides an overarching frame for the detail of what comes below. The positioning of the footprint and the hands indicate where action and gestures were made in relation to what was said and by whom.
How the transcript is used	
Relation to the body text	The transcript is integrated into an analysis of around 3,000 words. This section begins by describing the rules of hopscotch and the cause of the altercation between Carla and Diana. The initial part of the analysis focuses on Carla's speech: her choice of lexis, how she substitutes words whilst retaining a common syntactic structure in order to introduce contrast, her vocal emphasis (with a chart showing intonation patterns), her invocation of the rules of the game through use of a pronoun, how she constructs mutual orientation to the topic and how she promotes a frame for modes beyond speech. Goodwin then builds on this examination of speech by introducing other embodied modes of communication.
Argument	In relation to the transcript being examined here, Goodwin argues that Carla's hand gesture is not redundant in replicating what is done in speech, but that by positioning her body in front of Diana and 'thrust(ing) her gestures towards Diana's face', she uses her body to secure the attention of the deemed 'transgressor' and hence to structure the encounter as a direct challenge (Goodwin, 2000: 1499). Goodwin argues that 'talk and gesture mutually elaborate each other' in this encounter (Goodwin, 2000: 1499).

Example 3: Computer-generated image and writing (Mavers)

The publication	
Publication details	Mavers, D. (2009) 'Student text-making as semiotic work', *Journal of Early Childhood Literacy, 9*(2): 145-155. DOI: 10.1177/1468798409105584 This paper can be accessed online at: http://ecl.sagepub.co./content/9/2/141
The author	Diane Mavers is interested in the variety of ways in which children interpret and produce meaning, in particular how what they draw and write, and say and do, relate to pedagogic interactions around curricular entities and classroom materials in primary and early years education.
Focus of the publication	Even when graphic texts are fleetingly here and gone, children invest 'semiotic work' in their drawing and writing. This paper is concerned with the relationship between how children interpret their teacher's instructions and how they respond in drawing and writing. It argues that features of what the children drew and wrote did not derive only from what the teacher said, but also, and in some instances primarily, from what they saw of what he did.

	Redesign from the embodiment of the instructions to graphic representation entailed processes of remaking across modes.
Methods of data collection	The materials gathered include: video recording in the classroom (what the teacher did and said in framing the task, how the children went about text making and subsequent whole-class interactions), photographs of the graphic products (the dry-wipe whiteboard texts made by the children), interviews with the teacher (with a focus on planning and issues in assessment), group interviews with all of the children in the class (including information regarding their prior experience, as well as their reflections on what they did and why), and lesson planning.
The transcript	



Visualizer	Speech
(a) places the bar magnets on a small board	okay (.)
(b) touches each bar magnet and adjusts them slightly	two bar magnets (..)
(c)	now looking back to what our aim for today was (.) Tom (..) okay (..) we will learn that forces act between two magnets (..)
(d) touches each bar magnet and adjusts them slightly	so there are our two magnets okay (..)
(e)	what do you think (.) think about this (.) don't put your hands up for now (.)
(f) touches each bar magnet and adjusts them slightly	if I (..) move them
(g) brings fingers together above the magnets	closer together (..)
(h)	then let go (..) what do you think would happen to the magnets?

Figure 7 Multimodal transcript of the teacher's framing of the task (page146)

Re-printed from Mavers, D. (2009) 'Student text-making as semiotic work' *Journal of Early*

Childhood Literacy, 9(2): 145-155, with permission from SAGE Publications and the consent of the author.	
Subject matter	This transcript is a re-presentation of a teacher's framing of a classroom task (hypothesizing what will happen in a scientific investigation into magnetic attraction and repulsion).
Location	The transcript is located in the analysis of the paper's body text.
Relation to other transcripts	This is the only transcript in the paper. With a focus on the relationship between teaching and what children inscribe in response, it informs the analysis of the children's subsequent hypotheses.
Multimodal configuration of the transcript	
Description	This transcript consists of two columns and eight numbered rows: • To the left of the first column, following numerical ordering, is an image of what was done on the visualiser and seen on the class screen, followed by a written description on the right. • The second column documents the teacher's speech.
Image	The images in this transcript represent what the children saw on the class screen: objects (bar magnets on a dry-wipe whiteboard), and actions on and gestures around them (hands). They are not an exact replication of what the class saw, in the sense that the particularities of the teacher's hands are lost and relative size is reconfigured. Three-dimensionality re-presents the visuality of the magnets and shading shows that they were bi-coloured, although the blue and red of the actual objects are remade as shades of grey in the journal publication. Each image represents a moment in time.
Symbols	The double-headed arrows suggest side-to-side movement and the single-headed, curved three-dimensional arrow, in conjunction with the shifted position and orientation of the hands, indicates an inward gesture.
Writing	The single words heading each column – 'Visualiser' (a digital display technology) and 'Speech' – name what was projected by the display equipment and what the teacher said. Letters of the alphabet list eight moments in the teacher's framing of the task. No timings are provided. Written text describes what the teacher did on the visualiser. Emboldening distinguishes these descriptions from the plain typography used for speech. Exclusion of punctuation suggests that action and gesture are not sentenced and creates continuousness with successive rows. The opening phrase of the transcript specifies what the image alone does not, namely that the two cubes represent bar magnets, that the rectangle represents a dry-wipe whiteboard and that the former were placed on the latter. The second statement, which appears three times, supplies motion absent from the image in the verbs 'touches' and 'adjusts'. The description 'brings fingers together' intensifies the image with its curved arrows as the key moment of the transcript. With the aim of suggesting speech, the transcription of what the teacher said is in lower case (bar the child's name) and, apart from the final question mark used to indicate a query, excludes punctuation marks. Single and double dots inside brackets signal briefer and more extended pausing.
Layout	The transcript is tabular. Two vertical columns represent what the class saw and what was said; horizontality shows what was uttered simultaneously with what was done. This allows the transcript 'reader' to follow the actional or the spoken, or to track the relationships between them. In the first column, two sets of identical images and writing appearing three times (cells b, d, and f are

<table>
<tr><td></td><td>repeated, as are cells c, e, and h) draw attention to patterning. The body text makes explicit these episodes of visual likeness.</td></tr>
<tr><td colspan="2"><b>How the transcript is used</b></td></tr>
<tr><td><b>Relation to the body text</b></td><td>The body text (approximately 750 words in this section) provides a commentary on the immediately succeeding transcript by describing the teacher’s instructions and picking out significant features for the analysis. Certain written details are repeated, if reworded, such as touching and slightly adjusting the bar magnets (Mavers, 2009: 145), some features are made explicit (e.g. ‘adjacent positioning and horizontal orientation of the two bar magnets’), some are omitted (e.g. specification of the dry-wipe whiteboard shown as a rectangle in the transcript) and other facts are added, namely the explanation that the demonstration stipulated experimental conditions and procedures (Mavers, 2009: 145). The transcript is related to photographic shots of six dry-wipe whiteboard texts (Mavers, 2009: 148) and what the children said in the interviews (e.g. Mavers, 2009: 149), which underpin the remainder of the analysis. The implicitness of the shading, positioning and orientation of the magnets represented in the transcript becomes pivotal in subsequent analysis of the children’s drawings.</td></tr>
<tr><td><b>Argument</b></td><td>The analysis examines the relationship between what the children saw on the screen display projected from a visualiser and what they heard in the teacher’s verbal framing of task instructions, and the texts they subsequently made. This entailed tracing their distribution of meaning across drawing and writing, and the interrelationships between them, in relation to prior pedagogic interaction. The children represented experimental conditions, method, prediction and theorization. Their hypotheses were both a response to the habituated practices of the classroom and to the requirements and representational practices of the curricular subject science.</td></tr>
<tr><td colspan="2"><b>Author’s response</b></td></tr>
<tr><td colspan="2">Transcribing what went on demanded numerous viewings of the video clip, including tracking movements frame by frame. Created electronically using shapes, symbols and wordart, I developed the images due to the unsatisfactory quality of the video stills and because my attempts at drawing were laughable. My division of the transcript logs eight key moments. As multimodal interaction is so intricate, attending to the ‘plenitude’ (Mavers, 2011) of what is communicated may not be relevant, never mind do-able. Certain modes are excluded (e.g. gaze, facial expression) because these were employed to monitor and gain attention rather than to contribute to the explanation of the task. Placing the images to the extreme left (i.e. ‘first’ in the left to right reading directionality of English) foregrounds the visual, and hence mimics the visual dominance of the large screen at the front of the class, whilst the teacher’s speech came from the children’s right as a kind of voice-over. The fleeting becomes frozen as a semi-permanent inscribed record. What the class saw was ongoing over time. The spatio-temporality of the original modes of communication are reconfigured as graphic spatiality on the page or the page-like screen. As the motion of actions and gestures on and around the magnets is lost in the images, I decided also to describe the teacher’s movements in writing. I initially followed the conventions of Conversation Analysis in order to suggest the sounds and rhythms of speech, but this became overly complex and did not contribute to the argument I was making, so I removed this detail. Layout suggest more or less insistently how the data are ‘read’, but as control is devolved in the act of ‘reading’ the transcript, choices regarding how to engage with the re-presented subject matter pass to the ‘reader’.

This is just one way in which I have transcribed video materials. Elsewhere I have used writing only in both tables and vignettes, and have experimented with different typographic styles in order to represent different research participants in the same transcript (Mavers, 2011). Rather than narrowing down to one means of transcription, alternative approaches in re-presenting the</td></tr>
</table>

richness of social interaction are suited to different rhetorical purposes. In a recent article where I examine children’s talk and actions, I have not included any moment-by-moment transcripts. Rather, I have described the sounds and rhythms of speech and the teacher’s and children’s movements in the body text. Arguably, this is a form of written transcription. Interest in how researchers go about documenting video materials in their publications provides a helpful ground for critical reflection on the forms and functions of different modes of re-presentation and for continuing to experiment with possibilities.

Example 4: Video stills superimposed with writing (Norris)

The publication	
Publication details	Norris, S. (2011) ‘Three hierarchical positions of deitic gesture in relation to spoken language: a multimodal interaction analysis’, *Visual Communication, 10*(2): 129-147. DOI: 10.1177/1470357211398439 This paper can be accessed online at: http://vcj.sagepub.com/content/10/2/129
The author	Sigrid Norris investigates how people communicate in everyday social interaction. Her work in multimodal interaction analysis has its basis in mediated discourse analysis.
Focus of the publication	Arguing against the taken-for-granted view that spoken language is always primary in social interaction, this paper investigates shifts in the ‘hierarchical positions’ of gesture in relation to speech. ‘Modal density’ (Norris, 2011: 132) is a term that refers to the concentration of modes in social interaction. Norris argues that, within modal density, the ‘modal configurations’ of ‘high-level actions’ are not fixed, but change from moment to moment (Norris, 2011: 130-134). The analysis examines three instances of social interaction: firstly where gesture is subordinate, secondly where there is an equal relationship in the ‘modal aggregate’ (Norris, 2011: 139) of gesture, object handling and spoken language, and thirdly where gesture takes on a superordinate role.
Methods of data collection	The data were gathered through video recording. The researcher participated as the person being explained to.
The transcript	

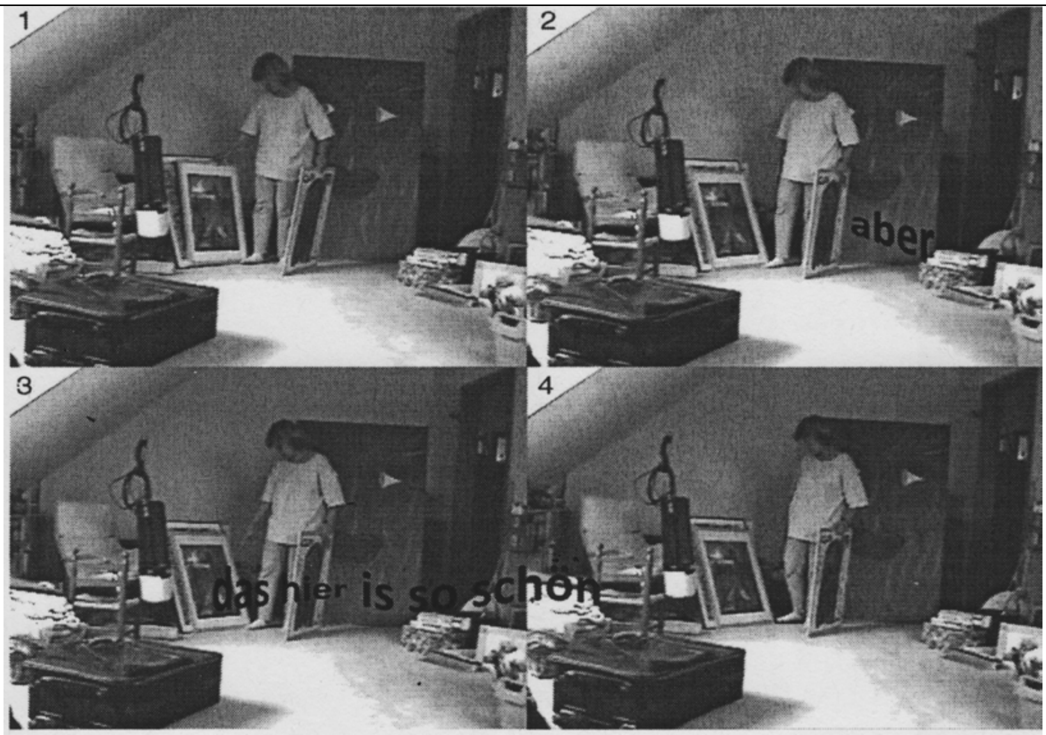


Figure 4: Deitic gesture sub-ordinated to spoken language (page135)
Re-printed from Norris, S. (2011) 'Three hierarchical positions of deitic gesture in relation to spoken language: a multimodal interaction analysis', *Visual Communication, 10*(2): 129-147, with permission from SAGE Publications and the consent of the author.

Subject matter	This transcript re-presents Sandra picking out from a stack a painting that is 'so schön (really nice)'.
Location	The transcript is located in the analysis of the paper's body text.
Relation to other transcripts	All five transcripts included in this paper appear in the analysis in the body of the paper rather than as an appendix, giving immediate access to the visuality and sequentiality of the social interaction. The transcript examined here is the first in the article. In building the analysis, it is repeated two pages later, where the modes under focus are specified (Norris, 2011: 137).
Multimodal configuration of the transcript	
Description	This transcript consists of four video stills with superimposed written transcripts of speech.
Video stills	The four stills that constitute the basis of the transcript are selections from the moving image of the video recording. Chosen shots exemplify the argument being developed rather than representing embodied actions over evenly spaced time periods. Across the stills, the setting and aspects of Sandra's bodily stance remain constant. Distinguishing shifts in her actions and gestures requires some effort on the part of the 'reader', a factor that implicitly underlines the argument put forward in the paper concerning what tends to pass by unseen. Provision of the 'raw' image stills allows alternative or additional interpretation on the part if the 'reader' (e.g. the right leg – straight, slightly bent, more fully bent and then retracted – might constitute an aspect of communication with the researcher).
Writing (alphabetical)	What was said is remade in lower case writing. Undulating letters of varying size suggests intonation and emphasis, although this is not specified in the paper. A transparent background to the transcribed utterances and positioning in the centre rather than at the top or bottom of the images has the effect of withholding the foregrounding of speech.
Writing	Temporality and sequentiality are shown in the numbering of the four stills.

(numerical)	Timings are not provided, although they are elsewhere in Norris's work.
Layout	Four equally sized stills are put together as a block in a left to right, top to bottom sequence. By positioning the writing across the stills, Norris shows what was said when in relation to Sandra's handling of and gesture relating to the paintings.
How the transcript is used	
Relation to the body text	The analysis relating to this transcript (approximately 1,250 words) supplies descriptions of movements that might be gleaned from the video stills, such as 'moving the paintings' and 'pointing out a particular framed picture' (Norris, 2011: 136). The body text notes the layout of the room and specifies Sandra's posture, positioning and gaze in relation to the paintings, whilst the video stills provide detail not given in the analysis, such as the shape and decoration of the room, the furniture and objects in it and how they were arranged, and Sandra's appearance (e.g. gender, ethnicity, clothing, hairstyle). The actional and gestural shifts across the four video stills of the transcript are not immediately obvious visually, and so the descriptions of the body text are crucial for understanding the argument, such as 'moving it slightly towards herself, before moving it back to the stack of others as illustrated in image 2 of Figure 4' (Norris, 2011: 136).
Argument	This first section of the analysis argues that, in this instance, deitic gesture is subordinate to spoken language. Norris contends that what is uttered can be understood without the gesture, but not the gesture without the utterance (Norris, 2011: 143) – and that both modal expressions are interpreted in relation to other modal expressions such as gaze, proxemics and object handling.

Discussion

It is not that transcript designs consisting of writing or image, or combinations of them, are more or less accurate, or better or worse. All transcripts are partial, in the sense that they are a reduction of the video footage. They are also a reshaping (see Ochs 1979; Roberts 1997). Epistemologically, choice of and meaning making with modal resources frames, either deliberately or incidentally, what is included and excluded, and therefore what can be known. A transcript is not a 'replica' of reality, as if its meaning is independent of the researcher or 'reader'. Nor is it merely 'description'. No transcript is value-free. Framed by the research question(s) and the transcriber's analytical focus (see Erickson, 1986), there is an argument to be developed and an audience to be convinced. Through a reconfiguration of modes the researcher gives meaning to the world. As such, transcription does not precede analysis, but is part of it. This is not distortion, but a process of making material into data.

Weighing up, experimenting with, deciding between and reflecting critically on methods of transcribing video footage on the page or the page-like screen are important methodologically, analytically and rhetorically. Questions one might ask oneself include:

Methodology

- Why is multimodal transcription important in my work?
- How do the descriptive and explanatory terms of multimodal transcription cohere with my theoretical and methodological framework?

Purpose

- What is the purpose of the transcript?
- Why did I select this episode for transcription?
- What is the point I wish to make?

Modes of transcription

- Which modes of the original expression or interaction will/did I transcribe? Why?
- Which modes will/did I use for transcribing the video extract? Why?
- What is the function of each mode of transcription?
- What if I had chosen a different mode for transcription?
- What layout will/have I create(d)? Why?

Reshaping

- What is and what is not included with regard to participants, objects, setting and embodied modes? Why / why not?
- What is foregrounded or backgrounded?
- How does my transcript reshape the original expression?
- How does my transcript shape how the 'reader' understands the extract?

Reflection

- What have I achieved?
- What surprised me?
- What might I improve next time?

This paper is not at all exhaustive. Consecutive images, rather like a comic strip, have been integrated into transcription of speech (Lindwall and Ekström, 2012), and images and symbols are combined in 'analytical' transcripts (Britsch, 2009). Music notation is being used as a basis for transcribing the gaze and actions of instrumental players (e.g. Falthin, 2012; Ideland, 2012). Laban notation and maps (e.g. Hackett, 2012) are ways of recording movement. Gaze, which is a key mode in turn-taking in face-to-face communication, is lost in synchronous online discussion (Sindoni, 2012). As methods for transcribing video continue to emerge, this is not a stationary topic. A focus for future work might be to compile a collection of multimodal transcripts for systematic analysis and comparison within and across disciplines. Nor does this paper address other important topics inherent in transcription, such as sampling and units of analysis. Other areas for further investigation include use of digital transcription tools (e.g. Baldry and Thibault, 2006; Bezemer, 2012) and the use of multimodal transcription in quantitative analysis. In pinpointing and understanding moments of decision-making from an array of possible choices, it would also be interesting to observe processes of transcription and to interview transcribers about how they went about it and why. With the prospect of ever-increasing opportunities offered by digital technologies, both for the collection and dissemination of materials, much is yet to come. This paper has endeavoured to suggest some issues for consideration in remaking multimodal expression and interaction captured in video footage as graphic transcripts.

References

Andrews, R., Borg, E., Boyd Davis, S., Domingo, M. and England, J. (eds) (2012) *Handbook of Digital Dissertations and Theses*. London: Sage.

Baldry, A. and Thibault, P.J. (2006) *Multimodal Transcription and Text Analysis: A Multimodal Toolkit and Coursebook with Associated On-line Course.* Sheffield: Equinox.

Bezemer, J. (2012) 'How to transcribe multimodal interaction?' Working paper available online at: http://eprints.ncrm.ac.uk/2299/1/transcriptionchapterforreader1_(2).pdf

Britsch, S. (2009) 'Differential discourses: the contribution of visual analysis to defining scientific literacy in the early years classroom.' *Visual Communication 8*(2): 207-228.

Erickson, F. (1986) 'Qualitative methods in research on teaching', in M.C. Wittrock (ed) *Handbook of Research on Teaching* (pp. 119-161). New York: Macmillan.

Falthin, A. (2012) *Music Notation as a Transcription Tool*. Paper presented at the 6th International Conference on Multimodality, Institute of Education, University of London (August, 2012).

Flewitt, R. (2006) 'Using video to investigate pre-school classroom interaction: education research assumptions and methodological practices.' *Visual Communication 5*(1): 25-50.

Geertz, C. (1973) *The Interpretation of Cultures*. New York: Basic Books.

Goodwin, C. (2000) 'Action and embodiment within situated human interaction.' *Journal of Pragmatics 32*(10): 1489-1522.

Goodwin, M.H. (2009) 'Constructing inequality as situated practice.' In U.T. Kissman (ed) *Video Interaction Analysis: Methods and Methodology* (pp. 41-58). Frankfurt-am-Main: Peter Lang.

Hackett, A. (2012) *Running, Walking and Dancing as Multimodal Communication of Young Children.* Paper presented at the 6th International Conference on Multimodality, Institute of Education, University of London (August, 2012).

Ideland, J. (2012) *Playing Drums or Hitting Pads.* Paper presented at the 6th International Conference on Multimodality, Institute of Education, University of London (August, 2012).

Jefferson, G. (1984) ‘Transcript notation.’ In J.M. Atkinson and J.C. Heritage (eds) *The Structures of Social Action: Studies in Conversation Analysis* (pp. ix-xvi). Cambridge: Cambridge University Press.

Kress, G. (1997) *Before Writing: Rethinking the Paths to Literacy*. London: Routledge.

Lancaster, L. (2007) ‘Representing the ways of the world: how children under three start to use syntax in graphic signs.’ *Journal of Early Childhood Literacy 7*(2): 123-154.

Lindwall, O. and Ekström, A. (2012) ‘Instruction-in-interaction: the teaching and learning of a manual skill.’ *Human Studies 35*(1): 27-49.

Mavers, D. (2009) ‘Student text-making as semiotic work.’ *Journal of Early Childhood Literacy 9*(2): 145-155.

Mavers, D. (2011) *The Remarkable in the Unremarkable: Children’s Drawing and Writing*. New York: Routledge.

Newfield, D. (2009) *Transmodal Semiosis in Classrooms: Case Studies from South Africa*. Unpublished PhD thesis, Institute of Education, University of London.

Norris, S. (2011) ‘Three hierarchical positions of deitic gesture in relation to spoken language: a multimodal interaction analysis.’ *Visual Communication 10*(2): 129-147.

Ochs, E. (1979) ‘Transcription as theory.’ In E. Ochs and B.B. Schieffelin (eds) Developmental Pragmatics (pp. 43-72). New York: Academic Press.

Roberts, C. (1997) ‘Transcribing talk: issues of representation.’ *TESOL Quarterly 31*(1): 167-72.

Sacks, H., Schegloff, E. and Jefferson, G. (1974) ‘A simplest systematics for the organization of turn-taking for conversation.’ *Language 50*: 696-735.

Sindoni, M.G. (2012) *Writing, Screenshot, or Drawing? Multimodal Transcription of Spontaneous Web-Based* Interactions. Institute of Education, University of London (August, 2012).

Further reading

Bezemer, J. and Mavers, D. (2011) ‘Multimodal transcription as academic practice: a social semiotic perspective.’ *International Journal of Social Research Methodology 14*(3): 191-206.

Davidson, C.R. (2009) ‘Transcription: imperatives for qualitative research.’ *International Journal of Qualitative Methods 8*(2): 1-52.

Flewitt, R., Hampel, R., Hauck, M. and Lancaster, L. (2009) ‘What are multimodal data and transcription?’ In C. Jewitt (ed) *The Routledge Handbook of Multimodal Analysis* (pp. 40-53). London: Routledge.

Heath, C., Hindmarsh, J.and Luff, P. (2010) *Video in Qualitative Research: Analysing Social Interaction in Everyday Life*. Los Angeles: Sage.

Mittelberg, I. (2007) ‘Methodology for multimodality: one way of working with speech and gesture data.’ In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson and M.J. Spivey (eds) *Methods in Cognitive Linguistics* (pp. 225-248). Amsterdam: John Benjamins.

Norris, S. (2002) ‘The implication of visual research for discourse analysis: transcription beyond language.’ *Visual Communication 1*(1): 97-121.

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