

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Journal of Pragmatics 46 (2013) 1–7

**journal of
PRAGMATICS**

www.elsevier.com/locate/pragma

Introduction

Multimodal interaction from a conversation analytic perspective

This special issue of the *Journal of Pragmatics* has its origins in the International Conference on Conversation Analysis 10 (ICCA10), which took place in Mannheim (Germany) in July 2010. More than 650 scholars attended the conference, whose theme was “multimodal interaction”. This volume includes papers based on the four plenary talks given at ICCA10 and four additional contributions related to the conference theme.

All of the papers in this volume take a conversation analytic approach to multimodal interaction. Multimodal interaction was a timely theme to be chosen for the only international conference focusing exclusively on conversation analytic research. Research based on video-data and dealing with aspects of bodily conduct in interaction, in coordination with talk and beyond talk, has been very much on the rise for at least the last decade in Conversation Analysis (henceforth “CA”) and CA-related approaches (cf. [Stivers and Sidnell, 2005](#)). However, research using video and dealing with multimodal aspects of interaction has been present in CA already early, see e.g., Sacks’ and Schegloff’s analysis of gestural “home position”, which was published in [2002](#), but is based on a paper dating already back to 1975.

The study of multimodal interaction within a methodological framework which heavily draws on (and in turn adds to) the stock of knowledge of CA has mainly been advanced by the plenary speakers of ICCA10, whose papers are present in this volume: Charles Goodwin, Christian Heath and Paul Luff, Lorenza Mondada and Jürgen Streeck. These scholars have paved the way for a multimodal CA. Their writings have been most influential for everyone who deals with multimodal interaction in a CA-framework.

Starting in the mid-1970s, the Goodwins ([C. Goodwin, 1979, 1980, 1981](#); [M.H. Goodwin, 1980](#)) have gained groundbreaking insights into how the organization of gaze and mutual monitoring is basic for participation in social interaction and how it matters for turn-taking and turn-construction. Their studies have made clear that interaction does not only inform the production of turns because of their embeddedness in structures of sequential organization. Moreover, turn-design is negotiated continuously, because in face-to-face interaction, recipients take part in the ongoing production of speaker’s turns by their concurrent bodily responses and other activities. Basic to the Goodwins’ approach is the notion of ‘participation’, borrowed from [Goffman \(1981\)](#). However, it is accommodated to a CA-framework by insisting that ‘participation’ means embodied situated actions of participants, involving dynamic (re-)negotiation and reconfiguration of spatial, attentional, epistemic and affective alignments of (multiple) participants ([Goodwin and Goodwin, 2004](#); [Goodwin, 2007a](#)). Although providing major contributions to the study of “talk-in-interaction” (cf. [Schegloff, 1991](#)), their research reaches well beyond talk in still other respects. In particular, [Goodwin \(2003, 2007b\)](#) has pointed out the relevance of semiotically rich structures of the material environment for the organization of “environmentally coupled gestures” and other embodied activities in social interaction involving cultural artifacts (e.g., [Goodwin, 2000](#)).

Christian Heath was probably the first and definitely the most influential researcher in bringing the rigor CA’s methodology of detailed sequential analysis to bear on ethnomethodologically inspired workplace studies ([Luff et al., 2000](#)) by using video analysis ([Heath et al., 2010](#)). For a large array of settings as diverse as doctor–patient-interaction ([Heath, 1986](#)), underground control rooms ([Heath and Luff, 1992](#)), galleries ([vom Lehn et al., 2005](#)), and fine art auctions (this volume), Heath and his collaborators, namely, Paul Luff, Jon Hindmarsh and Dirk vom Lehn, have studied how technology, in particular computers, and cultural objects are used and interpreted by actors in cooperative work ([Luff et al., 2000](#)). They have shown that making cognition publicly available in activities and designing them so as to be perceived and understood by others lies at the heart of the coordination of collaborative action ([Heath et al., 2002](#)). Their research has averted (not only) conversation analysts to how attention, monitoring, (peripheral) perception and practical inference provide for the infrastructure of social coordination. Their studies reveal an unprecedented richness of details of the precise timing of tacit activities in coordinated work, which also crucially inform talk-in-interaction. Heath and his

colleagues have developed a unique approach to organizational practice: They show how the social division of labor is enacted in embodied interaction and how it is governed by entitlements and obligations to participate in coordinative work which are relative to role-related duties, but require individual appropriation, creative reasoning, situated interpretation, and fine-grained online coordination of participants.

Lorenza Mondada has provided extensive analyses of a diversity of institutional and ordinary settings, exploring the multimodal resources mobilized by participants to organize their actions from a conversation analytic sequential perspective. She has taken issue with long-standing objectivist views on deixis and space in interaction, which are still widespread in pragmatics (Mondada, 2002). She has shown how the *origo*, the point of departure for successful deictic spatial reference, and interactionally relevant, commonly oriented-to space are achievements which are not prior to interaction, but need to be accomplished in interaction themselves. Mondada (2009) has analyzed the ways in which the coordination of joint attention and the construction of joint interactional spaces are prerequisites for turn-production and reference in interaction. Her work invites us to reconsider classical topics of CA-research like self-repair (Mondada, 2009), turn-taking (Mondada, 2006, 2007a, in this issue) and conversational openings (Mondada, 2010; Mondada and Schmitt, 2010) from a multimodal perspective by showing how non-vocal actions and contingencies of space, (mutual) visibility and body movement inform the production of conversational structures step by step. In her studies, the multimodal approach thus both reveals functions of conversational practice which had not been noticed before, and uncovers multimodal underpinnings on which talk-in-interaction rests. Her current research focuses on asynchronous temporalities of multiple modalities in coordinating multi-activities (Mondada, 2012). Mondada's research is distinctive by her skilled work with collections, applying this core methodology of CA to the study of the complexities of multimodal phenomena in sophisticated ways (Mondada, 2008, 2011).

Jürgen Streeck's work is centered around the organization of gesture within embodied interaction. His studies on varieties of the detailed sequential relationship between gestures, talk and turn-taking are basic to our understanding of the workings of projection in interaction (Streeck and Hartge, 1992; Streeck, 1993, 2009a). His praxeological approach goes beyond the semiotic and cognitive approach, which prevails in research on gesture. Instead, he shows how gesture is social action and how it is used to organize interactional trajectories. Gestures are part of skilled bodily practices (Streeck, 2009b). They exhibit embodied knowledge, which precedes or transcends explicit formulation by talk. Streeck reminds us that in addition to the vocal-auditive and the visio-spatial modalities, there are still other modalities which matter to social interaction, most prominently touch (Streeck, in this issue). Streeck links insights from detailed multimodal and sequential analysis of situated action in a CA-mode to philosophical theorizing about meaning, agency, knowledge, 'being-in-the-world' and the body, drawing on and extending insights from phenomenologists like Heidegger, Merleau-Ponty, and Gibson. His focus increasingly is on the individual, the "multimodal man" – a title of a forthcoming book by Jürgen Streeck – and his/her multiple ways of appropriating the world and dwelling in it by ways of embodied multimodal conduct.

The turn to multimodal interaction is a consequential move for a discipline which aims at a comprehensive understanding of human interaction and which sets as its goal to uncover the practices by which social interaction is produced. While CA and its offspring Interactional Linguistics has already come a long way in showing how minute details of language use matter to the organization of interaction, we are now witnessing an increasing awareness of the relevance of the full array of visible and audible resources to social interaction and of its sensitivity to spatial environments populated with material objects. This combines with an increasing interest of scholars from CA in social settings where interaction is only partially governed by talk or where talk sometimes doesn't even play any role at all. The rise of digital video technology with miniaturization of affordable and reliable recording devices has helped a lot to facilitate this line of research.

Still, at least for some researchers in CA, "multimodal interaction" is a disputed topic. The terms "multimodal interaction" and a fortiori "multimodality" were not originally coined within a CA-context. They have been used by different communities in very different ways for very diverse kinds of phenomena (see, e.g., Jewitt, 2009; Kress and van Leeuwen, 2001; Norris, 2004). In some ways and from some perspectives, "multimodal interaction" is not easily to be accommodated within the framework of conversation analytic research. First, terminology may seem to be misleading: There are some who question if it makes sense phenomenologically to speak of bodily interaction in terms of "multiple modalities", as if they were separate channels in need to be put together. This criticism reminds us to be sensitive to holistic, gestalt-like properties of action-formation and to the synesthetic nature of experience (see Streeck, in this issue) and to be careful not to reify "modalities" as analytic entities. Thus tasks of multimodal coordination need to be studied as they arise from a participants' perspective, in their praxis, not based on a priori assumptions about distinct modalities to be assembled. Second, "multimodality" is a label which is already worn out and has become most fuzzy by its use in various strands of semiotics, discourse and media analysis. So, there may be the danger that under the heading of "multimodal interaction" a lot of things will gather which do not have to do much with CA.

These objections challenge us to determine necessary properties of a conversation analytic approach to multimodal interaction, i.e., an approach in line with CA's analytic mentality. This requires research to be naturalistic, data-driven

inquiry. It needs to attend to all details of multimodal conduct which are accountably relevant for the participants' interaction and by reflecting the way in which they orient to them. The goal would be to uncover (generic and context-sensitive) multimodal practices for the production of situated interactional and social order. Research on multimodal interaction within a CA-framework is faced with basic methodological and theoretical questions (which, as is always the case in CA-research, directly resonate with properties of the interactional events studied). I will shortly point out four of these and refer to components of answers to be found in the research of Goodwin, Heath, Mondada and Streeck.

1. How does simultaneity link up with sequentiality?

This question includes two issues: (a) the temporalities of several multimodal resources which are mobilized to construct some action and (b) concurrent actions by several participants, which can take place simultaneously.

(Ad a) In multimodal interaction, several modal resources are mobilized to build an action. However, what may seem to be clear-cut actions and strictly ordered sequences of actions when looking at talk only, are often multi-layered phenomena with temporalities which are not always isochronic. Therefore, complexities of the relationship between sequentiality of activities and simultaneity of different modalities crucially impinge on the identification of actions. A case in point is a multimodal perspective on openings and closings of interactions (Mondada and Schmitt, 2010; Mondada, forthcoming): Openings and closings are incrementally produced, temporally extended phenomena; activities performed in one modality often do not simply coincide with those performed in another modality in terms of boundaries of actions. Rather, actions are prepared, foreshadowed, or pursued in one modality, while in another modality, participants still or already orient to some other business. Asynchronicities between modalities do not seem to be just accidental, but they constitute systematic coordinative practices where each modality has its own distinctive place in the temporality of the ongoing production of interactional structure.

(Ad b) Already the early studies by C. Goodwin (1979, 1980, 1981) made it very obvious that turn-construction is informed not only by sequential context, but also by simultaneous, multimodal activities of other participants. Interactivity thus is not only a matter of consecutive responding actions, but is already implicated in the ongoing production of turns mediated by monitoring interlocutors simultaneous activities and responses. Recipients' and other participants' concurrent activities may, in addition to sequentially organized actions, be an essential property of the organization of interactional exchanges. This challenges us to abandon the model of 'speaker-hearer', which still seems to underlie a lot of research in CA focusing on sequential organization by only attending to verbal activities. Instead we need to work with a model of 'participation' which takes into account the simultaneous multimodal activities of all participants in an interactional event. We need to see how *all* participants' activities contribute to bring off the event in the precise multimodal and sequential shape it gets.

2. How can the CA-style of fine grained detailed analysis of the particulars of interactional episodes be brought to bear on what can be seen and heard on a video recording?

Firstly, this is a question about data representation, and in particular, about transcription. It is commonplace that transcription is necessarily selective (Duranti, 2006) and depends on theoretical standpoints (Ochs, 1979). However, there are standards for transcription of talk-in-interaction captured by conventions by Jefferson (2004) and GAT2 (Selting et al., 2011) which solidly rest on research about which features of vocal production are regularly oriented to by participants when producing and interpreting interactional structures. This is not the case for visual transcription. There is no evident generic convention which defines which phenomena need to be transcribed and in which granularity. The visual richness of the video and requirements on the readability of transcripts obviously make the notion of "comprehensive video transcription" fictitious. The production of a multimodal transcript therefore is a thoroughly reflexive and interpretive accomplishment (Mondada, 2007b): Which phenomena are captured in which way depends on what participants make relevant in their activities and on the research question and analytical interests which in turn need to be adapted to the participants' orientations. The same applies to the selection of stills to be included in transcripts and analyses. In this way, however, multimodal transcripts are much more a product of detailed analysis than its precondition. Much more radically than verbal transcripts, they are means to enable the reader to see what matters analytically, but they do not provide an uninterested objectivist view on the scene recorded, which would provide for a neutral ground to test the analysis in a strict sense. This does not mean that multimodal transcription is simply idiosyncratic. There are emerging conventions to be used for the representation of fairly formal temporal parameters, which apply to a large range of multimodal phenomena (like onset, duration, apex, coordination with talk; see Mondada, 2009) and which (along with the systematic use of stills) allow for recovering relevant features of the visual record.

Secondly, because of the simultaneous complexities of multimodal interaction, studies are mostly restricted to in-depths case studies. Multimodal interaction poses challenges for setting up and working through collections of data (Mondada, 2008). The analysis of multimodal collections requires to consider a larger range of formal audible and visual features. The simultaneous complexities of video data imply that what may look like similar cases when

considering only talk, may differ considerably in terms of the use of other modalities. The number of simultaneously potentially relevant properties of action multiplies when dealing with multimodal data. Thus the analyst is led to probe how some putative practice might hinge on the realization of each of the modal resources (and perhaps also properties of the material surroundings) involved in sequences of activities. Issues of accounting for selective attention to specific modalities (while neglecting others) and of arriving at adequate orders of abstraction of phenomena across several modalities for instances to be included in a collection become much more acute than with verbal data.

3. How can we account for bodily activities?

CA has shown that the sequential organization of interaction provides an “architecture of intersubjectivity” (Heritage, 1984:254; Schegloff, 1992): Speakers display their understanding of a prior turn in the next turn (and are taken to be doing so by their partners) and they take a stance on how they are understood by others in third-positioned turns. For us as analysts this means that participants’ displays of understanding in consecutive turns offer a “proof procedure” (Sacks et al., 1974:728) for the analysis of interaction. Video data can enhance our understanding of participants’ understandings, because the analyst can draw on the participants’ bodily displays of how they interpret others and how their own actions are coordinated with and oriented to other events, objects and participants. Still, what is to be seen on a video does not always clarify participants’ activities. Video data offer a wealth of minute bodily activities, objects involved in activities, features of spatial surroundings participants orient to, (technical) devices with semiotic properties (like computer-screens, control panels, written materials) and events going on in the background of some focused interaction which in some way or another seem to impinge on it. Often, the meaning of such phenomena and their import for the interaction at hand does not become clear from close interactional analysis of participants’ displays alone. This is already because bodily displays are implicit and therefore often very ambiguous. Sometimes they can even only be noticed if the analyst is averted beforehand to how some property of bodily action might matter for interactional coordination and sense-making. Close analysis of visible and audible multimodal interaction organization therefore often has to be complemented by ethnography. This becomes particularly evident in studies such as by Goodwin (1994, 2003) on professional vision of archeologists and their use of professional semiotic artifacts like the Munsell chart, which embody a whole theory of professional knowledge; another example are Heath and Luff’s studies on London Underground control rooms (Heath and Luff, 1992), which rest on a profound understanding of requirements and dependencies of the organization of the traffic on the Underground, the local division of labor, and information properties of control panels and displays. Ethnography and access to professional background knowledge is necessary to understand situated displays the way participants do, to make sense of indexical references peculiar to professional practice, to decipher the significance of objects, to recover references, understand grounds for evaluations, etc. With video data we become aware how social organization (of workplaces, institutions, etc.) and cultural understandings (of objects, rights and obligations associated with categories of participants) are immediately consequential for interaction and how they are (re-)produced by it. With respect to the actors involved, studies based on video show that interactional organization rests on an embodied cognitive infrastructure of visual monitoring and skilled bodily action.

4. How do basic notions of CA have to be adapted to and perhaps reworked within a multimodal framework?

Sequencing and construction of turns is sensitive to spatial context, participants’ body movements and configurations and the affordances of objects in interaction, which are often semiotically laden. Mondada (in this issue) shows how turn-taking within a multi-party context is organized by gesture and the body movements in space, which at the same time poses constraints and specific requirements on organizational activities. Repair-practices are another point in question: Mondada (2009) has shown how self-repairs are used to delay vocal production in fine coordination with the bodily activities performed to accomplish a shared interactional space as a prerequisite for the incipient collective action. Well-known interactional practices thus may be revealed to be organized by bodily resources, be sensitive to properties of the material surroundings and have motivations and functions which were not known hitherto and become only evident when turning to video data. These uses could not have been discovered or would even have been misapprehended by analysing only audio-data.

Video analysis will not only offer new insights into the interactional organization of practices which have already been known. Video-based findings may also lead us to rethink basic notions of Conversation Analysis. A case in point is the question: Are there embodied, multimodal turns? It is well known that second pair parts of an action sequence which are made conditionally relevant by a first pair part, like answering a yes/no-interrogative, may (under some conditions) be performed without talking, or, what is more, it may not be a verbal action which counts as the conditionally relevant response, like in many cases of fulfilling a request. This invites us to ask how sequential organization of turns relates to sequential organization of actions. We might be tempted to think of multimodal turns which could also be produced by kinesic behavior only. But if we do so, then basic properties of what we know about turn-taking-organization and what it accounts for, e.g., the occurrence and resolution of overlap, participants’ orientation to the floor and to a principle of “one speaker at a time”, the composition of turns out of distinct turn-constructive units, etc., do not seem to sit well with “turns” without talk. Still, other properties well-known from studies of turn-organization like completion points,

projection, conditional relevance etc. are fundamental for kinesic action as well. Thus, multimodal video analysis requires us to consider how sequencing of actions relates to the affordances of turn-taking and turn-construction, how vocal turns and non-vocal multimodal conduct might follow different orders of organization, implying different requirements and providing different resources for participants, which, of course, are mostly deeply intertwined and synthetically integrated in embodied conduct.

All papers in this volume address issues of how talk-in-interaction is organized in the context of embodied multimodal conduct. *Charles Goodwin* takes an anthropological perspective to action formation in interaction. He discusses how different modal resources are assembled and sequentially re-assembled to build situated action. He underscores that human interaction is accomplished by the cooperative transformation and “lamination” of contextually, i.e., sequentially and historically, provided social resources to build composite structures of action. This achievement is unique to humans and, Goodwin argues, it is most basic for cultural traditions and development. The construction of single actions and turns rests on both past and simultaneous contributions and activities by several participants. *Christian Heath* and *Paul Luff* deal with contracting in auctions by “the fall of the hammer”, a prime example of how sequential organization of embodied action systematically involves semiotically rich objects. They show how turn-design and turn-transition are coordinated in perfectly timed coordination with embodied conduct. The auctioneer produces the fine-grained details of delicately timed action to be seen and interpreted by potential bidders, thereby exhibiting its sensitivity to the multi-party participation framework in which it occurs, reflexively encouraging (possible) participation. *Lorenza Mondada*'s paper advances our understanding of embodied practices of turn-allocation. She studies practices of a chairman in a participatory democratic meeting showing how he uses gestures and body movement to take note of bids for the floor, queues, allocates and defends speaking rights in a complex spatial configuration of a multi-party situation. Mondada works out how this situated organization of turn-taking embodies a particularized, practical understanding of how to facilitate egalitarian democratic participation in a political debate. While the turn to multimodal analyses almost always has lead researchers to attend to the kinesic-visual mode in addition to vocal-acoustic mode, *Jürgen Streeck* shows the relevance of a non-visual modality, i.e., touch and tactile, kinesthetic experience for the organization of interaction. Drawing on phenomenological and anthropological concepts, Streeck pleads for a more comprehensive understanding of embodiment in interaction. His focus is on how embodied knowledge is displayed and used in episodes of action and how meanings are conveyed by embodied conduct. Moreover, he shows how intersubjective coordination, e.g., by collective walking, may be accomplished without and beyond talk. *Arnulf Deppermann* looks at the design of turn-beginnings. He argues that turn-beginnings are both informed by their placement within sequential structures of interaction and by requirements of the multimodal coordination of participation in face-to-face interaction. The paper is to show that turn-construction is sensitive to an interactionally motivated order of tasks, which arise when intersubjective coordination and cooperation is to be achieved. Tasks include accomplishing shared attention, displaying uptake and understanding, dealing with expectations established by prior talk and projecting properties of the incipient turn. Drawing on German data, Deppermann shows how the design of turn-beginnings can be accounted for by the use of multimodal practices which are used to systematically deal with these tasks. *Marjorie Harness Goodwin* and *Asta Cekaite* deal with the multimodal accomplishment of directive response sequences in social space. They show how when dealing with the project of getting children to bed, embodied (inter-)action extends well beyond the scope of the verbal “core” activities, amounting to trajectories of negotiated and recycled communicative projects involving bodily action and the organization of domestic space. The authors attend to the detailed interplay of verbal action with bodily activities, showing how stance, affect, insistence and moral facets of the requests provide interactional relevancies of their own, which impinge on the precise design of request response sequences. *Florence Oloff* studies dropping out of overlap and withdrawal from a trajectory of action as an embodied phenomenon. She shows that to stop talking (“drop out”) does not necessarily amount to abandon the claim to speaking rights related to some action to be accomplished (“withdrawal”). Speakers use gestural and other bodily displays as visible activities to show whether they are still claiming or abandoning the turn. Oloff's paper is an exercise in probing the relevance of differing temporalities of different modalities for conversational action, showing that only attention to all modalities of embodied conduct and their temporal coordination allows to capture adequately how an action (here: withdrawal of a turn) is accomplished. The volume closes with *Eiko Yasui*'s study on the use of gestures in the process of the interactional negotiation of proposals. Gestures are shown to be an important resource in the cooperative production, clarification and negotiation of ideas which are worked out collectively. Yasui argues that repetition of gestures can be used as a means of displaying understanding and affiliation with the partner's proposal, while disaffiliation, however topically aligned, is conveyed by modified uptake of the partner's prior gestures.

Acknowledgement

I thank Lorenza Mondada for helping me to clarify some of the issues raised in this introduction.

References

- Duranti, Alessandro, 2006. Transcripts, like shadows on a wall. *Mind, Culture, and Activity* 13 (4), 301–310.
- Goffman, Erving, 1981. Footing. In *His: Forms of Talk*. Blackwell, Oxford, pp. 124–159.
- Goodwin, Charles, 1979. The interactive construction of a sentence in natural conversation. In: Psathas, George (Ed.), *Everyday Language: Studies in Ethnomethodology*. Irvington, New York, pp. 97–121.
- Goodwin, Charles, 1980a. Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry* 50 (3/4), 272–302.
- Goodwin, Charles, 1981. *Conversational Organization*. Academic, New York.
- Goodwin, Charles, 1994. Professional vision. *American Anthropologist* 96 (3), 606–633.
- Goodwin, Charles, 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32, 1489–1522.
- Goodwin, Charles, 2003. Pointing as situated practice. In: Kita, Sotaro (Ed.), *Pointing: Where Language, Culture and Cognition Meet*. Lawrence Erlbaum, Mahwah, NJ, pp. 217–241.
- Goodwin, Charles, 2007a. Participation, stance, and affect in the organization of activities. *Discourse and Society* 18 (1), 53–73.
- Goodwin, Charles, 2007b. Environmentally coupled gestures. In: Duncan, Susan, Cassell, Justine, Levy, Elena (Eds.), *Gesture and the Dynamic Dimensions of Language*. John Benjamins, Amsterdam/Philadelphia, pp. 195–212.
- Goodwin, Charles, Goodwin, Marjorie Harness, 2004. Participation. In: Duranti, Alessandro (Ed.), *A Companion to Linguistic Anthropology*. Blackwell, Malden, MA, pp. 222–244.
- Goodwin, Marjorie Harness, 1980b. Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry* 50 (3–4), 303–317.
- Heath, Christian, 1986. *Body Movement and Speech in Medical Interaction*. Cambridge University Press, Cambridge.
- Heath, Christian, Hindmarsh, Jon, Luff, Paul, 2010. *Video Analysis and Qualitative Research*. Sage, London.
- Heath, Christian, Luff, Paul, 1992. Collaboration and control. Crisis management and multimedia technology in London Underground line control rooms. *Computer Supported Cooperative Work* 1, 69–94.
- Heath, Christian, Sanchez Svensson, Marcus, Hindmarsh, Jon, Luff, Paul, vom Lehn, Dirk, 2002. Configuring awareness. *Computer Supported Cooperative Work* 11 (3/4), 317–347.
- Heritage, John, 1984. *Garfinkel and Ethnomethodology*. Polity, Oxford.
- Jefferson, Gail, 2004. Glossary of transcript symbols with an introduction. In: Lerner, Gene H. (Ed.), *Conversation Analysis: Studies from the First Generation*. John Benjamins, Philadelphia, pp. 13–23.
- Jewitt, Carey (Ed.), 2009. *The Routledge Handbook of Multimodal Analysis*. Routledge, London.
- Kress, Gunther, van Leeuwen, Theo, 2001. *Multimodal Discourse*. Hodder Arnold, London.
- Luff, Paul, Heath, Christian, Hindmarsh, Jon (Eds.), 2000. *Workplace Studies: Recovering Work Practice and Informing Systems Design*. Cambridge University Press, Cambridge.
- Mondada, Lorenza, 2002. Die Indexikalität der Referenz in der sozialen Interaktion: die diskursive Konstruktionen von “ich” und “hier”. *Zeitschrift für Literaturwissenschaft und Linguistik* 125, 79–113.
- Mondada, Lorenza, 2006. Participants’ online analysis and multimodal practices: projecting the end of the turn and the closing of the sequence. *Discourse Studies* 8, 117–129.
- Mondada, Lorenza, 2007a. Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies* 9 (2), 195–226.
- Mondada, Lorenza, 2007b. Commentary: transcript variations and the indexicality of transcribing practices. *Discourse Studies* 9 (6), 809–821.
- Mondada, Lorenza, 2008. L’analyse de «collections» de phénomènes multimodaux en linguistique interactionnelle. À propos de l’organisation systématique des ressources gestuelles en début de tour. *Cahiers de Praxématique* 50, 23–70.
- Mondada, Lorenza, 2009. Emergent focused interactions in public places: a systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics* 41, 1977–1997.
- Mondada, Lorenza, 2010. Eröffnung und Vor-Eröffnung in technisch vermittelter Interaktion: Videokonferenzen. In: Mondada, Lorenza, Schmitt, Reinhold (Eds.), *Situationseröffnungen: Zur multimodalen Herstellung fokussierter Interaktion*. Narr, Tübingen, pp. 277–334.
- Mondada, Lorenza, 2011. The organization of concurrent courses of action in surgical demonstrations. In: Streeck, Jürgen, Goodwin, Charles, LeBaron, Curtis (Eds.), *Embodied Interaction, Language and Body in the Material World*. Cambridge University Press, Cambridge, pp. 207–226.
- Mondada, Lorenza, 2012. Talking and driving: multi-activity in the car. *Semiotica* 191 (1/4), 223–256.
- Mondada, Lorenza. Multimodal completions. In: Deppermann, Arnulf, Günthner, Susanne (Eds.), *Temporality in Interaction*. Benjamins, Amsterdam, forthcoming.
- Mondada, Lorenza, Schmitt, Reinhold, 2010. Zur Multimodalität von Situationseröffnungen. In: Mondada, Lorenza, Schmitt, Reinhold (Eds.), *Situationseröffnungen: Zur multimodalen Herstellung fokussierter Interaktion*. Narr, Tübingen, pp. 7–52.
- Norris, Sigrid, 2004. *Analyzing Multimodal Interaction*. Routledge, London.
- Ochs, Elinor, 1979. Transcription as theory. In: Ochs, Elinor, Schieffelin, Bambi (Eds.), *Developmental Pragmatics*. Academic Press, New York, pp. 43–72.
- Sacks, Harvey, Schegloff, Emanuel, 2002. Home position. *Gesture* 2 (2), 133–146.
- Sacks, Harvey, Schegloff, Emanuel A., Jefferson, Gail, 1974. A simplest systematics for the organisation of turn-taking in conversation. *Language* 50 (4), 696–735.
- Schegloff, Emanuel A., 1991. Reflections on talk and social structure. In: Boden, Deirdre, Zimmerman, Don H. (Eds.), *Talk and Social Structure*. Polity Press, Cambridge, pp. 44–70.
- Schegloff, Emanuel A., 1992. Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology* 97 (5), 1295–1345.

- Selting, Margret et al., 2011. A system for transcribing talk-in-interaction: GAT 2. Translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. *Gesprächsforschung* 12, 1–51. <http://www.gespraechsforschung-ozs.de/heft2011/px-gat2-englisch.pdf>, (accessed 22.10.12).
- Stivers, Tanya, Sidnell, Jack, 2005. Introduction: multi-modal interaction. *Semiotica* 156 (1/4), 1–20.
- Streeck, Jürgen, 1993. Gesture as communication. I: Its coordination with gaze and speech. *Communication Monographs* 60, 275–299.
- Streeck, Jürgen, 2009a. Forward-gesturing. *Discourse Processes* 4 (2–3), 161–179.
- Streeck, Jürgen, 2009b. *Gesturecraft*. Benjamins, Amsterdam.
- Streeck, Jürgen, Hartge, Ulrike, 1992. Previews: gestures at the transition place. In: Auer, Peter, di Luzio, Aldo (Eds.), *The Contextualization of Language*. Benjamins, Amsterdam, pp. 135–158.
- vom Lehn, Dirk, Heath, Christian, Hindmarsh, Jon, 2005. Examining exhibits: interaction in museums and galleries. *Communication and Cognition* 38 (3/4), 229–247.

Arnulf Deppermann*

Institut für Deutsche Sprache, PF 10 61 21, D-68016 Mannheim, Germany

*Tel.: +49 06211581309

E-mail address: deppermann@ids-mannheim.de