Qualitative Research

– Offprint –

Videography: analysing video data as a 'focused' ethnographic and hermeneutical exercise

Hubert Knoblauch
Technical University Berlin, Germany

Bernt Schnettler
University Bayreuth, Germany

Abstract

Visual research methods are becoming increasingly important for qualitative studies. Within this dynamically expanding field, methods for analysing 'natural' video recordings have developed considerably over the past decades. In this article we discuss methodological aspects of general importance for any analysis of this type of video data. Being a fundamentally interpretive method, our first argument is that sequential analysis is always a *hermeneutic* endeavour, which requires methodical understanding. The second refers to data collection. We stress that, in addition to sequential analysis, the *ethnographic dimension* of video analysis should be taken into account methodologically. Video analysis requires, thirdly, a systematic account of the subjectivity, both of the actors analysed as well as of the analysts. Our arguments are grounded in extensive data from several studies, including the communicative genre of powerpoint presentations, commemoration rituals and public events. Selected data fragments are presented here to support our claims. Building upon this expertise, we propose further improvement of video analysis methodology by reflecting on our own practice of analysing video in data sessions (i.e. the 'video analysis of video analysis').

Over the past decades, video analysis has emerged into a powerful new tool for qualitative research. Particularly, it has been productive in generating excitingly detailed insight for inquiry in areas including workplace-studies, science and technology studies, or

Corresponding author:
Bernt Schnettler, University of Bayreuth, GW II, 95440 Bayreuth, Germany.
Email: schnettler@uni-bayreuth.de

education – to name but a few. One of its most outstanding properties consists in the unprecedented access it provides to the minutiae of social interactions in real time. Video, in short, offers a 'microscope' for an in-depth study of the on-going production of situated social order. The methods for analysing video data have evolved quickly, especially in the last few years. Mainly, this has been accomplished as a by-product of research in substantial fields, so that the main methodological advancements are firmly rooted in extensive research experiences with video data. In recent years, the corpus of texts specializing in the methodology of video analysis has been rapidly growing.

Social actions and interactions constitute the major focus of interest in this type of video analysis. So, one may even dare to claim that the paradigmatic case for interpretive video analysis is what Goffman (1983) has called 'interaction order'. Most often, within the interaction order, these video analyses orient towards 'focused' forms of interaction. According to Goffman (1963), this is a form of interaction in which the participants share a common focus of attention. In the simplest case this is constituted by two actors. But focused interactions may also extend to larger social occasions, such as meetings, stage events and demonstrations. Whereas the interactive core in such situations may be produced by spatial and material items (such as body formations, stages, megaphones), other technological mediations of interaction may also move into the analytical center of attention.

Undoubtedly, the methodology of interpretive video analysis has been chiefly influenced by researchers trained in the ethnomethodological tradition of conversation analysis. The elaborate practice of 'sequential analysis' worked out in this tradition is the methodological cornerstone of the emerging video interaction analysis. It proves most helpful for the in-depth examination of audio-visual recordings. It is not an exaggeration to claim that sequential analysis forms the core of video analysis. Its methodological principles and the corresponding research practice have been described in detail by Heath et al. (2010). Put in short terms, the analytical exercise of rigorously reconstructing the exact sequential organization of a more or less complex stretch of interaction matches with the ethnomethodological premise of social order as being locally produced, 'moment-by-moment'. This exercise results in a remarkable training of sociological attention to the fine-grained details of interaction and its intricacies. In other words, video interaction analysis is not a version of content analysis that exclusively takes into account those aspects documented on the video-tape. In order to make sense of the recorded interaction – which is indeed the main object of analysis – the *context* it is embedded in has to be considered systematically. The role of contextual knowledge, however, is a controversial point among video researchers, particularly those related to Conversation Analysis.

Our methodological arguments are the following: (1) video interaction analysis is a fundamentally interpretive method. Therefore, analysing video data has to be considered a *hermeneutic* activity. (2) Video recordings are only tools. The analytical focus lies on the interaction taking place in a certain social situation. Making sense of what has been tape-recorded essentially depends upon additional contextual knowledge. This insight is provided by focused *ethnography*. (3) With the aim of further improving the existing video analytical methods, we need to observe our own practices of video analysis and reflect them methodically. In addition to principles borrowed from established

interpretive methods (of verbal or textual data), this reflection may allow us to generate unique new methodical principles specific for video data. Therefore, in the last section of this article we will proceed to the question of how the research practice itself can yield methodological insights in the sense of what might be called a 'video analysis of video analysis' or second order video analysis.

Video interaction analysis as interpretation

Why do we consider video analysis a *hermeneutical* activity? Obviously, we thereby want to emphasize that this method is firmly rooted in the interpretive tradition in the social sciences. The main task is, very generally speaking, to understand 'what is going on'. But that what is going on is not exclusively a category of observation. More specifically, we recall one of the fundamental methodological assumptions of any interpretive social science: actions and interactions are not only to be observed – rather, actions are guided by meanings any observer must try to account for, not only in principle but in each instance (see Schutz, 1962; Weber, 1984 [1921]).

Remembering this basic principle of the interpretive paradigm allows us to distinguish interpretive video analysis from other methodological approaches dealing with video data that have become recently quite popular. Those are, indeed, more or less orientated towards a kind of 'standardized' methodology and mainly positioned in research areas dealing with experimental studies of audio-visual conduct. However, within the growing body of research using video in the social sciences, there are even some who claim to be 'qualitative', although they follow a methodology much more compatible with classical quantitative research. This more or less 'standardized' video analysis starts from theoretical assumptions that are subsequently 'operationalized' into observational categories. According to these categories, (larger) stretches of video recordings are rated applying pre-established codes; for example, as something identified as 'supportive' or 'non-supportive', 'aggressive' or 'non-aggressive behaviour' (Mittenecker, 1987). Commonly, in this kind of research, the reasons for subsuming empirical instances to codes are not being stated so that the process of interpretation remains completely concealed. The codes, instead, are habitualized (tested by way of 'inter-coder reliability') or even automated into software programmes (see Dinkelaker and Herrle, 2009; Koch and Zumbach, 2002).

These approaches may be appropriate for certain research purposes. However, their methodology is distinct from interpretive video analyses for which explicit acts of interpretation form an essential part. We start from the assumption that action can only be explained if we understand its meaning. Along this line, Schutz (1962) suggested to distinguish between the meanings actors link with their actions and the ways how we, as observers, conceive of these meanings and conceptualize them scientifically. He calls this distinction 'constructs of the first order' versus 'second order constructs'. The world of everyday-life is a reality already interpreted by those we study. Any scientific understanding has to be grounded on the everyday-life understanding – notwithstanding that scientific concepts have to fulfil specific requirements like unambiguity, logical consistency, etc. Schutz called this grounding the 'postulate of adequacy'. It is important to notice that adequacy is not limited to the contents of interpretations. It also applies to the

forms or 'methods' of understanding. If we closely inspect *the ways* in which people in everyday-life make sense, we might extract procedures and heuristic principles usable for scientific understanding. Obviously, these principles have to be de-pragmatized and methodically controlled. This has been elaborated into a proper social theory of interpretation by Soeffner (1989; 1996) who uses the term *hermeneutic* in order to emphasize that understanding is a kind of artful practice to reconstruct sense-making. Sociological hermeneutics as a method of sociological understanding and explaining has been influenced substantially by Soeffner (1999: 48), who characterizes it as a 'specifically historical self-reflexive epistemological style rooted in the conviction that there is no irrevocable, a-historically certain knowledge, no finally-settled social theory'. The origin for this line of theory is Weber's *Verstehende Soziologie*, which received its phenomenological deepening from Schutz and its profile of a 'new' sociology of knowledge from Berger and Luckmann (1966). The aim of this sociological approach is to reconstruct the 'social construction of reality'.

With his notion of 'double hermeneutics' Giddens (1984) added the idea that scientific constructs themselves do not exist completely isolated from first order constructs. Instead, both do interact with each other. As particularly ethnomethodological authors have stressed, this interaction is not to be considered as problematic. Rather, it can provide a resource for understanding the subject matter under scrutiny (Sacks et al., 1973). Social scientific analysis and its specific categories (such as 'turns', 'moves', 'sequences', etc.) rest on the hermeneutic assumption that analysts dispose of knowledge of the culture in which the actions occur and make use of this knowledge in order to understand what is going on by way of a hermeneutic circle. We will return to the relevance of hermeneutic knowledge in the final section of the article.

In the following section we discuss some of the basic methodological premises of video analysis. They are related to the fact that audio-visual recordings entail both *sequential* and *simultaneous* aspects that need to be interpreted. We will shortly sketch the basic principles of sequential analysis before turning to ways in which the simultaneous aspects of audio-visual data can be addressed through focused ethnography. We shall do so with particular respect to examples from a larger study on powerpoint presentations.

Sequentiality: the on-going orchestration of communicative modalities

One of the most exiting properties of video recordings for research on interaction is its capacity to register on-going social activities in a very detailed way that preserves its sequential organization. Therefore, sequential analysis is the appropriate method to scrutinize audio-visually recorded real time interaction. One should remind the non-German speaking audience that since the early 1970s hermeneutics developed a particular concept of sequential analysis. Here, it refers to a research practice that seeks to interpret data excerpts step-by-step, seeking to unfold all possible interpretations or 'reading versions' (*Lesarten*). This can be understood as a systematic excise in applied sociological imagination. Sequential analysis in sociological hermeneutics is an artful

interpretation practice situated on the 'second order level'. We are following here the somewhat different concept of sequential analysis developed within ethnomethodology. This notion of sequentiality is located on the 'first order level' and it treats the social context in which the sequence is imbedded differently. The ethnomethodological idea of sequentiality may be best understood by the classical example from Conversation Analysis in which Schegloff (1968) examines the openings of telephone calls to a 'disaster centre'. Concluding his first inspection, Schegloff assumed that there must be a 'distribution rule of first utterances'. Complying with this rule, the one who receives the phone call speaks first. Then, however, he is confronted with the following 'deviant case':

Excerpt 1 (quoted from Schegloff, 1968: 1079):

> (Police make call)
> Receiver is lifted, and there is one second pause
> *Police*: Hello.
> *Other*: American Red Cross.
> *Police*: Hello, this is Police Headquarters . . . er, Officer Stratton [etc.]

Obviously, the distribution rule of first utterances does not hold in this case because there is a one second pause after the phone call starts. Then, it is the caller who talks first. This seems unusual. But the fact that the caller is talking first is not really a violation of the rule. Rather, the person called leaves a pause, which is taken as a kind of answer, to which the 'hello' in the second turn replies. Observe that this 'hello' is not only a salutation. It also accounts for the lacking response. The general insight to be drawn from this example is *that actions are rendered meaningful in the context and by virtue of other actions*, and that this context is constituted by the sequence of this action.

For a proper understanding of this type of sequential analysis, it is most important that the 'turn' is not a 'speech act'. In the case cited above, it is not even a human action in the strict sense but rather a technologically mediated action – the phone ringing constitutes a 'turn'. Consequently, non-linguistic cues can function as turns. This may not surprise since acoustically coordinated action depends essentially on the temporal ordering of sound – otherwise it becomes 'noise' (Goffman, 1981). How then, we may ask, can audio-visual conduct be analysed sequentially?

To answer that question, we will inspect an example from a research project on the communicative genre of powerpoint presentations (Schnettler and Knoblauch, 2007). Its data corpus includes recordings of 196 presentations, lasting between two minutes and two hours and mounting to a total of 6047 minutes of footage. Data was collected between 2004 and 2006.[1] We choose a fragment from this research because its setting is familiar and does not require extensive contextual information. The following video fragment (2a) has been recorded in a non-academic administrative organization on the occasion of a small group meeting in which an invited expert instructs staff by using a powerpoint presentation. We will concentrate on the initial sequence of this presentation. It is delivered in German, but visually supported with English slides[2]:
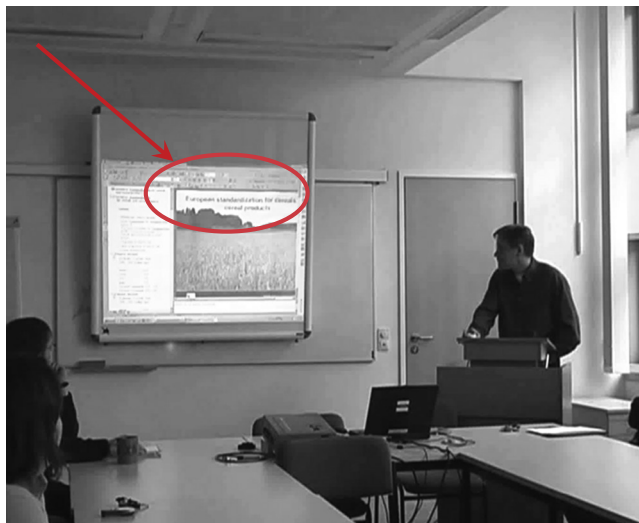
Fragment 2a. Typewriter on powerpoint

Excerpt 2b. Typewriter on powerpoint – sequential organization of verbal and visual conduct

1	speaker	uh ich darf Sie also dann jetzt herzlich begrüßen.
		am I may welcome you here now cordially
2		(0,8)
3	audience 1:	O: ⌐aah *wow*
4	audience 2:	uhh └
5	audience 3:	hmmm ⌐mmm
		└hmm
6		(2,0)
7	speaker:	ja, europäische Standardisierung (.) für Getreide
		yes, European standards for cereals

> **While the letters of the following sentence appear subsequently, the sound effect of a mechanical typewriter is produced by the computer: 'European standardization for cereals and cereal products']**

Our attention concentrates on the first part of the fragment (sequences 1–6). After the speaker starts to welcome the audience cordially (line 1), an inscription appears on top of the screen as the title of the projected slide, reading 'European standardization for cereals and cereal products'. Each single letter pops up subsequently, accompanied by an audio effect that sounds like the rattle of a traditional typewriter (line 2). Shortly after this special effect starts, members of the audience begin to express their astonishment and appreciation (lines 3–5). Since the speaker's introductory sentence (line 1) is a ritualized welcome and does not provide any reason for this reaction, we can infer that the audience's audible reactions ('aah', 'uuh', 'hmm') are oriented towards the clattering and may be appreciative of this technical gimmick. Prosodically there is a slightly ironic tone

going with the first appreciation in a particular way (line 3), and indeed this interpretation as ironic is picked up by the next turn (line 4) in which someone else starts to laugh – a laughter that is obviously not followed by a next turn of the speaker but yet by someone else joining in the laugh (line 5). This whole sequence is performed while the sound effect is going on and ends exactly with its termination. Then the speaker leaves the slot open (line 6) – as is often the case when technologies 'act' – so that the turns of the audience members who become here primary speakers can be understood as oriented exactly to the technology. The speaker himself, however, does not share this orientation (e.g. by smiling back so as to acknowledge somehow the authorship of the gimmick) but rather turns from the paper orientation to the slide and repeats the title that has been produced by the typewriter sounds.

This example illustrates that the sequential organization of interactive 'turns' cannot be deduced from the structure of verbal interaction only, as Conversational Analysis used to claim. Video provides access to the complex interplay of communicative activities that take place in different modalities. We see how what appears to be a 'nonverbal' turn (the sound of the typewriter) is designed to engender another 'nonverbal' effect (the appreciation tokens) and how these, retroactively, bestow sense on the prior turn – the sound effect is appreciated by the audience members, instead of being, say, understood as an improper interruption of the speaker's talk caused by the technology.

Such sequential organization of turns has been analysed in much detail (e.g. by Goodwin, 2000; Heath, 1997; Heath and Hindmarsh, 2002), and it is well described in Heath et al. (2010). We can, therefore, immediately proceed to a more complex example, again recorded during a powerpoint presentation. This data extract further illustrates the ways in which visual and verbal activities interplay with each other.

Given the popular cultural critique claiming that powerpoint slides oversimplify information, in our analysis of powerpoint we decided to examine how slides are factually used in powerpoint presentations. This is the case in video fragment (3a), recorded on a high profile scientific conference with some 50 persons in the audience:
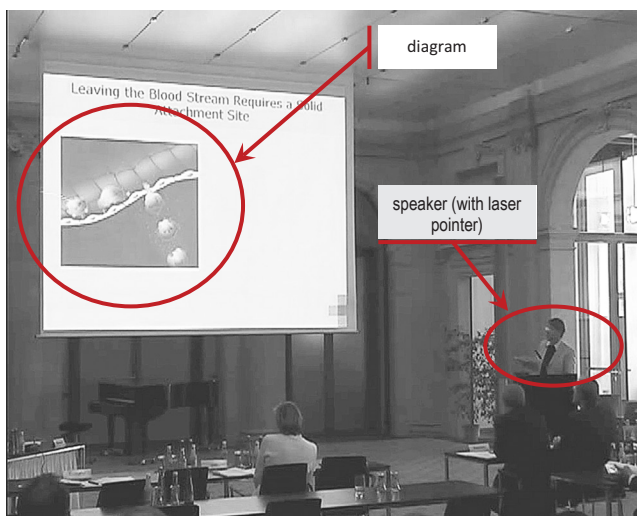
Fragment 3a. The diagram

In this presentation the speaker comments upon an elaborated diagrammatic form of slides (see Excerpt 3b). While showing the slides, he delivers the following words (the underlined text indicates where the speaker uses a laser pointing device to highlight particular items):

Excerpt 3b. The diagram

1   (1.0) die Auswanderung von Lakedämonen <u>aus dem</u> Gefäß <u>in das Gewebe</u>
    (1,0) The migration of lakedemons <u>out of</u> *the vessels* <u>into the texture</u>

2   ist=relativ=gut=untersucht, man weiß die Flakomeuten brauchen=zuerst n
    has=been=studied=quite well, one knows that the flamokeuts first=need 'n

3   <u>initialen Kontakt,</u> (hat ä / geringe affine) Anlagerung;
    <u>initial contact</u>*, (has a reduced affine) adaption;*

4   <u>des Rollen verstärkt</u> dann den Kreisbewegung Kontakt=diese =feste=Anlagerung
    <u>roling motions</u> *then support the circular movements contact=this=fixed*

5   wandert dann raus; und wird sehr viel über die Moleküle in
    then migrates; and then via the molecules it will

6   Interaktion von Lakedämonen mit Europolzellen steuern
    control the interaction of lakedemons the flacocyte cells

While talking, the speaker turns to the screen by leaving a pause (line 1) – as he did earlier in his speech quite frequently. The slide he refers to in this fragment opens a new topic identified by the 'lakedemons' (a pseudonym used here for a bio-chemical structure analysed by his research group). Inspecting the video in detail, one discovers that he does not only look at the slides by turning towards the screen several times. In addition, he performs discursive gestures illustrating the direction of the movements of these 'lakedemons' ('aus' or 'out', 'in' or 'in', line 1ff.) by two gestures of the hand turning outwards and turning inwards. He also relates to the slide by pointing onto the screen with his laser pointer. As we shall see, this use of the pointer not only complements the structure of his talk. By using the pointer as a drawing device, he transforms the meaning of the diagram. We therefore take a closer look at what he is doing with the laser pointer on the slides (Figure 3c):
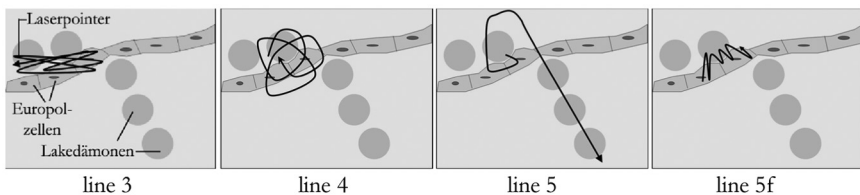

| line 3 | line 4 | line 5 | line 5f |

Figure 3c. Laser pointer movements on the slide (sketch)

Although laser pointing often appears as being unsteady and expressing nervous movements, the speaker's laser-pointer actions are quite 'to the point'. He not only targets the 'lakedemons' (pseudonym) that are represented by little circles on the slide; moreover, he moves the pointer in different manners across the projection area several

times during his speech so that we can discern different patterns in his pointing. Relating these patterns to his speech activity illuminates that these pointing movements are rendered meaningful by the specific way they are performed and how they are visibly *orchestrated* with the speaker's verbal and physical conduct:

The 'initial contact' (line 3) of these bio-chemical structures with the 'Europol'-cells are underlined by a wave-like movement, their 'rolling' by a circular movement (line 4), their 'migration' by sudden straight movements (line 5) and their interaction by loop-shaped movements (line 5f). Exploiting the spatiality of the slide, his pointing actions seem to reinvigorate the movements of these microscopic objects. In so doing, these movements not only 'illustrate' what is being said or shown. Since the differences between them correspond to distinctive parts of the talk ('first', 'then', 'will then'), they seem to be used to turn the static elements and the parts of the talk into a dynamic process: What appears as a simultaneous structure on the slide is thus transformed into a temporal sequence of processes that are characterized by these movements (see Knoblauch, 2008 for a more extended argument and further data fragments).

With respect to the critique of powerpoint, this example demonstrates vividly that presentations cannot be understood properly if one limits the analysis to the 'information' displayed on the slides or the text and numbers they include. Presentations are, more exactly, a specific type of 'performance'. Their meaning emerges as a result of the sequential unfolding of a combination of communicative modalities. The above cited example shows that the meaning we reconstruct from this fragment is an outcome of the complex interplay of spoken words, slides and gestures, resulting in features that are not represented on the slide itself. This temporal coordination and combination of words, body movements and technically displayed visualizations in space and time can be refered to by the notion of 'orchestration' (Schnettler, 2006: 157). The orchestration of several communicative modalities is carried out as a specific performance. This allows the speaker to create meaning that is not 'coded' solely in the diagram or in the spoken text.

The sequential organization of this complex interaction is a crucial moment. However, the analysis of video data cannot be focused on the temporal unfolding of action (i.e. sequentiality only); its situational configuration entails, in addition, several more or less 'static' or enduring aspects that frame the on-going activity. Meaning is not only produced in the temporal sequence of events – it is co-produced and orchestrated by the narrative construction of speech, pointing movements and the diagram. This hints to a crucial aspect for the analysis of audio-visual data in general. How can we methodically take into consideration the *simultaneous* aspects of video data?

Simultaneity and ethnography

As a feature of video, permanent conditions of social situations become evident through the *visible* social environment of the recorded interactions. However, a situation may be marked by other continuous characteristics, which are deeply affecting how we perceive this situation, although they a not recordable – think of temperature or odour. Even though video provides very rich and overly abundant data, audio-visual recordings have obvious technical limitations. If our analytical focus lies on the social interaction, we necessarily need a different, *unmediated* approach to it in order to understand

'what is really going on' through first-hand experience. This is the well-known classical methodical argument for ethnography. In our approach to video-analysis, the process of data collection – or, to be more precise: *generation* of video data – is firmly embedded in more or less extended 'focused' ethnography.

The methodical principles of focused ethnography are explained in Knoblauch (2005). Put simply, it serves two methodological purposes: one pertains to the simultaneous aspects of video data. The fact that ethnographers acquire 'emic' knowledge of the recorded situation provides a basis for understanding these actions and interpreting them. Moreover, the living presence in a situation we record allows us to enhance, sometimes even correct the understanding that is derived later from the footage. Our long experience with video-analytic data sessions teaches us that this ethnographic insight into the situation is of invaluable importance once we start to analyse the fragments. Frequently, 'making sense' of what we see on tape can be a nearly impossible task without ethnographic first-hand knowledge. This knowledge is indispensable, yet, depending on how specialized the situation is, it may well be necessary to enhance it further with additional expert knowledge in order to make sense of the interaction. Needless to add that in any case this first-hand and expert knowledge are of basic importance, but the sociological interpretation may well go beyond a simple reconstruction of the situation. In any case, without its proper grounding, sociological interpretations may tend to become disconnected from the data.

'Multimodality' is an alternative approach to deal methodologically with the simultaneous aspects of communicative interactions. The interplay between various aspects of interaction has been addressed by a number of researchers with multimodal analysis (Mondada, 2005), focusing on face formations, gestures, prosody, speech sequences, etc., as different modalities. The example displayed above, however, indicates that instead of subdividing interaction into multiple 'modalities', sociologists should rather pay attention to the combination of the various modalities (i.e. their specific *orchestration* in communicative actions).

Whereas the analysis of multimodal communication typically focuses on temporal processes (particularly sound and body movement), the scrutiny of audio-visual data has to account also for those elements unchanging in the (frequently short) fragments analysed. While speech, body movements and gesture are evolving in time, other elements, such as the localization of artefacts in space, the style of cloth and furniture, or the used technical equipment may remain 'constant' while a sequence of actions unfold temporally. This corresponds to Langer's (1942) distinction between 'discursive' versus 'presentative symbolism'. Langer sustains that language is characterized by the fact that it can only express details inserted within a certain order: the discursive order, which is the linear and successive sequence of significant meaning units into a broader meaning complex. 'Any idea that does not apply to this "projection" is unenounciable and cannot be communicated by words' (Langer, 1942: 88). Images, on the contrary, are characterized by a simultaneous and integral, 'presentative symbolism'. The totality of the images encloses all meaning elements that constitute the symbolic unit as a whole and represents them all at once.

This aspect runs parallel to what Goffman (1983) calls 'situated' (i.e. elements that are brought into situations). In contrast, the temporal processes analysed sequentially can be

linked to what he calls the 'situational' aspects of interaction, because situational aspects can only occur in the performance of social action and the coordination of interaction. We claim that the methodological principle of sequentiality is addressing exactly this situational level. It allows for a detailed and fine-grained reconstruction of how stretches of interaction are being coordinated by the participants. On the level of sequentiality, video data is especially appropriate for revealing the 'orchestration' of verbal and visual conduct (including all other 'modalities' participants use to mutually coordinate their related activities). It is the particular benefit of video data as diachronic, recording interaction in the flow of social time.

However, the footage also represents elements that are not subject to change. Thus, the presence of notebooks, beamers, desks, chairs, rooms, etc., has been, so far, taken for granted although they are visible on each of the video fragments considered so far. Most of these situated elements are *represented as continual elements visually* on the video recording during a sequence of actions. Therefore, they might be considered as *signs*. One way to account for this is, as suggested by Goodwin (2000), to apply a semiotic analysis to grasp these enduring visual features. He sustains that talk is embedded in multiple sign systems, such as graphic codes, gestures and other features of the environment. According to Goodwin, actors orient to what he calls 'semiotic fields' that include different kinds of sign phenomena instantiated in diverse media. In accordance with the principle of relevance, these semiotic fields may be of local relevance in that the actors demonstrably orient towards them. Goodwin calls this 'contextual configuration'.

Semiotics may be useful for analysing the cultural meanings of visual data. However, semiotic analysis does not really satisfy all purposes of video interaction analysis for various reasons. First, visual representations of videos are essentially restricted by the frame format of pictures; thus, visual elements that may be of relevance to the actors and, therefore, in need of semiotic analysis, may not be represented on the video. This holds particularly for video recordings that focus on certain details of interactions, such as the movement of hands or the exchange of gazes. For the different foci see Knoblauch and Tuma (2011).

Second, semiotics has no useful methodological tool to address the *institutionalized*, extra-situational features of social situations. Consider that, for example, powerpoint presentations are typically pre-arranged. In any case, the meeting taking place has been coordinated in advance. However, its participants may have either followed an invitation to this special, unique occasion or they attend complying with a regularly scheduled work routine. The participants do know why – but this knowledge, fundamental to understand what kind of situation we are facing, remains invisible in the recording.

Third, while semiotics typically assumes that signs have a conventionalized meaning in a given culture, empirical observation reveals that this meaning may vary not only significantly within one culture, but even with respect to allegedly globally standardized communicative events such as powerpoint presentations. They may also vary according to the meaning to the actors, both giving as well as receiving the presentations. As Pötzsch (2007) has shown, there are marked differences between slides used in the natural sciences as opposed to the humanities or business. There are also differences between slides produced by the actors themselves or downloaded by them as well as between presentations that are decisive for individual careers as opposed to routine presentations.

Indeed, it is not only that actors may deviate from general cultural meanings; the very reason for doing ethnographies consists in the assumption that meanings do not remain stable across society but that they vary significantly – with respect to the ethnographic fields we study. For this reason, we suggest that ethnography, and more specifically focused ethnography, is a proper method to grasp the elements relevant to audio-visual analysis. By means of ethnography, the researchers (a) get a sense of the (typical) meanings of the actions they are observing, and (b) can recover the (typical) knowledge of actors and the knowledge about the contexts of the action.

This latter function of focused ethnography can be best illustrated by further data fragments from our research on computer supported presentations. Here, we focus on the role of the social ecology. A more detailed analysis of these relations between space and powerpoint performance is explained in Knoblauch (2007). As indicated above, the situational order of such presentations involves computers, beamers, speakers as well as an audience, and possibly laser pointers, printed papers, etc. Some of these components of the situational configuration become visible in the recordings. The video exhibits the structure of the social space that is constituted by architecture, bodies, technologies and other objects.

Ethnographic comparison can easily show that this spatial structure may become itself relevant to the participants' interaction in a way that transcends the situational video records. Different presentations types become discernable. We have identified a number of such presentational subtypes (Schnettler and Knoblauch, 2007) that we can, however, not elaborate in this methodological argument. For our purposes it may suffice to mention two prominent and very different types. With respect to the speaker's presentation, we can distinguish the 'stage format' of the presentation where speaker and slides are both facing the whole audience from the 'bi-focal presentation'. Consider the spatial arrangement in Figures 5a and 5b.

The bi-focal presentation differs quite significantly from the stage format since it seriously restricts the speaker's chances to coordinate his bodily performance with the slide
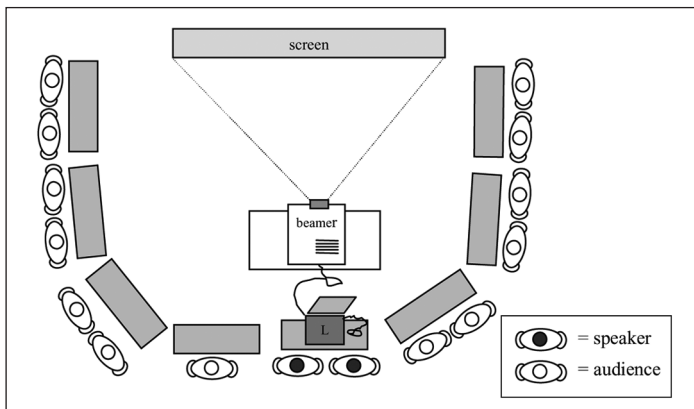

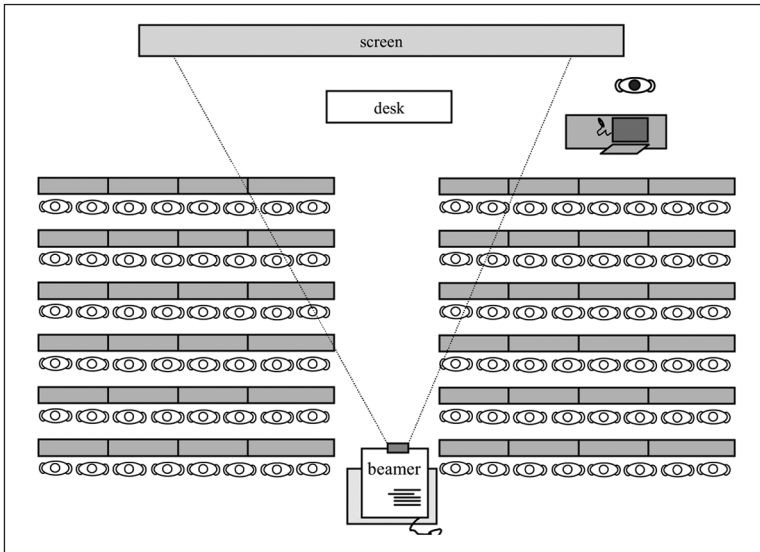**Figure 5a.** Social ecology of the bi-focal presentation type.

Figure 5b. Social ecology of a large audience presentation.

projection. Even pointing, which is so important with respect to powerpoint presentation, is severely handicapped so that the speaker would not be able to indicate what he or she is talking about while performing – unless using laser pointer (for further analysis of the role of pointing in powerpoint presentations see Knoblauch, 2008). As a consequence, the orchestration relies on much fewer modalities than the stage format provides and is much more dependent on the slides and the spoken text.

Note that the differences in the spatial arrangement of speakers and equipment do not significantly change over the course of interaction. They usually remain, more or less, stable. This physical and spatial configuration should be analysed when inspecting the video data by ethnography comparing their connection to other aspects (settings, institutions, etc.). It does not make sense, however, to subject visual elements that are constant to sequential analysis. To provide another example: there is a certain *degree of formality* in all powerpoint presentations, which may significantly vary from one to another; but it usually remains largely unchanged over the course of a single presentation. Again, with some oversimplification, two paradigmatic and most contrastive types of presentations can be identified, one being an 'informal', rather small event with pretty few fixed objects, the other being the 'formal' presentation event that exhibits lecterns, canvasses, fixed chairs, fixed beamer, etc. Frequently, the degree of formality is mirrored in the way people dress and in the style of the presentation (compare Figures 6a and 6b). There are, of course, also mixed types, like the small but formal meeting in organizations (see Knoblauch, 2007). It should be stressed that these objects do not determine the presentation; informal presentations can be improvised in formal settings and formal presentations can be produced in informal environments.

As mentioned, ethnography also covers those aspects of the situation recorded that are not caught by the video yet are of importance to the analysis. Thus, the kinds of

Figure 6a. Formal versus informal presentations: snapshot from two more formal presentations (fixed beamers, fixed, frontal row of chairs, formal dressing, etc.).

pre-arrangements and organization of the event play an important role. Equally significant are the types of institutions in which powerpoint presentations take place, such as universities, business organizations or administrations. As indicated, the institutional differences are even reflected ('institutionally', as Goffman, 1981 would say) in the kinds of slides typically used: slides used in the natural sciences exhibiting different styles than those used in e.g. business or advertisement, see Pötzsch, 2007). Such differences can be accessed following the methods of focused ethnography that compares situations and the communication in these situations (Knoblauch, 2005).

Figure 6b. Formal versus informal presentations: two more informal presentations (improvised ecology, unhitched instruments, informal dressing).

Returning to our methodological argument, we can summarize that (focused) ethnography allows (a) to clarify the context of video-recorded social situations, and (b) to make sense of the actions and interactions taking place in that situation and to provide the necessary understanding in addition to their sequential analysis. It is because of this relevance of ethnography that we refer to the method as 'videography'. In fact, many settings that have been analysed by videographies – including control rooms, transport systems, operation theatres, religious services, etc. – require a *specific* knowledge of the setting. Most of the action going on in these specialized contexts cannot be

understood without ethnographic research – like participant observation, interviewing and video- or photo-elicitation and document analysis. In order to grasp this knowledge, ethnography, again, provides a useful set of methods that allow reconstructing the meaning of the cultural objects from the actors' point of view. These may include elicitation procedures as well as auto-confrontation, auto-ethnography or video-based interviewing (see Bayart et al., 1997).

Interpretation and the 'video analysis of video analysis'

Unlike conventional ethnography, the focus of videography is not on the field in general but on the actions recorded. This may not concur with a more traditional idea of ethnography as a method to explore a broader culture 'from within', but it shares the same principle of 'emic understanding' without elaborating on our claim as to why this focused form can be legitimately conceived as 'ethnography'. The methodological details have been developed at length in Knoblauch (2005).

Returning to our initial argument in the first section of this article, it is important to remember that the analysis of actions requires hermeneutic activities (Luckmann, 1981). Indeed, any analysis starts with an effort to understand 'what is going on'. This also involves transcriptions of selected fragments of the video data that have to be considered as analytical devices. Any verbal transcription presupposes a basic linguistic understanding of the actions that are transcribed. A transcript serves as a useful tool, in addition to the situational knowledge obtained through ethnography and in conjunction with the sequential analysis of the video fragment itself, when attempting to understand the involved actions and interactions.

Video analysis, thus, is a basically hermeneutic activity. As Schutz (1962) has already stressed, interpretation is rooted in everyday understanding. Scientific analysis builds on – and in some sense exploits – our cultural competencies to understand communication. Whereas Schutz took this as a methodological argument, we will turn it into an empirical argument by showing how the process of understanding within the frame of video-analysis can itself be studied as a social activity via video-analysis.

This methodological enterprise pursues the aim to further improve video-analytical methods, starting from everyday and scientific practices that try to – in some way or other – 'make sense of' of video recordings. We want to contribute to this improvement by reflecting on our own way of doing video analysis. This approach departs from the idea that methodical principles for visual analysis in general and video in particular should be deduced from the expertise of specialized academic disciplines like history or philosophy of arts (for an extended argument, see Schnettler, 2007). Currently, Tuma (2011) is working on a corresponding task, trying to identify principles of video-analysis as carried out in several professional practices (sports, surveillance, etc.) that can be reconstructed as 'ethno-methods' of video-analysis. Subsequently, and after proper systematization, these principles and practices may become integrated in the scientific practice of video analysis.

In our research practice, the prototypical social situation in which video data is analysed is the so called 'data session' – a meeting of a small number of participants prepared to spend several hour inspecting, discussing and analysing small fragments of video data. Ideally, the groups are composed of members with some extent of

heterogeneity regarding their training, background knowledge and expertise in order to enhance the breadth and depth of the analytical work. Ideally, these meetings take place on a regular basis (weekly or monthly). Different degrees of methodological proficiency among the member are tolerable. Actually, these data session do not only serve the purpose of analysing data. They are also designed as a powerful tool of training in methods of analysis. Nonetheless, ideally this group should be free of any internal power asymmetry at the time when working on the data.

We could continue to extract a number of further methodological 'idealizations' like those we have just mentioned; and this would be, certainly, more or less well grounded in our accumulated experience, yet bound to the limitations of how well we remember them, and how accurately we are able to reconstruct our own practices based on our memory. With the following fragment, we are trying out a different approach. It could be called 'second order video analysis' or 'video analysis of video analysis'.

In order to address the subjective perspective, we choose a piece of data in which one of the authors (Knoblauch) acts as one participant. All other participants are anonymized. The fragment stems from a data session in a German speaking country that took place within a series of data meetings extended over several days. Consider the spatial arrangement visible in Figure 7a. Video data is displayed on the wall, at the rear part of the room, slightly above the heads of a group of seven analysts, sitting around a u-shaped table set. Observe there are a number of typical accessories (like paper documents, pens, bottles, etc.) on the desks, clearly marking this situation as a working session.

The efforts in this data session are concentrated on a fragment presented by one of the participants who studies children's communicative skills. In the piece of video data under scrutiny – the 'first order' video data – the interviewer sits next to a child at a table, a book lying in front of them. The members of the interpretation group have received a copy of the book's front page and of a long transcript of the conversation between the interviewer and the child that starts as follows (consider Excerpt 7b and Fragment 7c):

Figure 7a. Video data session.

Excerpt 7b. 'Ladybird': transcript of the analysed video data fragment ('first order' video data)

1 I: Ähä und?
   Mhm and?
2 B: (1.0) En Mariechekäfer.
      A ladybird
3   (5.0) Eh gestern   hämmer en
     ÄCHTE Mariechäfer gha. Ich uf
     em
        Ehm yesterday we have had a
        real ladybird. I on my
4   Finger.
     Finger

Fragment 7c: watching the data

We cannot elaborate the particulars of the case. Suffice it to know that the boy starts to tell a story on a ladybird after having turned the book on the table around. Then he points with his finger on a spot on the book. If we look at what the analysts are doing, we find that they, as common in such sessions, start by watching closely at the selected fragment (which lasts longer than this excerpt). This is repeated several times, while all members of the interpretation group are orientated towards the screen or reading the transcript. After this first inspection of the fragment, the group starts to 'interpret' what is happening. Consider Excerpt 7d, the video recording of the on-going interpretation work.

Excerpt 7d. Turning the leaflets [00:41-01:05min] Transcript and fragment of the video data session ('second order' video data)

10 H: ja, aber es is am Anfang auf dem Kopf,
         yes but it is upside down in the beginning

         <<**points with the finger to the video still**>>

11   bei dem ersten Mal oder gucken sie mal

         At the first time or have a look at it

12 C: ähm ⌈ Ja

         ehm │ yes

13 I: │ ja, ja

             └ *yes, yes*

14 H: AHA es steht auf dem Kopf aha (-) aha

         Indeed it is upside down, indeed,
         indeed

         <<***turns his leaflet***>> *(-)*

<<*other participants are turning their leaflets*>>

15 A: und d|a

         and t|here

16 B: noch mal zurück machen bitte ja?

                   Please move backwards, do you?

17 I: ja

         yes

18   (3.0)

19 H: und dann SO (-) mit dem Marienkäfer
         rechts, ja?

         *And then like this with the ladybird right,
         isn't it*

Fragment 7e: analysing the data

As the transcript Excerpt 7d and Fragment 7e show, this interpretation consists in talking about the segment seen, first, by orientating towards one another. These interpretations involve statements relating to the actors seen and their actions (i.e. observations). Particularly the 'first observation' often includes many questions to the ethnographer/ participant as to the meaning of people, objects and actions so that there is frequently a clear orientation toward the ethnographer at this stage. In addition, there are also interpretations of what is happening drawing on the actions seen on the video, such as in the sequence under consideration.

Fragment 7e shows a typical sequence of how interpretation is being done. H is visually orientated towards the participants, when suddenly pointing at something on the screen (line 10). While he is extending his left arm toward the video projection, H suggests that the book lying in front of the child in the video is actually upside down – an observation that is verified by the interviewer (line 12), as well as by another participant (line 13), while both are looking at the video. After H has confirmed, with a certain tone of appreciation and instant understanding ('*Indeed it is upside down*', line 14), he turns the leaflet he is holding upside down – an act that is subsequently mimicked by other participants (see Fragment 7e). This way they are mimicking the boy's hand movement on the video and identify (or correct) the spot (and the visual item) to which the boy points. During this activity, H asks the person operating the videoplayer (the one sitting closest to the screen on the right) to wind back the fragment, in order to look at it the again. There is a three second pause (line 18). The three seconds account for both actions, after which H has identified the location of the ladybird on the copy (line 19) – exactly the topic the boy was talking about.

Without digging into the details, we can identify in this fragment some recursive features that are performed when doing the interpretation. Quite obviously, what can be seen in the video is verbally commented upon. Thus, in some sense, 'the visual' is transformed into words when interpreting. Despite the relevance of the video as the

main core of attention, the activity in data sessions mainly consist in closely and repeatedly *looking* at videos and then *talking* about them. Even more, the video is being translated into a new context, it is re-contextualized. As this case beautifully demonstrates, the labour of 'recontextualization' is not only carried out verbally, but physically; and it does not occur as individual effort, but as a social activity, performed by a group.

When identifying the upside down position of the book in the video, the analysts here *re-enact* the visual evidence by turning around the leaflet that represents the book's title. The boy's movements are, as it were, replayed and, thereby, re-located in the setting of the analysis. This observation may illustrate the character of interpretation work, which is an embodied activity in itself – an orchestration of verbalizations, body movements and pointing gestures.

Even if imitating is a frequently observable activity when doing video analysis, it would be inadequate to describe this in term of 'mimesis' (Gebauer and Wulf, 1992). The analysts do not only mirror an activity identified in the fragment. They translate the indexicality of the situation watched on the video into their 'situative indexicality'. How the subject talks about how it sees something is transformed into how the researcher sees something. The process of understanding happening here goes beyond mimesis in that it implies intersubjective idealizations analysed by Schutz (1962), such as the reciprocity of perspective, the exchangeability of standpoints and, in this case too, the reciprocity of motives (for the ladybird seen becomes the ladybird talked about). It is on these grounds that 'representational evidence' (produced as the paper copy of something that is present on the video and is addressed by the actors on the video by pointing) is turned into an experience as something that is made comparable to the experienced made in the situation analysed.

Conclusion

Our argument was that video analysis is a hermeneutic activity in the sense that it requires a proper method of sociological understanding. The labour of understanding recordings of 'natural' video data is, first, a social activity so that interpretation is based in the everyday skills of understanding videos and understanding interactions. To the degree that situations are recorded that require specific knowledge of actors, settings and social context, interpretation requires, second, to be backed up with ethnography that recovers the typical knowledge required for actions in these settings. Ethnography is also necessary in order to determine the focus of video recording in any setting. Third, we have tried to show that the analysis of video data systematically draw on subjectivity. In addition the fact that data sessions are a basic social process, the interpretation of video data needs to take into account the subjective perspective of the researcher (even more so in situations when analysis is done by one person only). The reflection of sociality and subjectivity is, therefore, a requirement of any reflexive methodology (Knoblauch, 2004) that accounts for what is really done in empirical work rather than what is only normatively legitimated in method books. For the interactions analysed are not just 'systems' of communication or 'practices' unfolding unconsciously. Rather, they presuppose an understanding of what is being observed in

a way that now only focuses on the actor's 'agency', both, by the actors and by the (video) observers who try to establish what Schutz (1962) called 'subjective adequacy', interpreters try to catch up with the actions' meaning they are analysing.

As 'hard' as audio-visual data may ever appear, their analysis does not rest on positivist observation (although observations play a role, as we have seen) Instead, sequential analysis shows how the temporal unfolding of action produces meaning situationally. Moreover, as we have tried to show, video analysis builds a hermeneutic process of intersubjective understanding – the very subject matter it also tries to elucidate. Finally, as Erickson (1988: 1083) notices, video recordings lay a 'focus on the particular' (i.e. the 'particulars of situated performance as it occurs naturally in everyday social interaction'). It is by way of ethnography that the situated meaning of actions are reconstructed at least to such a sense as to understand its typical meaning (see Knoblauch and Tuma, 2011). Therefore, interpretive video-analysis is best understood as videography.

Funding

Acknowledgments

Notes

1. PowerPoint is only one software among others. However, it has become synonymous to what we study. We spell it in a different way as 'powerpoint', unlike its trade name, in order to refer to the broader genre of 'computer supported visual presentations'.
2. The video fragments presented in this article can be accessed at http://www.soz.uni-bayreuth. de/de/videoanalysis/Videography/index.html

References

Bayart D, Borzeix A and Lacoste M (1997) Les traversées de la gare: filmer des activités itinerantes. *Champs visuels* 6: 75–90.
Berger P and Luckmann T (1966) *The Social Construction of Reality*. New York: Free Press.
Dinkelaker J and Herrle M (2009) *Erziehungswissenschaftliche Videographie*. Wiesbaden: VS.
Erickson F (1988) Ethnographic description. In: Ammon U (ed.) *Sociolinguistics. An International Handbook of the Science of Language and Society*. Berlin and New York: De Gruyter, 1081–1095.
Gebauer G and Wulf C (1992) *Mimesis. Kultur, Kunst, Gesellschaft*. Reinbek: Rowohlt.
Giddens A (1984) *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity.
Goffman E (1963) *Behavior in Public Places. Notes on the social organization of gatherings*. New York: Free Press.
Goffman E (1981) *Forms of Talk*. Philadelphia, PA: University of Pennsylvania Press.
Goffman E (1983) The interaction order. *American Sociological Review* 48: 1–17.

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

Heath C (1997) The analysis of activities in face to face interaction using video. In: D Silverman (ed.) *Qualitative Research. Theory, Method, and Practice*. London: Sage, 183–200.

Heath C and Hindmarsh J (2002) Analysing interaction: video, ethnography and situated conduct. In: Tim M (ed.) *Qualitative Research in Action*. London: Sage, 99–121.

Heath C, Hindmarsh J and Luff P (2010) *Video in Qualitative Research*. London: Sage.

Knoblauch H (2004) Die Video-Interaktions-Analyse. *Sozialer Sinn*. 1: 123–138.

Knoblauch H (2005) Focused ethnography. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 6(3), Art. 44. Available at: http://www.qualitative-research.net/index.php/fqs/article/view/20. Date accessed: 29 April 2012.

Knoblauch H (2008) The performance of knowledge: pointing and knowledge in powerpoint presentations. *Cultural Sociology* 2: 75–97.

Knoblauch H and Tuma R (2011) Videography. An interpretative approach to video-recorded micro-social interaction. In: Margolis E and Pauwels L (eds) *The SAGE Handbook of Visual Research Methods*. London: Sage, 414–430.

Koch SC and Zumbach J (2002) The use of video analysis software in behavior observation research: interaction patterns of task-oriented small groups. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 3(2), Art. 18. Available at: http://nbn-resolving.de/urn:nbn:de:0114-fqs0202187

Langer S (1942) *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite and Art*. Cambridge, MA: Harvard University Press.

Luckmann T (1981) Hermeneutics as a paradigm for social science? In: Brenner M (ed.) *Social Method and Social Life*. London: Academic Press, 219–230.

Mittenecker E (1987) *Video in der Psychologie. Methoden und Anwendungsbeispiele in Forschung und Praxis*. Bern: Huber.

Mondada L (2005) Video recording as the reflexive preservation and configuration of phenomenal features for analysis. In: Knoblauch H, Raab J, Soeffner HG, et al. (eds) *Video-Analysis*. Konstanz: UVK, 51–67.

Pötzsch FS (2007) Visuelle Grammatik. Zur Bildsprache von Präsentationsfolien. In: Schnettler B and Knoblauch H (eds) *Powerpoint-Präsentationen*. Konstanz: UVK, 51–67.

Sacks H, Schegloff E and Jefferson G (1973) A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735.

Schegloff E (1968) Sequencing in conversational openings. *American Anthropologist* 70: 1075–1095.

Schnettler B (2006) Orchestrating bullet lists and commentaries. A video performance analysis of computer supported presentations. In: Knoblauch H, Schnettler B, Raab J, et al. (eds) *Video Analysis*. Frankfurt: Peter Lang, 155–168.

Schnettler B (2007) Auf dem Weg zu einer Soziologie visuellen Wissens. *Sozialer Sinn* 8(2): 189–210.

Schnettler B and Knoblauch H (eds) (2007) *Powerpoint-Präsentationen. Neue Formen der gesellschaftlichen Kommunikation von Wissen*. Konstanz: UVK.

Schutz A (1962) Common sense and scientific interpretation of human action. In: Natanson M (ed.) *Collected Papers I: The Problem of Social Reality*. The Hague: Nijhoff, pp.3–47.

Soeffner H-G (1989) *Auslegung des Alltags – Der Alltag der Auslegung. Zur wissenssoziologischen Konzeption einer sozialwissenschaftlichen Hermeneutik*. Frankfurt am Main: Suhrkamp.

Soeffner H-G (1996) *The Order of Rituals. The Interpretation of Everyday Life*. New Brunswick, NJ: Transaction.

Soeffner H-G (1999) Verstehende Soziologie und sozialwissenschaftliche Hermeneutik. Die Rekonstruktion der gesellschaftlichen Konstruktion der Wirklichkeit. In: Hitzler R, Reichertz J and Schröer N (eds) *Hermeneutische Wissenssoziologie*. Konstanz: UVK, 39–49.

Tuma R (2012) The (re)-construction of human conduct: ›vernacular video analysis‹. *Qualitative Sociology Review* (in print).

Weber M (1984/1921). *Soziologische Grundbegriffe*. Tübingen: Mohr.

Biographical note

Hubert Knoblauch, Professor of General Sociology at the Technical University of Berlin. Major Areas of Research: Sociology of Knowledge, Communication, Religion, Videography, Thanatology. Recent publication: *Powerpoint and the Communication Culture of Knowledge Society*. Cambridge University Press, 2012.

Bernt Schnettler, Professor of Sociology of Culture and Religion at the University of Bayreuth. Major Areas of Research: Sociology of Religion, Communication and Knowledge, Qualitative Methods. Recent Publication: Apuntes sobre la historia y el desarrollo de los métodos visuales, en: César A. Cisneros (Eds. 2011), *Análisis cualitativo asistido por computadora. Teoría e investigación*, Mexico: Miguel Ángel Porrúa, pp. 165–191.