Methods in Cognitive Linguistics

Edited by

Monica Gonzalez-Marquez
Irene Mittelberg
Cornell University
Seana Coulson
UCSD
Michael J. Spivey
Cornell University

John Benjamins Publishing Company
Amsterdam/Philadelphia
2007

Methodology for multimodality

One way of working with speech and gesture data

Irene Mittelberg

1. Introduction

Multimodality, besides entailing at least two modes of expression, also engages at least two modes of interpretation. For instance, when language and gesture coincide, the body of the speaker is as much involved in expressing thoughts, attitudes, or emotions, as are the linguistic articulators. Simultaneously, while listening, the addressee/observer draws on both her sense of audition and vision to make sense of what the interlocutor is conveying. Compared to the various other types of multimodal communication that have come to predominate contemporary media landscapes, this kind of fusion of two modes of expression, speech and gesture, is perhaps the most basic one. Yet given the complexities of human communicative behavior, researchers wishing to investigate its specific aspects face a variety of methodological questions and challenges. These may range from data access to combining methods that do justice to both the semiotic idiosyncrasies of each modality and to the ways in which the different modes interact in a given context.

This chapter provides an example of the empirical steps involved in working with multimodal data consisting of spontaneous speech and its accompanying gestures. It is meant as a companion to the introductory chapter by Mittelberg, Farmer, and Waugh (this volume), which offers a general overview of approaches to discourse and corpus analysis. It is important to realize that the routine presented below is only one of many ways of transcribing, coding, and analyzing co-speech gesture (see also Sweetser this volume). While special attention is paid to gesture-speech co-expression, transcribing naturally occurring speech data, which represents the logical first step in multimodal data transcription, will also be discussed. Overall, the chapter reflects and advocates the trend, noticeable already for some time in various areas of linguistic inquiry, to account for bodily semiotics in the recording and analysis of spoken discourse (e.g., Cienki 1998; Clark 1996; De Stefani fc.; Duranti 1997; Goodwin 1986; Kendon 2004; Lantolf & Thorne 2006; Levinson 2003; McNeill 1992, 2005; Müller 1998, 2004a/b; Schegloff 1984; Streeck 1993, 1994; Sweetser 1998; Tao 1999).

Despite this substantial body of work and the ubiquity of gesture in human communication, manual gestures and other types of bodily semiotics, such as gaze and head

movements, have received considerably less scholarly attention than language itself. One of the pragmatic reasons for this is that it took the advent of film and video technology to actually capture ephemeral body movements evolving in space and time. While portraying visual imagery, gestures are not simply visual signs, but dynamic visuo-spatial, or "motor signs" (Jakobson 1987:474ff.). Spontaneous co-speech gestures are, for the most part at least, polysemous signs that are inseparable from the human body and its physical and sociocultural environments. They thus differ considerably from visual signs that one can contemplate and analyze also after they are taken out of their original context of production. If we, for instance, compare gestures to static sculptures or pictorial signs captured on paper or canvas, we realize that they are extremely fluid figures vanishing as quickly as they take shape. Only with the help of an artificial medium can we truthfully represent and examine co-speech gestures as holistic gestalts of locally-anchored meaning-making.

Having access to advanced video technology has evidently opened up whole new possibilities to the language analyst whose work with usage data had usually been confined to audio recordings. These new ways of looking at language might not be exactly comparable to the revolution ignited by the invention of motion pictures and the altered perception of the world it engendered. Yet the evolution of the film medium has been a precondition for the study of gesture as it is practiced in today's digital era: moving images have made the reproduction of bodily motion in real-time possible, and constitute valuable visual resources. Crucially, video materials facilitate the repeated viewing of the same motion/speech event, even in slow motion and frame-by-frame in frozen bits and pieces, which is crucial for gesture analysis. The opening paragraph of Solnit's book on Eadweard Muybridge, who in the late nineteenth century succeeded in capturing high-speed motion photographically, thus preparing the ground for movies in the U.S., succinctly depicts the tight interrelation between perception, media, art, science, and consciousness:

> In the spring of 1872 a man photographed a horse. The resulting photograph does not survive, but from this first encounter of a camera-bearing man with a fast-moving horse sprang a series of increasingly successful experiments that produced thousands of extant images. The photographs are well known, but they are most significant as the bridge to a new art that would transform the world. By the end of the 1870s, these experiments had led to the photographer's invention of the essentials of motion-picture technology. He had captured aspects of motion whose speed had made them as invisible as the moons of Jupiter before the telescope, and he had found a way to set them back in motion. It was as though he had grasped time itself, made it stand still, and then made it run again, over and over. Time was at his command as it had never been at anyone's before. A new world had opened up for science, art, for entertainment, for consciousness, and an old world had retreated farther. (Solnit 2003:3)

Working with co-speech gesture provides insights into these dimensions of the new world, especially in terms of temporality and motion through space. Indeed, while closely watching video sequences, the analyst might find herself mesmerized by the poetry arising from the exactly timed interplay between rhythmic imagery and speech prosody (cf. McNeill (2005) and Furuyama (2000b) on the poetic function, Jakobson (1960) in gesture). At the same time, working with video data puts additional demands on the researcher who needs

to be both a careful listener and an astute observer. Gesture studies thus do not only put the speaker's body into the picture, they also engage the linguist's eye, behind the camera and in front of the computer screen (cf. Goodwin 2001; Settekorn 1993, 1996, 2003). Investigating linguistic and visual action in this way is not without theoretical repercussions. It lends a certain dynamism to language study which is also reflected in views of language and grammar that bring to light the dynamic side of cognitive and semiotic processes (e.g., Fauconnier & Turner 2002; Jakobson 1987; Jakobson & Pomorska 1980; Hopper 1998; McNeill 2005; Müller 2004b; Peirce 1960; Waugh et al. this volume).

As is the case with approaches to discourse concerning the verbal side of communication, there exists to date no unified empirical method used for gesture analysis. This lack of methodological unity can in part be explained by the fact that gesture researchers come from a variety of disciplines with their own prominent research questions and methodological traditions: among others, psychology, anthropology, linguistics, human development, and communication. From a psycholinguistic perspective, methods of coding and annotating co-speech gesture have been primarily shaped by McNeill (e.g., McNeill 1992, 2000, 2005; McNeill et al. 2001), Duncan (Duncan 2003; McNeill 2005), and Goldin-Meadow (2003) (additional approaches will be discussed below, see also Sweetser this volume; and Mittelberg et al. this volume).[1]

The methodological approach presented here is based on Mittelberg's (2006) study of metaphor and metonymy in gestural representations of grammar. It has particularly been inspired by the methods of transcription, coding, and analysis developed by members of the McNeill Lab at the University of Chicago (McNeill 1992), Cornelia Müller (1998, 2004a) at the Freie Universität Berlin, and Rebecca Webb (1996) at the University of Rochester.[2] The corpus built for the purpose of this research comprises twenty-four hours of naturalistic academic discourse and co-speech gestures produced by four linguists (all native speakers of American English) videotaped during linguistics courses at two major American universities. Given the constraints on space, we cannot reenact the entire research enterprise in great detail, nor can we go much into theoretical motivations or content analysis. While the empirical methods were developed based on both theoretical interests and the nature of the data, the focus here will primarily be on methodological considerations and practical issues of general interest. The intent of the chapter is to convey an idea of some of the decisions one needs to make, along with the different phases of the necessary hands-on work. Along the way, I will point to additional options and more comprehensive sources that might be useful and inspiring to anyone who is considering the option of collecting and exploiting multimodal usage data. Also, the procedure

1. For a concise overview of the field see Kendon (1997); for more comprehensive accounts see Calbris (1990), McNeill (1992, 2000, 2005), Müller (1998), Streeck (1993, 1994), and the recently published monographs by Beattie (2003), Kendon (2004), and Müller (2004b). The interested reader is referred to the journal GESTURE (A. Kendon & C. Müller, Eds.) and to the website of the *International Society for Gesture Studies*: www.gesturestudies.com. For information on the *Berlin Gesture Center* visit www.berlingesturecenter.de.

2. For coding and transcription conventions see McNeill (1992, 2005); an up-to-date gesture coding manual, prepared by S. Duncan, can be found on-line at www.mcneilllab.uchicago.edu

I will walk the reader through here is not to be regarded as completely well-rounded from a methodological point of view. As one often experiences with this kind of work, what works and what doesn't work becomes evident during the process or even after the fact. So, although the methodological approach has been adjusted and refined a number of times, it still represents a work in progress. The reader is invited to evaluate the methods developed for the study discussed below and to determine what appears to be relevant in the context of her or his specific research interests.

The chapter is structured as follows: after addressing some preliminary issues regarding research interests, genre, and data elicitation (Section 2), I will describe procedures of video recording and editing (Sections 3 and 4), discourse transcription (Section 5), and annotation of gesture-speech synchrony (Section 6). Finally, coding parameters used to describe physical gesture features (Section 7) as well as ways of accounting for semantic and pragmatic functions of gestures will be discussed (Section 8).

2. Preliminary considerations: Motivations, subject matter, and genre

Before setting out to collect data, the researcher obviously needs to carefully delimit the domain of inquiry. One source of motivation could be to find non-linguistic evidence for conceptual phenomena one has worked on before, such as metaphors for specific ideas or emotions. This was the case for the present study, which turned out to be a continuation of the author's previous work on linguistic and pictorial metaphors for grammar (Mittelberg 2002). One might also get inspired by existing gesture research: gesture and language development (Goldin-Meadow 2003; McNeill 1992, 2005), second language acquisition and teaching (Gullberg 1998; Lantolf & Thorne 2006; McCafferty 2004; Negueruela et al. 2004), conversation analysis (e.g., De Stefani fc., Goodwin 1986, 2001; Schegloff 1984; Streeck 1993, 1994), and (cross-)cultural investigations (e.g., Efron 1972; Kendon 1995, 2004; McNeill 2005; Müller 1998) are only some of the manifold possibilities. Within these areas, one might consider investigating the different forms and functions of one particular gesture (cf. Müller's (2004a) study of the palm-up open-hand gesture), a gesture family (cf. Kendon 2004), or pointing practices (Fricke 2002; Kita 2003). Other possibilities include micro-analyses of the close relationship between linguistic and gestural co-expression of ideas in an unfolding discourse (Kendon 2000; Kita 2000; McNeill & Duncan 2000; McNeill 2005; McNeill et al. 2001). These decisions will determine not only the environment one needs to explore, but also the kind of genre the data will represent. It will also affect the choice of parameters and the level of detail that will ultimately shape the ensuing procedures of transcription and analysis. One should thus keep in mind that genre and context strongly influence discourse pragmatic factors that in turn are likely to motivate a speaker's linguistic choices and the kinds of gestures that will be made.

A look at previous gesture research attests to the tight interrelation of these factors (subject matter, genres, context, communicative behavior, etc.) and conveys an idea of how they are linked to different methods of eliciting and collecting multimodal data. One can broadly distinguish between naturalistic environments (e.g., authentic talk-in-interaction), experimental conditions, and quasi-experimental conditions.[3] With respect to the latter method, McNeill and colleagues based a significant portion of their investigations on particular visual stimuli, e.g., films and animated cartoons (such as "Canary Row"; see McNeill 1992, 2000, 2005; McNeill & Levy 1982). Immediately after watching the film, participants are asked to recount the story from memory to a listener. This kind of storytelling technique, based on the same visual stimulus (without much verbal content) for all participants, has several advantages: it can be used with speakers of different age groups, including children, speakers of different languages, speakers with certain impairments (e.g., brain injuries), etc. It also provides a common denominator in terms of visual imagery, semantic content, sequence of events, and narrative structure. This work has especially shed light on iconic gestures that render the speakers' mental images of concrete objects, actions, and settings as they were seen in the cartoon. Being aware of what the participants try to convey in words and gestural imagery, the researcher does not rely exclusively on the speech content to determine the meaning of a gesture. Moreover, she can identify patterns within and across the different renditions of a single scene. When comparing narrations performed in different languages, this sort of data allows one to discern tendencies in the distribution of semantic features across modalities, thus revealing aspects of information management and how gestural action seems to be intertwined with the phrasal organization of the concurrent speech (see also Kendon 2004: 113). For instance, there is interesting work on the ways in which the various semantic aspects of complex motion events (involving path and/or manner) are encoded in speech and/or gesture. These kinds of investigations have yielded compelling insights into typological differences and linguistic relativity (cf. Duncan (2003) on verbal aspect; cf. Kita & Özyürek (2003), Lantolf & Thorne 2006; McNeill (1992, 2000, 2005); McNeill & Duncan (2000); and Müller (1998) on gesture research illuminating aspects of thinking-for-speaking (Slobin 1987, 1996) and different patterns of speech-gesture synchrony in verb-framed vs. satellite-framed languages (Talmy 1985).

Narratives elicited in quasi-experimental settings such as the one described above are an especially well-represented genre in gesture research. Such set-ups allow for the control of environmental factors such as physical setting and participant constellation. Participants are typically asked to sit down in a chair in front of a dark, uni-colored background so that arm and hand movements can be easily discerned when analyzing the video data. As pointed out earlier, there is also a fair amount of work done on different types of conversational data (e.g., Müller 1998, 2004a/b; Kendon 2004; Sweetser this volume; Streeck 1993; Tabensky 2001). Except for an increase in speaker number, the speaker constellation in talk-in-interaction can be similar to the one just described, i.e., in that the interlocutors are seated and do not move around. However, the range of potential research contexts is obviously vast; there is a host of scenarios (two-party or multi-party interaction) in which multimodal communicative practices can be studied (e.g., service encounters, workplace activities, classroom interaction, etc.). Work exploring diverse in-

3. The focus of discussion will be on discourse-based work; for experimental work on gesture see, for instance, Beattie (2003), Goldin-Meadow (2003), Emmorey & Casey (2001), Parrill (2003), and contributions in Emmorey & Reilly (1995) and in McNeill (2000).

teractional settings has shown that the activities in which participants engage shape the interplay between language and bodily communication in a dynamic fashion (cf. contributions in De Stefani fc.; Streeck 2002). Yet different dynamics can be observed in small group discussions centered around situated cognitive activities such as solving math problems (Goldin-Meadow 2003; McNeill 1992; Smith 2003), reviewing architectural models (LeBaron & Streeck 2000), interpreting archeological excavations (Goodwin 2003), giving presentations in science classrooms (Kress et al. 2001; Ochs et al. 1996; Roth 2003), or giving instructions in expert-laymen exchanges (Furuyama 2000a; Haviland 2000; Streeck 2002; Williams 2004).

Expository prose, such as academic talks and lectures about specific topics, is another genre that has received attention from gesture researchers (e.g., Sweetser 1998 this volume; Parrill & Sweetser 2004; Webb 1996) and represents the kind of data on which the present methods discussion is based (see also Mittelberg 2006). When asking the subjects of the study for permission to videotape their lectures (in regards to appropriate protocols in the use of human subjects for data collection, please see Gonzalez-Marquez et al. and Wilcox & Morford, both in this volume), the purpose of the project was framed in very general terms and the fact that gesture was of interest was not revealed so that the speakers would talk and act as naturally as possible. The conditions were thus in no way controlled. The focus was on the teacher giving lectures in linguistics courses, and not on teacher-student interaction, which would certainly be a fruitful next step of inquiry. The reason for choosing this focus came from the realization that for an initial study of metaphoric and metonymic gestural representations of grammar, it would suffice (and already be a complex enterprise) to analyze the teachers' discourse and gestures. Introductory courses seemed to be the right level, as teachers could be expected to make pedagogical efforts when introducing new technical terms, concepts, and theories. Also, the courses were selected such that the data would cover, in addition to general grammatical phenomena, several views of grammar and linguistic theory: generative grammar, emergent grammar (in the context of teaching second languages), and relational grammar. Since linguistic theories are themselves built on specific sets of metaphors, and since gestures are assumed to depict aspects of the source domains of metaphorical mappings, it was hypothesized that the framework talked about in a given instance would influence the kinds of gestures produced to illustrate the speech content.

3. Equipment, physical setting, speaker activities, and videotaping

We will now turn to more practical, hands-on issues involved in video data collection. The following technical equipment was used for audio-video recording: a digital camera (Sony Handycam DCR-TRV900 NTSC), a tripod, and tapes (mini DV cassettes). As the camera has an excellent internal microphone, and the classrooms in which the courses were held were relatively small, no additional microphones were used. However, to achieve superior sound quality of recordings done in relatively big rooms, it is recommended to use wireless microphones and ask the speakers to wear them on the body.

In teaching contexts, speakers usually move about the classroom, write on blackboards, whiteboards, or overhead transparencies, point to information on boards and screens, interact with the audience by turning and/or walking towards students who ask questions, and so forth. These practical issues are not trivial, for when it comes to collecting data, factors such as physical environment, speaker constellation (number of speakers, actions they perform, etc.), and the use of artifacts and space will influence how the video camera needs to be set up in order to ensure that all the physical elements, and especially the speaker's gestures, will be captured as completely as possible. In the case of the study reported here, the video camera was mounted on a tripod placed in the back of the classroom, usually in the middle of the back wall. This choice was motivated by two considerations: first, the full range of potential speaker movements needed to be covered, and second, the investigator tried to be as unintrusive as possible. Once the taping is completed, it is recommended to make back-up copies of each tape and work with the copies, not the originals. Importantly, using digital video technology has the advantage that the data can directly be transferred into the computer to cut and edit the sequences of interest (see Goodwin 1993 for more details and helpful tips regarding video formats, taping procedures, quality of sound and lighting, labeling and storing tapes, software, etc.; see also McNeill et al. 2001 and McNeill 2005 for an example of work with advanced, semi-automated technology).

4. Assessing and editing video data

In a first approach, one might want to view the data several times in order to get an idea of the idiosyncracies of each speaker's linguistic and gestural expression (e.g., is she or he left-handed or right-handed?), the general speech content (in the case of this study: morphology, syntax, etc.), the use of tools in the environment (e.g., overhead, blackboard, chalk, markers, pointers, etc.), and other factors of interest. For purposes of documentation, it is useful to keep a tape content log with information about what each tape contains and other aspects one deems important. Such a log differs from a transcript in that its primary function is to provide a table of contents for each tape, indicating where (using time stamps) specific episodes can be found and whether/why they are of interest. This comes in handy if, in the course of the analysis, one notices particularly interesting things that did not seem to be relevant at first. A content log will also help the researcher to quickly locate specific instances in the entire corpus.

After preliminary data screening, the investigator might choose to do a micro-analysis of a single sequence, or to work with a number of segments representing different speakers, examining the occurrence of one or several gestures across speakers, subject matters, or contexts. Opting for the latter approach, the corpus was assessed from a thematic point of view, selecting and capturing episodes in which gestural representations of grammatical phenomena occurred. Gaze, facial expressions, and movements of the head and torso were not taken into consideration. The goal was to determine how speakers linguistically and gesturally represented particular linguistic units (morphemes, words, phrases, etc.), categories (verb classes, semantic roles, etc.), and structures (clauses, sentences, etc.), syn-

tactic operations (active-passive transformation, subordination, reiteration, etc.), as well as theoretical views of language and grammar more generally.

The software used to edit the video material was the professional editing program FinalCutPro (designed for Macs). After marking the starting and exit time code of each segment, the selected sequences were cut, captured, and saved as separate files (altogether approximately 120 clips), with each clip named according to the speaker and grammatical point talked about in the segment. When editing the clips, one needs to take care not to cut off a speaker in the middle of a sentence and also to capture gestural movements from their onset all the way to their completion, trying to avoid making a cut before the hands fully retract to a rest or neutral position. Working with individual video files has the advantage that copies can be easily made, and that the clips can then be categorized according to speaker, content, gesture type, underlying metaphorical concept, and so forth. In order to facilitate data access for the (often traveling) investigator, advisors, and other gesture researchers interested in viewing the data, all clips (converted to i-movie files) were made available online.

5. Discourse transcription

Transcribing naturally occurring discourse entails many decisions that should ideally, in each case, be guided by the specific research interests and the nature of the data. In essence, the task is to transpose the flow of verbally conveyed information into a visual medium, or record, that one can work on and share with other researchers: "Discourse transcription can be defined as the process of creating a written representation of a speech event as to make it accessible to discourse research" (Du Bois et al. 1993:45). There are various established transcription conventions developed by discourse linguists that may serve as a model (Du Bois et al. 1993, for methods used in the various approaches to discourse analysis see Cameron 2001 and contributions in Atkinson & Heritage 1984; Edwards & Lampert 1993; Jaworski & Coupland 1999). Nonetheless, it is fairly unlikely that one can simply adopt a method that was originally developed for a particular purpose and set of data without making substantial adjustments. In the case of the study discussed here, the outline of discourse transcription provided by Du Bois and colleagues (Du Bois et al. 1993) was adopted and subsequently adjusted. A crucial distinction to make here is to decide what kind of information needs to be represented in addition to the speech content (e.g., pauses, intonation, lengthening, non-vocal noises, etc.) and what kind of information seems to be irrelevant (e.g., detailed information on primary and secondary pitch). The reasons for deploying this particular convention were that the investigator had previously used it to transcribe speech events and that it is widely used, in various adaptations, by conversation analysts and gesture researchers (e.g., the McNeill Lab).

When deciding which transcription method to use, and how to modify it, one of the central questions concerns the degree of detail needed in order to account for what is going on in the data. Generally, one distinguishes between *broad* and *narrow* transcriptions. According to Du Bois et al. (1993:46), *broad* transcriptions usually contain the following types of information: topic of a segment, speaker labels, time stamp of at least the beginning of the segment, and the words spoken (including all truncated words, false starts,

self-repair, and vocalizations such as *um, mhm, uh, oh*, etc.). The stream of speech gets decomposed into intonation units (Chafe 1987). When working with conversational data, the speakers' turns, as well as speech overlap, also need to be indicated. Broad transcriptions additionally include intonation contour information (pitch direction, such as rising and/or falling), hesitations, laughter, pauses, truncated words, and uncertain hearings. When doing a *narrow* transcription, the transcriber also includes notations of breathing, accent, prosodic lengthening, tone, and other vocal noises. In view of all the choices one needs to make, it becomes evident that a finished transcript already reveals a lot about the researcher's interests and foci in terms of both theory and analysis (Ochs 1979; see also Waugh et al. this volume).

Whether or not one is interested in investigating bodily communication, it seems worthwhile to exploit the latest recording technologies and videotape talk-in-interaction, or any other kind of communicative event one wishes to analyze. Visual information about the physical setting, environmental surroundings, the speakers' postures, gestures, facial expressions, lip movements, and the object-oriented actions and social interactions participants engage in facilitates the process of transcribing speech considerably. With the help of visual cues, one can get a better grasp not only of what is actually happening in a given speech situation, but also of the details pertaining to the speech delivery. When transcribing spoken discourse solely from audiotapes, the lack of visual input often creates the impression that one does not get it all, and can even mislead the processing and understanding of what is said and referred to (cf. Duranti 1997:144f.).

Before looking at a sample transcript from the corpus, it should be mentioned that in the present study only the selected speech/gesture segments were transcribed and that all transcriptions were conducted and/or verified by two transcribers. Each clip was first viewed/listened to in its entirety to get an impression of what was talked about and what actions were performed. The next step then was to listen to the sequence very carefully a few times and write down the utterance word by word. This can first be done without much attention to internal structure and prosody. Subsequently, however, it is necessary to identify intonation contours in order to divide the stream of speech into single intonation units (each of which is represented on a separate line as shown below; cf. Chafe 1987, 1998; Du Bois et al. 1993). The transcript shown below represents only the verbal part of a short sequence in which the speaker explains the fact that there are, from an emergent grammar point of view, no *a-priori* grammatical categories; rather, linguistic form is shaped by discourse function. The name of the sequence is indicated in double parentheses.

(1) ((*no a-priori nouns or verbs*))
... *There*fore, _
you *can't* a-priori for instance, _
de*fi=ne*, \
(...) even a *noun* from a *verb*, \
(...) because, \
(..) *verbs* becomin- become nouns,_
(..) and *nouns* can become verbs, /
(..) depending on *how*_
they are *used* in the *dis*course. \

As the procedure of transcribing cannot be described at great length here, I will provide a list with the notational conventions used in the discourse transcripts (adapted from Du Bois et al. 1993; for a fuller account see Du Bois et al. 1992). The information presented below roughly reflects the different steps that go into creating such audio transcripts.

Intonation units:
- each intonation unit appears on a separate line
- truncated intonation units show a double hyphen at the place where the speaker breaks off the intonation unit before completing its projected contour: - -
- truncated words (end of projected word remains unuttered) within intonation units show a single hyphen: -
- transitional continuity of an intonation unit is continuing: ,
- falling terminal pitch movement in an intonation unit: \
- rising terminal pitch movement in an intonation unit: /
- the direction of the terminal pitch movement is level: _

Pauses:
- short unfilled pause: (..)
- medium unfilled pause: (...)
- long unfilled pause indicating duration in minutes and seconds: (....5)
- the sound quality of a filled pause is indicated in round brackets: (*ehm*)

Vocal noises:
- the symbol for laughter is: @ (one symbol for each 'syllable' of the laughter)

Stress and prosodic lengthening:
- a stressed syllable, word or sequence of words are highlighted in *italics*
- preceding segment is lengthened prosodically: "defi=ne"
- prosody was not a major concern here; see the manual prepared by S. Duncan for transcripts where prosody is registered with changes in font size (McNeill 2005:262, 275ff.)

Unintelligible speech:
- a pair of angle brackets filled with the capital letter 'X': <XXXXX>

It is advised to keep a copy of just the transcribed discourse without any gesture annotation. That way, fine details pertaining to the speech delivery, noticed only later in the process of gesture analysis, can be added. Also, if one later decides to carry out a narrower transcription of a specific episode, one can always go back and add more specific information.

6. Transcribing gesture-speech synchrony

The accurate description of human behavior and social action is admittedly a challenge, and one might wonder whether verbal descriptions are an adequate means of representing visual information and communicative action. Finding a systematic way to document the dynamics of body motion and actions, unfolding in space and time, is already a difficult task. Integrating information on gesture movements with the concurrent speech, prosody and other paralinguistic aspects of talk is even more complex. Gesture researchers have suggested various schemes for how to graphically capture not only the close temporal relationship between speech and co-speech gesture, but also the kinetic features of gestures (e.g., Calbris 1990; Duranti 1997:144–154; Kendon 2004; McNeill 1992: Chapter 3, appendix; McNeill 2005:259–286; McNeill et al. 2001; Müller 1998:175–199, 284ff.; Parrill & Sweetser 2004; Webb 1996). In such comprehensive transcripts, the course of gestural movements (onset, peak, duration, hold, end, etc.) is typically translated into (typo)graphic representations, superimposed onto the written speech transcript (in the form of brackets, different font sizes, highlighting, underlining, glosses, etc.). This process becomes even more intricate if the researcher tries to account for the repeated occurrence of particular gestural features over a longer stretch of discourse, thus identifying gestural patterns of gestural imagery and anaphoric reference (e.g., "growth points" and "catchments" as introduced in McNeill & Duncan 2000; McNeill 2005; McNeill et al. 2001).[4]

Let us now imagine that we are sitting at a computer with a speech transcript in front of us and the corresponding video sequence opened up in the viewer. To arrive at a fine-tuned analysis, we need to view the same clip, and especially chunks of it, over and over again. Using a navigating device (such as Contour ShuttlePro Controller, a tool for video and audio editing) ensures easy going back and forth in slow motion and frame-by-frame, a procedure absolutely necessary to determine where and how exactly a gesture unfolds in relation to the concurrent speech. Before starting the fine-tuned documentation of each gesture, using the already prepared speech transcript, one might want to consider detailing a prose description of the entire sequence. Even though such a description is an optional part of the transcription procedure, it has the merit of requiring the analyst to watch the video data very carefully and to verbalize the global movement of each gesture, how hands interact with artifacts, as well as the ways in which final phases of one gesture may feed into the next one, etc. Capturing these fleeting phenomena in prose has several advantages: prose descriptions complement the obviously extremely condensed information provided by the gesture-annotated transcripts; one can use the descriptions later on to recall and relate the details of a sequence (e.g., when the video material is not handy); and this kind of textual information can build the basis for, and save time regarding, the eventual write-up of the analysis.

The next step in the gesture coding procedure consists of identifying single gestures within a video episode, aiming at accounting for the exact mechanics of speech-gesture synchrony. Due to the fluid nature of gestural movements and their tendency to overlap, it may be difficult at times to determine the boundaries between individual gestural move-

4. In the study discussed here no additional software was used to analyze the speech and body movements. There are, however, various software options that were particularly developed for these purposes; see, for example, the semi-automated methods developed by Francis Quek, provided in McNeill 2005: appendix; McNeill et al. 2001.

ments – that is, to trace a gesture from the moment the articulators (here hands/arms) begin to depart from a position of rest or relaxation until the moment when they return to rest. Kendon (2004: 111) calls such a full "movement excursion" a *gesture unit*. As mentioned above, it requires the repeated slow-motion viewing of chunks of an episode to find the exact moments in speech, i.e., the linguistic segments on which a gesture starts to unfold (the so-called *preparation*), manifests its peak of effort and clearest shape (the *stroke*), may exhibit a sort of hold (the *post-stroke hold*, according to Kita 1993), and then finally comes back to a relaxed position (*retraction* – be it punctual, held, or continued (cf. Müller 1998: 286f. for more variants)). These different movement *phases* make up a *gesture phrase* (Kendon 2004: 110ff.). Importantly, the stroke (if applicable, together with the post-stroke hold) is assumed to express the crucial part of a gesture's meaning, and is the obligatory component of a gesture phrase. Some of the other types of phases mentioned above are optional (for a fuller discussion and transcription examples, see McNeill 1992: 83f., 2005: Appendix; Kendon 2004: Chapter 7). Regarding the question of how exactly the phases of a gesture phrase and speech segments correlate, McNeill (2005: 35) maintains that "[the] *preparation* for the gesture precedes the co-expressive linguistic segment; the *stroke* coincides with this segment; and any pre- and poststroke holds ensure that this speech-stroke synchrony is preserved (Nobe 1996)." As one can imagine, the broad spectrum of gestural actions cannot always be easily accounted for, and one is likely to find hand movements that defy segmentation and classification.

Now, coming back to the issue of annotation, a gesture-phrase may be represented as follows (again, this list does not contain all possible fine-grained dimensions one could account for):

- The linguistic segments that co-occur with a gesture phrase are set off by [square brackets]. Each gesture is assigned an identification number.
- The syllable/word/words on which the stroke of the gesture falls is marked in bold type face. If not otherwise stated, the preparation phase is understood as the movement leading up to the stroke (i.e., from the onset of the gesture (opening square bracket) until the stroke itself (word(s) marked in bold); the retraction phase is understood as the movement between the end of the stroke phase and the rest position (closing square bracket).
- The duration of a gesture hold (cessation of movement while preserving form and location) is indicated by underlining the co-occurring speech segments. Differences in the execution of hold phases may be captured as follows: solid underline indicates no incorporated movement; dotted underline indicates some movement such as superimposed beats.

The transcript below is an example of an annotated transcript. In this sequence, the speaker demonstrates the word order change implied in transforming an active sentence into a passive sentence by referring to the subject-object inversion as a 'flip-flop.' She produces a gesture starting out with both forearms held vertically and aligned with her shoulders, palms facing her body and fingertips pointing straight up. Then she crosses her arms over her chest (G1, Figure 1). The gestures are numbered, and glosses provide a very

brief description of each gesture (abbreviations used below: bh 'both hands,' pcoh 'palms facing center, open hands').

(2) ((*flip-flop passive*))

G1
Arms held vertically, then cross over chest
[The passive basically **flip**flops _

G1
crossing-gesture being held
the subject and object of the sentence, \
a=nd _

G1 still being held
(...) what we find out by forming this particular passive], _

G2
bh, pcoh-box, being held, hands move up and down
is [that the **string** 'John's sister' forms a constituent, \

G2
box-gesture still being held, move up and down
namely the object of the verb, \

G2
still being held, move up and down hands retract to fist
(..) and that's an object noun phrase] in fact. \

Following the stroke of the first gesture (G1), which coincides with the mention of the verb "flipflops," the arms-crossed gesture is being held while the speaker keeps explaining the syntactic operation involved (the underline indicates the duration of the gesture hold). She turns her torso leftward towards the overhead screen behind her, walks briefly towards the screen, then turns back and finally faces the audience again. Subsequently, the speaker's arms open up again and merge right into the next gesture (G2): both open hands being held more than shoulder width apart and facing each other, seemingly holding an (imaginary) object, which here represents "a constituent." There are several dimensions of meaning interacting in this gesture: the physical action represents, iconically, the switching around of two things. At the same time, it stands, metaphorically, for the abstract action of 'flip-flopping' two elements in a sentence. This is an example of a frequently found gesture reflecting the metaphorical concept IDEAS ARE OBJECTS (Lakoff & Johnson 1980, 1999). In the data, variants of this gesture were found to refer to linguistic entities of different complexity (categories, words, phrases, sentences, discourses, etc.) in terms of bounded spaces.

7. Physical gesture features

After focusing on aspects of gesture-speech synchrony and the internal structure of gesture phrases, we will now turn to the task of documenting the physical features of gestural communication. In the gesture literature, the most widely used coding parameters are *hand presence* (left and/or right hand), *hand shape, palm orientation, movement* (trajectory

and manner), and the *location* in gesture space where a gesture is performed. Throughout the multi-layered process of gesture annotation and analysis, there is a danger of mixing the domain of form, i.e., what one actually observes with the naked eye, and the domain of meaning, i.e., the semantic and pragmatic functions a particular combination of physical features may assume in a given moment. Taking the material side of the semiotic processes as the point of departure meant, in the case of the present study, moving from concrete forms (gesture) to abstract concepts and structures (grammatical phenomena). Below, I briefly discuss how each parameter was dealt with in the present study, and point to additional work that might be of special interests to cognitive linguists.

7.1 Hand shape and palm orientation

In order to categorize and represent *hand shapes*, a data-driven typology of manual signs was developed. Another possibility would be to adopt the form inventory of a signed language such as American Sign Language (ASL) (cf. McNeill 1992: 86–88; Webb 1996). First, the data were searched for hand shapes and arm configurations recurring across speakers and contexts, and then each identified form was assigned a name. For example, one of the most frequently used hand shapes in the data is an open hand with the palm turned upwards, thus building a sort of surface. Here it seemed worthwhile to build on conventions introduced by Müller (2004a) in her study of forms and functions of the palm-up open hand gesture (or, "puoh").[5] Each variant of the open hand gesture was given an abbreviation such as "puoh" indicating the *orientation of the palm*, a short name evoking the degree of openness of the hand ("tray," "cup," etc.), as well as an indication of which hand performed the gesture. So for instance, "puoh-tray-lh" stands for a flat palm-up open hand, evoking the shape of a tray, produced with the left hand.

Other gestures were observed that were simultaneously performed with two hands, thus exhibiting an internal structure (cf. Kendon 2004: 104, 275ff.). For example, "pcoh-box-bh" stands for another frequently observed gesture consisting of two hands held apart, with both palms being held vertically and facing each other as if they were holding an object (such as a shoe box) between them (i.e., "pcoh" stands for "palm-center open hand," with "center" denoting the direction that the palm is facing). Numerous other hand shapes were identified which cannot be discussed here (for details see Mittelberg 2006).

7.2 Location in gesture space

Space is the natural habitat of gestures. Speakers may exploit the space around them to depict the location of objects, people, places, events, and concepts, as well as the spatial relationships among things, a task usually more difficult to master with purely linguistic means. The range, organization, and preferred use of a person's *gesture space* is conditioned by factors such as age, gender, cultural background, and personal style (cf. Calbris 1990; Emmorey & Casey 2001; Goldin-Meadow 2003; Kendon 1997, 2004; McNeill 1992, 2005; Müller 1998). Also, gesture space is relative to, and constituted by, the position and posture of the speaker-gesturer who, in each communicative act, sets up the coordinates of gesture space around her. Here, several factors may come into play: the dimensions and movements of the speaker's body and gestural articulators, artifacts and physical setting, and, if applicable, also the interpersonal space spanning between herself and her interlocutor(s) (see McNeill (2005: 159ff.) on how participants may (re)shape shared gesture space).

The *location* of a gesture can be described from various angles: relative to the gesturer's body, relative to previously or subsequently produced gestures, or relative to the addressee's gesture space. One of the central interests of the present study was the spatialization of abstract information, i.e., ways in which the speakers' use of the different dimensions of gesture space could be linked to spatial metaphor, and, in particular, to particular theoretical models of grammar. This is where different diagrammatic representations of linguistic structure come to light. For example, horizontal lines running from the left to the right of the speaker may depict a linear sentence model; tree diagrams (generative grammar), by contrast, exploit horizontal, vertical, and diagonal axes, and so forth. Such gestural diagrams appear to serve as virtual grids in gesture space providing slots where imaginary objects, standing in for abstract grammatical units, can be placed and subsequently referred to (cf. Mittelberg fc./a on different types of diagrammatic iconicity in gesture (Peirce 1955)).

It should be mentioned that for the purpose of documenting the exact locations where gestures occur, gesture researchers have developed virtual grids dividing the space around the speaker into compartmentalized regions. The system McNeill (1992, 2005) has proposed shows a shallow disc consisting of concentric squares superimposed on a drawing of a seated person (reflecting the experimental set-up described earlier). With the help of such a system the density of occurrence of gestures in specific sectors of gesture space can be represented, thus revealing where certain gesture types tend to occur, e.g., close to certain body parts (head, chest, etc.), in the center versus the periphery of gesture space, on upper vs. lower planes, etc. To give an example, McNeill (1992: 88) identified a correlation between gesture type and location of occurrence. In cartoon narrations by American university students, iconic gestures (depicting concrete objects) were predominately performed in the immediate center, metaphoric gestures (depicting abstract entities) in the lower center, and pointing gestures in the periphery (for definitions of the different gesture categories used for the study cf. McNeill (2005: 38ff.)). In a further step, these findings were supplemented by a frequency analysis of gesture types in correlation with different discourse contexts: narrative clauses (promoting the development of the story, i.e., the plot line) versus extranarrative clauses (descriptions of the setting and characters, summaries, comments, etc.). McNeill (1992: 92ff.) found that whereas iconic gestures and abstract pointing gestures occur primarily in narrative clauses, metaphoric gestures appear chiefly in extranarrative clauses (for details regarding the relationship between discourse demands and saliency of certain gesture dimensions see McNeill 2005: 42ff.).

From a cognitive linguistics perspective, the relationship between space, embodied cognition, and gesture is particularly fascinating and constitutes a promising domain of

5. In the literature, this kind of open hand gesture is also referred to as "palm presentation" gesture (Kendon 2004: 264) or "a gestural version of the conduit metaphor" (McNeill 2005: 45; cf.: Parrill 2003).

inquiry. The interested reader is referred to work on dynamic representations of spatial concepts, spatial arrays, frames of reference, spatial portrayals of time, and mental spaces (e.g., Haviland 2000; Levinson 1997, 2003; Nuñez & Sweetser 2006; Sweetser this volume). Additional insights may be gained from consulting work on the logic and use of space in signed languages (e.g., Dudis 2004; Emmorey & Reilly 1995; Liddell 2003; Milford & Wilcox this volume; Taub 2001).

7.3 Movement

Gestures typically involve some *movement* through space. As opposed to the (unmarked) flux of hand movements, a gesture hold can be regarded as a marked instance (Waugh 1982). There are various types of gestural movements. For example, a movement of a hand can result in the evocation of a form (such as the size and shape of a table), it may be influenced by the object involved in the action that gets imitated by the hand movement (such as writing something down with a pen). A gesture may also simply imitate a manual action (such as waving at somebody) or the manner and/or speed of a movement executed by a person or an object. The hand movements observed in the present data also exhibit several intrinsic logics: movements evoking linear traces (along horizontal, vertical, or diagonal axes) or nonlinear traces (imitating the shape of a wave, circle, or arch); pointing gestures whose direction and range depend on the location of the object or person pointed at; object-oriented actions such as placing something; and genuine motor actions such as two hands rotating around each other. In keeping with the notational conventions used for hand shapes, the prominent movement patterns observed in the data were given labels that inform about its trajectory and manner. For example, "hori-trace-rh" signifies a horizontal line traced in the air with the right hand, and "wrist-rota-lh" refers to a wrist rotation performed with the left hand.

As a result of the steps described above, a set of image-schematic and motor patterns emerged from the data, some of which evoke geometric figures (circles, triangles, squares, etc.) and/or image schemas proposed in the literature (Johnson 1987; Lakoff 1987; Mandler 1996; see also contributions in Hampe 2005, especially Cienki's experimental study on image schemas and gesture). Examples of image schemas evoked in the gesture modality include support, containment, source-path-goal, balance, scale, iteration, front-back, part-whole, link, and forced motion (see Mittelberg fc./b). The next task was to find instantiations of each of the identified shapes and movement patterns across speakers. This was not possible for all categories, but it was possible for most and thus allowed the researcher to discern semantic and pragmatic commonalities and differences, taking into consideration the concurrent speech.

8. Semantic and pragmatic functions of gestures

Depending on one's research interests and theoretical approach, the lens one decides to use to detect meaning in gestures will naturally influence what one actually sees. After attending to the material side of gestures, the eye of the linguist finally becomes engaged in

the task of visual interpretation which is, compared to our natural way of perceiving fluid gestural movements, aided and intensified by the repeated viewing of segments. Watching gestures in slow motion and frame-by-frame reveals nuances of bi-modally achieved meaning-making one might otherwise miss. Gestures convey not necessarily the same semantic aspects as the speech segments with which they coincide. It has been shown that spontaneous gestures may reinforce or complement what is expressed verbally in various ways (cf. Kendon 2000, 2004; McNeill & Duncan 2000; McNeill 2005).

Whether one chooses to adopt an existing gesture classification system or decides to design one's own, it is important to realize that in a single act of gestural signification, as in most sign processes, multiple semiotic modes tend to interact (i.e., iconicity, indexicality, symbolization, according to Peirce 1995, 1960). Many gesture researchers have realized that multifunctional signs such as co-speech gestures hardly fit into discrete categories. The gesture categories originally proposed by McNeill (1992) are thus not to be understood as mutually exclusive (i.e., *beats*, *iconics*, *metaphorics*, and *emblems*). As a matter of fact, McNeill (2005:41) has recently emphasized that in one and the same gesture different semiotic and functional dimensions may layer to various degrees and that speaking of categories may be misleading: "We should speak instead of *dimensions* and say *iconicity, metaphoricity, deixis, 'temporal highlighting'* (beats), *social interactivity*, or some other equally unmellifluous (but accurate) terms conveying dimensionality" (italics in the original; for examples and overviews of established paradigms and latest developments cf. Calbris 1990; Efron 1972; Kendon 2004; McNeill 1992, 2000, 2005; Müller 1998, 2004a/b; Parrill & Sweetser 2004; inter alia).

In her functional gesture typology, Müller (1998) distinguishes gestures whose primary function is "referential" (i.e., depicting objects, attributes, actions, etc.) from those that are rather "discursive" (pertaining to the structure of the discourse), or "performative" (i.e., gestures that have a similar function as speech acts). Depending on the specific nature of the referent of the gestural sign, Müller further distinguishes referential gestures that represent concrete entities from those that represent abstract entities (see also Cienki 2005; and Müller 2004a/b). Müller also stresses the point that one and the same gesture may fulfill several functions at the same time (see also Kendon 2004).

The issue of layering semiotic and functional modes comes clearly to the fore when one investigates abstract subject matters and thus enters the domain of metaphor. Metaphoric gestures, referring to abstract concepts and relations, are said to be iconic in the sense that they may depict aspects of the concrete source domain of the underlying metaphorical mapping. As demonstrated above, abstract grammatical categories such as a "constituent" may be gesturally depicted in the form of a manipulatable object, such as an imaginary box seemingly held between the speaker's hands. Although the speech in this example is not metaphorical but technical in nature, the image schema OBJECT surfaces in this manual constellation, hence reflecting the metaphorical concept IDEAS ARE OBJECTS or CATEGORIES ARE CONTAINERS (Lakoff & Johnson 1980, 1999).[6] Proposing a dynamic metaphor theory, Müller (2004b) suggests that gestures may also awaken

6. To represent this conceptual-semiotic linkage in the gesture-annotated transcript, one can, for instance, add two separate lines below the speech line: one describing the iconic (and/or deictic) qualities

metaphors that are supposed to be dead, or inactive in our conceptual system. In other words, while a linguistic expression based on an inactive metaphor might be perceived as literal, the metaphorical understanding still materializes in the accompanying gesture (for an overview of recent advances in research on metaphor and gesture see contributions in Cienki & Müller fc.).

To account for the specific meaning-making processes in the data of the study discussed above, it was necessary to develop a theoretical framework combining several different strands: traditional semiotic theories (Jakobson 1956, 1964, 1987; Peirce 1955, 1960), gesture research (e.g., Bouvet 1997, 2001; Calbris 2003; Cienki 1998; Fricke 2004; Kendon 2004; McNeill 1992, 2000; Müller 1998, 2004a; Sweetser 1998; Webb 1996), and cognitive metaphor and metonymy (e.g., Barcelona 2000; Dirven & Pörings 2002; Gibbs 1994; Goosens 1995; Johnson 1987; Lakoff & Johnson 1980, 1999; Lakoff 1987, 1993; Langacker 1993; Panther & Thornburg 2004; Radden & Kövecses 1999; Sweetser 1990; Taub 2001; Wilcox 2004). Central to this approach is Peirce's (1955) view of the sign as a dynamic process in which the different referential modes layer to various degrees. It further is based on the assumption that in any sign-object relationship one of the interacting modes is predominant, thus causing the receiving mind to interpret it as predominantly iconic, indexical, or symbolic (Peirce 1960), presupposing a relative hierarchy of the various modes (cf. Jakobson 1964, 1966, 1974). Furthermore, the notion of salience played a role in determining those qualities of a gestural gestalt that contribute most significantly to its meaning and function. For example, in certain cases, the movement proves to be more salient with respect to the meaning of a gesture than the particular shape of the hand performing the movement (e.g., in certain gestures that trace the dimension of a sentence in the air, from left to right of the speaker, whereby it does not matter whether the hand tracing is a relaxed flat hand or whether the fingers are configurated in a particular way); in other cases, the hand shape is more salient than the contextual movements (e.g., in the case of the emblematic ring gesture, also known as the OK-sign); and in yet other cases, both dimensions are significant (e.g., a push performed with an open flat palm facing the addressee, thus building a barrier and evoking the idea of 'stop' or 'rejection'). Based on the set of geometric and image-schematic patterns that emerged from the data, the next step then was to examine how the minimal information provided by the hand shapes and movements played into the iconic, metaphoric, and metonymic meaning construction in the ongoing discourse (cf. Mittelberg 2006, fc.).

In light of the different perspectives gesture researchers have been exploring to enhance our understanding of situated practices of language use and its cognitive and sociocultural underpinnings, it becomes evident that gestures may provide valuable non-verbal evidence for, among other things, spatial cognition and the interplay between iconicity, metaphor, and metonymy (see also Sweetser this volume).

of the gesture and one indicating what the pictorial content stands for metaphorically (for transcript examples see McNeill 1992: 94ff.; Sweetser 1998; Taub 2001).

9. Concluding remarks

There are many more things to be said and considered here, but I hope that this chapter has provided the reader with a picture of the kinds of decisions and methodological steps that shape empirical work with multimodal usage data. As has become evident, there are diverse approaches and methods to choose from, and the investigator will have to see which ones appear most suitable for a specific research project and also how they might need to be adjusted to exploit as effectively as possible the particular kind of data with which she or he wishes to work.

While the focus here has been on transcribing and describing the material side of multimodal communication, and not on content analysis or theoretical issues, the body of research pointed to throughout the chapter suggests that investigations into the gesture modality may offer a window on the bodily foundation of meaning (see also Sweetser this volume). Working with co-speech gesture data allows the researcher to witness, even if it is just for a moment, how embodied structures become visible in the form of metaphors in speakers' hands, diagrams drawn in the air, or pointing gestures drawing attention to imaginary events being set up in gesture space. Essentially, co-speech gestures are *dynamic* semiotic acts. Compared to, for example, static pictorial metaphors in cartoons and advertisement, metaphorical gestures are strikingly spontaneous and may convey figurative understandings in a comparatively unreflective fashion, even if the concurrent speech is perceived as literal (Müller 2004b).

Overall, it seems that multimodal utterances can only be studied empirically: gestures are inseparable from the speaker's body, and the speaker's body and its communicative practices are always anchored in the physical, interpersonal, and cultural context of the speech event (Kendon 2004). So, no matter how closely we look at instances of bodily semiotics, we may not be able to distinguish the gesturer from the gesture, or to "tell the dancer from the dance" (Kramsch 2002: 1).

Acknowledgments

I would like to thank Linda Waugh, Thomas Farmer, Adam Kendon and Elwys De Stefani for valuable comments on earlier versions of this paper and Cornelia Müller and Rebecca Webb for generously sharing their experience with gesture coding and transcription. I am grateful to Richard Feldman and Andrew Page from the Cornell Language Resource Center for technical support and producing video materials. I thank Allegra Giovine, Joel Ossher, and Daniel Steinberg for their input regarding the methodology presented here and for helping with gesture annotation and analysis. I also thank the Cornell Cognitive Studies Program for financial support.

References

Aitkinson, H. M. & Heritage, J. (Eds.). (1984). *Structures as social action: Studies in conversation analysis.* Cambridge: Cambridge University Press.

Barcelona, A. (Ed.). (2000). *Metaphor and metonymy at the crossroads: A cognitive perspective.* Berlin/New York: Mouton de Gruyter.

Beattie, G. (2003). *Visible thought: A new psychology of body language.* Hove: Routledge.

Bouvet, D. (1997). *Le corps et la métaphore dans les langues gestuelles: A la recherche des modes de production des signes.* Paris: L'Harmattan.

Bouvet, D. (2001). *La dimension corporelle de la parole. Les marques posturo-mimo-gestuelles de la parole, leurs aspects métonymiques et métaphoriques, et leur rôle au cours d'un récit.* Paris: Peeters.

Calbris, G. (1990). *The semiotics of French gestures.* Bloomington: University of Indiana Press.

Calbris, G. (2003). From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture, 3*(1), 19–46.

Cameron, D. (2001). *Working with spoken discourse.* London/Thousand Oaks: Sage.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse.* Amsterdam/Philadelphia: John Benjamins.

Chafe, W. (1998). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure, volume 1* (pp. 93–111). Mahwah, NJ: Erlbaum.

Cienki, A. (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J.-P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 189–204). Stanford: CSLI Publications.

Cienki, A. (2005). Image schemas and gesture. In B. Hampe (Ed.), in cooperation with J. Grady, *From perception to meaning: Image schemas in Cognitive Linguistics* (pp. 421–442). Berlin/New York: Mouton de Gruyter.

Cienki, A. & Müller, C. (Eds.). (Fc.). *Metaphor and gesture.* Amsterdam/ New York: John Benjamins.

Clark, H. (1996). *Using language.* Cambridge: Cambridge University Press.

De Stefani, E. (Ed.). (Fc.). Les données vidéo dans la recherche linguistique Looking at language. [Video data in linguistic research. /Regards sur la langue.] Special issue of the *Bulletin Suisse de linguistique appliquée VALS/ASLA.*

Dirven, R. & Pörings, R. (2002). *Metaphor and metonymy in comparison and contrast.* Berlin/New York: Mouton de Gruyter.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1992). Discourse transcription. *Santa Barbara Papers in Linguistics, 4,* 1–225.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–87). Mahwah, NJ: Lawrence Erlbaum Assoc.

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics, 15*(2), 223–238.

Duncan, S. (2003). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture, 2*(2), 183–206.

Duranti, A. (1997). *Linguistic anthropology.* Cambridge: Cambridge University Press.

Edwards, J. A. & Lampert, M. D. (Eds.). (1993). *Talking data: Transcription and coding in discourse research.* Hillsdale, NJ: Lawrence Erlbaum.

Efron, D. (1972). *Gesture, race, and culture.* Preface by Paul Ekman. The Hague: Mouton and Co.

Emmorey, K. & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture, 1*(1), 35–50.

Emmorey, K. & Reilly, J. S. (Ed.). (1995). *Language, gesture, and space.* Hillsdale, NJ: Erlbaum.

Fauconnier, G. & Turner, M. (2002). *The way we think: Conceptual integration and the mind's hidden complexities.* New York: Basis Books.

Fricke, E. (2002). *Origo,* pointing, and speech: The impact of co-speech gestures on linguistic deixis theory. *Gesture, 2*(2), 207–226.

Fricke, E. (2004). Origo, Geste und Raum – Lokaldeixis im Deutschen. Doctoral Dissertation, Technische Universität Berlin.

Furuyama, N. (2000a). Gestural interaction between the instructor and the learner in origami instructions. In D. McNeill (Ed.), *Language and gesture* (pp. 99–117). Cambridge: Cambridge University Press.

Furuyama, N. (2000b). The semiotic making of speech and gesture and how it contributes to the structuring of discourse. *Cognitive Studies, 7*(1), 71–77.

Gibbs, R. W., Jr. (1994). *The poetics of mind: Figurative thought, language, and understanding.* Cambridge, UK: Cambridge University Press.

Goldin-Meadow, S. (2003). *Hearing gesture: How hands help us think.* Cambridge, MA/London: The Belknap Press of Harvard University Press.

Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation. *Semiotica, 62*(1–2), 29–49.

Goodwin, C. (1993). Recording human interaction in natural settings. *Pragmatics, 3*(2), 181–209.

Goodwin, C. (2001). Practices of seeing visual analysis: An ethnomethodological approach. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of Visual Analysis* (pp. 157–182). London: Sage Publications.

Goodwin, C. (2003). Pointing as situated practice. In S. Kita (Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 217–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Goossens, L. (1995). Metaphtonymy: The interaction of metaphor and metonymy in figurative expressions for linguistic action. In L. Goosens, P. Pauwels, B. Rudzka-Ostyn, A.-M. Simon-Vandenbergen, & J. Vanparys (Eds.), *By word of mouth. Metaphor, Metonymy and Linguistic action in cognitive perspective* (pp. 159–174). Amsterdam/Philadelphia: John Benjamins.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse. A study of learners of French and Swedish.* Lund University Press.

Hampe, B. (Ed.). (2005). *From perception to meaning: Image schemas in cognitive linguistics.* Berlin: Mouton de Gruyter.

Haviland, J. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13–46). Cambridge: Cambridge University Press.

Hopper, P. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure.* Mahwah, NJ: Lawrence Erlbaum.

Jakobson, R. (1990 [1956]). Two aspects of language and two types of aphasic disturbances. In L. R. Waugh & M. Monville-Burston (Eds.), *Roman Jakobson, On Language* (pp. 115–133). Cambridge/London: Harvard University Press.

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in Language* (pp. 350–377). Cambridge, MA: MIT Press.

Jakobson, R. (1971[1964]). Visual and auditory Signs. Phonetica XI, 216–220; reprinted in *Selected Writings II* (pp. 334–337). The Hague: Mouton.

Jakobson, R. (1990 [1966]). Quests for the essence of language. In L. R. Waugh & M. Monville-Burston (Eds.), *Roman Jakobson, On Language* (pp. 407–421). Cambridge/ London: Harvard University Press.

Jakobson, R. (1990 [1974]). Linguistics in relation to other sciences. In L. R. Waugh & M. Monville-Burston (Eds.), *Roman Jakobson, On Language* (pp. 451–488). Cambridge/ London: Harvard University Press.

Jakobson, R. (1987). On the Relation between Auditory and Visual Signs. In K. Pomorska & S. Rudy (Eds.), *Roman Jakobson, Language in literature* (pp. 467–473). Cambridge/London: Belknap Press of Harvard University Press.

Jakobson, R. & Pomorska, K. (1980). *Dialogues.* Paris: Flammarion. Eng. Trans. 1983. Cambridge, MA: MIT Press.

Jaworski, A. & Coupland, N. (Eds.). (1999). *The discourse reader.* London/New York: Routledge.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. *Journal of Pragmatics, 23*(3), 247–279.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology, 26*, 109–128.

Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kita, S. (1993). Language and thought interface: A study of spontaneous gestures and Japanese mimetics. Doctoral Dissertation, University of Chicago.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet.* Mahwah, NJ: Lawrence Erlbaum Assoc.

Kita, S. & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16–32.

Kramsch, C. (2002). 'How can we tell the dancer form the dance?' Introduction to C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 1–30). London/New York: Continuum.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom.* London/New York: Continuum.

Lakoff, G. (1987). *Women, fire, and dangerous Things. What categories reveal about the mind.* Chicago: University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–251) (2nd ed.). Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic Books.

Langacker, R. W. (1993). Reference-point constructions. *Cognitive Linguistics, 4*, 1–38.

Lantolf, J. P. & Thorne, S. (Ed.). (2006). *Sociocultural theory and the genesis of second language development.* Oxford: Oxford University Press.

LeBaron, C. & Streeck, J. (2000). Gestures, knowledge, and the world. In D. McNeill (Ed.), *Language and gesture* (pp. 118–138). Cambridge: Cambridge University Press.

Levinson, J. (1997) Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology, 7*(1), 93–131.

Levinson, J. (2003) *Space in language and cognition: Explorations in cognitive diversity.* Cambridge: Cambridge University Press.

Liddell, S. (2003). *Grammar, gesture, and meaning in American Sign Language.* Cambridge: Cambridge University Press.

Mandler, J. (1996). Preverbal representation and language. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 365–384). Cambridge, MA: MIT Press.

McCafferty, S. G. (2004). Space for cognition: Gesture and second language learning. *International Journal for Applied Linguistics, 14*(1), 148–165.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: Chicago University Press.

McNeill, D. (Ed.). (2000). *Language and gesture.* Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought.* Chicago: Chicago University Press.

McNeill, D. & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.

McNeill, D. & Levy, E. T. (1982). Conceptual representations in language activity and gesture. In R. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 271–295). Chichester, England: Wiley & Sons.

McNeill, D., Quek, F., McCullough, K.-E., Duncan, S. D., Furuyama, N., Bryll, R., Ma, X.-F., & Ansari, R. (2001). Catchments, prosody and discourse. *Gesture, 1*(1), 9–33.

Mittelberg, I. (2002). The visual memory of grammar: Iconographical and metaphorical insights. *Metaphorik.de, 2/2002*, 69–89.

Mittelberg, I. (2006). Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar. Doctoral Dissertation, Cornell University.

Mittelberg, I. (fc./a). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and gesture.* Amsterdam/New York: John Benjamins.

Mittelberg, I. (fc./b). Geometric patterns in gesture space. In V. Evans & P. Chilton (Eds.), *Language, cognition, and space: The state of the art and new directions.* London: Equinox.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich.* Berlin: Berlin Verlag A. Spitz.

Müller, C. (2004a). Forms and uses of the palm up open hand: A case of a gesture family? In R. Posner & C. Müller (Eds.), *The semantics and pragmatics of everyday gestures. The Berlin conference* (pp. 233–256). Berlin: Weidler Verlag.

Müller, C. (2004b). *Metaphors. Dead and alive, sleeping and waking. A cognitive approach to metaphors in language use.* Habilschrift, Freie Universität Berlin.

Negueruela, E., Lantolf, J. P., Jordan S. R., & Gelabert, J. (2004). The "private function of gesture in second language speaking activity: A study of motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics, 14*(1), 113–147.

Nobe, S. (1996). *Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/threshold model of gesture production.* Doctoral Dissertation, University of Chicago.

Núñez, R. E. & Sweetser, E. E. (2006). Aymara, where the future is behind you: Convergent evidence from language and gesture in the cross-linguistic comparison of spatial construals of time. *Cognitive Science, 30*, 1–49.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schiefflin (Eds.), *Developmental pragmatics.* New York: Academic Press. Reprinted in A. Jaworski & N. Coupland (1999) (Eds.), *The Discourse reader* (pp. 167–182). London/New York: Routledge.

Ochs, E., Gonzalez, P., & Jacoby, S. (1996). "When I come down I'm in the domain state": Grammar and graphic representation in the interpretive activity of physics. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 329–369). Cambridge: Cambridge University Press.

Panther, K.-U. & Thornburg, L. L. (2004). The role of conceptual metonymy in meaning construction. *Metaphorik.de, 06/2004*, 91–113.

Parrill, F. (2003). Intuitions and violations of good form in metaphoric conduit gestures. Paper presented at the theme session "Metaphor and Gesture," 8th International Cognitive Linguistics Conference, Logroño, Spain, July 23.

Parrill, F. & Sweetser, E. E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture, 4*(2), 197–219.

Peirce, C. S. (1955). Logic as semiotic: The theory of signs (1893–1920). In J. Bucher (Ed.), *Philosophical writings of Peirce* (pp. 98–119). New York: Dover.

Peirce, C. S. (1960). *Collected papers of Charles Sanders Peirce (1931–1958). Vol. I.: Principles of philosophy, Vol. II: Elements of logic.* C. Hartshorne & P. Weiss (Eds.). Cambridge: The Belknap Press of Harvard University Press.

Radden, G. & Kövecses, Z. (1999). Towards a theory of metonymy. In K.-U. Panther & G. Radden (Eds.), *Metonymy in language and thought* (pp. 17–59). Amsterdam/ Philadelphia: John Benjamins.

Roth, W.-A. (2003). From epistemic (ergotic) actions to scientific discourse: The bridging function of gestures. *Pragmatics & Cognition, 11*(1), 141–170.

Schegloff, E. A. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (Ed.), *Structures of social action* (pp. 266–296). Cambridge: Cambridge University Press.

Settekorn, W. (1993). Das Auge und der Körper des Linguisten. Für eine Medien- Technik- und Sozialgeschichte (nicht nur) der Sprachwissenschaft, *Linguistik und Literaturwissenschaft, 90/91*, 195–234.

Settekorn, W. (1996). Vor-Film. Zur Geschichte der Inkorporation von Sprache, Bild und Raum. In H. Segeberg (Ed.), *Die Mobilisierung des Sehens. Zur Vor- und Frühgeschichte des Films in Literatur und Kunst* (pp. 13–43). München: Fink.

Settekorn, W. (2003). Representation: The influence of history, culture, and (mass) media on the evolution of a concept central to cognitive science. Presentation given at the EMCL Workshop at Cornell University.

Slobin, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (Eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistic Society* (pp. 435–445). Berkeley: Berkeley Linguistics Society.

Slobin, D. (1996). From 'thought and language' to 'thinking for speaking'. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 97–114). Cambridge: Cambridge University Press.

Smith, N. (2003). Gesture and beyond. Honors Thesis, Program in Cognitive Science, University of California at Berkeley.

Solnit, R. (2003). *River of shadows: Eadweard Muybridge and the technological Wild West.* London/New York: Penguin.

Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs, 60*, 275–299.

Streeck, J. (1994). Gesture as communication II: The audience as co-author. *Research on Language and Social Interaction, 3*, 239–276.

Streeck, J. (2002). A body and its gestures. *Gesture, 2*(1), 19–44.

Sweetser, E. E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure.* Cambridge: Cambridge University Press.

Sweetser, E. E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. *Actes du 16e Congrès International des Linguistes* (CD-ROM), Elsevier.

Tabensky, A. (2001). Gesture and speech rephrasing in conversation. *Gesture, 1*(2), 213–235.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon* (pp. 57–149). Cambridge: Cambridge University Press.

Tao, H. (1999). Body movement and participant alignment in Mandarin conversational interactions. *Papers from the 35th regional meeting of the Chicago Linguistic Society, Vol. II: The panels* (pp. 125–139). Chicago: Chicago Linguistics Society.

Taub, S. (2001). *Language from the body: Iconicity and metaphor in American Sign Language.* Cambridge: Cambridge University Press.

Waugh, L. (1982). Marked and unmarked – a choice between unequals in semiotic structure. *Semiotica, 39, 3/4*, 299–318.

Webb, R. (1996). Linguistic features of metaphoric gestures. Doctoral Dissertation, University of Rochester.

Wilcox, P. P. (2004). A cognitive key: Metonymic and metaphorical mappings in ASL. *Cognitive Linguistics, 15*(2), 197–222.

Williams, R. (2004). *Making meaning from a clock: Material artifacts and conceptual blending in time-telling instruction.* Doctoral Dissertation, University of California at San Diego.

PART IV

Behavioral research