

A Universal of Human Interaction?
**Manual Movement as
Interactional Practice
in Spoken and Signed
Conversation**

Paul Cibulka



UNIVERSITY OF
GOTHENBURG

**A Universal of Human Interaction? –
Manual Movement as Interactional Practice in
Spoken and Signed Conversation**

A Universal of Human Interaction?
**Manual Movement as
Interactional Practice
in Spoken and Signed
Conversation**

Paul Cibulka



UNIVERSITY OF GOTHENBURG

Doctoral dissertation in linguistics
University of Gothenburg
2016-06-03

© Paul Cibulka, 2016

Cover photo:

“Ο κόσμος αλλιώς”

by spapax ([flickr.com/photos/spapax/](https://www.flickr.com/photos/spapax/))

used under CC BY-NC 2.0 (creativecommons.org/licenses/by-nc/2.0/)

Printed by Ineko AB

ISBN 978-91-628-9828-1 (print)
978-91-628-9829-8 (digital)

e-publication available at:
<http://hdl.handle.net/2077/42407>

Distribution:
Department of Philosophy,
Linguistics and Theory of Science
University of Gothenburg
Box 200, SE-405 30 Gothenburg

Ph.D. dissertation at the University of Gothenburg, 2016

Title	A universal of human interaction? – Manual movement as interactional practice in spoken and signed conversation
Author	Paul Cibulka
Language	English
Department	Department of Philosophy, Linguistics and Theory of Science
ISBN	978-91-628-9828-1 (printed version) 978-91-628-9829-8 (digital version)
e-publication	http://hdl.handle.net/2077/42407

When humans interact, they may make use of a range of resources, such as head movements, facial expressions, manual movement, body posture and speech. It is assumed that participants both produce and perceive this stream of information in a differentiated way: Some segments are attended to as belonging to the content of the discourse while others are rather backgrounded and may serve to regulate the interaction in terms of speakership and turn-taking.

This thesis is an anthology comprised of four studies that all touch upon the role of these backgrounded segments of behaviour in both spoken and signed interaction. In particular, I analyse manual movement phases as well as self-touching behaviour in the area of the face and the head. It is found that participants may tweak individual movement phases (such as withholding the retraction to a stable rest position or transforming the manual movement into a self-touch) that provide an *in situ* interpretation of the sequential structure (e.g., that a given line of action is complete) and may occasion the emergence of hierarchically structured levels of degrees of involvement (e.g., it may indicate suspension of a given line of action).

As a result, I suggest that speakership is best understood as a continuum, rather than a binary concept (i.e. speaker and listener). It turns out that all roles within this spectrum are not static ones but have to be enacted and performed in order to be perceived as such.

I show that participants in signed and spoken conversation exploit the same resources, i.e. segments of manual movement, as part of the same practices in order to regulate speakership and turn-taking. Some of these resources are apt to be ascribed a linguistic status as part of the system of a given sign language (e.g., Swedish Sign Language and American Sign Language), while in spoken language they are often regarded as an add-on to vocal resources. This a priori divide between what counts as sign and gesture respectively obscures areas of overlap (cf. Kendon 2008) and, in the light of the results, it is suggested to treat them in the same way.

Furthermore, with regard to the geographic distance and linguistic diversity of the languages herein analysed (Swedish Sign Language, Japanese and German), I discuss whether the use of manual movement phases as interactional practice may be considered a universal in human interaction.

List of studies

STUDY 1

When the hands do not go home:
A micro-study of the role of gesture
phases in sequence suspension and closure

STUDY 2

On how to do things with holds:
Manual movement phases as part of
interactional practices in signed conversation

STUDY 3

Exploring common ground in gesture and sign –
Interactional workings of manual
holds in spoken and signed interaction

STUDY 4

Self-touching behaviour in social interaction:
Shifting in and out of speakership

CONTENTS

PART 1: THESIS FRAME

Chapter 1: Introduction	3
Chapter 2: Gathering data and making sense of them	13
Chapter 3: Key concepts in social interaction and visible behaviour	25
Chapter 4: Summarising the studies	47
Chapter 5: Roundup discussion	53

PART 2: STUDIES

Study 1: When the hands do not go home	77
Study 2: On how to do things with holds	111
Study 3: Exploring common ground in gesture and sign	145
Study 4: Self-touching behaviour in social interaction	177
In lieu of an Afterword	215

Part 1

Thesis Frame

CHAPTER I

Introduction

The erudite reader would expect, at this point, an introductory sentence that summarises the content and aims of the present work. I chose to borrow one from Adam Kendon:

“Willingly or not, humans, when in co-presence, continuously inform one another about their intentions, interests, feelings and ideas by means of visible bodily action.” (Kendon, 2004, p. 1)

This sentence very well fits what is presented in this thesis, in the sense that (1) I focus on situations when humans are in co-presence, (2) I presuppose that the body is at all times available to others for interpretation and (3) humans employ visible behaviour knowing that it can be interpreted by others.

Gesture research generally focusses on overtly communicational bodily behaviour, i.e. segments of behaviour such as hand movement as part of an utterance in order to convey meaning. But my own work instead zooms in on more backgrounded – nonetheless potentially communicational – bodily behaviour. I look into the constituents of action gestalts, i.e. small segments of bodily movement that together comprise a bigger whole, but are meaningful in their own right by virtue of being embedded within this very context.

The point of departure for this thesis is the assumption that language in all its embodied glory – in addition to other forms of visible action that are generally not considered part of language – is mobilised primarily to the end of

producing actions as part of social interaction. The thesis is concerned in particular with the systematicity of the employment of such visible resources in social interaction.

But why is it important to scrutinise human behaviour on such a micro-level? It tells us something about how we interact with one another, and human interaction lies, as I believe, at the foundation of sociality. There are all kinds of communities on the planet. We speak different languages, have different habits and values, wear different clothes, work in different jobs, learn different things. Notwithstanding all these differences, there are certain things that are shared, certain patterns that exhibit consistency across different communities. These shared patterns manifest in certain interactional practices – ways of connecting and cooperating with one another – some of which are dealt with in this thesis. These practices are mundane and ordinary, and should thus be obvious, but perhaps they are too obvious to be noticed as readily analysable entities. My job has been to excavate these entities and explicate what we usually take for granted.

The practices dealt with herein are strong candidates for being universal across members of different communication communities that are otherwise considered unrelated. These findings may offer fresh perspectives on how we perceive ourselves and each other as part of a world full of contrasts and diversity and relate to the question of whether, at the core, we are perhaps not so different from one another after all.

1.1 Preliminaries

Here I will provide some clarifications regarding terminology used herein.

First, lexical items in signed languages are referred to as *signs*, just like lexical items in spoken languages are referred to as *words*. This is also how I use the term *sign*. I do not use it in order to denote semiotic signs in a Saussurian way. Second, terminology such as *speech*, *speaker*, *speakership*, *listener* and *talk* reflect a bias towards the analysis of vocal resources in research on human interaction. However, I use these terms regardless of whether a participant uses vocal or visible resources, such as in the case of signed conversation. A *speaker* may thus be a participant who is involved in the activity of producing an utterance,

be it vocal or visible or both. Just as *asking* or *answering* are not necessarily understood as activities that involve vocal resources, *speaking* is understood as utterance production.

I have nonetheless adopted the well-established term *spoken language* for languages that utilise vocal resources as their primary modality. The term contrasts with *signed* language, but also with *written* language, constituting yet another terminological headache: Written and spoken language are, in general terms, two manifestations of a given language. Signed languages are, however, largely unrelated to this given language (unless we are talking about planned languages such as manually coded English, which indeed is a signed manifestation of the English language).

In using the term *spoken* to exclusively denote *vocally produced* one runs the risk of suggesting that participants in signed conversation do not *speak*. But I maintain that what participants do when they produce utterances is not inherently different depending on which modality they use to produce these utterances. Hence I use the term *spoken language* only because it is well-established, albeit reluctantly as I do not concur with the idea that the activity of *speaking* should be reserved for participants who produce utterances vocally.

This is, however, merely a terminological issue. For the sake of comparison, other languages offer more accurate ways of terming languages in which sound is the modality that interactants chiefly attend to. Examples are the German term *Lautsprache* (“sound language”) and the Japanese term *onsei gengo* (“voice language”), both of which contrast with signed language, but not with written language. It would be worth considering to revise the way we refer to such languages in English.

1.2 Overall aims of the thesis

The studies are concerned with the description of visible behaviour, as it occurs in interaction. The notion of visible behaviour, in the broad sense, includes all kinds of human behaviour that are visibly perceivable, such as walking, sitting, postures, head movement, and also involvements or practical actions such as reading a book, changing a flat tyre or opening a door. Each of these kinds of behaviour allow a percipient – a person who observes or happens to see an

action – to assess what a participant is doing and as such they may constitute recognisable social actions upon which a percipient can comment or produce a responsive action.

In the specific context of this thesis the analysis revolves around the role of *manual movements phases* in both spoken and signed conversation. A movement phase is a discernible segment that is part of a greater whole. For instance, the turning of a book page can be thought of as consisting of three phases: (1) a phase during with the hand transitions from a relaxed state and moves towards the book, followed by (2) the turning of the page whereafter (3) the hand is withdrawn to a relaxed position. Manual movement employed as (part of) utterances in interaction can be broken down in the same way. See Chapter 3 for a detailed account.

The overarching aspects that I want to highlight in this chapter are (1) the role of manual movement phases in reflecting or creating layers of interaction that are hierarchically ordered, (2) the suggestion that speakership – when viewed in the context of visible behaviour – is best understood as a non-binary, gradual concept along a continuum of possible involvements (3) the integration of both what is counted as gesture and as sign into a single framework of kinesic expression and (4) findings that may constitute universals in human communication, both across languages and cultures, and regardless whether conversation is spoken or signed.

1.3 Manual movement phases as part of interaction

The first aspect revolves around hierarchically arranged layers in interaction that are reflected or occasioned by participants' use of movement phases. Participants can display boundaries in interaction and may move back and forth between the layers. A strong point made in the thesis is that manual movement phases are not only dependent on the structure and contents of an utterance (as in co-speech gesture, e.g. McNeill, 1992), but that they may be mobilised in response to, or in orchestration with, interactional units. This holds true for both spoken and signed conversation.

Phases of manual movement may serve as some sort of delimiter within sequential organisation, i.e. the orderly buildup of an utterance in relation to a previous utterance. A basic sequence, sometimes referred to as “chunk of actions” or “bundle of actions” in the individual studies, consists of an initiating action (such as, most stereotypically, an enquiry) by a first speaker, a responsive action (such as a reply) by a second speaker and often a sequence closing third, that is, an action which acknowledges receipt of the responsive action, by the first speaker.

As shown in the studies, this structure is relevant in visible behaviour: The kinesic structure of a gesture (Study 1 and 3), a sign (Studies 2 and 3) and to some extent of self-touching behaviour (Study 4) is not only tied to the utterance structure as a whole, but also regularly reflects or occasions parts of a sequential structure, i.e. interrelated chains of utterances. For instance, when the speaker of an initiating action mobilises a gesture (in the case of spoken conversation), they often hold the gesture in mid-air in coordination with the responsive action and retract the hand(s) when the sequence closing third is produced. Manual movement is, thus, also related to sequences as a whole.

The position in space where a gesture or sign is held is variable. The limb may be fully extended or it may be in a half-way retracted position. Since traditional descriptions (see Chapter 3) of movement phases do not capture this difference in position, I figured it would be helpful to have terms at hand that denote a relative position of the limbs. I introduce the term *stage* to denote the place in space in front of the gesturer or signer where manual movements are performed in a fashion to be seen and noticed. The stage contrasts with *home position*, or simply *home* (Sacks & Schegloff, 2002), which is a position of relaxation. Home and stage, thus, can be used to describe the position of the hand(s) in space: In a basic manual excursion the limbs travel from home to stage where the main part of the movement is performed, after which they travel back home. A manual hold, then, may occur in any location between the stage and home, and (as argued in Studies 1, 2 and 3) these various positions appear to relate to a variable degree of claim of speakership: The closer to the stage the hold occurs, the stronger the claim of speakership.

1.4 Speakership as a non-binary concept

As noted in the previous section, participants can make claims of speakership in varying degrees by virtue of bodily conduct. This questions, by extension, the concept of speakership as a binary one (i.e. speaker versus listener). I instead argue that it is best viewed as a continuum of possible involvements ranging from “actively speaking” to “having no involvement at all”, with speakership, imminent speakership, suspended speakership and so forth as particular sorts of involvement.

Note that the term *speakership* as it is used here encompasses, but is not equal to, *having the turn* or *having the floor*. Sacks et al. (1974) define a turn at talk as a unit in conversation that participants in interaction orient to: One participant can have the turn at a time. Participants are able to foresee when a speaker’s turn is about to end, so that speaker change usually occurs seamlessly. The understanding of what a turn constitutes, however, is based on assessing the vocal resources that participants employ. But participants orient not only to vocal, but also to visible resources when turn-taking is viewed as a multimodal phenomenon (cf. Oloff, 2013).

As a result, the boundaries of where speakership starts and ends become less clearly defined in comparison with the boundaries of a turn at talk.¹ It is argued that conversational statuses such as currently speaking, incipient speakership, temporarily suspended speakership and so forth are various roles that participants actively enact, in employing various resources (including visible resources such as manual movement and self-touching behaviour), in order to be perceived as such.

1.5 Manual movement in spoken and signed conversation

Another aspect concerns the comparison of two modes of communication – spoken and signed conversation. Even though these two modes are different from one another in many ways, parallels can be found in particular in the organisation of manual movement phases as part of interactional practices.

¹ It is for this reason that I gradually refrained from using the term *turn*, given that it is chiefly defined in terms of vocal resources. The term *utterance*, which is the preferred alternative, does not seem to encompass such a modal preference.

The prevalence of a strong distinction between what is gesture and what is sign has, in my opinion, produced misleading assumptions about the role of manual movement (cf. Kendon, 2008, 2014). Whereas a sign is considered part of a linguistic system, gesture is apt to be considered “non-verbal”, i.e. not part of language. It is sometimes said that gestures may function like a word, in particular regarding what is referred to as emblems, such as the thumbs-up gesture or the OK gesture. McNeill (2005) assigns them a semi-linguistic status, arguing that “the OK gesture, like a word, is constrained to assume a certain ‘phonological’ shape. Yet these constraints are limited and don’t by any means amount to a full language” (2005, p. 9). For an example, see Figure 1 (taken from Andrén, 2010, p. 58), illustrating an emblem, the hush-gesture, in its phonologically correct (left) and partially correct form.



Figure 1: Correct (left) and partially incorrect forms of an emblem (taken from Andrén, 2010, p. 58)

Even though gestures may have similar properties to a word, we rarely find a gesture being classed as word.² Peculiarly, when it comes to signs, the words of a sign language (Liddell, 2003, p. 1), there seems to be no doubt that these are entities of linguistic nature.

² This may also be due to the fact that the concept of a “word” seems to be tacitly reserved for vocally produced entities in spoken languages, possibly reflecting a linguistic preference of arbitrary conventions, rather than iconic representations.

A contribution of this thesis is towards mitigating this discrepancy by pointing to rather obvious yet largely overlooked similarities. It also constitutes a contribution towards what Kendon (2008) calls a *comparative semiotics of kinesic expression*. Kendon proposes to leave behind the distinction between gesture and sign, suggesting that we need comparative studies of how visible action is used in utterance construction, taking into consideration what other modalities (such as the vocal portion of an utterance) are available (2008, p. 359f.). The thesis investigates the mechanics of kinesics and their role in the deployment of interactional practices as part of signed conversation in comparison to those in spoken conversation. If we assume that these mechanics are inherent to sign language as a linguistic system (cf. Bergman [1982] who argues that sustained gaze and prolonged hold phases are part of forming a question in sign language grammar), then we also need to revisit certain visible components in spoken conversation and we need to examine whether these mechanics should be considered part of spoken language, as a linguistic system. As I will discuss in the final chapter, it is worth reevaluating our views on whether some aspects of visible behaviour such as certain segments of manual movement should be classed as the one or the other.

1.6 Universals in human communication

Some of the interactional practices herein analysed exhibit a systematicity across different communication communities to an extent that suggests that they may constitute universals in human communication. I draw on conversation data in a range of languages, and the participants are primarily treated as interactants within a specific local context, without stipulating that there be *a priori* differences between different language communities, cultures or nations. Nor do I believe that what participants do in spoken conversation fundamentally differs from what they do in signed conversation. I am purely focussing on participant practices, rather than dealing with imposed identities of participants as members of certain categories, communities or groups, such as nations, languages, native or non-native speakers or deafness. As a result, I

found communicational practices that are shared across different communities, which, in turn, constitutes a contribution towards demystifying specific communities.

These findings allude to the existence of universals in human communicative behaviour, such as manual holds as part of interactional practices. It is empirically established that these practices are used in conversation in Swedish Sign Language as well as in Japanese. These constitute two distinct communication communities, in the sense that they are typically spoken in geographically distant locations, and also in the sense that the former is a signed and the latter a spoken language. Even though the number of participants and field work sites is too small to make stable assumptions on universality, it is nonetheless intriguing to find the same manual movement patterns as part of the same interactional practices, which encourages further investigation of the matter.

1.7 What this thesis frame is about

The remainder of the thesis frame will set the stage for the individual studies, introduce and discuss the method and terminology, and bind the studies together in a coherent way. In Chapter 2 I give an account of the video data and corpora subjected to analysis, and I illustrate the way I dealt with representations of videotaped interaction. In Chapter 3 I introduce terminology and review literature related to gesture, sign, visible behaviour and human interaction in general. Chapter 4 provides a summary and further discussion of the individual studies. In Chapter 5 I draw conclusions and discuss implications.

CHAPTER 2

Gathering data and making sense of them

This chapter describes how data were gathered and how transcripts of stretches of interaction were produced. Making sense of the data is crucial, given that the studies are intrinsically data-driven. Transcriptions are a way of making the data accessible on paper to the reader, but, on the other hand, the very act of producing transcriptions and annotations, along with repeated viewings, is a process that helps with identifying phenomena and finding patterns and orderliness in the data, and is thus of immense benefit to the analyst.

The act of transcribing, i.e. rendering what is being said or done in interaction into written form, is an analytic activity that draws on the transcriber's knowledge about the social world they are part of. Due to the simple fact that it is impossible to include the sheer variety of participant behaviour in the transcription, the transcriber needs to be selective, and also the choice of including certain aspects and disregarding others is up to the analyst. The final transcripts may thus differ tremendously depending on the choices on transcriber's part and on the purpose of the analysis.

A further issue, then, is the choice of how the data should be presented to the audience. The dilemma, however, is that studies such as the ones presented as part of this thesis are commonly printed on static, rather two-dimensional paper. When attempting to bring motion pictures into a paper-friendly form,

one will notice that there are a myriad of ways to achieve this, some of which will be dealt with in this chapter. Also, this choice has an impact on how the audience digests it (Mondada, 2007a).

This chapter thus gives an account of what kind of data the analyses in the respective studies are based upon as well as how they were gathered and annotated. Furthermore, I will describe the steps I took in order to create transcripts on paper from videotaped data while attempting to preserve accessibility to the reader.

2.1 Overview of the data

The studies in this thesis draw on data partly made available by other researchers and partly gathered by myself. Specifically, I made use of five corpora: (1) Natural Conversation Corpus Japanese from Kobe University, (2) Natural Conversation Corpus German from Kobe University (3) the Sakura corpus from TalkBank (MacWhinney, 2007), (4) the Swedish Sign Language corpus (Mesch, Wallin, Nilsson, & Bergman, 2012) and (5) data gathered by myself. The data from corpora were chosen with respect to their authenticity, i.e. I chose datasets in which the flow of conversation is constrained as little as possible and which exhibited a balanced exchange of utterances, rather than situations in which the participants were asked to retell a story or otherwise experimentally induced talk.

The different corpora listed here feature conversation in a number of languages: German, Japanese, Swedish and Swedish Sign Language. However, I do not take a contrastive approach to the data, i.e. I do not use the language as a variable. Also, the analyses are not meant to represent the data as a whole. The different corpora are rather used as a database which I draw on in order to illustrate specific interactional features.

2.1.1 Natural Conversation Corpus Japanese

This corpus was recorded as part of a research project³ at Kobe University, Japan. It consists of videotaped dyadic conversation among students (24 sessions of approximately 60 minutes each). The participants were instructed to talk freely, without being assigned any topics. I am indebted to Toshiyuki Sadanobu for granting me access to the corpus along with transcriptions. I was not involved in the recording of this dataset.

2.1.2 Natural Conversation Corpus German

I gathered this corpus myself as part of a research project at Kobe University.⁴ It consists of videotaped conversation in German among students and friends. The conversation was mostly dyadic, but sometimes included three or more participants. The recordings took place partly in Japan, partly in Germany. The corpus has a length of approximately ten hours.

2.1.3 Sakura Corpus

This corpus was originally recorded as part of several graduation theses at Aichi Shukutoku University, Japan. It comprises 18 sessions of videotaped conversation among university students in groups of four. The participants were provided with the topic of conversation prior to the recording, but were otherwise allowed to talk freely and deviate from the initial topic. The corpus has an overall duration of seven hours and 30 minutes. The video files plus

3 Grants-in-Aid for Scientific Research A 16202006 “Developing teaching materials for education of spoken Japanese based on contrastive studies among Japanese, English and Chinese” from the Ministry of Education, Culture, Sports, Science and Technology (Japan) between 2004 and 2006. For more information about the project see: <https://kaken.nii.ac.jp/d/p/16202006.en.html>

4 Grants-in-Aid for Scientific Research A 19202013 “Speech Grammar based on Speaker’s Characters” from the Ministry of Education, Culture, Sports, Science and Technology (Japan) between 2007 and 2010. For more information about the project see: <https://kaken.nii.ac.jp/d/p/19202013.en.html>

detailed transcriptions have been made available to researchers as part of the TalkBank project (MacWhinney, 2007).⁵ I was not involved in recording this dataset.

2.1.4 Swedish Sign Language Corpus

The Swedish Sign Language Corpus comprises free dyadic conversation and was recorded and annotated as part of a research project at Stockholm University (Mesch, Rohdell, & Wallin, 2014; Mesch et al., 2012).⁶ Apart from free conversation, the corpus also features recordings of narrations and retelling tasks. These were not utilised in this thesis, as free conversation yields a more frequent back and forth in terms of turn-taking, which is more typical of everyday conversation. The part of the corpus that was subjected to analysis has a total length of one hour and 50 minutes.

2.1.5 Data that I gathered myself

The collection of my own data took place in Japan and Sweden. The sites of data collection are very diverse: home parties, dinner at a restaurant, an after-work gathering, guests at a guest house gathering in the living room, a barbecue in a park, classroom interaction, among others.

Even though the amount of data might appear huge, it should be kept in mind that a great deal of it is not visible in the resulting studies. Parts of the data have been annotated in detail, were presented and discussed in data sessions, or have in other ways provided me with inspiration. I did not necessarily go through all of the data for each type of analysis, but all of the data presented herein, albeit only visible in the studies to some extent, have thus been an integral part of the thesis as a whole.

5 For more information about the corpus see: http://search.language-archives.org/record.html?id=talkbank_org_CABank-Sakura

The corpus can be downloaded from: <http://talkbank.org/cabank/>

6 The corpus and annotation files with Swedish translations are available for download from the Stockholm University website: www.ling.su.se/teckenspraksrkorpus

Participants' informed consent

Prior to the start of the recording, as a rule, the participants were asked whether they consent to being recorded and if so, whether they were willing to fill out a consent form. Sometimes before, but mostly after the recording, the participants were asked for their written consent. This was done using a consent form that provides the participants with information about the intended use, manner of storage and handling of personal information. If the participants agreed, they were asked to write their name and date of birth and provide their signature or name seal.

Field work sites

As for field work sites, I tried to find occasions in which participants have already gathered somewhere for some reason, and to do the recording as a background activity – i.e. the recording was not the main reason for the gathering. The part of the dataset collected in Japan mainly consists of friends and co-workers gathering at locations such as restaurants and private homes and has an approximate total length of eleven hours.

Also, I participated in an intermediate course in Swedish Sign Language held in small groups of two to eight participants. I took the chance to ask the teacher and the participants for their consent to videotape some of the sessions, which resulted in a corpus of approximately eight hours of conversation data in the sign language class. The teacher being deaf and the participants being hearing resulted in mixed use of spoken and signed language, with a great deal of language-blending and language-switching. A part of this dataset was used in Study 3.

Equipment

For videotaping I used digital cameras with video recording as an add-on function, which turned out to be sufficient for my purposes. The camera was either furnished with a tripod or a strong clip, which proved especially useful when attaching the camera to the branches of a tree or to the backrest of a

chair. Built-in microphones in digital cameras tend to be of questionable quality. In many cases I remedied this by using an audio recorder – in addition to the camera – which I placed in the vicinity of the participants.

Conversion and editing

After the recordings, the resulting video (from the digital camera) and audio files (from the audio recorder) were synchronised and stored in a single file. For video editing, including this process, I used Avidemux⁷ as well as Handbrake⁸, both of which are open-source software.

2.2 Data transcription and data presentation

The activity of transcribing is in itself an activity that presupposes knowledge of what is going on within a given stretch of interaction. These are skills that we learn through being part of a given communication community. Transcribing is thus an interpretative and analytic activity and researchers very often disagree on one or another method of transcription.

Every interactional analyst is faced with the question of how to design the transcript. This choice has two layers that are relevant for (1) analysis and (2) presentation. By the former I mean that the choice of the method of transcription has an impact on how we, as analysts, view the data, i.e. which phenomena or aspects in the data appear salient to us. The latter, presentation, is about how we present the data to an audience, in contexts such as a journal publication or a conference. These layers should be regarded as two separate aspects of transcription, and, in my case, they are constituted by two separate transcripts. It should also be kept in mind that transcripts alone should not be considered data. They are secondary products merely reflecting parts of the primary data – the recorded video or audio stream (cf. Mondada, 2007a). Producing these secondary data allows the analyst to practice defamiliarising and detaching themselves from the interaction in the primary data, but also to consciously

⁷ <http://fixounet.free.fr/avidemux/>

⁸ <https://handbrake.fr/>

alter the way they view the interaction. This can be achieved by choosing to highlight certain segments of behaviour that are to be analysed, thus allowing reassessment of the primary data.

Producing transcripts that let the primary data appear in a different light, which, in turn, affects the method of transcription, is a back-and-forth process between data and transcription. The use of annotation or transcription software, such as ELAN⁹ (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) or CLAN¹⁰, allows the user to link primary and secondary data, contributing to an increased efficiency of this process.

At the same time, as different analytic software tools have different affordances and offer a variety of ways to represent the primary data, the choice of a tool has an impact on how the analyst may view the data. Tools such as CLAN and Transana¹¹ enable the analyst to produce transcripts in a fashion that resembles text on a sheet of paper. This resemblance makes this sort of transcript easily accessible. Also, turn ends are usually represented by line breaks, which facilitates an analysis focussing on turn-taking and sequentiality. However, seen that they are text-based, they compromise for a precise description of overlapping or simultaneously occurring segments of behaviour.

Time-axis-based tools such as ELAN or ANVIL¹², on the other hand, represent time as an endless timeline and allow the user to add an infinite number of parallel rows (referred to as *tiers* in ELAN) on which annotations can be placed. This method of transcription highlights the fact that several interactional phenomena occur at the same time and thus fosters a description not only of behaviour that can be represented by text, but also of visible behaviour that can be easily annotated.

The second issue regarding the way data excerpts should be presented, again, is very different from the practice of transcribing. While in transcribing it is central to discover, systematise and make collections of phenomena, when it

9 <https://tla.mpi.nl/tools/tla-tools/elan/>

10 <http://childes.psy.cmu.edu/clan/>

11 <http://www.transana.org/>

12 <http://www.anvil-software.org/>

comes to presentation, the focus shifts to viewer-friendliness, zooming in on the phenomena relevant for the respective study and at the same time blocking out what is less relevant.

One of the most widespread systems of transcription is the Jeffersonian transcription notation (Jefferson, 1996; Sacks et al., 1974; Schegloff, 2007). This system has been developed and modified over time and according to the purpose of study may accommodate gaze direction (Goodwin, 1981), manual movement phases (Hosoma, 2009), manual and bodily movements (Mondada, 2007b, 2009) or non-vocal lip-movement (Oloff, 2013).

The more complexity and analytical depth a transcript offers, the more difficult the viewer will find it to decipher. A way to drastically make an excerpt of conversation more accessible is to transform the interaction into comic strips, such as in Andrén (2010, 2012), Ivarsson (2010) and Laurier (2014). Doing so has several advantages: A comic is a format that the majority of readers should be familiar with, and so a long appendix listing transcription conventions is not necessary. Also, this format has been developed for the very purpose of depicting visuals, so it perfectly suits an analysis in which visible behaviour is central. They are thus ideal in order to give the viewer a rough idea of the content of an interaction.

A major downside of using comic strips as excerpts, however, is that the representation of time is blurred: Every frame in a comic strip excerpt constitutes a specific stage in interaction that is considered essential by the analyst. However, it is rarely transparent how much time has passed between the frames, nor is it evident from what specific moment in interaction the frame originates. Furthermore, it would be a rather complicated undertaking to mark overlap and speech gaps, not to mention the timing of bodily movement.

My own analyses rely to a high degree on the accurate rendering of time (be it absolute or in relation to vocal speech), while at the same time the excerpts are intended to be easily accessible. For this reason, I devised an alternative way of rendering stretches of interaction as excerpts. The following section is dedicated to a documentation of the procedures I have taken.

2.3 Workflow for producing transcripts

The question of in what way one should transform video data into an excerpt on a static sheet of paper is a practical issue for all researchers dealing with the analysis of interaction. Even though there appear to have been attempts to develop a tool aiding the conversion from ELAN annotations to text and screenshots (Clair-Antoine Veyrier, 2013, personal communication), I am not aware of such a tool that is functional. This time-consuming process thus remains a manual task. In this subsection I will describe my own workflow for this task. We start out with a stretch of interaction readily annotated in ELAN, with the tiers comprising all relevant information. For spoken conversation, this may comprise tiers for vocal behaviour, a word-by-word gloss and a translation of vocal behaviour and gesture phases. For signed conversation, the tiers would include, for example, manual movement phases, a word-by-word gloss for signs, and a translation. In the case of spoken conversation, I find it most intuitive when visible behaviour is aligned to the vocal behaviour tier (see Figure 3), because viewers should be able to infer the approximate length of a movement when put into relation to a stretch of talk. In transcripts of signed conversation a time axis gives the viewer a rough idea of the length of a specific sign or movement phase (see Figure 5).

In this example I will use a transcript from spoken conversation in Japanese. The goal is to create a transcript in which visible behaviour is aligned to the vocal behaviour tier. I start out by taking screenshots of points in interaction that will be included in the final transcript and by marking in the speech tier where they originate. Next, in the speech tier I mark the points in time of onsets and endpoints of the kinds of visible behaviour that are to be included (Figure 2). The resulting vocal/visible behaviour tier hybrid looks like this:

/uchi /mo: #/ha*:*-* /haha#oya ga (.) mo/rota/ra# /tsukeru (.) itsu/mo.

In my own personal convention, slashes represent gesture phase onsets (the very last one representing the endpoint of a retraction phase) and hashes represent the timing of screenshots. Other symbols (such as ‘(.)’ and ‘**’) are part of vocal speech transcription.

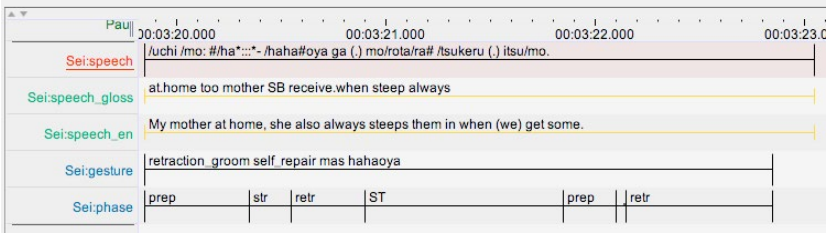


Figure 2: Annotated stretch of interaction in *ELAN*

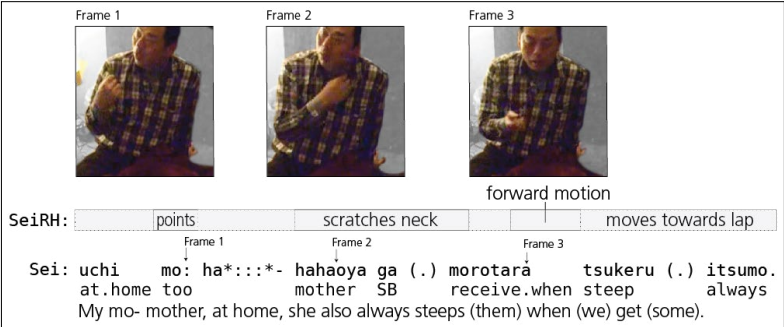


Figure 3: Finished transcript of spoken conversation (appears as Excerpt i in Study 4)

This vocal/visible behaviour tier hybrid, along with the word-to-word gloss and the translation, is copied into an empty Adobe Illustrator document, where the visible behaviour is then manually added in an annotation-like fashion, while paying attention to the alignment between vocal and visible components. Also, arrows are added above the visible behaviour tier, indicating the point in time where screenshots originated.

The screenshots are edited in Adobe Photoshop in such way that the participants' bodily and manual configuration are easily discernible. The edited screenshots are then placed in the Illustrator document, resulting in a transcript as shown in Figure 3.

For transcripts of signed interaction, I have a slightly different procedure. I take a screenshot from *ELAN* of the part of the interaction that is to be included in the transcript. This screenshot is inserted into an empty Illustrator document and used as a template for creating rectangle objects representing manual movement phases (Figure 4).

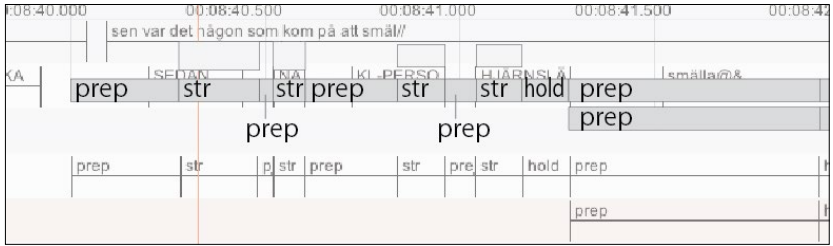


Figure 4: Producing transcripts from signed conversation

After this step, I add tiers for the word-by-word gloss, the translation and a time axis that is also alined to the underlying template from ELAN. Screenshots of the participants are then placed into the Illustrator document. A portion of the resulting transcript is depicted in Figure 5.

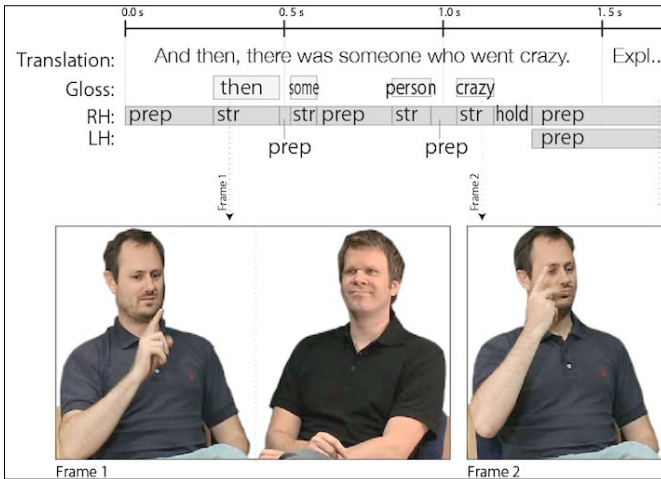


Figure 5: Finished transcript (appears as Excerpt 1 in Study 2)

This process is time-consuming, and will hopefully not be necessary any longer once we either have a functional conversion tool at our disposal or leave behind paper as a medium for the presentation of video data.

CHAPTER 3

Key concepts in social interaction and visible behaviour

This chapter is largely dedicated to an overview of the analytical terminology used herein, including my own novel terminology introduced in the studies. I will also discuss analytic concepts with regard to how they relate to individual studies of this thesis.

3.1 The description of manual movement phases

When we use the words *gesture* or *sign* in everyday conversation, we usually mean a temporally delimited stretch of manual movement that we recognise as a meaning-laden entity. But when we look at manual movement as (part of) utterance within social interaction, we will find that timing and alignment matter in all stages of manual movement; from the point in time when the hand starts to travel towards the place of articulation until the moment when the hand reaches a relaxed position.

Adam Kendon's notation for what he refers to as *gesture phases* is the most widely recognised notation for describing these stages of manual movement. This notation is adopted in this thesis, albeit in a slightly extended form. Also, I refer to these phases as *manual movement phases*, a more general term that does not run risk to be understood as excluding movement phases seen in signed conversation.

In manual movement, Kendon (2004, p. 111) states, the hands undertake a *movement excursion*. Within this excursion, the hands move from a relaxed position towards the place of articulation, a transitional movement referred to as the *preparation phase*. The main, expressive part of the manual movement is called the *stroke*, and it is this segment of the movement excursion that is recognised as “the gesture” in every day terms, i.e. the segment that comes to be seen as carrying meaning. The segment during which the hand is brought back to some relaxed position is generally called the *retraction* (cf. McNeill, 1992), even though Kendon (2004) himself refers to it as *recovery*. Finally, a segment in which the hand stops and freezes within a movement excursion is called a *hold*.

Kita et al. (1997) have added some further distinctions, which were subsequently adopted by Kendon: namely, the *pre-stroke hold* (a hold phase that occurs before the stroke) and the *post-stroke hold* (a hold phase that occurs after the stroke). Also, Kendon sets forth the notion of the *nucleus*, a segment composed of the stroke and any post-stroke hold following it, stating that the nucleus is “interpreted as that part of the action that carries the expression or meaning” (2004, p. 112).

This notation of manual movement phases can be applied to manual movement as part of both spoken and signed conversation. It should be noted that, confusingly, similar notions are prominent in the field of sign language phonology. In this field, the main focus of investigation has been on movement phases that constitute a lexical item – that is, movement segments that help to describe the composition of the stroke phase of a sign. According to Liddell and Johnston (1989), any given sign is seen as composed of a sequence of movements and holds. For instance, a sign in which the hands move from one position to another position is said to be composed of the sequence *hold – movement – hold*. The notions of hold and movement in this line of research denote segments that are internal to the production of a given sign, i.e. they are part of what is referred to as a stroke in Kendon’s terminology. The notion of movement and hold are solely intended for the phonological description of a sign. Movement and hold segments that occur for reasons other than the

production of lexical items (and which thus are not necessarily part of the sign) are not taken into account. For the sake of clarity it should be noted that this is not the way these terms are understood in this thesis.

Finally, the position of relaxation that the hands assume before and after a movement excursion is called *rest position* (Kendon, 2004; McNeill, 1992) or, as in this thesis, *home position* or simply *home* (Sacks & Schegloff, 2002).

3.2 Extended terminological apparatus

In the course of examining and annotating video data, I frequently came across segments of manual movement that were of a disputable status. In particular, I was quite frequently faced with segments of non-movement that could be classed as neither holds nor home. They could not fully be classed as holds given that (1) they are often preceded by a retraction, even though the hand does not go all the way to a relaxed home position, and that (2) the manual configuration, i.e. the hand shape formed by the digits, that has been apparent during the stroke is released – that is, some *local* relaxation has taken place. On the other hand, they do not qualify for home position, given that the arm and hand are still maintained in position somewhat away from the body. I termed this sort of segment *provisional home position* (as in Study 1), given that it is a bodily configuration primarily assumed by participants who are momentarily not producing talk. However, in later studies (Study 2, 3) I refer to this segment as a *half–hold*. This, then, contrasts with the canonical Kendonian notion of a hold, in which the manual configuration is preserved. I call this segment a *full–hold* (Study 2, 3) and, in cases in spoken conversation where such a hold extends over the boundaries of the vocal portion of an utterance, a *prolonged hold* (Study 1).

Provisional home position and half–hold essentially refer to the same phenomenon: a segment of non-movement in an intermediate position that foreshadows change in either direction. The shift in terminology was made in order to make it more consistent and give a clearer idea of what the segments are about. In addition, the terms full–hold and half–hold are better suited to

reflect the idea that segments of non-movement are not binary (i.e. either a hold or home position), but can occur in stages anywhere between the location where the stroke occurs and home.

There are a few researchers who make similar distinctions. Kendon (1980) uses the term *partial recovery* in order to denote a transitional phase in which “the hand does not return all the way to the position it was in” (1980, p. 213) before the preparation.¹³ In Baker’s (1977) early work on signed conversation, she proposes a set of terms – *full-rest*, *half-rest* and *quarter-rest* – that are fine-tuned descriptions of possible positions in which the hands are in some way sustained by the body or a surface, whereas my own terms denote manual states in which the limbs are supported solely by the arms in mid-air.

Similarly, André (2012, p. 151) uses the term *intermediate home position*, which is similar to my own. Also, DeStefani (2007) distinguishes between maintaining a gesture (*maintien gestuel*), corresponding to full-holds, and suspending a gesture (*suspension gestuelle*), corresponding to half-holds. Furthermore, in her work on practices of dropping out or withdrawing from an overlap (i.e. two or more participants speaking at the same time), Oloff (2013) finds that this practice is organised in a number of stages from partial to full withdrawal. She refers to intermediate stages as “being on standby”, during which participants show that they are anticipating the next possible slot for gaining speakership. These segments of manual non-movement occur with different degrees of muscular tension/relaxation and in different positions relative to the body of the participant and the location where the stroke has taken place. While there is a term for the position of the hands when they are relaxed (home position), there is no established term denoting the location in space where the hands are showing, regardless of what kind of movement phase they are involved in. In the studies I developed the term *stage* to denote this location, in analogy to its everyday meaning as a place where people perform in public in order to

13 Note, however, that Kendon’s (1980) term denotes a movement phase that results in this position. My own term, in contrast, denotes the manual configuration (position and hand shape) resulting from such a partial recovery.

be seen and noticed. Thus, when the hands are on stage, they are performing in some sense, regardless of whether they are enacting, moulding, drawing, sketching etc., and regardless of whether they are moving or motionless.

It should be kept in mind that the stage and home are considered two ends of a continuum rather than a binary distinction, i.e. the hands may stop to move on stage or in any location on their way between stage and home. The hands may also take a detour to touching the face or the head, as shown in Study 4 regarding self-touching behaviour. This can also be considered part of the stage-home continuum, given that keeping the hand in the area around the face or the head often qualifies as a provisional position, a segment during which the hands are neither performing nor relaxed.

McNeill (1992) devised a scheme for coding the position in space where a gesture occurs, called the *gesture space* (see Figure 5). It is rendered as a two-dimensional area with an x-axis (for horizontal movements) and a y-axis (for vertical movements). In a way, the terms stage and home complement this scheme with a z-axis for movements away from and back to the speaker.

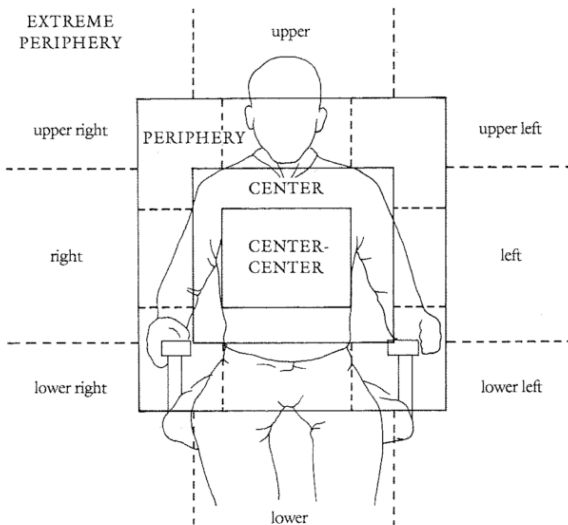


Figure 5: Gesture space according to McNeill (1992, p. 89)

Finally, there is another term concerning movement phases, namely, the *prolonged stroke*, introduced in Study 2 on signed conversation. It denotes a segment of movement on the stage in which the hand shape, position and movement of the sign are preserved for a markedly extended period of time. Note that a prolonged stroke cannot be applied to all manual actions, given that some do not feature motion as part of their stroke (such as pointing). McNeill (2005) refers to such motionless strokes as *stroke hold*, pointing out that “such movements are strokes in the sense of meaning and effort but occur with motionless hands” (2005, p. 32). In cases where such a stroke hold is preserved for an extended amount of time, they are virtually equivalent to post-stroke holds, i.e. it is not possible to delimit where (and if) the stroke ends and where the post-stroke hold begins.

On the other hand, in manual expressions that feature movement, it is possible to preserve the movement pattern for an extended amount of time by slowing down the movement or repeating it more often than usual. Only in such cases is it possible to speak of prolonged strokes.

3.3 Participant practices involving manual movement

Janet Bavelas is one of the more famous proponents of the idea that gesture is not always part of utterance construction, but that it also serves interactive functions and that gesture “may be tailored for a particular addressee in a particular conversation” (Bavelas, 1994, p. 206). What she and her colleagues refer to as *interactional gestures* resemble to what I deal with in the studies, in the sense that they “assist the dialogue itself rather than serving semantic or syntactic functions” (Bavelas, Chovil, Coates, & Roe, 1995, p. 404). The practices investigated in this thesis are employed as (part of) social actions, i.e. by definition, they can only occur within interaction. Bavelas’s studies focus on conventionalised gestures, i.e. on the stroke phase of manual excursions, that are employed as social practices, whereas my own studies primarily zoom in on phases of manual excursions other than the stroke and on how they may constitute social actions in their own right.

Along these lines, other researchers have shown that manual movement is related to turn-taking – e.g., it may foreshadow an utterance or indicate that an action has ended. In this subsection I review literature suggesting that participants employ manual movement not only as part of utterances, but also in order to regulate the interaction itself. The literature reviewed here deals primarily with manual movement as it occurs in spoken conversation.

Sciubba (2010) takes bodily movement into account in relation to the organisation of turn-taking and reports two practices: (1) when participants end, or are about to end, an utterance, they may withdraw or “undo” whatever visible behaviour has been prominent during the utterance and (2) participants may show their imminent speakership by virtue of visible action that foreshadows a vocal utterance. She predominantly analyses manual movement, but also points out that relaxing the body posture, closing the eyes and smiling may be employed in order to show that an action is about to end. She also observes that participants who are about to produce an utterance may indicate this by moving their hands towards the stage and keeping them there (i.e. employing a pre-stroke hold), often at times when another participant is still speaking. Very similarly, Streeck (2009) discusses the role of manual movement in foreshadowing further actions to come, a practice that he calls *forward-gesturing*. He points out that participants can project, i.e. foreshadow (see section 3.7), incipient speakership by moving their hands towards the stage. Manual movement can also display a participant’s stance (or propositional attitude) in what they are going to say. He observes that such actions involving manual movement may appear well before the participant in question produces a vocal utterance.

Streeck also presents a case in which the hands are held on stage after utterance completion “as unfilled time passes during the conversation and its progression is suspended” (p. 175). Such held hand shapes manifest within the interactional structure of a sequence (i.e. a chain of interrelated social actions; see 3.6), namely that the production of an initiating action requires a responsive action in order to be completed. He concludes that (1) human interaction is forward-oriented and that (2) multimodal resources are deployed in order to foreshadow foregrounded behaviour through backgrounded modalities.

Clark (2005) analyses sustained pointing gestures used for directing the attention of another participant to something or someone and remarks that participants “often maintained their touching, giving it a time course that they exploited for communicative purposes” (p. 511).

He identifies three phases within this practice: (1) initiation: “I now want you to attend to this”, (2) maintenance: “I continue to want you to attend to this”, (3) termination: “I now consider your attention to this to be complete” (p. 511). In the case of pointing gestures, the initiation phase corresponds to moving the hand towards the stage, the maintenance phase to holding the pointing hand on stage, and the termination phase to the retraction towards home. Clark points out that the termination only occurs when the recipient of the gesture produced some kind of acknowledgement (such as “m-hm”). Thus, the retraction phase of manual movement may in many cases be a matter of whether the recipient has registered the action.

As Clark argues, the termination of a manual visible action, i.e. the retraction phase, has a potential of gaining an *in situ* meaning. Related to this, Andrén (2012) finds that this movement can be tweaked in a way that emphasises its meaning of completion: Participants employ stylised manners of retraction, *marked retractions*, that mark the completion of an activity. The notion of marked retractions captures the fact that these transitional movements may have a stroke-like or gesture-like character in their performance. Transitional phases, such as the retraction phase, are not limited to being practical movements, during which the hand “merely” returns to home, but they can be read as having an additional layer of meaning: Retracting the hand implies that the activity in which the hand has been involved or where it originates has finished, and performing the retraction in a marked manner puts emphasis on the fact that the activity is complete.

This is also illustrated by Hosoma (2009), who investigates hold phases in manual movement and relates them to the production and structure of the actions by other participants. He refers to gestures that feature a hold phase extending partly over the boundaries of the vocal portion of the utterance as *grand gestures* (presumably due to the fact that they delimit larger units) and in a later work *extended gesture unit* (2011). His point of departure is a

model of discourse structure consisting of three stages: initiation, response, feedback. He finds that manual movements employed by the initiating party are at times held until a recipient produces a response, or even until the initiating party gives feedback. Furthermore, he shows that (1) a manual hold by an initiating party can be mitigated, i.e. moved slightly towards home, as a reaction to silence, i.e. to the fact that a response is noticeably absent, and that (2) a manual hold by an initiating party can be retracted as a reaction to a negative response. Given that a manual hold originating in an initiating action can extend over a second action while referring back to the first, this study challenges the notion of an utterance as a clearly delimited and readily analysable entity. It also suggests that manual movement is not just related to producing gestural meaning itself, but also constitutes a display of understanding of a whole interactional unit, so that the retraction phase may show the completion of such a unit rather than merely the termination of the gesture itself.

Phases of manual movement may also pertain across longer stretches of talk by different speakers, as illustrated in Mondada (2007b). She identifies various resources that participants make use of in order to negotiate turn allocation, i.e. an indication of which participant will be allotted speakership. One of these resources is constituted by *persisting pointing gestures*. Mondada states that “speakership [...] is not strictly limited within turn (verbal) boundaries, but [...] interactively shaped” (p. 215). Persisting pointing may imply a categorisation of the sequence as being an insertion or momentary suspension (p. 216). Participants design the span of their actions and indicate whether they have been responded to in an adequate way, as they are responsible for the sequence they have initiated.

Thus, visible actions, such as pointing, also have an interactional dimension which is relevant for turn-taking. The setting investigated in the study – meetings in which the participants work with cartographic material – has a strong focus on deictic activities, which, in turn, legitimises an analysis of pointing gestures as foreshadowing speakership. As a consequence, different settings in which different activities are relevant could make available other resources for the organisation and projection of turn-taking.

Andrén (2011) approaches the issue of extended units of manual action from a slightly different angle: He investigates children's pointing stroke endpoints set in relation to the kind of response the interlocutor (the parent) gives. The data are coded according to the relative length of the hold phase (short, between turns, sustained), the degree of response (no response, minimal response, expanded response) and, in cases of sustained pointing gestures, according to the communicative effort (plain hold, renewed stroke, upgraded renewal). He finds that (1) the parent tentatively gave more elaborate responses when the child employed sustained pointing gestures and that (2) in cases of sustained pointing, receiving a (satisfactory) response is the main condition in order to withdraw the pointing. His findings suggest that what the children are seeking is not solely to achieve shared reference to certain things around them, but also to receive evaluations or comments on the object that is being pointed to. Also, given that the children's communicative apparatus appears to comprise this practice, it should be considered a quite fundamental one to human interaction in general.

3.4 Hierarchically structured levels of activity

It was said above that participants in interaction may indicate their engagement or involvement in the conversation by virtue of manual movement. Involvement is a term borrowed from Goffman (1963) who discusses involvements in various activities, such as walking, knitting, smoking, reading or humming. Involvements can be divided into main involvements, which absorb the major part of an individual's attention, and side involvements, which an individual can pursue on the side without threatening the maintenance of the main involvement (p. 43). Furthermore, Goffman distinguishes between dominant and subordinate involvements. A subordinate involvement can be continued until a social occasion obliges an individual to cease that activity and focus on something else. For example, "while waiting to see an official, an individual may converse with a friend [or] read a magazine [...], only until his turn is called" (p. 44). The individual then is obliged to set aside the activity, even if it is unfinished.

I proposed the term of the stage, which is opposed to home, and argued that these two positions should be considered respective ends of a continuum. Depending on where along this continuum a segment of non-movement occurs, we find varying degrees of claim of speakership or displays of involvement in the interaction. Such segments occurring in an intermediate position between the stage and home can often be observed in contexts where the projected trajectory of an activity or line of action is interrupted, paused or suspended due to another activity that is imitated by another participant.

Participants can thus show which activity they are involved in and which one of these is ranked higher for the time being. In such cases we find hierarchically structured levels that participants deal with, and they manage to do so by orchestrating and neatly adjusting visible resources. This is not only achieved by virtue of manual movement but includes other kinds of visible behaviour, such as body posture (as discussed briefly in Study 1).

This is, however, not a novel idea. Schefflen (1964) argues that bodily behaviour, in this case body posture, is linked to interaction and reflects both the relationship between the interactants as well as what kind of activity they are currently engaged in. As a preliminary observation, he points out that “human behavior can be communicative whether or not it is intended to communicate” (1964, p. 318). He argues that, on the one hand, there seem to be postural shifts that are indeed read as social actions, e.g., shifts in which the body moves away from the other participant(s) appear to indicate completion and temporary disengagement (p. 324). On the other hand, Schefflen points out that that posture shifts have no meaning in themselves but only when viewed in relation to their context of occurrence: “social investigators must avoid a glossary in which one kind of postural shift ‘means’ this and another kind means that” (ibid).

In his analysis he identifies several hierarchically arranged levels of shifts, from subtle to gross movements and with tentatively rising durations: (1) the point (turnings of the head, gestures), (2) the position (shifts in which at least half the body is involved) and (3) the presentation (sitting down, standing up or leaving the room). Each of these levels of structural units corresponds to minor or major interactional shifts, turns or changes of roles.

Based on Schefflen's findings, Kendon (1972) presents a more fine-grained analysis of phases of bodily movement in relation to utterance structure. He analyses a longer stretch of talk by a single speaker and divides it into prosodic phrases, locutions, locution groups and locution clusters. These units are delimited by speech-based cues such as change in pitch and change in loudness and are hierarchically ordered. He finds that patterns of body motion are associated with these units. For instance, where the speaker "changes from one cluster to the next, there is movement in the trunk and legs which does not occur within clusters" (1972, p. 192). Also, he finds that "the larger the speech unit, the greater the difference in the form of movement and the body parts involved" (p. 205). According to Kendon, speech units, such as locutions, as well as what he refers to as "very high level units, such as discourse or listening" (p. 205) are part of the same continuum.

These different hierarchical levels can be split between different foci or activities that go on at the same time, as shown by Schegloff (1998). He analyses a bodily configuration that he refers to as *body torque*. When a participant's body is *in torque*, the upper body faces into a different direction than the lower body. The respective orientations of the upper and lower body are associated with different activities or involvements, the lower body facing towards the main activity and the upper body to an activity that is being inserted into or interruptive to the main activity. Thus, the configuration of the body provides an indication about the ranking of several ongoing activities. In contrast to the sort of bodily shifts presented in Schefflen (1964) and Kendon (1972), respectively, body torque projects change and is interpreted as a temporary shift, associated with an activity that is categorised as inserted into the main activity. When several activities intersect, it is up to the individual participant to decide how to proceed. Raymond and Lerner (2014) point out two practices for adjusting action. Participants involved in a practical activity (e.g., a clerk managing payments at the check-out counter or a participant involved in preparing food) who are faced with the additional involvement of interacting with other participants overtly employ forms of adjusting actions to produce a different relation between these involvements. They do so by (1) momentarily suspending one activity in order to pursue the other or (2) retarding an activity, i.e.

slowing down the progressivity of one activity while showing some engagement in the additional activity. Participants thus show that they sustain “a visible commitment to an erstwhile ongoing course of action while pursuing a second course of action” (p. 243).

3.5 Perceptual layers in interaction

Participants are capable of disassembling a multimodal action into its constituents, i.e. vocal components (such as verbal content, prosody, pitch, loudness or speed) and visible components (such as body posture, gaze, head movement or manual movement; cf. Kendon, 1978, 2004). They can also engage in meta-communication about these constituents, for instance, by commenting on someone’s prominent hand movements, notable accent or intense gaze. Participants are also able to differentiate between actions such as manual movement which are employed as part of a proposition and those which are purely practical, i.e. unrelated to the ongoing discourse.

Even though it should be clear that participants are not constantly actively engaged in scrutinising another participant’s ensemble of actions, given the steady flow of information pouring out of an interlocutor, percipients still have to separate the wheat from the chaff. That is, they are faced with the task of differentiating between what segments or portions in the flow of an utterance or action are accountably produced as social actions, to be recognised, and thus relevant within the local contexts, and what segments can be disregarded.

The stream of information is broken down into those portions that are perceived as being part of what a participant is conveying (i.e. the content) and those that are unrelated to it. Goffman (1974) refers to these portions as tracks that flow side by side in interaction. He distinguishes between the *main attentional track*, that is, the content of an activity that participants routinely attend to, and the *disattend track*, which comprises any activity that is systematically blocked out or “blotted out” as Goffman puts it. He also points out that there is a “stream of signs which is itself excluded from the content of the activity but which serves as a means of regulating it, bounding, articulating, and

qualifying its various components and phases” (Goffman, 1974, p. 210).¹⁴ I refer to this as the *regulatory track*, and segments of behaviour that are associated with this track are called *regulatory behaviour*.

Kendon (1978) picks up this idea and focuses on the question which components within the flow of signs in interaction are apt to be considered central, from a participant’s perspective, and refers to the ability to view social actions as an intermeshment between several mutually attuned systems as *differential attention*. He argues that the flow of behaviour is perceived by participants as a number of different systems of action that are separable, and that some of these systems (in Kendon’s view, vocalisations and speech) have a special status within an *attentional hierarchy* (Kendon 1978, p. 309).

Suffice it to say that, from a participant’s point of view, behaviour is neither produced nor perceived equally, but in a differentiated fashion. There are segments of behaviour that stand out as significant or central, other segments that are not counted as part of the interaction, and still other segments that are not part of the content *per se* but serve to regulate the flow or the back and forth within a given interaction.

Regulatory behaviour is thus not *a priori* backgrounded, but can be backgrounded or foregrounded depending on how a participant performs, enacts and packages it. Foregrounded utterances such as “Hold on a second!” are employed in order to put the ongoing conversation on hold. They not part of the content, but rather part of regulating the flow of the interaction and evidently being attended to. Another example is constituted by what are referred to as interactional gestures (Bavelas et al., 1995; Bavelas, 1994), i.e. foregrounded gestures that are employed solely for regulatory purposes.

On the other hand, we find segments of behaviour that are apt not to be noticed and whose regulatory meaning emerges locally within the context. The phenomena analysed in this thesis are situated within this sphere. For

¹⁴ Goffman labels this track the *directional track*. He is somewhat inconclusive regarding the positioning of this track: He places it outside the content of the activity (1974, p. 210), but further below mentions that directional cues cannot be blotted out, “for these must be kept in mind enough so that they can do their work” (p. 214). He does not seem to distinguish between foregrounded, i.e. attended, segments of behaviour that serve for purely directional/regulatory purposes (such as interactional gestures mentioned further below) and backgrounded regulatory behaviour (such as manual movement phases).

instance, acts such as stroking the hair or touching the face are usually perceived as a self-directed, private involvement unrelated to the ongoing interaction (cf. Ekman & Friesen, 1969). But such acts may be ascribed an interactional meaning by virtue of its sequential positioning and at times exhibits traits of utterance-dedicated visible action in certain contexts, such as conveying uncertainty.

3.6 Adjacency and sequentiality

In the individual studies I analyse (phases of) manual moment in relation to what else is happening in the unfolding interaction. Participants in interaction produce social actions (e.g., in the form of utterances), and these are understood in relation to one another, i.e. in a *sequentially* ordered manner. I argue that participants orient not only to the completion of each action individually (for instance, utterance completion), but also to the completion of “lines of actions” or “chunks of actions”, as I refer to them in the studies. This roughly corresponds to the notions of *adjacency pairs*. This section is dedicated to introducing and discussing these concepts.

3.6.1 Adjacency pairs

In social interaction, actions follow one another and are interdependent. Although there are cases in which several actions or involvements are ongoing at the same time, let us for now focus on the idea that actions occur in sequence. In spoken conversation, these actions may be mainly vocally produced utterances, they may consist of ensembles of a vocal portion and visible portion, or they may be purely visible (such as waving a hand, walking into a shop or putting on a jacket, to name a few).

According to conversational analysts, the basic structure of actions is the *adjacency pair*, based on the observation that actions, minimally, come in pairs: an initiating action, referred to as *first pair part*, uttered by the first speaker, and a responsive action, referred to as *second pair part*, uttered by the second speaker (Sacks et al., 1974). These actions, as a minimum, constitute a sequence. Additionally, sequences can be closed with a closing action, referred to as *sequence closing third*, uttered by the speaker of the initiating action. The

most stereotypical example is an enquiry such as “What lightbulbs do you use at home?”, followed by a response such as “Generic forty watts lightbulbs”, in turn followed by an acknowledgement as a closing action such as “Oh, okay”. The concept of adjacency is omnipresent in conversation. An initiating action makes relevant the production of a responsive action, and when no responsive action appears to be forthcoming, i.e. when it is *recognisably absent*, participants might assume that the recipient failed to perceive it, so they may re-launch the initiating action, comment on or sanction the fact that no response is given (“Hey, don’t ignore me!”) or laugh.

Any utterance or action that exceeds two members is considered sequence expansion (Schegloff, 2007). Expansions are constructed in relation to a *base sequence*, i.e. the adjacency pair that is most central to the sequence. Its constituents are referred to as *base first/second pair part*. There are several types of expansions: *pre-expansions* (before the base sequence), *insertion expansions* (between base first and second pair part) and *post-expansions* (after the base sequence).

Note that, according to Schegloff (2007), the minimal structure of a sequence consists of two members. A sequence closing third (such as “Okay” and the like) is considered a post-expansion, a concept that is not always intuitive. Some studies that take into account visible behaviour as part of interaction employ sequential formats that include three members rather than two. As mentioned above, Clark (2005) refers to these as initiation, maintenance and termination, and Hosoma (2005) adopts the three stages of initiation, response and feedback. Also, my own analyses suggest that participants orient to a sequential format with three members: In prolonged holds, the retraction phase is often associated with giving feedback, sometimes in orchestration with nods or vocal feedback engendering closure (Study 1). See also Linell (2009, p. 184 ff.) who discusses the limitations of the notion of the adjacency pair in a similar way.

3.6.2 Insertion sequences: temporarily changing trajectories

Insertion sequences are sandwiched within the base sequence, i.e. between the base first and the base second pair part. Insertions are interruptive to the main line of action or trajectory in the sense that the production of the second pair part is delayed. Some of the analyses in this thesis consider participant practices that reflect or occasion shifts back and forth between an insertion and the main trajectory to which the participants will eventually return. These participant practices manifest within the hierarchical structure that can be found in visible bodily behaviour (encompassing posture, manual movement, and the orientation of the torso, head and gaze) discussed above.

A prominent example of insertions are repair sequences (described further below). A sequence squeezed in between the base first and second pair part implies that the constituents of an adjacency pair are not always adjacent to one another, so that adjacency is a typical, but not a required property of adjacency pairs. Imagine, for instance, that a friend of mine asks me whether I want to go to see a film in the evening. I could respond by accepting or declining directly, but I could also play hard to get and say “I don’t know, do they have good films?”, thereby expanding on the initiating action (the offer) and postponing my responsive action towards the base first pair part. When convinced about the enterprise, which could be five seconds or five minutes later, I might go ahead and say “Okay, let’s go”, thereby referring back to the base first pair part. Repair sequences are a particular kind of insertions. In conversation, participants are faced with various kinds of *trouble*, such as mishearing, misunderstandings, not being able to find the right words and the like. Repair can be analysed in phases: the initiation and the completion of repair. During the former a participant indicates the trouble and during the latter a participant gives a candidate completion.

These two stages may be carried out by the same participant, or one participant may initiate the repair, and another may complete it, or vice versa. We can then speak of a participant repairing their own talk as doing *self-repair* and a participant repairing another participant’s talk as doing *other-repair*. Furthermore, *self-initiation* denotes cases in which the a participant initiates the repair on

their own utterance, and *other-initiation* implies that a participant initiates repair on another participant's utterance (Liddicoat, 2007; Schegloff, 2007). The combination of these two distinctions thus yields four categories:

- (1) self-initiated self-repair: a participant initiates and completes repair on their own talk
- (2) self-initiated other-repair: a participant initiates repair on their own talk, whereafter another participant completes it
- (3) other-initiated self-repair: a participant initiates repair on another participant's talk and the latter completes it
- (4) other-initiated other-repair: a participant initiates and completes repair on another participant's talk

An inherent feature of repair sequences is that they constitute an insertion that is interruptive to the original trajectory or agenda of the interaction. Participants thus need to deviate from this original trajectory to deal with the trouble, temporarily prioritising the newly emerging trajectory over the original one. Upon repair completion, participants shift back to the original agenda.

Studies 2 and 3 attempt to shed light on this issue. The studies deal with the ways insertion sequences, such as repair sequences, are handled in signed conversation, also in comparison to spoken conversation. It was found that the individual stages of shifting back and forth between main agenda and insertion are closely intertwined with manual movement phases in both spoken and signed conversation. Also, Study 4 presents instances in which the boundaries of a repair sequence are associated with segments of self-touching behaviour (e.g., touching the hair), which at the same time gives a fresh perspective on seemingly self-directed or private segments of behaviour, and suggests that they are employed systematically. This sort of manual movement is regulatory to the flow of conversation in a seen but unnoticed fashion.

Research on the organisation of repair in spoken conversation has generated a substantial corpus of literature (e.g., Fox, Hayashi, & Jasperson, 1996; Schegloff, Jefferson, & Sacks, 1977; Seo & Koshik, 2010; Shimotani, 2007; Wong, 2000). Also, very recently there have been studies investigating how

repair is organised in signed conversation (Floyd, Manrique, Rossi, & Torreira, 2014; Groeber & Pochon-Berger, 2014; Kikuchi, Bono, & Otsuka, 2011; Kikuchi & Bono, 2012; Manrique & Enfield, 2015).

3.7 Projection: foreseeing what will happen

Participant action can be foreshadowed in various ways. For instance, if I observe a participant who is indoors putting on their jacket and shoes, I may predict that they are on their way out. When someone on the street comes walking straight up to me in a direct trajectory, I will assume that they have some business with me.

An action can always be seen as a preparatory move towards the next, i.e. a given action foreshadows the next. This foreshadowing is referred to as *projection*. The next step within a chain of actions is thus *projectable* by virtue of assessing the local context of the preceding action.

I once heard an anecdote about somebody who had a party going on at his place, and at some point he became tired and wanted to go to bed. One could think of a few strategies in order to hint to the guests that it is time to leave, such as asking them to help with tidying up, telling them that you have a long day ahead of you or, a rather blunt solution, thanking them for a marvellous evening.

The host, however, chose a quite peculiar yet effective strategy to make the guests leave: He went to the bathroom to retrieve his toothbrush and began to, publicly, brush his teeth. Regardless of how this strategy was received, it should have been clear to the guests that this was not an invitation to open another bottle of Bordeaux.

It should be kept in mind, though, that an action such as brushing one's teeth does not have an intrinsic interactional meaning when viewed in isolation (of course, it has meaning in itself given that it has the purpose of dental hygiene, but this is not the kind of meaning that is of importance here). What kind of meaning is ascribed to such actions chiefly depends on the local context within which they are produced. Social actions gain meaning due to the very fact that they are part of a repertoire of social practices that participants take for granted (cf. Kendon, 1990, p. 15). By extension, projection depends on prior

familiarity with these social practices, i.e. a next action can only be projected provided we have either gone through, observed or logically induced a given chain of actions or a sequence. We thus know that dental hygiene, the context described above, most likely constitutes a preparatory action for retiring to bed which, in turn, should effectively discourage further party-related activities.

Projection is also possible in smaller units, such as the syntactic and prosodic structure of an utterance, which is one of the resources participants draw on to foresee about when a speaker's utterance is about to end (Auer, 2005). The structure of a manual movement excursion also constitutes a resource that allows projection. Participants are familiar with the recurrent pattern of a manual excursion: the hand is brought towards the stage, performs a stroke (the main part of a manual movement) and is subsequently brought back to home. A movement towards the stage thus makes relevant a stroke as the next step, and hands on stage make relevant a movement back towards a relaxed home position.

In both spoken and signed conversation, hands-on-stage is associated with speakership, as pointed out in Study 1. Thus, a movement towards the stage projects a manual excursion and thereby speakership. As set forth in several studies, including this thesis (Andrén, 2011; Cibulka, 2015; Hosoma, 2009; Sidnell, 2005), participants in spoken conversation regularly keep their hands on stage even when they are not talking, specifically after the end of the vocal portions of their utterances. They also make use of provisional home positions or half-holds, displaying that they are on standby and anticipating the next possible slot for resumption of their contribution.

This can be viewed as an extension of speakership, and different positions between stage and home project different degrees of speakership. In the thesis I raise the question of whether speakership constitutes a binary concept (i.e. speaker or non-speaker) or whether it is best viewed as a continuum ranging from assuming a speaker role to different degrees of claims of speakership and assuming a non-speaker role (Studies 1 and 2).

It should be kept in mind that a projected trajectory is by no means predetermined. Participants may alter or adjust their actions, but they will often be constrained by the relevancies set up by their prior actions (Linell, 2009,

p. 180). Manual movement phases are to some extent projectable given that a manual excursion follows a recurrent pattern, from home to the stage and back home. When a participant in spoken conversation lifts their hand in preparation for a manual movement in coordination with speech, it projects one or several strokes followed by a retraction back home. Such a projected excursion may at times be interrupted, for instance due to another participant speaking in overlap. In such cases, the vocal portion of the utterance so far can easily be brought to a premature end by producing a cut-off, i.e. in most cases a glottal stop. When a manual movement excursion so far is prematurely brought to an end, participants regularly choose to adjust their action by transforming the manual movement into a practical or self-directed action, such as self-touching (as shown in Study 4). Self-touching behaviour is perceived to be self-directed and to satisfy personal needs. It is “by definition unrelated to the ongoing dialogue” (Ekman & Friesen, 1969). This makes self-touchings a inconspicuous and less accountable way of fading out a prematurely discontinued action. Perhaps it is precisely due to the very fact that self-touching behaviour is regarded as “out of frame”, i.e. unrelated to the content of an activity, that it manifests in the form of a display of disengagement from the activity. Self-touching behaviour is thus a way for participants to adjust their already launched visible action so far and transform it into another action on the fly.

CHAPTER 4

Summarising the studies

STUDY I

When the hands do not go home: A micro-study of the role of gesture phases in sequence suspension and closure

This study, along with Studies 2 and 3 summarised below, deals with the organisation of manual movement phases as segments of behaviour that gain interactional meaning by virtue of their local context of production. In particular, I zoom in on phases of non-movement, namely the *prolonged hold* and the *provisional home position*. The impetus for pursuing the interactional working of manual movement phases emerged in the course of annotating movement phases in my video data during which I was faced with segments that did not fit the canonical description used in the majority of gesture-related studies (e.g., Kendon, 2004; McNeill, 1992).

In the study I argue that movement phases may be exploited by participants in order to show how a pursued trajectory or line of action is adjusted according to the unfolding interaction. Participants can indicate that a projected line of action is maintained by holding the hand shape and position, that it is suspended (i.e. set aside for the moment) by slightly retracting the hands and relaxing the digits, and that it is discontinued or abandoned by moving the hands to a relaxed position. Also, through constant monitoring participants may adjust the location of the segment of non-movement as the interaction

unfolds in time. These locations ranging from stage to home are viewed as a continuum, in which various possible locations reflect various degrees of involvement or claim of speakership.

This study shifts attention to more backgrounded phenomena in interaction, i.e. segments of behaviour that, from a participant perspective, are not part of the content of a given utterance and that are not routinely attended to, proposing a way to deal with them in an analysis of visible behaviour.

STUDY 2

On how to do things with holds: Manual movement phases as part of interactional practices in signed conversation

This study essentially argues the same way as Study 1 above, except that it deals with signed conversation. Unsurprisingly, manual movement in signed conversation has been studied extensively. However, the majority of the studies aim at analysing the linguistic structures of various sign languages and tend to neglect segments of behaviour that are motivated as part of interactional practices in conversation. This study thus shifts focus from manual movement in signed conversation purely seen as a means of conveying utterance content to manual movement segments as a locus for situated meaning that emerges locally within the context.

Furthermore, as pointed out by Groeber and Pochon-Berger (2014), the transcription of signed conversation has often been reduced to the stroke phase, rendering the timing imprecise as it obscures the actual length of each movement phase within a given manual excursion. Transcribing only strokes thus restricts analyses focussing on the back and forth between participants on a micro-level. The analysis makes use of video recordings from the Swedish Sign Language corpus gathered by Mesch and colleagues (Mesch et al., 2012). In the study I argue that participants in signed conversation, very much like in spoken conversation, make use of manual movement phases (in particular,

prolonged strokes, full-holds and half-holds) as part of regulating speakership. By virtue of these, participants can indicate repair initiation, urge for a response or resolve overlapping talk.

Although there are differences between spoken and signed conversation in terms of frequency, complexity and speed of manual movement, I did not find any notable differences regarding interactional practices involving the segments of behaviour analysed. This finding led me to launch Study 3 summarised below, a comparative study of manual movement.

STUDY 3

Exploring common ground in gesture and sign – Interactional workings of manual holds in spoken and signed interaction

Visible components as part of spoken languages are apt to be regarded as an add-on to vocal components, in the light of the fact, among others, that participants in an audio-only telephone conversation do not seem to do be troubled by not being able to see each other's visible behaviour. Many spoken languages possess a writing system that enable us to strip utterances from otherwise meaningful behaviour such as prosody, speech volume, speed and visible components. This division has led to a divide between these aspects, namely what counts as verbal and non-verbal respectively.

In signed conversation, on the other hand, the participants' bodies are on display at all times. Unlike in spoken languages, it is not possible to disassemble linguistic content and interactional meaning (which may be due to the fact that no widely accepted writing system for sign languages exists). In sign language research, this has led to integrating all kinds of visible behaviour as potentially part of a given sign language as a system, such as raising and frowning of the eyebrows as interrogative markers, which otherwise would be treated as non-verbal (i.e. as not belonging to the language system) when encountered in spoken conversation. This has created a gap between what is regarded as gesture and sign respectively (cf. Kendon, 2008).

In this study I examine whether this sharp distinction is sensible at all times. I present a comparative analysis of virtually the same phenomena in similar sequential contexts in spoken (Japanese) and signed (Swedish Sign Language) conversation respectively. It is found that participants in both spoken and signed conversation employ the same interactional practices that involve full-holds and half-holds. The distinction between gesture and sign turns out to be not only irrelevant but also impractical for these kinds of segment of behaviour.

STUDY 4

Self-touching behaviour in social interaction: Shifting in and out of speakership

Scratching one's head, rubbing one's nose, adjusting one's glasses: Everyone exhibits some self-touching behaviour from time to time, regardless of whether we are engaged in conversation or not. It has been maintained that such behaviour occurs without awareness, intention or interactional purpose and such acts are sometimes regarded as involuntary materialisations of emotions or internal states such as nervousness. Thus, self-touching behaviour has been regarded peripheral to interaction, not being part of what a participant is trying to convey and, as a result, either implicitly or explicitly been excluded from the scope of gesture-related studies.

Freedman and Hoffman (1967, p. 533) maintain that self-touching behaviour is "manifestly unrelated to the ongoing dialogue". This begs the question of whether self-touching behaviour may be regarded as a practice within interaction simply because – somewhat paradoxically – it is not part of the conversation. That is, the very fact that a self-touching behaviour is to a great extent considered, by interactants, a self-directed act unrelated to whatever is ongoing in a given interaction assists participants in establishing themselves as momentarily detached.

In the analysis, we limited ourselves to self-touchings in the area around the face and the head for practical reasons. As noted by Goodwin (1986), the face is the area where participants gaze to in face-to-face interaction, and self-touching behaviour in this area thus interferes with the other participants' line of

regard. Also, self-touchings in the area of the face and the head imply that the limbs have to move upward, usually away from the stage and home – that is, the limbs perform a less economical movement, yet participants do it for some reason and with a particular timing with regard to the progressivity of interaction.

We found that there are a few interactional environments in which self-touching behaviour is apt to occur, namely (1) conversational repair (either the initiating or the repairing party may exhibit self-touching in order to frame the discourse as “dealing with trouble”), (2) after discontinued courses of action (in cases in which a projected course of action, for whatever reason, cannot be brought to completion, participants may employ self-touching as [part of] an adjusting action – that is, as a way out), (3) at activity end points and (4) as an utterance-dedicated action (such as scratching the back of the head as part of conveying uncertainty). Self-touching behaviour is thus best viewed as residing on a scale, from backgrounded (unnoticed yet consequential) to foregrounded instances (as part of utterances, such as in conveying troublesomeness).

Another interesting observation not part of the study will be presented here. The data analysed were a recording of an event featuring speedruns,¹⁵ i.e. a very quick play-through of a video game (in this case, a game called *Super Mario Bros. 3*, a classic platform game first released in 1988 with individual levels that have to be completed in order to advance). The recording of the speedrun shows a split screen, with the game screen on one side and the speedrunner plus approximately 20 spectators behind him on the other side (see Figure 7). I noticed that, first, several spectators in the audience performed some kind of movement, often self-touching behaviour, every time the player died in the game (for instance though touching enemies or falling into a pit) during the game, most of the times in orchestration with a compassionate “Oh!” in unison. I noticed further that participants employ self-touching behaviour, albeit to a lesser extent, every time the player completed a level, and I suspect that it is related to the fact that the end of a level creates an intermission that con-

15 The recording is titled “Super Mario Bros 3 speedrun 100% by tjp7154 (Awesome Games Done Quick 2013)”. The video has been made available by the Speed Demos Archive and can be watched on YouTube (youtu.be/watch?v=HUju5czXNow), as of February 2016.

trasts with the tense, restless and exited run through the course of a level. It is during these intermissions that the speedrunner himself is granted a brief moment of rest and relaxation, and it is thus not surprising that he utilises this moment in order to adjust his glasses or wipe off sweat from his hands.



Figure 7: Recording of a speedrun event. Note that three people in the audience are involved in self-touching behaviour after a level has been completed.

This kind of regularity suggests that self-touching behaviour is related (albeit perhaps not directly) to some kind of delimitation marker between units, such as in this case upon completion of a level in a game. I assume that this also holds true for the completion of interactional units, whether adjacency pairs, lines of action or others.

CHAPTER 5

Roundup discussion

5.1 Manual action, speakership and involvement

One of the suggestions presented herein has been that closeness of the hands to the stage is associated with speakership and that moving the hands towards home is associated with withdrawal from speakership. Speakership can be understood as one among many possible kinds of involvement that participants may have in interaction. It is a quite large involvement in the sense that it absorbs a great deal of the speaker's resources, compared to, for instance, observing an interaction with minimal engagement.

When a participant in spoken conversation produces an enquiry in combination with manual movement, however, the hands are often kept on stage for an extended period of time, extending beyond the boundaries of the vocal portion of the utterance. In signed conversation, participants regularly hold the last sign in position when they produce an enquiry. This segment of non-movement is referred to as a prolonged hold (Study 1) or full-hold (Study 2 and 3). In such cases, the hand is retracted to home during or after the response by the recipient of the enquiry. Figure 8 shows a schematic representation of the sequential environment that is typical of the production of prolonged holds/full-holds.

Holding the hand on stage, the participant refers back to the action during which it originated, engendering the participant's understanding that a second action is needed in order to finish the line of action so far. The production of an initiating action, such as an enquiry, makes relevant the production of a responsive action, and participants show whether these relevancies are met. This

may best be understood as a visible extension of speakership; the participant in question continues to be involved in the interaction as they still are in control of the projected line of action.

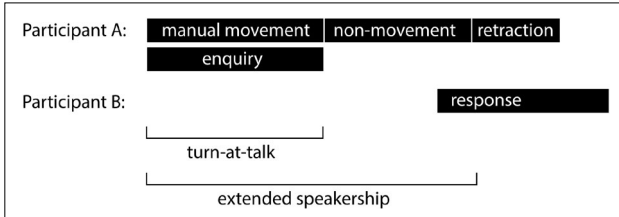


Figure 8: Schematic representation of an environment for prolonged holds

Participants orient to turns-at-talk as units which only allow one speaker at a time (Sacks et al., 1974), but at the same time there appears to be a second layer by virtue of which participants visibly refer to these actions and bundle them into chunks. This second layer of visible components of social interaction is apparently not governed by the one-speaker-at-a-time rule. Face-to-face interaction turns out to consist not only of sequentially organised actions (as the notion of the adjacency pair entails), but also of concurrently occurring tiers of action such as gaze, manual movement and body posture that – among other things – serve to show the degree of involvement a participant has.

Mondada (2007b) discusses pointing gestures that pertain over an extended period of time and notes that rights and obligations that are typical of speakership do not stop when a vocal utterance ends, but participants remain responsible for the sequence they have initiated.

I have argued that various possible positions of manual non-movement in midair should be considered a continuum, ranging from clearly on stage to intermediate positions to home – each respective position reflecting a different degree of involvement or claim of speakership. Participants may upgrade or downgrade their respective involvement or claim of speakership by moving their limbs back and forth within the continuum between stage and home. Such intermediate positions are often occasioned by overlapping talk. Participants who have dropped out of overlap may display by employing

provisional home positions/half-holds that speakership is held in abeyance and that they are ready to resume at the next slot. See Figure 9 for a typical environment for provisional home position/half-holds in spoken conversation.

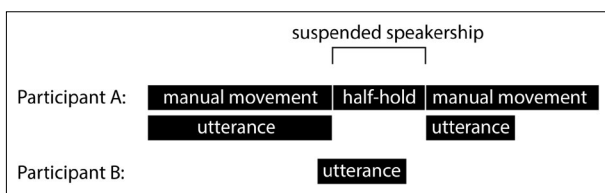


Figure 9: Schematic representation of a participant suspending speakership

Participants may thus suspend their speakership without completely abandoning it. In these cases, the hand that is held in midair is part of occasioning a hierarchical ranking among different parallel lines of action and thus different possible orientations. In the schematic representation, we find the emergence of a new potential trajectory (B's utterance) that competes with the existing projected line of action (A's utterance so far).

Alternatively, participants may abandon speakership by withdrawing their hands from the stage (see Figure 10). It is also possible for participants to employ self-touching behaviour in order to show that they are temporarily suspending speakership or discontinuing speakership.

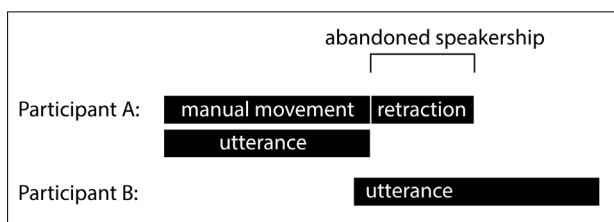


Figure 10: Schematic representation of a participant abandoning speakership

Speakership may be extended, suspended or abandoned. As a consequence, speakership is by no means a binary concept, but is best considered a kind of involvement along a continuum. Both speakership, non-speakership and

everything in between appear to be interactional roles that, by mobilising various bodily resources (in particular manual movement), need to be performed or enacted in order to be perceived as such.

5.2 Gesture as part of language

Much of my analysis revolves around participant practices in the use of gesture with spoken language and in sign languages in the form of signs. In Study 3 I illustrate parallels between these two modes of communication, and in this subsection I am going to discuss a little more about the nature of both gesture and sign.

The terms gesture and sign are often used in a mutually exclusive way, so that a manual action is understood as either gesture or sign, which implies some kind of essential difference between the two. This choice in terminology largely seems to depend upon whether we are speaking of spoken or signed conversation. For instance, we speak of co-speech gesture and sign phonology, but rarely about co-speech sign or gesture phonology (although the latter would not be far-fetched; see Figure 1 in Chapter 1).

Also, the terms bear different connotations: Whereas a sign is regarded as a precise and well-defined entity, speech-coordinated gesture is often thought of as idiosyncratic and made up on the spot (McNeill, 2005). This perceptual gap is related to the fact that a sign is considered part of language (namely, sign languages), while gesture is perceived as an add-on subordinate to language (in spoken languages).

But what is language anyway? Language can be understood in two different yet not necessarily contradicting, ways:

- (1) language as a system
- (2) language as social action

In the former view, language constitutes a closed system that can be analysed in terms of syntax, lexicon, semantics and so forth, most prominently represented by Chomskyan linguistics. This view allows researchers to analyse the rules of language in an isolated, abstracted manner, detached from context.

When language, on the other hand, is understood as situated social practice, we must presuppose that there is an exchange of actions, i.e. inter-action, by virtue of language.

Chomsky (1965) distinguishes between the speaker's competence (constituting the subject of linguistic enquiry) and the actual performance during which the speaker is affected by "grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic)" (Chomsky, 1965, p. 3). Language is thus primarily viewed as an abstract entity represented as text, stripped from indications of speed, prosody, voice quality, and visible components, focussing on what is considered verbal. This focus on verbal components has been a typical feature of formalist grammar, and it has been pointed out that it belongs to a tradition whose view of language is biased by written language (Linell, 1982). What has been the subject of study in linguistics – phonology, morphology, and syntax – are those aspects of language that can be written down (cf. Kendon, 2008). Visible behaviour such as gesture is thus marginalised and filtered out from the scope of the enquiry, resulting in an incomplete representation of speaker competence. When such a disembodied picture of language is applied to sign languages, it is evident that there are shortcomings. In signed conversation, one cannot draw a clear line that separates verbal from non-verbal, given that "verbal" and "non-verbal" components are both delivered as visible behaviour, whereas in spoken conversation visible components can be conveniently filtered away. As the signer's body is available at all times in signed conversation, every aspect, including those that are typically not deemed verbal in case of spoken conversation (such as eyebrow behaviour or head movement), is inevitably in play. The lack of an embodied linguistic theory forces the analyst to either incorporate those aspects of signed conversation into the linguistic model of sign language or to overlook it (cf. Kendon, 2014). As a result we find a great deal of features that are apparent in both spoken and signed conversation, but that are ascribed linguistic status in sign languages, whereas they have hardly been considered to be part of spoken languages due to the prevailing written language bias.

An example for this discrepancy is eyebrow behaviour. Valli and Lucas (2000) state that in American Sign Language, when “someone asks a yes-no question, the eyebrows are raised, the eyes are widened, and the head and body may tilt forward” (p. 139) and when “someone asks a Wh-question, the eyebrows squint and the head tilts” (p. 140). Bergman (1982, p. 56) makes similar observations for Swedish Sign Language.

In spoken languages we also find eyebrow work. When the word “What?” is uttered in rising intonation with the eyebrows raised, it is most generally understood as a way of saying “Pardon me?”. When uttered with frowned eyebrows, it comes off as “What the heck (did you just say)?”. While the former is used as a repair initiator, the latter entails an assessment, which presupposes an understanding of the utterance that it is directed at. This example is based on my own introspection, but Poggi et al. (2010) note that frowning the eyebrows is used to communicate a question in spoken conversation (p. 36).

Spoken and signed languages thus have in common the fact that eyebrow behaviour can operate on an utterance. The difference on a descriptive level is that it is represented as part of grammar in sign languages, whereas for spoken languages such descriptions are a rare sight. Along these lines, Johnston (2013) points out that pointing has been under-analysed for spoken languages, while, for sign languages, it is claimed to be on par with pronouns.

It is said that signs are to sign languages what words are to spoken languages (Liddell, 2003, p. 1). While words in spoken languages are produced in the vocal tract, signs in sign language are produced using the hands and the arms. But this is where the analogy ends. Signs performed with the hands are referred to as manual signs, while there is a range of non-manual signs that are produced using the torso, face, mouth and head. The above mentioned eyebrow behaviour seen in American Sign Language, for instance, is classed as such a non-manual sign (Valli & Lucas, 2000). Other examples for non-manual signs are mouth configurations used to modify a verb or a noun, as well as head nods used to mark phrase boundaries or for back-channelling.

When it comes to the description of spoken languages, it would be quite odd to refer to a nod, raised eyebrows or the wrinkling of the nose as a word (even though they have similar workings to those described for some sign

languages).¹⁶ In the case of sign language linguistics, however, they are referred to as signs. We thus find a linguistic preference for vocally produced components over visible ones in spoken languages and a disproportionate over-analysis of non-manual signs in sign languages.

In both spoken and signed conversation holding the hand in position on stage engenders an understanding on the part of the speaker that the initiated line of action is yet to be completed (Study 3), such as in the case of enquiries waiting for a response. It has been pointed out that this constitutes a grammatical feature for Swedish Sign Language. Bergman (1982, p. 54) maintains that the last sign in an interrogative sentence is held for a longer period of time than in a declarative sentence. If participants in spoken conversation employ the same practices as participants in signed interaction, why are they treated and classified in different ways?

One reason for this unbalanced approach towards gesture is the aforementioned written language bias. Also, it seems that we, as ordinary interactants, solely understand the term “grammar” in the context of written or vocally produced language. Grammar is part of the language we speak and those we learn as second language. We rarely speak of a grammar of gestural expression, even though this would not be impossible, in the light of the example of eyebrow behaviour in spoken conversation. We thus not only find a written language bias, but also a vocal language bias, meaning that spoken languages are thought of as consisting of vocal resources only.

This view is also in part reflected in conversation analytic approaches to interaction. For Schegloff (2002, p. 288), the primordial site of sociality is constituted in talk-in-interaction and especially conversation. If social interaction is to

16 On the other hand, back-channelling vocalisations such as “uh huh” would be more likely to pass as a word. The definition for a word seems, again, to be biased towards entities in spoken conversation that have a written counterpart.

be understood as an exchange of turns-at-talk as part of conversation, analysts are prone to rank interactional phenomena that are not constituted by talk (alone) as subordinate to talk.¹⁷

In sign language research it is often maintained that the structure of spoken languages is inherently linear or sequential, given that only one speech sound can be produced at a time (cf. Pfau, Steinbach, & Woll, 2012, p. 34). This contrasts with sign language, where the unique combination of the aspects of a sign – configuration, position and movement – are perceived simultaneously. A sign thus cannot be segmented in time order alone but must be aspectual (Stokoe, 2005, p. 20). It seems that sign languages are presented as visual languages that contrast with spoken languages as a purely vocal language, a view that is prone to result in misconceptions, as it neglects the visible dimension found in spoken conversation.

It is true that spoken languages chiefly rely on vocal resources, and it is vocal resources that are oriented to as the most significant component in spoken conversation (Kendon, 1978, p. 307). Also, spoken languages can be reduced to this vocal channel, as in telephone conversations. Experiences like this may contribute to the idea that spoken languages are indeed just spoken. But one must be careful not to jump to conclusions: The fact that this reduction is possible does not imply that everything else that participants do in spoken conversation can be swept under the carpet. Talking on the phone is something that small children have to learn as a skill in itself. When they talk on the phone they may, for example, often point to things in their surroundings that their recipient cannot see, as they have not yet grasped how the different modalities of spoken conversation (including visible behaviour) work in different ways. That is, they have not yet learned to separate them.

17 Indeed, transcripts of conversation are well designed in order to capture the peculiarities of talk in interaction and to present them to an audience in a straightforward way. However, visible components of conversation are usually represented as add-ons to the stream of vocal speech. Although I have endeavoured to represent modalities in an unprejudiced fashion, my own transcripts herein presented are not an exception to this, as I rely, after all, on written language to which I align the onset manual movement phases.

Indeed, there is research suggesting to treat visible bodily behaviour in spoken conversation as part of the system of the spoken languages – namely, what is referred to as *eyelid morphemes* in gaze communication, even though the authors concede that there are limitations (Poggi et al., 2010). Still, if we think about what requirements are set in order to be recognised as a native speaker of a spoken language, it is certainly not only a matter of knowledge about grammar and expression. Considering a speaker fully competent in a language implies taking into account their embodied performance, the use of specific gestural and facial expressions. A degree of native-like competence is only attainable if one acknowledges that this competence comes as a full package that includes everything that participants do within a given communication community.¹⁸

Along these lines, Liddell (2003) suggests that our understanding of what constitutes language might have been too narrow, pointing out the possibility that “spoken and signed languages both make use of multiple types of semiotic elements in the language signal” (p. 362). Kendon (2014) also makes some convincing points that speak in favour for revising the concept of language, stating that if participants in spoken and signed conversation have the same anatomical resources available, the participants in spoken conversation likewise can move their hands independently of one another and engage in actions of the face, and they can do all this while speaking (p. 3). He also points out that some vocal components demand the use of visible action in orchestration in order to make sense: When a speaker says “like this” in the course of describing the form or shape of an object, they engage in a “shape-sketch” depicting the outline of the object. Many deictic expressions (such as “over there”) are also incomplete without taking pointing actions into account.

18 Interestingly, the above mentioned grammatical feature of interrogatives involving eyebrow frowning in Swedish Sign Language and American Sign Language is not a universal feature to all sign languages: It does not appear to be obligatory in Auslan (Johnston & Schembri, 2007, p. 201). The fact that the use of such specialised visible behaviour varies from sign language to sign language suggests that there are differences in eyebrow behaviour across different spoken languages as well. My own intuition tells me that raising the eyebrows is not an obligatory feature to forming questions in a Japanese context. If this is true and such features are indeed codified differently depending on the communication community, we have reason to assume that, by extension, they are inherent to the spoken language in question.

The different connotations that the terms gesture and sign bear respectively do not imply that participants in spoken and signed conversation employ visible action in wholly different ways. This divide obscures that fact that they indeed engage in similar patterns of visible action (such as the use of manual movement phases) in both spoken and signed conversation. Kendon (2008) argues that the use of these terms tends to exaggerate the differences and obscure areas of overlap, and he suggests treating both under what he refers to as *comparative study of kinesic expression*, leaving behind the divide between gesture and sign that treats them as a priori different in each and every respect.

5.3 Some final words on possible universals

I have written a great deal about participant practices that involve the body and have hinted, here and there, that I did not find any notable differences across various communication communities with regard to how these practices are accomplished. The argument for potential universals in human bodily action is twofold. On the one hand, empirical evidence – both in this thesis and elsewhere – suggests that participant practices are shared in across different communication communities. On the other hand, one can make a conceptual argument: Given that humans generally have the same bodily resources at their disposal, it is not implausible to assume that they make use of them in similar ways when interacting with one another.

Some of the empirical evidence is provided by the studies through the comparison of conversation in two languages: Swedish Sign Language and Japanese. They are both geographically quite far apart – one predominantly spoken in Scandinavia and the other in East Asia – and linguistically mutually unrelated. Practices involving segments of extended (or prolonged) holds regularly occur among various communities, such as on the Caribbean island of Bequia (Sidnell, 2005) and in classroom interaction in Swiss German Sign Language (Groeber & Pochon-Berger, 2014). This sort of empirical evidence suggests that there are some segments of behaviour that appear to be shared across different communication communities and across spoken and signed languages.

While one may find it remarkable to observe the same segments of behaviour employed to the end of the same practices by individuals across the globe, one can also take a step back and formulate a conceptual argument. In face-to-face interaction, which I believe to be the most fundamental manifestation of human sociality, the participants' bodies are visible at all times. As a result, the human body is always, by definition, entangled in interaction in one way or another. Assuming that the kinesic mechanisms that govern manual movement are the same, participants in interaction are bound to creatively exploit these for a variety of purposes, due to the simple fact that these mechanisms are available and mutually shared. This may be the main reason why we find manual movement as part of expression and interaction in virtually every communication community – that is, as far as we know (cf. McNeill, 2005, p. 4). In short, the present corpus of work gives us clues about how the body is deployed as a resource in social interaction and explicates fundamental interactional patterns upon which our communication in all its facets is based. These patterns are ubiquitous and seem consistent across communication communities, languages and modalities. They are taken for granted, and for this very reason we, as ordinary interactants, rarely take note of them. It is these components characterised by the utmost simplicity and mundanity that hold together sociality at its inner core.

REFERENCES

- Andrén, Mats (2010). *Children's Gestures from 18 to 30 months*. Doctoral Dissertation, University of Lund, Sweden.
- Andrén, Mats (2011). Integrating Gestures. In Gale Stam & Mika Ishino (Eds.), *Integrating Gestures. The interdisciplinary nature of gesture* (Vol. 4, pp. 153–162). Amsterdam: Benjamins. doi:10.1075/g4
- Andrén, Mats (2012). The Social World Within Reach: Intersubjective Manifestations of Action Completion. *Cognitive Semiotics*, 4 (1), 138–165.
- Auer, Peter (2005). Projection in interaction and projection in grammar. *Interdisciplinary Journal for the Study of Discourse*, 25 (1), 7–36. doi:10.1515/text.2005.25.1.7
- Baker, Charlotte (1977). Regulators and turn-taking in American Sign Language discourse. In Lynn A. Friedmann (Ed.), *On the other hand: New perspectives on American Sign Language* (pp. 215–236). New York: Academic Press.
- Bavelas, Janet Beavin (1994). Gestures as Part of Speech: Methodological Implications. *Research on Language and Social Interaction*, 27 (3), 201–221. doi:10.1207/s15327973rlsi2703
- Bavelas, Janet Beavin, Nicole Chovil, Linda Coates, & Lori Roe (1995). Gestures Specialized for Dialogue. *Personality and Social Psychology Bulletin*, 21 (4), 394–405.
- Bergman, Brita (1982). Några satstyper i det svenska teckenspråket. In *Forskning om teckenspråk* (Vol. 11, pp. 49–63). Institutionen för lingvistik, Stockholms universitet.
- Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.

- Cibulka, Paul (2015). When the hands do not go home: A micro-study of the role of gesture phases in sequence suspension and closure. *Discourse Studies*, 17 (1), 3–24. doi:10.1177/1461445614557756
- Clark, Herbert H. (2005). Coordinating with each other in a material world. *Discourse Studies*, 7 (4-5), 507–525. doi:10.1177/1461445605054404
- DeStefani, Elwys (2007). La suspension du geste comme ressource interactionnelle. In *Proceedings of the 2nd ISGS Conference 'Interacting Bodies'*. Lyon.
- Ekman, Paul, & Wallace Friesen (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage and Coding. *Semiotica*, 1 , 49–98.
- Floyd, Simeon, Elizabeth Manrique, Giovanni Rossi, & Francisco Torreira (2014). Timing of Visual Bodily Behavior in Repair Sequences: Evidence From Three Languages. *Discourse Processes*, 52 , 1–30. doi:10.1080/0163853X.2014.992680
- Fox, Barbara A., Makoto Hayashi, & Robert Jasperson (1996). Resources and Repair: A Cross-Linguistic Study of Syntax and Repair. In Elinor Ochs, Emanuel A. Schegloff, & Sandra A. Thompson (Eds.), *Interaction and Grammar* (pp. 185–237). Cambridge: Cambridge University Press.
- Goffman, Erving (1963). *Behavior in Public Places – Notes on the Social Organization of Gatherings*. New York: The Free Press.
- Goffman, Erving (1974). *Frame Analysis – An Essay on the Organization of Experience*. New York: Harper and Row.
- Goodwin, Charles (1981). *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic Press.
- Goodwin, Charles (1986). Gestures as a resource for the organization of mutual orientation. *Semiotica*, 62 (1-2), 29–49. doi:10.1515/semi.1986.62.1-2.29

- Groeber, Simone, & Evelyne Pochon-Berger (2014). Turns and turn-taking in sign language interaction: A study of turn-final holds. *Journal of Pragmatics*, 65, 121–136. doi:10.1016/j.pragma.2013.08.012
- Hosoma, Hiromichi (2009). Washakootai wo koeru jesuchaa no jikankoozoo: Rinsetsu pea no baai [Sequential Structures of Gestures in Adjacency Pairs]. *Cognitive Studies*, 16 (1), 91–102.
- Hosoma, Hiromichi (2011). Extended gesture and adjacency pair. (p. 104). Book of Abstracts of the IEMCA Conference 2011.
- Ivarsson, J. (2010). Developing the construction sight: Architectural education and technological change. *Visual Communication*, 9 (2), 171–191. doi:10.1177/1470357210369883
- Jefferson, Gail (1996). A case of transcriptional stereotyping. *Journal of Pragmatics*, 26 (2), 159–170. doi:10.1016/0378-2166(96)00010-0
- Johnston, Trevor (2013). Towards a comparative semiotics of pointing actions in signed and spoken languages. *Gesture*, 13 (2), 109–142. doi:10.1075/gest.13.2.01joh
- Johnston, Trevor, & Adam Schembri (2007). *Australian Sign Language (Auslan) – An introduction to sign language linguistics*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511607479
- Kendon, Adam (1972). Some Relationships Between Body Motion and Speech. In Aron Wolfe Siegman & Benjamin Pope (Eds.), *Studies in Dyadic Communication* (pp. 177–210). New York and Oxford: Pergamon Press.
- Kendon, Adam (1978). Differential perception and attentional frame in face-to-face interaction: Two problems for investigation. *Semiotica*, 24 (3-4), 305–316.
- Kendon, Adam (1980). Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207–227). The Hague: Mouton.

- Kendon, Adam (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.
- Kendon, Adam (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, Adam (2008). Some reflections on the relationship between ‘gesture’ and ‘sign’. *Gesture*, 8 (3), 348–366. doi:10.1075/gest.8.3.05ken
- Kendon, Adam (2014). Semiotic diversity in utterance production and the concept of ‘language’. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369 (1651). doi:10.1098/rstb.2013.0293
- Kikuchi, Kouhei, & Mayumi Bono (2012). Organization of repair in Japanese sign language conversations. *Special Interest Group on Spoken Language Understanding and Dialogue Processing*, 65, 37–42.
- Kikuchi, Kouhei, Mayumi Bono, & Kazuhiro Otsuka (2011). Repair Organization in Sign Language Conversation. *Technical report of the Institute of Electronics, Information and Communication Engineers, HCS*, 110 (459), 61–65.
- Kita, Sotaro, Ingeborg van Gijn, & Harry van der Hulst (1997). Movement Phases in Signs and Co-Speech Gestures, and Their Transcriptions by Human Coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction* (pp. 23–35). London: Springer.
- Laurier, Eric (2014). Noticing: talk, gestures, movement and objects in video analysis. In R. Lee, N. Castree, R. Kitchen, V. Lawson, A. Paasi, & C. W. Withers (Eds.), *The Sage Handbook of Human Geography* (pp. 250–272). London: Sage.
- Liddell, Scott K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- Liddell, Scott K., & Robert E. Johnson (1989). American Sign Language: The phonological base. *Sign Language Studies*, 64, 195–277. doi:10.1353/sls.1989.0027

- Liddicoat, Anthony J. (2007). *An Introduction to Conversation Analysis*. London: Continuum.
- Linell, Per (1982). *The written language bias in linguistics*. Doctoral Dissertation, University of Linköping, Sweden.
- Linell, Per (2009). *Rethinking language, mind, and world dialogically: interactional and contextual theories of human sense-making*. Charlotte: Information Age.
- MacWhinney, Brian (2007). The TalkBank Project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and Digitizing Language Corpora: Synchronic Databases, Vol.1* (pp. 163–180). Houndmills: Palgrave-Macmillan.
- Manrique, Elizabeth, & N. J. Enfield (2015). Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. *Frontiers in Psychology*, 6 , 1326. doi:10.3389/fpsyg.2015.01326
- McNeill, David (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- McNeill, David (2005). *Gesture and Thought*. Chicago: University Of Chicago Press.
- Mesch, Johanna, Maya Rohdell, & Lars Wallin (2014). *Annotated files for the Swedish Sign Language Corpus. Version 2*. Sign Language Section, Department of Linguistics, Stockholm University.
- Mesch, Johanna, Lars Wallin, Anna-lena Nilsson, & Brita Bergman (2012). *Dataset. Swedish Sign Language Corpus project 2009–2011. Version 1*. Sign Language Section, Department of Linguistics, Stockholm University.
- Mondada, Lorenza (2007a). Commentary: transcript variations and the indexicality of transcribing practices. *Discourse Studies*, 9 (6), 809–821. doi:10.1177/1461445607082581

- Mondada, Lorenza (2007b). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies*, 9 (2), 194–225. doi:10.1177/1461445607075346
- Mondada, Lorenza (2009). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41 (10), 1977–1997. doi:10.1016/j.pragma.2008.09.019
- Oloff, Florence (2013). Embodied withdrawal after overlap resolution. *Journal of Pragmatics*, 46 (1), 139–156. doi:10.1016/j.pragma.2012.07.005
- Pfau, Roland, Markus Steinbach, & Bencie Woll (2012). *Sign language: An international handbook*. Berlin: De Gruyter Mouton.
- Poggi, Isabella, Francesca D’Errico, & Alessia Spagnolo (2010). The Embodied Morphemes of Gaze. In Stefan Kopp & Ipke Wachsmuth (Eds.), *Gesture in Embodied Communication and Human-Computer Interaction* (pp. 34–46). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-12553-9_4
- Raymond, Geoffrey, & Gene H. Lerner (2014). A body and its involvements: Adjusting action for dual involvements. In Pentti Haddington, Tiina Keisanen, Lorenza Mondada, & Maurice Nevile (Eds.), *Multi-activity in Social Interaction* (pp. 227–246). Amsterdam: Benjamins. doi:10.1075/z.187.08ray
- Sacks, Harvey, & Emanuel A. Schegloff (2002). Home position. *Gesture*, 2 (2), 133–146.
- Sacks, Harvey, Emanuel A. Schegloff, & Gail Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 (4), 696–735.
- Schefflen, Albert E. (1964). The Significance of Posture in Communication Systems. *Psychiatry*, 27 , 316–331.
- Schegloff, Emanuel A. (1998). Body Torque. *Social Research*, 65 (3), 535–596.

- Schegloff, Emanuel A. (2002). Accounts of Conduct in Interaction: Interruption, Overlap, and Turn-Taking. In Jonathan H. Turner (Ed.), *Handbook of Sociological Theory* (pp. 287–321). New York: Plenum. doi:10.1007/0-387-36274-6_15
- Schegloff, Emanuel A. (2007). *Sequence Organization in Interaction. Volume 1: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Schegloff, Emanuel A., Gail Jefferson, & Harvey Sacks (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, 53 (2), 361–382.
- Sciubba, Maria Eleonora (2010). I gesti interazionali come risorsa per la gestione dell'alternanza dei turni. In Massimo Pettorino, Antonella Giannini, & Francesca M. Dovetto (Eds.), *La comunicazione parlata 3* (pp. 565–587). Università degli Studi di Napoli l'Orientale.
- Seo, Mi-Suk, & Irene Koshik (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, 42 (8), 2219–2239.
- Shimotani, Maki (2007). A claim-of-reanalysis-token e?/e- within the sequence structure of other repair in Japanese conversation. In Naomi McGloin & Junko Mori (Eds.), *Japanese/Korean Linguistics* (Vol. 15, pp. 121–134).
- Sidnell, Jack (2005). Gesture in the pursuit and display of recognition : A Caribbean case study. *Semiotica*, 156 (1), 55–87.
- Stokoe, William C. (2005). Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *Journal of Deaf Studies and Deaf Education*, 10 (1), 3–37. doi:10.1093/deafed/eni001
- Streeck, Jürgen (2009). Forward-Gesturing. *Discourse Processes*, 46 (2-3), 161–179. doi:10.1080/01638530902728793
- Valli, Clayton, & Ceil Lucas (2000). *Linguistics of American Sign Language*. Washington: Gallaudet University Press.

- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann, & Han Sloetjes (2006). ELAN: a professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation* (pp. 1556–1559).
- Wong, Jean (2000). Delayed next turn repair initiation in native-nonnative speaker English conversation. *Applied Linguistics*, 21 (2), 244–267.

Part 2

Studies

**When the hands do not go home:
A micro-study of the role of
gesture phases in sequence
suspension and closure**

2015, *Discourse Studies*, 17 (1), 3-24.

STUDY 1

On how to do things with holds: Manual movement phases as part of interactional practices in signed conversation

2016, *Sign Language Studies*, 16 (4).

STUDY 2

Exploring common ground in gesture and sign – Interactional workings of manual holds in spoken and signed interaction

unpublished

STUDY 3

Self-touching behaviour in social interaction: Shifting in and out of speakership

unpublished

STUDY 4

In lieu of an afterword

THE THESIS is over – thank you for reading.
I hope it was clear (and not all too misleading).
Conventions are helpful, that's totally clear.
Neglecting them hurts, but won't end your career.
I think there are rules that are there to be broken,
So this part is written in verses that rhyme.
It is of advantage to be quite outspoken
About minor changes in this paradigm.

NIGH FIVE YEARS ago my new life here began.
I moved to this place all the way from Japan.
To buy an umbrella was the first thing to manage,
I firmly believed it would be an advantage.
For unbeknownst reasons my clothes still got soaked:
The side-winds compel one to be fully cloaked.
The most simple things have become utmost dear
In a climate of this latitude.
And days when the sky is deep-blue and all clear,
Have filled me with pure gratitude.

I'M THANKFUL to deities in charge of the sky
And also bow down to the mortals nearby,
Without whom the lines I wrote wouldn't make sense
And who proffered their aid at whatever expense.

THE PLAN WAS to research quite different questions,
But they soon transformed due to many suggestions.
Most comments and thoughts I obtained from my guru,
Your Süperweiser, that's his secret handle.
My texts were a mess, but he knew what I'm up to.
He knows to assess and then helps disentangle.

MY MOOD becomes brighter, it is a delight,
I hear the birds singing when she is in sight.
HER WORDS OF advise have been quite consequential:
Inspecting approaches that were to be sketched,
Deeming them valid or perhaps too far-fetched.
A nutshell would say this was highly essential.

AS IF READING this book alone wasn't enough,
This guy beefed it up with some fine-tuning stuff.
He alters and changes, amends and corrects,
A mastermind all filled with wisdom deluxe.
We're cooking with gas, yielding lasting effects.
This fellow is gold, and gold's worth a few bucks.

ONE SHOULD NOT belittle the input I'm gaining
Emerging from contacts I've been entertaining.
At many occasions I went to sweet places
That made me acquainted to many new faces.
I embarked to Denmark, I remarked the large parks,
Then I took the chance to advance towards France,
Rent-a-van in Japan, where I joined the monarchs,
My big grin in Berlin felt akin from my stance.

THE SHEER NUMBER of folks who made contribution
In a rather ambiguous fashion:
People who raised me, people who fed me,
Those who amaze me and those who might dread me,
For close ones around me there's no substitution,
For they give me my inner passion.
Friends I can count on, not due to my wealth,
It's them whom I owe my good psychical health.

PLEASE NOTE that this list here is not comprehensive,
It's not an acknowledgement section.
The number of people worth praise is extensive,
And were I to do so it'd be quite expensive,
Officials would show their objection.

AS A VERY last note, one must keep in mind
The Swedish arrangement turns out very kind.
They gave me their money (the amount was not modest),
They gave me computers, they gave me a suite,
They gave me a screen and – again – not the smallest,
The desk's ergonomic, and so is the seat.

MY BACKBONE, now capable to fully bend,
I have nothing to moan, but there is no end.
I just cannot help but announce the gains
That were implemented in many domains.
I've been in brave hands and they turned out benign.
I'd do it again, tell me where I should sign.