The interplay of gesture, gaze, and speech
Analyzing task-based interaction

Dissertation

zur Erlangung des akademischen Grades eines Doktors der Philosophie
der Philosophischen Fakultäten der Universität des Saarlandes
vorgelegt von Gerardine Marina Pereira aus St. Wendel
Saarbrücken, 2014

Dekan:	Prof. Dr. Ralf Bogner
Berichterstatter:	Prof. Dr. Neal R. Norrick
	Prof. Dr. Elke Teich
	Prof. em. David McNeill, PhD
Tag der letzten Prüfungsleistung:	04.12.2014

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

There are many people I have to thank for their help, patience and encouragement over the period it took to complete this dissertation. First and foremost, I would like to express my deepest gratitude to my advisor, Neal R. Norrick, for his support and guidance throughout the years, for his faith in being able to put me "out there in the real world", not only making me a better scholar but also a stronger person.

I would like to thank Elke Teich for agreeing to co-supervise this study and for inviting me to present my work in her research colloquium. The challenging questions and insightful comments I received from her and the other participants of the colloquium have improved this study in many ways.

I am deeply grateful to Kristy Beers-Fägersten for passing on to me her enthusiasm and passion for linguistics and for inspiring and encouraging me early on in my studies to pursue a career in linguistics.

I thank Ruth Page for her help in the organization and facilitation of a one-week research visit to Birmingham City University, where I collected my data. Many of her students participated in my study. Thank you to so many great gesturers.

I am indebted to the Fulbright Commission for awarding me a research fellowship to the University of Chicago in 2012. During the six months I spent in Chicago, many insightful conversations and discussions with David McNeill, Susan Duncan and other members of both the McNeill and the Goldin-Meadow labs have shaped my understanding of the study of gesture. I thank Kamala Russell for welcoming me to the lab and for cooking the very first dinner in Chicago for me. I also thank Sabine Jung for adding "variety and spice" to my research proposal for the Fulbright Fellowship.

I have benefitted from many interesting conversations with colleagues and friends at Saarland University. I am grateful to every single person who has accompanied me on my journey to the completion of this study. In particular, I thank Lucas Hütten, Isabel Schul, Selina Schmidt and Steffen Witt for transcribing some of the data, Stefan Diemer for his comments on parts of this dissertation, Carrie Ankerstein for the regular lunches and her professional expertise, and Henry Rademacher for being such a loyal "next door neighbor".

I thank all my amazing friends for their help and patience, in particular Eva Löber for always making me laugh, Ursula Zender for being able to relate to the challenges of writing a dissertation and for putting the whole process into perspective, and Heike Raber for her motivational speeches. Special thanks to Eva Michely, not only the most understanding and caring roommate one could ask for, but also an excellent proofreader.

My deepest gratitude to my family, in particular to my wonderful parents, Ronald and Luzia Pereira, who always encouraged me to pursue what makes me happy. This work would not have come into being without your love, your trust, and your unwavering belief in me.

1 INTRODUCTION

This study investigates spoken language data, gesture, and gaze to gain insights into how dyads of interactants shape and coordinate their actions to mutually accomplish a given task. Based on micro-level phenomena consisting of multiple interactive modes, the findings illuminate new implications for the theoretical and methodological approach to language. This study arrives at an understanding of language which views it as composed of speech, gesture, and other nonverbal modes.

The data presented in this study derive from a larger research project and twenty-one video-recordings of dyadic interactions will be analyzed in depth. This study provides a qualitative account of how speech, gesture, and other nonverbal phenomena establish and shape interactions and how they provide information about the cooperativeness of participants as well as the coordination and organization of the task. An utterance such as "is that Adventures with Aqua there, in this bit here?" (MOV00F) exemplifies the special nature of discourse focusing on the planning and the achievement of an activity. It demonstrates the necessity of an assumption of a speech-gesture synchrony. *Locational pro-terms* (Schegloff 1972:87) such as "here" and "there", which are referential expressions, have a varying meaning depending of the context of the situation. Moreover, they need a gesture to clearly indicate their referent. The gestures described in this study are understood as occurrences of hand movements which accompany or replace speech and become meaningful as they are used in context to solve the task and shape the interaction. Together, speech and gesture create and convey meaning (Goldin-Meadow 2003a). This study supplements existing research on gesture by analyzing experimentally elicited data and contributes to the linguistic understanding of the relationship between speech and gesture. It furthermore takes other nonverbal phenomena, in particular gaze, into account to highlight the significance of a multimodal approach to language.

Goodwin (2000:1490) criticizes pragmatics for taking language as its primary topic of analysis thereby treating everything that is not language as 'context'. Based on Lascarides and Stone's assumption, however, that people "intend their actions to be understood as coordinated ensembles" (2009:1), a multimodal approach to interactions and conversations is necessary to investigate speech and gesture.

In the literature on nonverbal phenomena, different models of gesture render different definitions of the terms *nonverbal communication* and *nonverbal behavior*. There is no standardized terminology, and therefore, nonverbal behavior and nonverbal communication are not always clearly separated from one another. Rossini presents an approach to language that interprets it "as either perception, introspection, self-control, self-orientation of thought, or output of a message" (2012:3). Such a message is multimodal suggesting that language and communication, in particular, consist of different mechanisms integrating gesture and other nonverbal cues into the study of language. McNeill (2012) investigates the evolution of human language stressing that both speech and gesture are essential to human communication. Kendon (1990) takes a structural approach to the study of interaction. This approach "maintains that communication in interaction is a continuous, multichannel process and it seeks to provide descriptions of the structural characteristics of the communication system employed in the interaction" (Kendon 1990:15). Rossini (2012) further addresses the terminological problem of a definition of nonverbal communication to differentiate it from nonverbal behavior (see Rossini 2012 for a full discussion of the problem; see also Krauss et al. 1996). Since the theoretical approaches are so manifold, there is no single definition of nonverbal communication and thus, a wide array of nonverbal behaviors account for it (Goldin-Meadow 2003a), including the way we comport and dress ourselves to either express group membership or to distance ourselves from others. Goldin-Meadow (2003a:1) writes that all these messages frame a conversation, but are not the conversation itself. Even though there are some recent models focusing on communication, terminological problems persist, often due to the lack of a unified definition and a clear understanding of what researchers mean by *communication*.

In discourse analysis and linguistic pragmatics, frameworks are related to communication; utterances are used to convey information and thereby people lead one another to interpretations of meanings and intentions (Schiffrin 1994:386). Many of these linguistic approaches view communication and language as spoken (and written) systems. However, in this study it is argued that human communication consists of both verbal and visual phenomena, i.e., speech, gesture, and gaze. As a whole, these phenomena contribute to the activity in a meaningful way, providing communicative and interactive information. The linguistic perspective on 'language' must therefore also be revised, making communication as defined

above an important part of human language. In other words, the study of language should be perceived as the study of speech and gesture.

The present study is embedded in the interactional sociolinguistics approach to discourse as well as David McNeill's (1992, 2005, 2012) approach to gesture. The study also draws on multimodal and interaction research as proposed by Streeck et al. (2011), Schmitt (2007), and Norris (2011) as they outline new analyses and methods relevant to discourse analysis. For the investigation of other nonverbal phenomena, in particular gaze, the work of Argyle and Cook (1976), Kendon (1990), and Bavelas and Gerwing (2007) will be considered.

1.1 PRECISE

Chapters 2 and 3 address the theoretical and methodological background to this study. Chapter 2 sets the scene: it begins with a survey of the history of the study of gesture and gaze, presents different approaches and definitions of relevant phenomena, and summarizes some state of the art research. The second part of the chapter is concerned with discourse analysis and different model of communication. In particular, interactional sociolinguistics as one approach to discourse as well as multimodality in interaction are discussed. Other relevant objectives of this study include cohesion and reference, common ground and stance taking, which are addressed in the remainder of Chapter 2. Chapter 3 presents the methodology section. It provides information about the data collection and the selection of data for the present purpose, ethical considerations, and the adaptation of transcription conventions. It concludes with an application of Deppermann and Schmitt's (2007) modes of expression and Kendon's (1990) F-formation system.

The introductory chapters are followed by five analysis chapters, which relate to one another, but can be read as individual studies. The first three of the five chapters are concerned with central study objectives in discourse analysis. Chapter 4 addresses the question of openings of unscripted events. Understanding the activity as a joint action, this chapter develops different activity roles and identifies forms of interactions based on verbal and nonverbal phenomena. Chapter 5 is concerned with an investigation of cohesion, in particular reference and co-referential chains as proposed by McNeill et al. (2010). By differentiating references to the object level from references to the meta and to the para level, it is demonstrated how participants form coalitions and how they align with one another. Chapter 5 also expands on the traditional conception of cohesion as a textual phenomenon, proposing that the creation of cohesion and

reference is a multimodal process. Chapter 6 investigates the linguistic feature of repetition and extends this notion to include gesture recurrence in order to establish a relationship to cohesion and to investigate how repetition is used as a means of grounding and updating common ground.

Chapters 7 and 8 are devoted to a closer investigation of both gesture and gaze. Chapter 7 focuses on a particular type of gesture, the deictic gesture. It expands on the traditional understanding of deictics and establishes different types as well as functions attributed to each type of deictic gesture as they occur in the task-based setting. Chapter 8 develops out of previous research on eye gaze in relation to conversation (Goodwin 1980; Bavelas et al. 2002a). The findings of the present investigation illustrate the special nature of gaze work in task-based interaction is conceptualized, focusing on gaze direction, gaze shifts, and mutual gaze. Chapter 9 summarizes the main findings of this study, provides implications for the study of language in general, and concludes with ideas for further research.

2 THEORETICAL FOUNDATIONS

In the following, I will present research in the areas of gesture study, gaze and facial expressions as well as discourse analysis. I will outline the development of gesture and gaze studies and present some state of the art research. I will also discuss important implications for the study of discourse based on new research methods and the availability of audio-visual data. At the end of this section, I will present other topics relevant to this study, including research on cohesion and reference as well as stance taking and common ground.

2.1 ON GESTURE AND GAZE

The area of gesture studies is a fairly new and dynamic field. Research on gesture is as manifold as the data: from narrative and film narrations, conservation and math tasks to conversation. As a result, there are various theoretical frameworks and approaches to the study of gesture. The history and development of the interest in gesture as well as current approaches and classifications of gesture will be outlined. An overview of the development of research in facial expressions and gaze work will also be presented.

2.1.1 Semiotics

Communication is an essential factor in pragmatics and in semiotics. Semiotics is broadly defined as "the general study of signs" (Nöth 2011:167). Semiotics as an approach to communication has often been neglected as a framework in pragmatics. People communicate meaning with language and the linguistic meaning is part of a sign system (Saeed 2003:5). Nöth (2011) explains that semiotics, as it is perceived by Morris (1938), is not restricted to verbal communication, but includes signs in various manifestations, for instance linguistic and nonlinguistic, acoustic and visual signs. Morris' classification of sign behavior into three factors maps onto three areas of linguistics: syntax, semantics, and pragmatics. This division was, according to Nöth (2011), influenced by the writings of Peirce (1981). He is considered the founder of pragmatism, but his Theory of Signs (see Misak 2004) bears relationship to pragmatic theory and the study of gesture. Pape (2004:117) conceives semiotics as the science of signs and as such, it aims at understanding the sign's properties, conditions, and rules about its production and usage. Pape explains that the sign should be analyzed across disciplines because different

signals, for example gestures, but also traces in the snow, belong to the general science of signs. Thought is another central element in Peirce's theory, because thought is "a process of sign-production and sign-interpretation" (Short 2004:214) and the relationship between thought, speech, and gesture is also of current interest to researchers such as McNeill (2005), for instance.

Based on Peirce's theory, there are three basic signs: the icon, the index, and the symbol. The icon bears similarity to what it represents whereas a symbol is a sign that holds a conventional relationship to the signified. The indexical sign could, for instance, be a pointing finger or a deictic word, any feature that directs the interpreter's attention toward an object, which "may also be a feeling, an experience, a cognition, a thought, an imagination, or even a fictional event" (Nöth 2011:185-186). Short (2004:222) views the discovery of the index as an extension to Peirce's earlier perspective of semiotics as the study of thought and language. Indexicals or deictic movements are studied in pragmatics, analyzing how deictic words create reference, and pointing gestures are a central feature in the study of gesture as well. The relationship of language and thought (the sign and its relationship to thought) is, for instance, also reflected in McNeill's (2005) work as well as in the emergence of the field of cognitive semiotics (Holenstein 2008, Cienki 2009).

2.1.2 The development of gesture studies

The study of gestures and facial expressions has had a long tradition over the centuries, from Classical Antiquity to the present. Kendon's (2004) book entitled *Gesture: Visible action as utterance* gives a detailed and thorough account of the history and the development of gesture studies to the present. In the following, some important facts in the history of the study of gesture will be outlined.

According to Kendon (2004), gestures were understood as part of the oratorical technique in the Greek and Roman tradition. Kendon (2004:17-18) states that Marcus Fabius Quintilianus discusses *motum*, also called *gestus*, in his work *Institutio oratoria* and that he defines the term *gesture* not only as hand and arm movement, but also includes posture, head movement and facial expressions as well as glance. The comportment of the body was important in the Middle Ages and it received even greater attention in the course of the sixteenth and seventeenth centuries. Gesture manuals for legal and religious ceremonies were developed because gestures were seen as a part of the rhetoric. Bodily movements and facial expressions were also depicted

in the pictorial arts. This trend continued into the eighteenth century: gestures were still considered as a part of rhetoric. Good style in speaking and conversation gained importance in England and on the European continent. Slowly, an idea of the universality of gestures developed. In art forms such as painting, "an elaborate set of conventions developed" (Kendon 2004:30), which the painter was urged to apply to depict the conduct and emotions of people. In the performing arts, actors employed a variety of gestures and movements to transport emotions to an audience. Gestures were also considered part of a basic education:

> [which] ensured that these general techniques, at least among the educated classes, were carried into daily life. It made it possible for the techniques of acting and oratory to draw upon the shared understanding of gesture – an understanding shared across all of Europe, it should be added, so that, to a considerable degree, language differences were transcended (Kendon 2004:33).

The idea of a universality of gestures carried into the nineteenth century, raising questions about the evolution of language and whether gesture might have preceded speech (Kendon 2004:43). New aspects of gestures were investigated in archaeology and prehistory, adding an ethnographic perspective to the study of gesture. Sign language, which was first studied in French by Abbé de L'Epée (1776) in the eighteenth century, was further developed by Tylor (1865). Tyler provided a survey of different sign languages, including German and North American Indian. De Jorio (2000, originally printed in 1832) focused on gestural expressions found in the city of Naples to show how gestures were culturally inherited, i.e., how gestures of the "common Neapolitans" were related to "the gestural practices of the ancient Greco-Roman inhabitants of the city" (Kendon 2004:45). However great the interest in gesture, there was a decline and even a disrepute of gesture and sign language at the end of the nineteenth century and this continued into the 20th century. New developments in psychology, for example behaviorism and psychoanalysis, as well as new trends in linguistics led to a decline in the interest in gestures.

In the 1970s, anthropologists such Birdwhistell (1970) and Bateson (1972) developed a new interest in nonverbal communication. Bateson (1972), for instance, used film material and photographs to investigate behaviors and rituals in Bali and New Guinea tribes and transferred these methods to later research when he studied family interactions. Bateson collaborated with psychiatrists (including Watzlawick) who studied the processes of psychotherapy as "a consequence of the social interaction between patient and therapist" (Kendon 2004:69). Bateson

also studied the behaviors of animals in order to understand how certain signals are interpreted. When monkeys engage in certain actions, for example in play, the question of how the monkeys can distinguish a playful bite from a bite in combat arose. Bateson noted that this requires communication about communication, which he termed *metacommunication* (Bateson 1972; see also Murray 1998). He drew comparisons between observations he had made about animal as well as human behavior to develop a conceptual frame for the observed behaviors, including both verbal and nonverbal. Bateson's notion of frame, for example the difference between the frame "this is play" and the frame "this is ritual" (Bateson 1972:182) influenced other researchers. It led to frame analysis (Goffman 1974) and is also reflected in Gumperz' (1982) contextualization cues as well as Tannen's (1986) conception of metamessage. In another project, the "Natural History of the Interview" project, Bateson and other researchers, such as the linguists Hockett and McQuown as well as the anthropologist Birdwhistell, investigated spoken and kinesic behaviors as they happened in social interaction (see Kendon 2004, Allan 2013). The results of these developments and investigations established a dichotomy between verbal and nonverbal communication, but did not render a framework for the study of gesture. It was not until the 1970s that the study of gesture gained new impetus becoming a study topic in anthropology, linguistics, and psychology.

Current gesture research continues to be a multidisciplinary field, including work in anthropology and ethnography (Enfield 2009), psychology and neuroscience as well as computer science and computational linguistics (Gibbon 2009), musicology (Gritten and King 2006, 2011) and other performing arts, communication, including face-to-face communication, and linguistics (Rossini 2012). Gestures are investigated in relation to other media taking a multimodal approach (Norris 2004, 2011; Schmitt 2007); moreover, research includes universal and cultural aspects of gesture (Kita 2003a), the role of gestures in child development (for example in deaf children, Goldin-Meadow 2003b) and their relation to language and thought. Sign languages (Davis, 2010; Haviland 2011) and the evolution of language (McNeill 2012) are further research areas (see International Society for Gesture Studies).

Rossini (2012) places emphasis on the relationship between speech and gesture and the fact that they both constitute human communication. Thus, the study of gesture has become a relevant topic in linguistics rendering a linguistic interpretation of gesture. Researchers such as Kendon, McNeill and Goldin-Meadow have demonstrated that conversation is produced through

both words and hand movements. I will argue that nonverbal phenomena such as gaze and body-orientation should be included when studying conversation because they contribute to the ongoing interaction and provide information about the level of involvement and cooperation between interactants. Only when these phenomena are taken into account can we paint a complete picture of the ongoing situation and the relationship between the interactants. Peräkylä and Sorjonen (2012), for instance, report on displays of emotion in everyday situations and investigate the relation between emotion and action in social interactions. Ekman explains that "[w]e can highly be informed by actions which were not made for the purpose of informing us" (1997:339) and thus, the successful coordination of an action and the successful cooperation between two or more people relies on facial signals, especially when someone intentionally looks at the other person, thereby asking for help, for instance.

2.1.3 Approaches to the study of gesture

According to McNeill (2005), there are four current approaches to gesture. Firstly, the functions of gestures are studied in the context of social interaction. Research in this area places an interpsychological point of view on the topic. Secondly, there is cognitive psychology. In this field, the origins of gesture and the interrelations of gesture with speaking in the real time mental processes of individuals are studied. Thirdly, gestures are examined with regard to modeling. In this area, computational models of gesture-speech performance are developed. Lastly, there is the study of sign languages (see Supalla 2002, 2003 for the American Sign Language), which grows out of the transition from gesticulation to sign. Signs represent one end of what McNeill labeled 'Kendon's continuum' (1992:37). This continuum will be outlined in the the following section on gesture classification.

Levelt et al. (1999) and McNeill (1992, 2005) take a psycholinguistic view on the role of gesture in language. Two main approaches result out of their theory: the lexical access theory, which proposes that gestures aide speakers in accessing an item in their mental lexicon, and the growth point theory, developed by McNeill. The growth point is "a dynamic unit of online verbal thinking"[1], suggesting that gestures result from a language-imagery dialectic. According to McNeill (2005:22), gestures carry meaning and frequently, they are synchronous with speech. Gestures are co-expressive and often closely timed with a speaker's words: "one characteristic of

[1] (McNeill: http://mcneilllab.uchicago.edu/writing/growth_points.html; date access: 06.01.2012)

hand gesture and facial displays is that their movements can match the speed of words." (Bavelas and Gerwing 2007:287). Current research in the study of gesture shows that gestures and facial expressions fulfill a communicative function and that speakers deploy them together with and instead of verbal utterances. Bavelas et al. (2002b:2) present two hypotheses regarding communicative gestures: firstly these gestures are used by speakers to communicate, the encoding hypothesis, and secondly the information in the speaker's gesture is understood by the recipient, the decoding hypothesis. Consequently, they argue for a theory that acknowledges the multitude of functions of gestures, opposed to a unidimensional theory such as the lexical access theory.

In conclusion, modern gesture studies are composed of a range of theories and approaches. They differ in focus as they either place speech above gesture or view gesture and speech as equal parts of utterances (see Gullberg and McCafferty 2008). The present study supplements modern gesture studies from a linguistics perspective, emphasizing that language must be conceived as speech and gesture as equal modes rather than placing one above the other.

2.1.4 Classification of nonverbal behavior and gesture

In the following I will briefly outline the main classifications of gestures with a particular interest in deictic gestures. I will also provide some early as well as state of the art research on facial displays and gaze work.

Ekman and Friesen (1969) distinguish five types of nonverbal behaviors. *Emblem*, a term earlier introduced by Efron (1941), replaces the term *gesture* to relate to signals that can be culture or group specific. Emblems have a dictionary translation, which means that they are used intentionally, with or without speech, and are most easily understood because they have "a quiet specific, agreed-upon meaning" (Ekman and Friesen 1969:64). The second type is called *illustrators* and, as the name already suggests, they illustrate something that is being said. Hence, *batons, ideographs, deictic movements, spatial movements, kinetographs* and *pictographs* are all forms of illustrators. They are tied to the content of speech and depict what is said. The third type of nonverbal behavior is called *affect displays*. Ekman and Friesen (1969) now move from the hand/arm as performing entity to the face: they argue that there are certain primary affect states (e.g. surprise, fear, sadness, etc.) which are informative and communicative (if displayed intentionally). Importantly, these facial affective displays "receive great attention and external

feedback from the other interactant" (Ekman and Friesen 1969:77) and thus need to be included in a descriptive analysis of dyadic interactions since they convey personal information and can be indicators of the level of involvement, cooperation, and agreement between people. *Regulators*, the fourth type, are smaller nonverbal acts which are similar to illustrators since they are also tied to the conversation. They help to regulate the flow of the conversation. A head nod, for instance, is a non-verbal signal, which is similar to a verbal back-channel such as *mhm*. The last type, *adaptors*, include learned, adaptive effects, which are usually unintentional and displayed without awareness. Adaptors can be differentiated for three types, *self-adaptor*, *alter-adaptor*, and *object adaptors* (Ekman and Friesen 1969:84).

An alternative representation of gestures is found in 'Kendon's continuum', which McNeill (1992:37) named as such in honor of Adam Kendon:

Gesticulation → Language-like Gestures → Pantomimes → Emblems → Sing Languages

Going from left to right, the presence of speech declines while language like properties increase; socially regulated sign replace idiosyncratic gestures. Kendon's continuum is helpful in differentiating different kinds of gestures. McNeill (1992) focuses his analysis on the gestures located on the left end of the spectrum, gesticulation, referring to idiosyncratic and spontaneous moves of the hands and arms. The classification of gestures proposed by McNeill (1992) is the following: *iconics, metaphorics, beats, cohesives* and *deictics*. By iconic gestures, McNeill means those gestures that "bear a close formal relationship to the semantic content of speech" (1992:12). They are pictorial and share this feature with metaphorics. Whereas iconics represent a concrete object, though, metaphorics refer to an abstract idea, having a topic and a vehicle. Beats ('batons' for Ekman and Friesen1969) differ from iconics and metaphorics because beats they usually have the same form and express the rhythm of the speech they accompany. Beats can highlight certain statements and fulfill a prosodic function. Cohesives do not represent an independent form of gesture: they can occur in the form of an iconic or beat gesture, for instance, but no matter their form, they express recurrent themes within a discourse and link discourse units. They demonstrate continuity displayed in a repeated gesture form, movement or gesture space. The final type of gesture is the deictic gesture, prototypically a pointing movement to a referent, but also pointing when there is nothing objectively present to point at (McNeill 1992, Goldin-Meadow 2003a).

A third and more specific classification was proposed by Bavelas et al. (1992). The illustrator class of gestures is divided into *topic* and *interactive gestures*, suggesting that beats, for instance, fall into the latter class. Topic gestures relate directly to the topic of the conversation whereas interactive gestures "refer instead to some aspect of the process of conversing with another person" (Bavelas et al. 1992:473). Thus, it is characteristic of interactive gestures that they address the interlocutor and that they are topic independent. These classification systems are only exemplary of a range of categories (see also Argyle 1975; Rossini 2012). The differences in terminology are not only due to the different approaches to the phenomena investigated but also due to the different data analyzed. It is often misleading to classify gestures into categories; rather, one should think of them in terms of dimensions because most gestures are multifaceted (McNeill 2005:38). This means that they can belong to several dimensions and fulfill more than one function at a time, for instance when a beat gesture is used as an interactive gesture.

2.1.5 Deictic gestures

Taking a core definition of deictic gestures, for example one due to Krauss et al. "(…) the 'meaning' of a deictic gesture is the act of indicating the things pointed to" (2000:263). Such a definition, however, covers only a very basic concept of a deictic gesture. Prototypically a deictic gesture is a pointing movement, which means that we use our index finger to point. Hence deictic gestures are also called pointing gestures. The index finger can be extended by various tools, for example a pen. Other tools like graspers and laparoscopic cameras are used in surgery to point out certain areas of interest, as Koschmann et al. (2010) describe in their study. In these cases, the gesture is often used "to indicate persons, objects, directions, or locations (…)" (Krauss et al. 2000:262). In their referential function, deictic gestures often co-occur with deictic utterances such as "here" or "this one". The findings of this study show that these references often cannot be understood without the co-speech deictic gesture. Thus, access to and visibility of the map used in this experiment play an important role in the identification of the referent and provide the space for the accompanying pointing gesture. The results also show that pointing can occur without a verbal specification. This is only possible when both participants see the referent and share a common understanding of it.

Deictic gestures not only index objects or people, but they also help one person to explain certain procedures to another person and, by a reciprocal or mutual point, the other person can reach and signal understanding (Koschmann et al. 2010:9-11). However, affiliation is not an inherent feature of a pointing movement since pointing can express disagreement or aggression, for instance. Children in many cultures are taught not to point at another person as this can be perceived as rude. When animals "point" their claws at their prey (see Desmond Morris' discussion on owls, for instance), this action will certainly result in an injury or even death. Pointing gestures moreover take various shapes, for example pointing with our thumb or even an extended hand, palm up. In some cultures, for instance in Laos (see Enfield 2001), pointing can be produced with protruded lips, with the head, and with the eyes or the eyebrows.

In a face-to-face interaction, interlocutors decide to point at each other. This could be the case when they want to identify someone, for example when a third (unacquainted) person enters the room. Sometimes we point at someone to put blame on this person or give him/her away. However, it is more likely that a gesture in a face-to-face context occurs to signal that the current speaker acknowledges something that the previous speaker has said. One can easily imagine an utterance such as "As Jane has just said" accompanied by a pointing gesture directed at Jane. Thus, the deictic movement has an interactive function (Bavelas et al. 1992).

The studies by Koschmann et al. (2010) and by Goodwin (2003) exemplify that pointing gestures can do more than just 'point'. The most obvious function is doubtlessly to direct the listener's (visual) attention to something. However, as Clark (2003:248) states, speakers can also use signals to move an object into the addressee's focus, for example when a customer places certain items he or she wants to buy on the counter in front of the salesperson. Clark highlights the difference between directing someone's attention to something and placing an object into the addressee's focus, thereby distinguishing 'pointing' from 'placing'. Yet another dimension, and this has been shown by Pereira (2011), is to signal (dis)agreement and to establish a shared focus by pointing. Speakers, and more importantly listeners, can both apply pointing movements to express these former notions. Thus, pointing gestures are sometimes used to indicate things that are not visible. Borrowing Goldin-Meadow's example, an utterance such as "I gave it to her yesterday" (2003a:7-8) can be accompanied by a pointing gesture. One can point at the person "her", but one can also point at a particular space in a room. This latter gesture would be abstract

since it does not indicate the person, but a space that has previously been associated with that person.

One special form of pointing is 'tracing'. Goodwin writes that "[…] the moving finger and the target of the point are brought into a dynamic relationship in which each is used to understand the other. The activity of pointing continues after reference per se has been accomplished" (2003:16). In the work of archaeologists, which Goodwin describes, tracing can be used to describe objects and their shape, and in so far, a trace can specify certain features that language alone cannot describe. Tracing in the present context is used to connect a starting point A and an end point B. The route that leads from A to B is outlined on the map and sometimes participants draw lines or put an X next to an animal exhibit once they have visited it. The map is a material object which can be used as a sense-making device. The traces—or *inscriptions* in Goodwin's terms (2003:16)—which participants leave on there are visible throughout the activity as well as upon completion, so that they represent prior arrangements of the interaction. To quote Streeck, there is a "continuum of symbolization" (1996:382), which means that objects and physical representation of objects can be used in a particular situation and they can become an "agreed-upon, publicly available code".

2.1.6 Facial displays and gaze work

Gaze can be directed toward a person or an object, just as gestures can: the speaker can look at an entity while verbally mentioning it and gaze can be used to direct someone else's attention to this entity as well. According to Goffman (1963:89), two or more participants, when involved in a mutual activity, engage one another in a focused interaction, in face engagements, for example in small talk as the simplest form. In children and even in adults face-to-face encounters can be nonverbal, for example when playing cards and making a move in a game (Goffman 1963). Gaze and mutual gaze are important components in face-to-face communication and in mutual activities because participants can coordinate and adjust their actions to one another. Eye contact plays an important role in the interaction with others because it conveys social presence (see Howarth and Anderson 2007) and creates engagement with another person. McNeill et al. (2010:148) apply the terms *gaze target* to mean the person or item looked at, and *gaze source* to refer to the person looking at another person or an entity. Kendon (1990) presents an F-formation system, which is based on the idea that pairs and groups of people cluster in certain patterns.

These patterns can change or they can be sustained. In the latter case, these patterns are considered *formations*. An F-formation is a formation which "arises when two or more people cooperate together to maintain a space between them to which they all have direct and exclusive access" (Kendon 1990:210). Within the F-formation, individual participants can direct their head movements and their facial displays to one another, i.e., they look at each other repeatedly. The F-formation thus includes spatial and orientational behavior.

The study of facial displays and gaze received substantial attention in the field of social psychology, the earliest research coming in the 1960s and 1970s. Researchers such as Ekman and Friesen (1969), Izard (1977), and Buck (1984) regarded facial expressions as being tied to emotions. The main interest was individual in focus, meaning that "the preferred setting for the study of the face as emotional expression is the individual alone" (Bavelas and Gerwing 2007:287). Monologic speech was favored over dialogic speech in order to investigate gestures and facial expressions with regard to the individual mental processes (Bavelas and Gerwing 2007). Ekman and Friesen write,

> If we are to understand fully any instance of a person's nonverbal behavior – that is any movement or position of the face and/or the body – we must discover how that behavior became part of a person's repertoire (…) (1969:49).

Any facial and bodily movement is considered nonverbal behavior, a notion which was established in antiquity and is still recognized today. However, Ekman and Friesen stress the importance of individual behavior as their focus of study, excluding interactions between two or several people. They question the interpretation and communicative aspect of nonverbal behavior when they write "If observers are able to interpret information accurately, can we infer that nonverbal behaviors are intentional efforts to communicate?" (Ekman and Friesen 1969:50). Most researchers nowadays acknowledge the communicative function of gestures. Deictic gestures, for instance, are a meaningful act through mutual contextualization and help "to establish a particular space as a shared focus for the organization of cognition and action." (Goodwin 2003:218). Peräkylä and Sorjonen (2012) present a collection of recent approaches to expressions of emotion in interaction, including facial expressions and laughter, for instance. Beukeboom (2009) analyzes the impact of listeners' affective expressions, including facial expressions and bodily posture, on speakers. In an experimental setting, participants tell a story summary of a film clip to two listeners. The storyteller and the listeners are unacquainted. The

listeners, however, have previously been instructed by the experimenter to display positive affective expressions, for example smiling and nodding, or negative expressions, such as frowns and serious facial expressions (Beukeboom 2009). The results of the study confirmed that a speaker's language use depends on the listener's affective expressions. These findings are similar to the results of a study by Goodwin (1980), who investigated natural speech, specifically restarts and pauses at turn beginning. Generally, people perceive smiling and nodding as encouraging. They are signs of agreement and understanding. In contrast to this, closed bodily postures and frowns display a low level of agreement and "speakers feel that a more careful, analytic and descriptive style of formulating information is called for and refrain from interpretative statements" (Beukeboom 2009:753). Nonverbal behavior is an integral part of interactions and communications between people and is present in various genres, such as story-telling and natural speech. As Cook writes, "'catching someone's eye' makes it almost obligatory to start an interaction" (1977:331). To look at someone shows that one is attending to them, demonstrating interest in the story or the conversation.

Argyle and Cook (1976) and Cook (1977) have given a full account of gaze and mutual gaze: The biological and cultural bases of gaze, its relationship to personal attitudes and emotions, and finally the measurement of gaze. Bavelas et al. (2002a:570) write that "the expansion of research on gaze seems to have been virtually limited to the period Cook described", acknowledging that the research on gaze expanded in the 60s and 70s, but criticizing that it was not continued after the end of the decade. Bavelas et al. (2002a) and Bavelas and Gerwing (2007) also view the exclusion of dyadic communication from the focus of analysis as a problem. They argue that gestures and facial expressions have a communicative and interactional function and that face-to-face dialogue is highly reciprocal (Bavelas et al. 2002a:567). In a more recent study, Rossano et al. (2009) consider gaze an interactional practice and they provide a statistical analysis of it. Their aim is to answer the question of whether gaze has a universal property across cultures. The researchers compare three cultures and languages, Italian, Yélî Dnye and Tenejapan Tzeltal, and their analysis yields two main findings: there is a strong tendency for speakers to uniformity in their gaze behavior, regardless of their culture. Recipients, however, display a clearer cultural difference. As such, there is a differential role of gaze behavior, which varies across cultures and is "a signal of active recipiency" (Rossano et al. 2009:212).

Other strands of research, such as computer linguistics and robotics, have investigated eye gaze and gesture. For example, Kanda, Ishiguro and Ishida (2001) looked at gaze control to gain insights into how humans' impressions of a robot are influenced by the robot's gaze. A similar study by Sidner et al. (2004) researched the impact of "engagement gestures" looking at how the interaction between human and robot changes depending on the display of gestures and the lack thereof. Staudte (2010) conducted eye-tracking experiments with a special focus on joint attention, i.e., how interlocutors coordinate visual attention, transferring results from face-to-face communication to human-robot interaction. Eye-tracking technology enables researchers to transfer observational results onto computer models. Bayliss et al. (2013), for instance, created a computer-based task to examine how people's eye movement behavior was influenced by a congruent/incongruent gaze behavior displayed by different faces on a computer screen.

2.2 DISCOURSE ANALYSIS AND MODELS OF COMMUNICATION

In the beginning, the lack of a unified definition of communication was addressed. Schiffrin writes that "[s]ince language is the central means by which people communicate with one another in everyday life, understanding communication is an important goal for linguists" (1994:138). Using discourse is one means of communicating and therefore the problem of defining 'communication' can be addressed from a discourse analytic perspective. Discourse analysis is a vast field and interactional sociolinguistics and Conversation Analysis (CA) are just two of many approaches to discourse. Schiffrin (1994:21) presents two paradigms, formal versus functional, as well as a third paradigm that intersects with the previous two paradigms. Depending on the paradigm, the definition of and approaches to discourse differ, resulting in a different understanding of communication as well. Whereas formalists focus on the structure of language, functionalists stress the interdependence of discourse, society and social phenomena (Schiffrin 1994:31). Schiffrin (1994:40) proposes a third paradigm, viewing discourse as utterances and providing a balance between the structural and the formal perspective on language. My own work intersects with these different paradigms and approaches to discourse. Much research on multimodality is influenced by CA (see for example Schmitt 2007, Streeck 2009, Mortensen 2012), but phenomena such as the negotiation of alignment and interactional frames fall into Interactional Sociolinguistics. In the following, I will outline three models of communication and argue that, for current purposes, the *interactional model* is best suited to the

study of speech and gesture. The models of communication vary depending on the approach taken to discourse: There is a *code model*, an *inferential model* and an *interactional model* (Schiffrin 1994:386). In the code model and the inferential model of communication, the principle of *intersubjectivity* plays a crucial role, i.e., people must have a certain shared knowledge to successfully communicate with one another and, at the same time, by communicating with each other, they establish shared knowledge. Thus, intersubjectivity has a dual role both allowing communication and achieving it (Schiffrin 1994:390). Within the code model, it is assumed that there is code which sender and recipient need to encode and decode and that the sender has a certain intention to convey a thought. The inferential model differs from the code model insofar as it assumes that an individual communicates when the following three intentions are activated (Schiffrin 1994:393):

(a) S's utterance of x to produce a certain response r in a certain audience A
(b) A to recognize S's intention (a)
(c) A's recognition of S's intention (a) to function as at least part of A's reason for A's response r.

Crucial to this definition of communication is the fact that an individual communicates when he/she makes his/her intentions clear to others. Consequently, communicative behavior can be separated from non-communicative behavior in this model. Gesture research conducted by Bavelas et al. (2002b), for instance, relates to this notion because in their view, gestures are used intentionally to communicate.

Summarizing so far, there are certain goals individuals want to achieve by communicating: one person needs to recognize the other person's intentions; the recipient needs to mirror the communicator's display of intentions by which intersubjectivity is achieved; shared knowledge of the same linguistic code, the same rules and the same principles of communication ensure the achievement of intersubjectivity (Schiffrin 1994:395). The third model, the interactional model, is the one applied in this study for several reasons, which I am going to explain in the following. Underlying communicating in this model is behavior, even behavior that is not intentionally communicated. More importantly, in any interactive situation, individuals communicate information. The context in which the situation occurs as well as the background information that individuals bring to the communication become essential components of such an assumption of communication. The advantage of the interactional model

is that not only intentional gestures, as proposed by Bavelas et al. (2002b), but also spontaneous nonverbal behavior can be included in the analysis of conversations and discourse. Nonverbal behavior in this respect refers to a particular behavior labeled 'gesture', movements of arms and hands which happen spontaneously when people talk. As McNeill writes, "people unwittingly display their inner thoughts and ways of understanding events of the world. These gestures are the person's memories and thoughts rendered visible" (1992:12).

The interactional model is assumed in the interactional sociolinguistic approach to discourse. Within this approach, *contextualization cues* function as features of language that frame messages and mark how they should be understood. In Schiffrin's terms, "any one utterance is assumed to be sequentially relevant to what came before (its local context) and to a general framework of understandings about a particular 'type' of situation (its global context)" (1994:407). Communication is therefore not merely our knowledge of grammatical rules and well-formed sentences, but it is an activity between two or more individuals involved in equally a verbal exchange (Gumperz 1982) and a nonverbal exchange.

2.2.1 Interactional sociolinguistics

Interactional sociolinguistics grew out of different disciplines, among them anthropology, sociology and linguistics (see Schiffrin 1994). Two of the most influential contributions come from the linguistic anthropologist John Gumperz and the sociologist Erving Goffman, the first researcher focusing on face-to-face encounters and the second one looking at social interaction. Goffman writes, "[s]ocial interaction can be identified narrowly as that which uniquely transpires in social situations, that is, environments in which two or more individuals are physically in one another's response presence" (1982:2). Thus, drawing on both contributions, interactional sociolinguistics focuses on situated meaning and it looks at how utterances are contextualized by the participants and the activity (Schiffrin 1994:109). Depending on the situation, there can be two or more alternative interpretations of the same utterance. According to Gumperz (1982:3), individuals need to be involved in a conversation to understand one another and they need to share linguistic and socio-cultural knowledge to maintain this involvement. Potentially conflicting interpretations of conversational inferences can lead to communicative difficulties, which can also arise when people from different social and cultural backgrounds come into contact. Communicative difficulties, and one can assume communication per se, are usually

based on people's culturally bound perception of similarities and differences in the world. These similarities and differences must be understood for communication to be successful.

Gumperz (1982:16) refers to a differentiation drawn by Saussure, distinguishing 'marginal features' of language from 'core features' of language. Marginal features include *intonation, speech* and *rhythm* as well as *lexical, phonetic* and *syntactic choices* and even though they may not change the meaning of a message, they nevertheless influence the expressive quality of it. In Gumperz' terminology these mechanisms are called *contextualization cues*. According to Gumperz, "a contextualization cue is any feature of linguistic form that contributes to the signalling of contextual presupposition" (1982:131). He names a range of realizations of such cues: "The code, dialect, style switching processes, […] prosodic phenomena […] as well as choice among lexical and syntactic options, formulaic expressions, conversational openings, closings and sequencing strategies […]" Gumperz (1982:131). It is highly important to stress the lack of nonverbal phenomena and cues in this definition. Schiffrin briefly acknowledges nonverbal phenomena when she writes:

> aspects of language and behavior (verbal and nonverbal signs) that relate what is said to the contextual knowledge (including knowledge of particular activity types […]) that contributes to the presuppositions necessary to the accurate inferencing of what is meant (including, but not limited to, the illocutionary force) (1994:99-100).

These two definitions reflect the development in research, in particular in the field of discourse analysis, and underline the need for further research in this area. It is not the intention of the author to say that early research has neglected nonverbal phenomena. Both Gumperz (1982) and Goffman (1982) acknowledge nonverbal phenomena. Goffman, for instance, states that "[e]motion, mood, cognition, bodily orientation, and muscular effort are intrinsically involved" (1982:3) in participants' cognitive states when they are engaged in the interaction order. Traditional discourse analysis, however, has excluded these nonverbal signals leading to a focus on verbal aspects of communication and interaction.

2.2.2 Interaction and multimodality

Both Schmitt's (2007) edited volume *Koordination* and Norris' (2011) monograph *Identity in (Inter)action* present studies in a recent research field on multimodality growing out of the possibilities introduced by new technologies. New methods of collecting data, such as the

replacement of audio recordings by video recordings, have also rendered new ways of analyzing data. "Traditional" discourse analysis, its tools and methods, which were used to analyze data, investigated audio recordings, thus focusing on verbal phenomena. Newer approaches are influenced by video recording technology and therefore study visual data. Norris (2011) proposes a multimodal (inter)action analytic approach in her work on identity (co)construction. She writes that "the framework of mediated discourse analysis encourages an integration of non-verbal modes of communication into a discourse study" (Norris 2011:3). Deppermann and Schmitt (2007:16-7) base their analytic methods on Goffman's (1982) proposal of a microanalysis of a face-to-face domain because it enables the integration of both the sequential analysis done in Conversation Analysis and, from a multimodal perspective, the inclusion of the aspect of simultaneity (Mondada 2007). Multimodality and multimodal interaction require a new perspective on communication, one that includes coordination as a central aspect of interaction. Deppermann and Schmitt (2007:40) define this study objective as a prerequisite of goal-oriented cooperation. Coordination is a central issue in the understanding of communication because it reflects upon how speaker and listener cooperate during a conversation. Coordination, in Deppermann and Schmitt's sense, must be taken into consideration in the current study because the participants are presented with a task which is based on route planning, containing a certain degree of coordination in itself. In this 'joint activity' (Clark 1996) participants in the study engage in a joint action to overcome potential difficulties while planning by mutually coordinating the task and collaborating to create mutual understanding.

2.2.3 Coordination and spatial organization

Interactants in this study use a variety of resources to coordinate and manage the task at hand, from co-constructions and repetition to collaborative pointing and mutual gaze. Deppermann and Schmitt present a framework for the study of coordination, placing nine different "Ausdrucksmodi" (2007:25), or modes of expression, in the center of analysis:

1. Voice ("Stimme")
2. Sound structure ("Lautstruktur")
3. Gesticulation ("Gestikulation")
4. Facial expressions ("Mimik")
5. Gaze ("Blick)

6. Position of the body (“Körperhaltung”)
7. Orientation of the body (“Körperorientierung”)
8. Position in the room (“Position im Raum”)
9. Ways of moving (“Bewegungsarten”)

Voice and Sound Structure

There were no correlations found in the data analyzed here. Hence, in this study voice and sound structure do not constitute applicable modes of expressions.

Gesticulation

Adam Kendon (1980) introduced the term *gesticulation* to refer to movements of the hand(s), the arm(s), and even the whole body. He extended the understanding of movement to include not only movements involved in speech production, for example the movement of the lips and the face while speaking, but also the movements of hand(s) and arm(s). This study focuses on a range of hand and bodily actions, putting a special focus on deictic gestures. Four types of deictic gestures are identified here: the single pointing gesture, the repeated pointing gestures, collaborative pointing, and tracing. These types of deictic gestures are similar in form, but fulfill different functions and occur at different moments in the interaction.

Facial expressions

There are positive and negative facial expressions, such as smiling versus frowning. This paper will deal with these kinds of facial behaviors in a broader context and analyzes their meaning and function in connection to gaze and the overall achievement of the task whenever relevant.

Gaze

Gaze will be of special interest in this study. According to Tiittula (2007:223), looking at someone and mutual perception (“Wahrnehmen”) are social activities. As such, they play a constitutive role in the formation and maintenance of interaction. Three major forms of gaze are investigated and their functions are established: first, the *gaze direction* of participants, i.e., the direction of gaze toward the map/task sheet or the direction of gaze toward the other person; second, *gaze shift*, i.e., shift of gaze from the map/task sheet to other person and back again or shift of gaze between the map and the task sheet; third, *mutual gaze*, i.e., when both people look at each other. It is the aim of this study to describe the various forms of gaze in a certain type of

situation, in a task-based interaction, in order to understand how the organization of gaze is coordinated with verbal behaviors and other forms of expression.

Position of the body

The position of the body is relatively fixed because participants sit at a table. At the beginning and at the end when the task reaches completion, the participants sit in an upright position and they usually face each other. During the course of the activity their upper bodies are usually bent over the table.

Orientation of the body

The neutral orientation of the body is toward the other person because the participants are sitting across from each other and there are limitations in the way they can move. The participants' body orientation is face-to-face at the beginning and end of the task. During the planning process, the orientation of the body can vary. Schmitt and Deppermann (2007:99) describe a side-by-side and a face-to-back constellation in their study. Throughout the planning process, the orientation of the body can shift, taking a side-by-side orientation as participants lean in and turn their bodies so their shoulders are almost next to each other.

Figure 1 Screenshot of face-to-face orientation

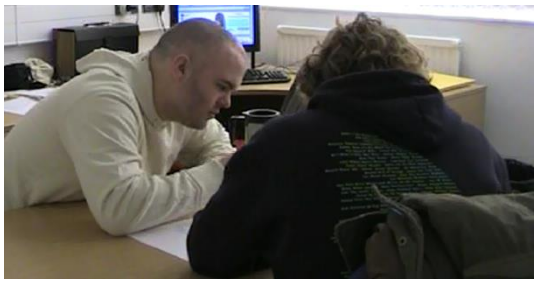

Figure 2 Screenshot of side-by-side orientation

Position in the room

The participants are sitting in the middle of the room. Both cameras are located on one side of the room opposite the window. One participant has the door to the other room behind him/her, facing the opposite wall, whereas the second person is facing the door, having the wall behind him/her.

Ways of moving

The participants are seated across a table from each other. They do not walk around the room, nor do they stand up fully. In some cases, participants will get up from their chairs to lean onto the table facilitating a closer look at the map/the instruction sheet.

Organization of the interaction space

As part of how the interaction space is organized and shaped, Kendon's (1990) F-formation system needs to be considered. The F-formation system is beneficial to the study of social encounters as "it provides a means by which the participants can maintain differential access to one another and it facilitates the maintenance of common focus of attention" (Kendon 1990:209). This system contains a spatial aspect of the individual's actions and Kendon emphasizes that a person's *lower body*, usually the feet, determines his/her location and orientation (1990:211). The *transactional segment* (Kendon 1990:211) is the space in front of a person where he/she carries out certain activities. Kendon provides different examples, such as the space between a television viewer seated on a sofa and the actual TV. In the space in front of the person, the activity of watching TV is carried out. In this space, gesticulations can occur, for instance when one holds the remote control and extends the arm toward the TV to change the channel. Within the transactional segment, the location and the orientation of the body frame the activity. Thus, when a person changes his/her body orientation or the location, this causes a shift in the activity and leads to a new frame of activity.

Applying Kendon's work to the current study, per definition the participants are not engaged in an F-formation because they can only display visible changes in body-orientation with their upper bodies and not with their feet. These limitations are due to the arrangement of the setting, the table, the cameras and so forth. However, there is still a transactional segment and a joint transactional space, the *o-space*, in which the participants "maintain joint jurisdiction and control" (Kendon 1990:211). Within the interaction space ("Interaktionsraum"; see Mondada

2007, Schmitt and Deppermann 2007), shifts in participants' body orientation can frame episodes within the interaction. For example, as soon as they begin the activity they lean forward. When there is a break in the planning process or when they reach a completion point they will retrieve their bodies from the shared space and when the session is finished completely, they will also change their location, i.e., get up to leave the room.

2.2.4 Cohesion

Two important components of the organization and resolution of the task are cohesion and reference. Cohesion has been treated from different perspectives in various strands of research, linguistic pragmatics being one of them. The most influential work on cohesion has been done by Halliday and Hasan, who published the monograph *Cohesion in English* in 1976. Halliday and Hasan differentiate between two major forms of cohesion, grammatical and lexical cohesion. Cohesion is defined as the "linguistic means whereby texture is achieved" (Halliday and Hasan 1976:293). Every text has texture, thereby differentiating it from something that is not a text. Cohesion can be expressed through grammar, for example by replacing a noun such as "a boy" with a personal pronoun "he". Cohesion can also be created through relationships between lexical items, for example by synonymous terms or collocations. The definition of cohesion can thus be extended to "the set of possibilities that exist in the language for making text hang together" (Halliday and Hasan 1976:18). Cohesion is also understood as a process that reflects the instantiation of relations within a text, one element presupposing the other. Even though Halliday and Hasan's (1976) work has been very influential, their study on cohesion only looked at text and not at real time conversations and gestures as they synchronize with speech.

In recent years, another form of cohesion has been analyzed: cohesion created through the use of gesture, viewing cohesion from the perspective of a speech-gesture synthesis. Levy and McNeill (1993) study referential cohesion in narrative and McNeill et al. (2010) place emphasis on floor control in military war gaming sessions. These two studies base their investigation on Halliday and Hasan's work (1976), but they present a notion of cohesion which included phenomena accounting for coherence as well. In McNeill and Levy (1993), the researchers present three forms of cohesion: one by space, one by handedness, and one by form. Cohesion by space means that gesture space is used to create cohesive links across narrative texts; regarding handedness, story tellers can use one or both hands to gesture and usually a complex gesture will

accompany a main clause; gesture forms, for instance when restarting a phrase, connect clauses to the crux of the story line. Parallel to Levy and McNeill's viewpoint, this study will highlight the importance of the assumption of a speech-gesture synthesis in the study of multimodal, interactive communication. I will adduce examples of cases where gestures and gaze co-occur with speech as well as cases where only one or the other mode is present. I will demonstrate the occurrence of modal shifts, by which I mean a shift from the verbal to the non-verbal mode and vice versa.

Cohesion must be differentiated from another phenomenon called coherence. Coherence as a conceptual phenomenon is partially "a reflection of how the content [of a discourse] comes together and is stored in the mind" (Dooley and Levinsohn 2001:21). Recipients of a message rely on world knowledge and expectations as well as cultural experience to construct a *mental representation* (Dooley and Levinsohn 2001:23). If the construction of such mental representations fails at some point, a discourse ceases to be coherent. Hence, "[a] text is said to be COHERENT if, for a certain hearer on a certain hearing/reading, he or she is able to fit its different elements into a single overall mental representation" (Dooley and Levinsohn 2001:23). Speakers therefore "plant linguistic signals in the text as clues to assist the hearers in coming up with an adequate mental representation" (Dooley and Levinsohn 2001:27). These linguistic signals are, in semantic terms, and according to Halliday and Hasan's viewpoint, considered cohesive ties. In this study, participants engage in activity based, interactional communicative acts. They need knowledge of a zoo and ideally, they should have been to a zoo before. Knowledge of how to read a map and of how to read and follow instructions is also required. Those are just some of the elements which are needed to produce with a single mental representation of the activity.

Schiffrin (1994:129) introduces the notion of sequential coherence, meaning that there are different interpretive frames for talk. These frames help us to understand how one utterance is followed by another. In Tannen's terms, these different utterances are considered metamessages, expressing "[w]hat is communicated about relationships—attitudes toward each other, the occasion, and what we are saying […]" (1986:29). Such metamessages are also found in meta level references, informing us about the state of the activity and the level of agreeing stance between the two participants. Two other constructs, contextual presupposition and situated inference, are part of the definition of sequential coherence. Schiffrin writes,

> Contextual presuppositions are a type of assumed background knowledge that allows the inferencing (during the course of an interaction) of two levels of meanings that are themselves related. One level is the communicative activity type […]. The second level is the particular illocutionary act that the speaker intends (1994:100).

The activity type is defined as "the basic socially significant unit of interaction in terms of which meaning is assessed" (Gumperz 1982:131). Gumperz further states that contextualization cues have implicit meanings. These meanings only become apparent when misunderstandings occur, leading to miscommunication. This notion is linked to the notion of expectations, as discussed by Tannen (1979). In the author's view, expectations are closely related to coherence. Tannen also approaches the notions of scripts, frames, and schemata to show that these notions "can be understood as structures of expectations based on past experience" (1979:179).

Linguistic signals alone do not suffice to conjure a mental representation. Speakers and listeners need non-linguistic signals and gestures in order for communication to be successful. A broader approach is therefore suggested, subsuming cohesion and coherence, to investigate how dyads of interactants shape and maintain the interaction, how they solve the task through verbal and non-verbal modes of expression.

2.2.5 Reference

Reference or referential cohesion (McNeill and Levy 1993; McNeill et al. 2010) can be established through words as well as through the contribution of gesture and gaze to discourse cohesion. Gestures, especially in the task-based environment, ensure a smoother communication by adding a separate layer of understanding to the interaction.

Halliday and Hasan (1976:37) define three kinds of reference, *personal*, *demonstrative*, and *comparative* reference. Reference is a special form of cohesion and its specific nature lies in the fact that "the information to be retrieved is the referential meaning, the identity of the particular thing or class of things that is being referred to" (Halliday and Hasan 1976:37). Looking at a short exchange between two participants in this study (excerpt taken from MOV00F), I will provide an analysis of the referential expressions according to Halliday and Hasan:

149.	Fiona	yeah so there's not really a thing you pass,
150.		is **that** Adventures with Aqua **there**,

151.		in **this** bit **here**?
152.		or is **that** a (blank)?
153.	Flavia	°I have no idea.°
154.		it doesn't look like it.
155.		(it's got lines coming out)
156.	Fiona	yeah.
157.	Flavia	I think it's like **this** (bit) **here**.

This excerpt contains examples of personal and of demonstrative reference. In line 153, Flavia uses the personal pronoun *I*, which can only refer to the speaker. *You* in line 149 is a generalized use of the pronoun, meaning 'any human individual' and constituting an example of an exophoric reference. This type of reference is a situational reference and it contrasts with textual or endophoric reference, as Halliday and Hasan (1976) state. The demonstratives are marked in bold in the transcript. Demonstrative reference can be distinguished in terms of proximity, between 'near' *this/these*, *here* and 'not near' *that/those*, *there* (Halliday and Hasan 1976:60). This closeness and distance is not only meant spatially, but also in terms of closeness or distance to the speaker. In lines 150, 151, and 157 in the transcript, *that* and *this* precede nouns, for instance. They are demonstrative adjectives and they modify the following noun. *That* in line 152, however, is an exophoric reference. It takes more than knowledge of semantic relations for Flavia to understand what Fiona means. *Here* and *there* in lines 150, 151 and 157 are similar to *that* because textual relations do not suffice to identify the referents. In order to understand and identify the referent and to agree on the same referent, the indication via a gesture or a gaze shift is important, as the results of this study show.

Clark and Wilkes-Gibbs (1986) use the term *reference* not to mean semantic reference, but speaker's reference, i.e., what the speaker does when he refers to a person or a thing. They propose a "collaborative model for the process of reference" (1986:3), contrasting written to spoken discourse, in particular the formulation of noun phrases in conversations. Their model is based on the assumption that when speaker A makes a reference to a thing X, it is A's intention to make X part of A's and addressee's B mutual knowledge. B must try to understand A's reference and must inform A of his/her understanding. Thus A and B "accept mutual responsibility for each definite reference" (Clark and Wilkes-Gibbs 1986:8). In a *basic exchange*,

which is the minimal unit of referring and accepting the reference, mutual acceptance can be reached through three processes: *initiating*, *refashioning*, and *evaluation* (Clark and Wilkes-Gibbs 1986:16-17). In initiating a direct reference, Clark and Wilkes-Gibbs differentiate six types of noun phrases by which the speaker offers an initial presentation to the listener. The example sentences are derived from the present data.

1. *Elementary* noun phrase: this type of noun phrase usually occurs in a basic exchange and can take this form, for instance: *the entrance on the left.*
2. *Episodic* noun phrase: as the term suggest, the noun phrase is uttered in episode or tone groups, for instance *the entrance on the left, with the big sign*.
3. *Installment* noun phrase: here there is an acceptance after each episode, such as
   A. the entrance on the left,
   B. okay.
   A. with the big sign,
4. *Provisional* noun phrase is a kind of phrase that is inadequate so that the speaker will expand on it, this time in a new clause and not as an expansion of the initial phrase. For example: *Then we go to the Pachyderm house. It's like elephants and stuff.*
5. *Dummy* noun phrase: this is uttered instead of a complete noun phrase, e.g. *thingamajig*.
6. *Proxy* noun phrase: this is a kind of joint production where the listener can present the final part of the noun phrase if he/she knows what it is:
   A. the entrance on uh ..
   B. on the left,
   A. on the left,
   B. okay.

Unacceptable noun phrases need to be refashioned, for which Clark and Wilkes-Gibbs propose three possibilities. Noun phrases can be repaired. The repair can happen via an expansion of the utterance, as it has been shown for a provisional noun phrase, or via a replacement of a noun phrase. Repairs can be self- or other-initiated as can be expansions and replacements. All these propositions need to be judged as acceptable or unacceptable and the evaluation can either happen in the form of an acceptance, a rejection, or a postponement. In summary, Clark and Wilkes-Gibbs' collaborative model provides a comprehensive analysis of how the process of

making a reference is a collaborative process that relies on establishing the belief that the listener has understood the speaker's reference.

In a broader context, this idea relies on the assumption that two or more interlocutors need to share common ground to understand the meaning of an utterance, for instance. In this study, it is proposed that two features of reference, referential repetition as well as meta level reference, account for the phenomenon of common ground. In one of my interactions, one of the participants says "Should we make a scale of this map" (MOV00C). This single utterance depicts how coherence and cohesion are connected: there is cohesion through the demonstrative "this", which, in fact, is not necessary since there is only one map and no need for disambiguation. Nevertheless, the speaker chooses to refer to a specific map and further suggests scaling it. In order to utter this suggestion, both the speaker and the listener must not only recognize and understand the individual lexical items as well as the structure of the utterance, namely the structure of a question, but, more importantly, they must know what it means to scale a map and ideally, they will have done it at least once before. The results of this study demonstrate that common ground is established and re-established through verbal and nonverbal elements as the interactive discourse and the activity progress. There is gesture recurrence and meta level reference, for instance, to reach mutual acceptance and to continue on the basis of this agreement.

2.2.6 Common ground

Common ground relates to the knowledge and the beliefs that two or more people share. Common ground is that kind of information that people take for granted, for example background knowledge. Common ground consists of the presuppositions speakers have when they speak (Stalnaker 2002), but also the assumptions they work to make shared (Clark 1996). For instance, in the example "the entrance on the left, with the big sign", the speaker makes an explicit reference to the entrance and specifies which entrance he/she means, thereby contextualizing and making sure speaker and addressee share the same presupposition.

Stalnaker states that what speakers presuppose "guides both what they choose to say and how they intend what they say to be interpreted" (2002:701). Stalnaker reminds us that speakers have intentions when they mean things and that they expect hearers to recognize their intentions to communicate. In Stalnaker's conception of common ground, presuppositions play a crucial

role as they relate to Grice's notion of *speaker meaning* (Grice 1989, originally printed in 1957). A presupposition is defined as "a proposition or inference whose truth is taken for granted in the utterance of a sentence" (Huang 2011:401). Huang further explains that Stalnaker's conception of presuppositions is a pragmatic one and as such, they "are treated as conditions on whether a sentence can be admitted into a context" (2011:407). In other words, the question is what speakers assume to be common ground when they use certain expressions in a conversation. Essential to this notion of common ground is common belief, determined by the individual beliefs of a group. Speaker and hearer share common beliefs and beliefs about common beliefs. Stalnaker explains that a speaker's presuppositions "can be identified with what the speaker believes to be common belief" (2002:707). Stalnaker further differentiates between two kinds of beliefs, the beliefs about a conversation topic and the beliefs about the current conversation itself. These beliefs are due to constant change, just as common ground can undergo change in conversations. In summary, Stalnaker assumes a relationship between presupposition, common belief, and common ground that reflects the dynamics of discourse. He writes, "the common ground could be identified with common belief, and that what a speaker was presupposing could be identified with the speaker's beliefs about common belief" (2002:715).

Shared knowledge is, in part, a product of communication. Common ground is repeatedly updated in dialogue (Clark and Krych 2004:76). In order to establish common ground, people need to work together to establish and agree on what they consider mutual knowledge. In a message to Fulbright Scholars published in 2012, Secretary Hillary Clinton addresses the notion of establishing common ground between nations and cultures when she says that it takes not only governments cooperating with each other, but "it [also] takes individuals working to find common ground". This statement nicely reflects the work that needs to be done when two or more people communicate with each other. Stalnaker (2002) introduces the idea of a defective context, which is a context in which a speaker's presupposition is presumptive and the speaker's beliefs are incongruent with the addressee's beliefs. The addressee can either correct the speaker or he/she can accommodate to the speaker's presupposition by accepting it as part of the common ground. Hence, common beliefs in this context can vary, but the common ground is accepted in order to facilitate successful communication (Stalnaker 2002:717-718).

Stalnaker's work is largely focused on the speaker and his/her presuppositions. As such, much of his reasoning develops out of a unidimensional perspective. However, as his example of

a defective context demonstrates, common ground needs to be viewed as product of a mutual effort by both the speaker and the listener. There needs to be a display of either acceptance or refusal to create common ground. There must also be a display of mutual understanding to ensure that common ground is successfully reached.

Clark, Schreuder, and Buttrick (1983:247) explain that common ground relies on three sources: *perceptual evidence*, *linguistics evidence*, and *community membership*. The perceptual evidence relates to what people jointly experience, in this case, the engagement in a dyadic task. What the participants say and hear, for example information about the process of the experiment and the consecutive planning carried out by them, accounts for the linguistic evidence. Lastly, their membership of a certain community, be it the university community or the Birmingham area community, makes up their beliefs of what is universally known.

Clark proposes that actions come in hierarchies, which he calls *action ladders* (1996:147-149). Each action ladder consists of four levels. On each level, a joint action takes place and grounding can happen on each level. The four levels are the following: *Level 1*. The speaker must bring the listener to attend to his/her voice or gesture; *Level 2*. The speaker must bring the listener to identify his/her words and/or gesture; to reach *Level 3*. The speaker must bring the listener to understand what he/she means; *Level 4*. The speaker must bring the listener to consider what he/she has said (Clark 1996; Clark and Krych 2004). Grounding can be successful, for example when addressees use neutral continuers such as *mhm*. Overlaps and interruptions, however, can also interfere with grounding. The success of grounding and the lack thereof relates to another concept, the phenomenon of how people express stance toward an utterance or a situation.

2.2.7 Stance taking

When people take a stance toward something, they position themselves in relation to others, their utterances and actions. Du Bois defines ‘stance’ as “a linguistically articulated form of social action whose meaning is to be construed within the broader scope of language, interaction, and sociocultural value” (2007:139). Du Bois further states that “the value of any stance utterance tends to be shaped by its framing through the collaborative act of co-participants in dialogic interaction” (2007:141). A stance utterance can either express agreement or disagreement and hence is used to align or disalign with someone respectively. Stance is also related to reference

and referential grounding as a given stance is directed at an object that we need to recognize in order to understand both the reference and the stance taken toward it.

DuBois' (2007) work on stance is usually cited for his development of the stance-triangle. Based on three types of stance, *evaluation*, *positioning*, and *alignment*, DuBois establishes the relationship between the three nodes of the triangle. If a speaker takes a stance, he/she evaluates something by which he/she positions the self and thus aligns with another person. The key entities, which DuBois places at each node of the triangle, are "the first subject, the second subject, and the (shared) stance object" (2007:164). The first and the second subject are connected to each other via the vector of alignment. Each subject also stands in a relationship to the stance object, which is represented by the evaluative vector. Stivers (2008) draws a distinction between alignment and affiliation. Alignment is shown to be a structural feature whereas affiliation fulfills a social role. In her analysis, Stivers (2008) focuses on vocal continuers and nods, which participants to storytelling events display either mid-story or toward the end of the story. Her results show that vocal continuers and nods are treated differently, both by the teller and the listener. By nodding, the listener can claim access to the teller's stance and thus suggest that the story will receive an affiliative uptake (Stivers 2008).

Davies and Harré (2007:49) differentiate two forms of positioning, depending on how a person positions himself/herself in a conversation: there is *interactive* and *reflexive* positioning. Interactive positioning describes the phenomenon of how one person is positioned in a conversation by someone else's, whereas reflexive positioning relates to how one positions himself/herself in a conversation. One way of positioning oneself in a conversation is via assessments. Pomerantz (1984) and Goodwin and Goodwin (1987) investigate the role of assessments in interactions and Raymond and Heritage (2006) establish how interlocutors create situation specific identities in the course of an action. Assessments in Pomerantz' (1984) terms are products of an interaction, relating to a referent to which a person claims access. To borrow Pomerantz' example, when someone speaks about the water temperature of a lake or the ocean saying "it's just wonderful" (1984:27), this assessment follows the person feeling and experiencing the water. The participants in this study do not physically attend a zoo, i.e., they do not experience their actions at the zoo in this particular moment. However, there are assessments which relate to prior experience and background knowledge, e.g., "this is the busiest zoo I've ever seen." (MOV00L, line 87). There are assessments which relate to the visual input provided

by the map, e.g. “this zoo is quite cool” (MOVOOA, line 111). There are also assessements which refer to prior activities within the experiment setting, e.g., “what a day we just had.” (MOV00A, line 247). Assessments in a sequential organization follow a preference structure. There is a preferred and a dispreferred next action following an initial action (Pomerantz 1984:64). If one interlocutor’s first assessment invites agreement, then the preferred next action will be agreement expressed by the other interlocutor. However, in the case of self-deprecation, as Pomerantz (1984) explains, agreement is dispreferred. In other words, disagreement is the preferred action. By providing or not providing a preferred action, the second interlocutor positions himself/herself in the course of action. Building upon the sequential organization of assessments, Goodwin and Goodwin (1987) view ‘assessment’ as relating to different organizational levels within an action. They include assessment signals such as evaluative adjectives, but also take nonsegmental signals such as intonation or nodding into account (Goodwin and Goodwin 1987). The term *assessment* can further refer to *assessment action*, which Goodwin and Goodwin understand as a form of speech act in which the “action being performed by an actor” (1987:8) is emphasized.

Raymond and Heritage’s work places the notion of actors performing actions into relation to *social structure* (2006:278). They define the multiple identities of a person as a key constituent in the mechanisms of actions in interaction. In the exemplary grandmother-granddaughter telephone conversation, they investigate how claims to knowledge (and lack thereof) are made visible through assessments, for instance. The grandmother’s “epistemic privileges” and her identity as a grandparent are sustained by the boundaries which are drawn in relation to knowledge and rights to knowledge (Raymond and Heritage 2006:700). Transferring these findings to the current study, the course of the actions and the participants’ epistemic social relations are not only visible in their utterances and nonverbal behaviors, but also in the ownership of the map and the task sheet. The organization and distribution of the map and the task sheet provide information about the state of the activity as well as the individual’s role within the activity. Different orientation states and gaze patterns emerge as a result of the orientational and interactional setting.

Penz (2007) introduces the notion of meta-communication as a means to establish common ground between two or more people. She writes that “speakers use metapragmatic comments to communicate on the discourse they are involved in and thus monitor the discourse”

(Penz 2007:263). Meta-comments and meta-references, as the results of this study show, are often accompanied by a gaze re-direction, for example from the map to the other participant, and sometimes even a shift in body-posture. Stance taking is thus not only a verbal phenomenon, but is expressed through body movement and facial displays as well.

Baynham (2011) analyzes narratives that are told during interview sessions. The findings of his study show that across different types of narratives, a shift in and out of performance is connected to how tellers position themselves to the current discourse and to the story content (Baynham 2011:70). By way of *small stories* (Baynham 2011:71; Bamberg 2006), interviewer and interviewee can express stance toward each other and their respective roles, and they can align to the topic of the interview. The notion of stance is closely related to grounding and intertwined with it are agreeing and also understanding. In the subsequent chapters, the notions of positioning, grounding, and stance taking, will be brought into relationship with reference and repetition, on the one hand, as well as gesture and gaze, on the other hand.

3 METHODS

The data presented and analyzed in this study derive from video-recorded data which were collected at Birmingham City University in Birmingham, England, in November 2009. The video material was recorded during a one-week research visit to Birmingham City University in co-operation with the university's English Department.

3.1 PARTICIPANTS

Forty-seven native speakers of English participated in the experiment. There was a 3:1 ratio of thirty-five female and twelve male students. The participants generally came from the Birmingham area, with the exception of one male Australian student. They were all students of English at Birmingham City University and participated voluntarily following an invitation from their professor. All participants gave written informed consent prior to taking part in the experiment, acknowledging the usage of video, audio and picture material for scientific research and publication.

3.2 PROCEDURES

The participants were divided into groups of four, and these groups were further split into pairs. Due to the uneven number of students (forty-seven), there were eleven groups of four people and one group of three. The last group is excluded from analysis. The participants were seated at a table across from each other, facing each other, and the experiments were recorded with two Canon Legria video cameras providing two vantage points to ensure that one camera each could be directed at an individual participant to capture both gestures and facial expressions. Figure 1 shows a simple layout of the seating arrangements in the room in which the recordings took place as well as the camera configuration.

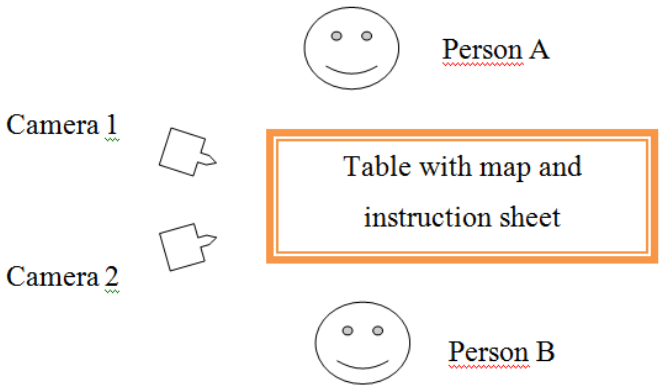

Figure 3 Setting: seating position, camera position, interaction space

The participants were given a map as well as a task sheet and were instructed to familiarize themselves with both sheets of paper. The participants were asked to read the instructions carefully and to locate every item mentioned on the map in order to explain in detail how to get there. No further instructions were given, unless participants had specific questions. Two pens were available in case participants wanted take notes or draw on the map, although this again was not explicitly stated.

3.3 TASK

The recorded dataset is two-fold: there is a so called *planning phase* and a following *telling phase*. The examples presented here all derive from the first part of the experiment. In the planning phase, dyads of participants were presented with a physical map of Brookfield Zoo in Chicago[2] and an instruction sheet, which indicates times (beginning and end of the day), different animal exhibits to visit as well as a list of other activities, such as seeing the dolphin show. The task is similar to a path planning task, such as the Traveling Salesman Problem (TSP) (Wiener and Tenbrink 2008), which is based on the idea that a salesman travels from location to location trying to avoid unnecessary detours to reach his destination on the shortest possible route. The TSP has been generalized and studied in computational mathematics, for instance. The Map Task (Anderson and Boyle 1994; Howarth and Anderson 2007) is another kind of task

[2] Map of Brookfield Zoo. http://www.czs.org/CZS/Brookfield/Zoo-Map/Brookfield-Zoo-map-2010 (01.11.2010)

designed for two participants, the Instruction Giver and the Instruction Follower, who collaborate to fill gaps of landmarks on their respective maps. The Instruction Giver explains the route to the Instruction Follower, who has to reproduce the route on his/her map. Such methods are also used in classroom interactions, for example when students acquire and practice the terms for directions and locations. In contrast to research that employs the Map Task, participants in this study have equal access to the task sheet and the map and both participants use the same task sheet and map while planning their outing at the zoo. Studies by Cohen and Harrison (1973), Klein (1982), Wunderlich and Reinelt (1982) and Kita (2003b) have used route planning and direction giving to elicit data. Goodwin (2003) analyzed related phenomena which occurred during the mapping of archaeological sites.

There are two versions of the instructions[3], similar in structure and layout, but different in the individual items and activities. Above the actual list shown below, there was a heading "A day at Brookfield Zoo" and a short written instruction, similar to the verbal one given by the experimenter: "Work together in pairs. Follow the instructions below to plan your day at the zoo." Some of the participants referred back to this description when they were uncertain about the proceedings during their task performance.

Start at the North Gate at 9 a.m. You go visit: - the Camels - the Habitat Africa - the Australia House Describe in detail how to get there. At 10.30 a.m. you go see the Dolphin Show at "Seven Seas". You have to pass "The Fragile Kingdom" to get there. After the show you want to go on the "Motor Safari" and meet your friends for lunch at "La	Start at the South Gate at 9 a.m. You go visit: - the Bear Grottos - the Fragile Kingdom - the Pachyderm House Describe in detail how to get there. At 11 a.m. you go to "The Living Coast" to see the feeding of the penguins. You have to pass the "Great Bear Wilderness" to get there. After the show you want to go on the "Motor Safari" and meet your friends at the "Roosevelt

[3] Pair A was given the first version, pair B the second version of the instructions. This was important for the second half of the experiment, during which one person each from pair A and B were re-grouped and asked to tell a story about their day at the zoo.

Gran Cocina" afterwards. Make sure to find the closest stops to hop on and off the ride. After lunch, you want to see - the Pachyderm House, - the Butterflies - and return to the North Gate at 4 p.m.	Fountain" afterwards. Make sure to find the closest stops to hop on and off the ride. From there, you go - to the Australia House - to the Butterflies - and return to the South Gate at 4 p.m.

Figure 4 Two versions of the task sheet

3.4 SELECTION OF DATA

Ninety-four videos, two for each session, were collected, varying in length from six to about twenty-five minutes. On average, however, the recordings are about eleven minutes. There are twenty-one videos of the first part of the experiment, the planning phase. These twenty-one recordings constitute the set of data analyzed in this study. Two out of the twenty-three videos had to be excluded due to insufficient quality. The data analyzed consist of twelve female dyadic interactions, three male-male interactions and six male-female interactions. The videos are labeled MOV00 followed by a letter, A, B, C and so forth. Moreover, all the names of the participants were made anonymous to protect their privacy. Hence, MOV00A represents pair A and they received names starting with the latter A as well. This is the complete list of names also indicating participants' seating positions at the table.

	Video	***Left***	***Right***
1.	*MOV00A*	*Anna*	*AJ*
2.	*MOV00B*	*Beth*	*Ben*
3.	*MOV00C*	*Clare*	*Cloe*
4.	*MOV00D*	*Dan*	*David*
5.	*MOV00F*	*Fiona*	*Flavia*
6.	*MOV00G*	*Gabriel*	*Gavin*
7.	*MOV00H*	*Helen*	*Harry*
8.	*MOV00I*	*Iris*	*Ian*
9.	*MOV00J*	*Janet*	*Jennifer*
10.	*MOV00K*	*Kara*	*Karin*
11.	*MOV00L*	*Laura*	*Larissa*

12.	*MOV00M*	*Madeline*	*Marina*
13.	*MOV00N*	*Norman*	*Neville*
14.	*MOV00O*	*Olga*	*Olivia*
15.	*MOV00Q*	*Quintina*	*Queena*
16.	*MOV00R*	*Rita*	*Rose*
17.	*MOV00S*	*Susan*	*Sabrina*
18.	*MOV00T*	*Tom*	*Tamara*
19.	*MOV00U*	*Uma*	*Ulrike*
20.	*MOV00W*	*Wendy*	*Wilma*
21.	*MOV00X*	*Xenia*	*Xandra*

Figure 5 List of participants' names according to recording

The current research project also uses visual prompts in the form of screenshots to demonstrate and explain certain phenomena. In consideration of ethical standards, I tried to select picture material which does not show participants' faces. However, in a multimodal setting and with the current goal of this study, case decisions must be made. The eyes and the face of the participant must be visible, for example to depict gaze orientation. The participants were informed of the usage of video and picture material and gave their written, informed consent.

3.5 TRANSCRIPTION CONVENTIONS

Transcriptions conventions of face-to-face interactions are manifold, varying, for instance, in terminology, design, and layout (see Kreuz and Riordan 2011). The transcripts for the spoken part of this study were produced in line with the conventions provided in the appendix. Speech is segmented into intonation units. The transcripts include verbal, prosodic, and paralinguistic features, providing a faithful representation of spoken language data. A typical piece of transcript looks like this (derived from MOV00A):

	Name	*Intonation Unit*
1.	Anna	but the car is pointing .. left,
2.		but we need to go right.
3.		so .. DO WE draw on this?
4.		((laughter))

The intonations units end with prosodic features and here we find continuing (line 1), falling (line 2) and rising (line 3) intonation at the end of the unit. There are short pauses in lines 1 and 3 and the words "do we" in line 3 are spoken louder than the surrounding discourse. Other aspects of an utterance, such as laughter, are also represented in the intonation unit.

The data were transcribed in several steps. First, a blind transcript was produced, which means that speech was transcribed based on sound only. In a second step, visual input was added, i.e., the video recordings were used to revise the transcripts of the spoken language data. The advantage of such an approach is that unclear passages can frequently be understood or corrected based on the visual input. To account for intertranscriber reliability, the transcripts were revised three times: a student assistant provided an initial, blind transcript and corrected this transcript based on the visual input. In the last phase, I compared and corrected the written representation to the recorded speech. The nonverbal behaviors were included in two separate steps after the verbal transcript was completed. Gestures were added first, followed by gaze. The video recordings were watched several times to ensure an exact and detailed representation of the nonverbal dimension.

For the purpose of this study, the transcription conventions and the format of the transcript itself must be extended to include particular kinds of nonverbal phenomena. According to Edwards, "the best choice of conventions in a given instance depends on the nature of the interaction, the theoretical framework, and the research question" (2001:321). In the original conventions, nonverbal behavior was captured and transcribed within the intonation unit. For the current purpose, nonverbal phenomena, in particular gesture and gaze, will be transcribed in the "gesture/gaze unit", parallel to the intonation unit.

Ochs (1979:51-61) addressed the issue of the placement of verbal and non-verbal phenomena in a transcript with regard to adult-child interaction. Ochs' discussion is a helpful basis for the present study as she addresses the issue of foregrounding speech over nonverbal phenomena in transcripts. Ochs presents different conventions which are used in psycholinguistics to represent children's behavior, both verbal and nonverbal. One such convention is to place nonverbal behavior in a separate unit from verbal behavior and Ochs states that there are advantages to this convention from a methodological point of view. Especially when there is a great amount of nonverbal data, a visual separation of these behaviors from verbal behaviors is meaningful. However, Ochs writes that "the transcriber heightens the

perception of these behaviors as distinct" (1979:54) when he/she places verbal and nonverbal phenomena into two different columns. She thereby addresses the issue of proximity. Proximity of speech and gesture might not be as applicable and timing not as obvious compared to transcripts where speech and gesture are represented in one line. However, due to reasons of readability, a separation into columns as a display on the page is meaningful, especially because individual participants are indicated and their behaviors are represented. In addition, nonverbal phenomena often require an explanation to fully understand their production and their meaning. In the present study, these explanations will be given as part of the analysis. When speech, gesture, and other nonverbal phenomena interoccur (Ochs 1979:58), for example when a gaze shift occurs in mid-intonation unit, this information will be provided as part of the description and analysis as well.

For the transcript itself a notation system which contains information about deictic gestures, their production, and information about gaze patterns, was developed and applied. The following example transcript depicts how the original transcription conventions were adapted to represent gesture and gaze.

	Name	*Intonation Unit*	*Gesture/gaze unit*
1.	AJ	start at the South Gate at nine am.	RIF point
2.	Anna	we're going to the Bear Grotto aren't we?	
3.	AJ	sure are.	gaze at TS
4.	Anna	so Bear Grotto is here,	RH point

AJ and Anna both take turns speaking to co-construct the routes. AJ uses his right index finger (RIF) to point. Anna holds the pen in her right hand (RH) and there is point to the map in line 4, which happens in synchrony with "Bear Grotto". Following Anna's question in line 2, AJ directs his gaze from the map (M) to the task sheet (TS).

In the transcripts, the following abbreviations will be used throughout: right hand is abbreviated RH, left hand LH, the right index finger is RIF, the left index finger is LIF, and both hands are indicated with BH. The map is M and the task sheet is TS. A single pointing gesture is labeled 'point'. Other forms of pointing are named in the transcript, for example repeated

pointing and tracing. Gaze is transcribed for gaze focus, namely to the map or the task sheet as well as gaze redirections to the interactant. For mutual gaze, the symbol "←" is used. It is included in the transcript once the second interactant looks at the gaze initiator, resulting in mutual gaze.

4 JOINT ACTION AND INTERACTION OPENINGS

According to Clark, language is used to form a *joint action* (1996:3) which is carried out by an ensemble of people. The participants of this study engaged in a joint action meaning that they "coordinate their individual processes" (Clark 1996:18) in such a way that they form a pair of people trying to achieve the same goal. In order to answer the question of how participants in this study shape joint actions, one must look at the beginning of the interactions, i.e., once the researcher has left the room and the two participants are to decide who is doing what. Possible openings could, for instance, be based on verbalizations of questions such as "Do you want to start?". Participants can also resort to nonverbal openings and I propose that this is reflected in the negotiation of the ownership of the map and the task sheet. Both verbal and nonverbal features of the activity openings are important indications of the creation of joint attention and the establishment of participants' *activity roles* (Clark 1996:33), which are such roles that people fulfill while taking part in an activity.

Goffman (1981) addresses the close relationship between bodily conduct and verbal openings, when two or more participants engage in a new activity with one another. He writes that

> […] a substantive, naturally bounded stretch of interaction comprising all that relevantly goes from the moment two (or more) individuals open such dealings between themselves and continuing until they finally close this activity out. The opening will typically be marked by the participants turning from their several disjointed orientations, moving together and bodily addressing one another […]. Typically, ritual brackets will also be found, such as greetings and farewells, these establishing and terminating open, official, joint engagement, that is, ratified participation. In summary, a "social encounter" (1981:130).

The most basic participant roles, which are often assumed in linguistics, are those of speaker and hearer. Goffman (1981; see also McCawley (1999) for a discussion of Goffman's work), however, points out that this global distinction does not suffice and that the role of hearer and speaker can each be further specified, depending on social encounters and context. There are three roles that fall into the speaker role: 1) the *animator*, which Goffman describes as a "sounding box" or "talking machine"; 2) the *author*, which is the person who selects his/her

words and the way in which they are encoded; and 3) the *principal*, the person who establishes his/her role by the words that are spoken (Goffman 1981:144). With regard to the hearer, he/she can either be the person who is addressed by the speaker, which is necessarily the case in a two-party conversation, or alternatively in multiparty conversations, some hearers can remain unadressed. Thus, Goffman differentiates the "addressed" from the "unaddressed" hearer (1981:133), which is closely related to the notion of intention. In public places, for instance, one person or several people can be the intended recipients of talk, whereas other people are considered bystanders, who eavesdrop or overhear a conversation (Goffman 1981:131-132). The present analysis of activity roles demonstrates that both participants can have equally important roles for the fulfillment of the task even though one person speaks while the other one takes notes, for instance. Both participants are intentionally involved in the activity and thus in this study the general terms *participant(s)* and *interactant(s)* will be used; specification of individual roles will be based on activity roles.

In Clark's terms, participants first have to be ratified as taking part in the activity, opposed to bystanders or nonparticipants, before they can ratify their *activity roles* (1996:32-33). The two participants pursue one dominant goal, a *domain goal* (Clark 1996:34), i.e., to complete the task they are presented with by following the instructions given to them. In this goal-oriented cooperation, they also fulfill what Clark termed *procedural* and *interpersonal goals*, negotiating the quickest and most efficient routes through the zoo, for instance, while maintaining the interactive frame with the other person. *Private goals*, such goals that are pursued out of personal reasons, are not relevant to the data presented here, nor is it likely that in an experimental context one person wants to deceive the other or turn it into a competition. Hence, only *public goals* will be considered. In contrast to some of Clark's examples of participant roles, such as guide-tourist, customer-server, musicians-audience, moreover, participants in this experiment have to assign roles to themselves. With regard to socioeconomic background, age and education, one can say that almost all participants enter the experiment at a fairly equal level. These and any other extra-linguistic criteria cannot be taken into account since this study does not take a sociolinguistic approach. It is, however, essential to investigate the interaction openings closely to see how roles within the planning process are distributed and how the interactions are structured. I propose that interactants have at least three choices of activity governance: they can be engaged in an egalitarian interaction in which they have equal access to the map and the task sheet maintaining

a cooperative frame throughout the planning process. Participants can also maintain the egalitarian frame, but distribute roles amongst them. For the third interactive form it is suggested that one person is in charge of the map or the instructions, while the other participant is fairly passive in the planning process. This also relates to the general dimensions that make up an activity. Clark (1996:31) lists the following dimensions: *scripted vs. unscripted, formal vs. informal, verbal vs. nonverbal, cooperative vs. competitive, egalitarian vs. autocratic*. The setting of the interactions was at the university, so in a rather formal environment. The students were invited by their lecturer to participate and the interactions were filmed. The activity type is scripted in so far as the participants are required to follow instructions. They also need certain knowledge about how to read a map and how to transfer the information from the task sheet onto the map. There is both verbalness as well as nonverbalness. The participants are cooperative on a general level concerning the dominant goal, but it remains to be investigated in how far they cooperate with each other on a microlevel since this is important for the overall organization and the completion of the task.

Common ground within an activity also needs to be negotiated and I propose that in the current study, not only verbal representations, but also nonverbal cues, namely the map and the instruction sheet, reflect how common ground is established. Borrowing Clark's term, *external representations* (1996:47), I suggest that the map and the task sheet provide information about the state of the activity, for example because they function as a *reference space* (Goodwin 2000:2015), which is a space that the interactants can refer to as they establish a shared orientation. The map and the task sheet can also be employed to signal the activity opening nonverbally. The actions taking place through the activity need to be understood as inclusive of speech as well as gesture and nonverbal phenomena. The activity closings, moreover, will demonstrate how common ground is not only achieved, but also asserted.

I will now present the results of the analysis of the distribution of the map and the task sheet. These findings derive from a purely visual investigation of the activity openings, i.e., the ratification of the ownership of the map and the task sheet. It is also hypothesized that the distribution of the map/task sheet first has a strong influence on the activity roles and second that these two factors (distribution and activity roles) further influence attention states. By attention states I mean the way in which participants pay attention to the same object and event or to a

different object/event. I will present one typical transcript each of three kinds of activity openings happening in the verbal and the nonverbal mode.

4.1 OPENING THE INTERACTION

When participants entered the room, they were asked to sit across a table from each other. At this point, the participants shifted out of their previous business (they waited in a separate room) and roles might have been ratified for the first time. However, the researcher cannot make any statements about it since the cameras were started afterwards. The map and the task sheet were placed in the middle of the table once the participants were seated and instructions were then given. The investigation of the beginnings will start at the point when the researcher has asked the participants if there were any questions and is about to leave room. At this point the two-party interaction begins.

Schegloff (1968) analyzed telephone conversations with a particular focus on how caller and receiver open the conversation. There are certain distribution rules at the beginning of a telephone conversation and the openings are usually geared at identifying and establishing the relationship of caller and answerer as well as the purpose of the call. Schegloff proposes that there is a sequence at the opening of telephone conversations that can be shaped as a question-answer (QA) or summons-answer (SA) pair. A summons, for instance, can be the ring of the telephone functioning as an attention-getting device and the person who is called answers this summons by picking up the receiver. Such a SA sequence only functions as a preface or preliminary to a further conversation (Schegloff 1968:1081). Phone conversations vary from face-to-face conversations in so far as people are not in the presence of each other and consequently gaze cues and shift of body orientation cannot be applied to receive someone's attention. Schegloff (1968:1082) gives the example of a greeting, which can take the form of a verbal item, e.g. to say someone's name, and a nonverbal item, e.g. a wave. Those two items taken together mark the activity as a greeting. Leaving out the gesture, however, might lead to a misinterpretation in the sense that the lexical item alone is now perceived as a summons.

Other types of discourse, such as narratives, fairytales and stories also have certain internal structures. Storytellers have different formula available to begin and end their narrative. When reading the formulaic preface "Once upon a time", for instance, the reader immediately knows that what follows is a fairytale. According to Labov (1972), who studied personal

narratives of near-death experiences, there are six elements that constitute a narrative: the *abstract* and the *orientation*, the *complicating action*, the *evaluation* followed by a *result* or *resolution* and finally the *coda*. The abstract and the orientation of a narrative usually answer the questions "Who?, What?, When?, Where?", so the narrator summarizes the content of the narrative and this fulfills a referential function. The coda is the ending or closing of a narrative in which the teller signals that the story is finished.

What telephone conversations and narratives have in common is their conventionalization and, in Clark's terms, a certain level *scriptedness*. As demonstrated, there are several structures and formula available, which both sender and receiver know and follow routinely. This is in contrast to the task at hand. There are some underlying conventions and the participants can borrow practices from other speech events. For example, they know how to read a map, how to follow instructions, and how to behave in a university setting. However, there is no clear script of how to open the interaction or to perform and solve the present task. It is also assumed that most of the students in this study do not participate in experiments on a regular basis. Opening the interaction in this context will be understood in a broader sense than Schegloff's definition of an opening due to the multimodal perspective of the study. The hypothesis is that participants have two options to begin the activity: 1) by a nonverbal cue, or 2) by a combination of a verbalization and a nonverbal cue. It is also argued that interaction openings are multifaceted and multimodal activities, which means that there is never a purely verbal opening.

4.2 OPENING THE INTERACTION NONVERBALLY

There are some challenges that participants face when they open the activity: first of all, they must negotiate their individual roles and second of all, they must find ways of solving and achieving the task together. Therefore, they must coordinate their individual actions and monitor the other interactant's actions at the same time. One of the findings of the analysis of the video-recordings is that there is no single solution nor is there a dominant pattern. Some dyads begin by silently reading through the whole instruction sheet, whereas others only read the first few points and then begin. Besides all this variation, it is argued that one essential external representation of the activity coordination and the negotiation of actions is the usage of the map and the task sheet. In this respect, they also become meaningful as they build action, i.e., "the accomplishment of particular concrete actions requires that these structures be deployed in conjunction with other

relevant meaning-making practices […]" (Goodwin 2000:1516). Interactants embed the map and the task sheet into the ongoing process, which means that the relevant actions necessary to accomplish the task could not be fulfilled without them.

With regard to the distribution of the map and the task sheet, there are two major patterns that occur: out of the twenty-one interactions, there are fifteen beginnings in which the two objects remain in the middle of the table. Hence, I propose that in these cases the verbal mode needs to be taken into account to understand how participants open the activity. In the six remaining interactions, roles are distributed in the nonverbal mode via the mediums of map and task sheet. This has obvious consequences for the ratification of participants' activity roles and the organization of the task because the map and the task sheet are not equally accessible to both interactants. The pattern is as follows: one person takes the map or the task sheet to start the interaction. This participant takes the activity in his/her own hands, signaling that they are in charge of either planning the routes or of reading the instructions. Parallel to phone conversations, these sequences can be understood as a summons-answer sequence, but occurring in the nonverbal mode. The summons here is the distribution of the map/task sheet, which means that it happens nonverbally through the visual channel. The other person then has to reciprocate. If one person, for example, takes the task sheet and starts reading the instructions out loud, the other person will take the map to orient toward it, or vice versa. These two screenshots demonstrate the distribution of the map and the task sheet: equally accessible versus distributed:

Figure 6 Screenshot of equal access to map and task sheet

Figure 7 Screenshot of distributed access to map and task sheet

Map and task sheet are thus *significant objects* (my translation; Schmitt and Deppermann 2007:111), similar to cameras and monitors used by film crews. According to Schmitt and Deppermann (2007), significant objects are those items that represent a spatio-temporal structure in an interaction; they gain importance through the realization of certain central activities and roles within in the activity. For instance, a video camera on a film set is exclusively operated and manipulated by a cameraman/camerawoman. The video camera marks the territory of the camera operator and is a predefined space with predefined roles and responsibilities. The monitor, in contrast to the camera, is not an object that belongs to or is operated by a particular person on the set and thus the space and the roles are not predefined. In summary, certain activities, spaces, and roles are defined and known to everyone on the set, whereas others need to be negotiated, as it is the case with a monitor. Map and task sheet are structurally similar to a monitor because their ownership must be negotiated and consequently reflects upon the activity roles. The map and the task sheet are external representations of the state of the activity as well as significant objects in Schmitt and Deppermann's sense because their distribution and co-usage with other practices is relevant to the organization of the activity.

Activity roles

There are three activity roles which can be established: first, there is the *Route Planner (RP)*, the participant who gives directions and plans the way through the zoo; second, there is the *Instruction Reader (IR)*, the person who reads the instructions out loud; third, there is the *Note Taker (NT)*, a role that can designate two activities: the person either writes down the planned routes and/or marks visited places on the map.

All of these activity roles are generally open to both participants and can also shift throughout the activity planning. That means that one person can start out planning, but then might ask the other person for help at some point, thereby inviting the other person to plan the route. The role of NT is not found in every interaction. Predominantly the interactants do not take notes. The realization of two (or even three) activity roles raises the question of how interactants organize their visual attention, for example how they create joint visual attention versus individual attention. Participants must monitor the other person's actions as well as the external representations while, at the same time, perform their individual actions relevant to the task.

4.3 MONITORING AND JOINT ATTENTION

Monitoring, as Schmitt and Deppermann (2007:120) explain, is a prerequisite for a smooth cooperation between interactants because it allows for a nonverbal organization of tasks. Clark and Krych (2004) conducted a study on how speakers (here directors) monitor their addressees (here builders) while they instruct them on how to assemble Lego models. They review two perspectives on speaking and listening in dialogue, a unilateral account and a bilateral account, and argue that speaking must be viewed as the latter of the two accounts, meaning that speaking and listening together constitute joint activities. Clark and Krych analyze four interactive conditions, which varied in terms of visibility: the builder's workspace was either visible or hidden and the partner's face was either visible or not. Clark and Krych (2004) found that the participants in their study were faster in fulfilling the task when the director could see the builder's workspace. Preventing monitoring also led to more errors in building the models. Hence, speakers monitor their conversation partners and the environment around them. People pay attention to voices, faces, workspaces, bodies, and shared scenes (cf. Clark and Krych 2004:64) when they interact with each other and, for certain conditions, such as the one presented in Clark and Krych's (2004) study and also in the present study, visual and auditory modalities secure a fast and smooth communication.

Joint action relies on joint attention, so the question is how participant create and engage in joint attention, i.e., how they attend to the same object or the same event and how the different interactive frames are reflected in the things participants attend to. Joint attention has been studied in the neurocognitive sciences and in child development, for instance. Moore and

Dunham (1995) study joint attention to objects and events in the world in infants as a means to build relationship and shared social meaning with others. Bakeman and Adamson (1984) study sequences of engagement in mother-infant interactions across different developmental stages. They present six different categories showing that "coordinated joint engagement" (Bakeman and Adamson 1984:1281) increases with age. Thus, the older the infant the more he or she will coordinate his/her actions and attention to the other person as well as the object the other person is attending to. Baron-Cohen (1995) proposes that there is a dyadic representation of how humans detect the eyes of another human, the so called *Eye Direction Detector (EDD).* In close relationship to EDD is the *Shared Attention Mechanism (SAM),* which functions as an identification of whether you and another human are both looking at the same object (Baron-Cohen 1995:49). Hence, this mechanism relies on a triadic representation and according to Baron-Cohen, there is an *embedded element* (1995:50) which differentiates EDD from SAM, enabling animals and humans alike to notice that they are attending to the same object the other animal/human is attending to. Emery (2000:588) draws a similar distinction between joint attention and shared attention. In joint attention, one individual follows another individual's gaze to attend to the same object, whereas shared attention, as Emery argues, is a more complex process in which both individuals have to be aware of each others' individual attention. Tomasello (1995:103) further states that adults have the ability to attend to things in their environment selectively and intentionally, which means that they can include or exclude other people.

From a multimodal perspective, Goodwin (2003), for instance, reports on a father-daughter interaction during which the father helps the daughter do her homework. The body position of both interactants as well as the homework sheet, similar to map and task sheet, inform us of how the participants build joint action, express stance and align with one another. The father makes repeated reference to the daughter's assignment sheet, not only verbally but also via pointing gestures and gaze. Thus he signals that he is attending to a certain place or object in his and the daughter's local environment. As Goodwin writes, "[…] participants frequently attend to multiple visual fields simultaneously, including both objects being worked with, and each other's bodies" (2007:56). In the three organizational patterns identified here, participants display different strategies of how they collaborate to achieve the task and how they shape their attention states, individual and joint.

1) Map and task sheet in the middle, roles equally distributed

The first example derives from recording MOV00N and represents a prototypical case of an egalitarian interaction, based on the fact that the map and the task sheet remain in the middle of the table. The interactants both take the roles of route planner and instruction reader, which is maintained as the interaction progresses. They are both reading silently and the transcript begins with a preface to the actual planning.

MOV00N, lines 1-10

	Name	*Intonation Unit*	*Gesture/gaze unit*
1.	Norman	°I'm terrible at maps°.	reading and gaze shift in between M and TS
2.	Neville	I'm not too bad,	
3.		((laugh))	
4.		but time will tell.	
5.	Norman	okay so I found the South Gate,	body shift
6.		start at the South Gate,	RIF point to South Gate
7.	Neville	yeah.	
8.	Norman	the Bear Grottos first,	
9.		(2.0)	
10.	Neville	mh there it is.	RIF point to Bear Grottos

In line 1, Norman says "I'm terrible at maps", thereby opening the conversation and informing his partner about his lack of expertise in reading maps. Neville responds by stating that he is not too bad at reading maps. I consider the verbalizations in lines 1 to 4 as a preface to the actual beginning of the route planning. This process happens parallel to the familiarization process with the map and the task sheet. Norman and Neville read the instructions silently and look at the map while speaking to each other. The actual preparation for the task that lies ahead of them takes place in the nonverbal mode. In line 5, Norman start the task by saying "okay so I found the South Gate" and turns his body toward the map. Lines 6 to 10 demonstrate how Norman and Neville co-construct the route planning, not only shifting in between activity roles, but also resorting to auditory and visual cues to organize the task. For example, they point at entities (lines 6 and 10) to establish a location on the map as a shared focus of attention and understanding for themselves and for one another.

2) *Map and task sheet in the middle, roles distributed*

This excerpt derives from MOV00M and represents an egalitarian interaction with distributed roles. Madeline and Marina keep the map and the task sheet in the middle of the table. The interaction starts in the nonverbal mode with both participants taking roughly five seconds to acquaint themselves with the map and the task sheet; their eye gaze is oriented toward the two object in the interaction space and it shifts between the two. The conversation then starts with Madeline taking the turn by reading out the instructions.

MOV00M, lines 1-22

	Name	*Intonation Unit*	*Gesture/gaze unit*
1.	Madeline	start at the North Gate .. [at 9am.]	
2.	Marina		RIF point to North Gate
3.		[so that's] there?	
4.	Madeline	so you go to visit the camels,	LIF point to North Gate
5.		where are the camels?	
6.		... camels.	LIF points
7.		and the Habita- the Habitat Africa,	RIF point to camels
8.		()	
9.	Marina	I think we're back (round) to Australia House,	
10.	Madeline	so do go over this?	
11.		let's go there,	RIF circle
12.		round this way,	
13.	Marina	((clears throat))	
14.		yeah,	
15.	Madeline	[and we can go over there.]	RH tracing
16.	Marina	[so we come throu:gh] the North Gate.	
17.	Madeline	yeah.	
18.	Marina	turn right?	RIF point
19.		(4.0)	
20.		[yeah,]	

21.	Madeline	[yeah,]	
22.	Marina	and visit the camels,	LIF point to camels

In lines 1 to 3, it becomes obvious that Madeline and Marina took a few moments at the very start to familiarize themselves with the map. This is reflected in Marina's prompt identification of the North Gate in line 2. While Madeline reads the instructions and thus visually focuses on the task sheet, Marina's gaze shifts between the task sheet and the map. She then directs Madeline's attention to the North Gate by pointing to it and Marina points to the North Gate as well in line 4 while she continues reading out loud. Both participants first focused on individual items, but were aware of the other person's attention as well. By the time both interactants have identified the North Gate, they have established joint attention to the North Gate, which was initiated by Marina's right index finger point, and they maintain this shared focus. The interactants' actions are tightly linked and there is verbal overlap, for example in lines 16 and 20-21. Starting in line 16, Marina repeats the route description from the North Gate to the first stop, the camels. This pattern frequently occurs at the beginning of the interactions and seems not only to be a strategy of acquainting oneself with the map, but also of creating a mutual understanding and an agreeing stance on the planning.

With regard to activity roles, the transcript demonstrates that Marina and Madeline are route planners, but only Marina is the instruction reader. This does not mean that Madeline has no access to the instructions. In fact, she could also read them out loud, but she does not. Clearly, right at the start, by taking the turn, Marina has decided that she will continue reading and Madeline does not challenge this.

3) *Map and task sheet distributed, roles distributed*

MOVOOK is an example of a non-egalitarian interaction in which one person, Karin, is the more forceful participant. She immediately turns the instruction sheet in her direction and leaves both hands on it while reading. This move signals ownership of the task sheet. Kara, the other participant, has to lean in to be able to read the instructions. When Karin has finished reading, she says "okay" and then turns her gaze toward the map. This is the first time Kara has full access to the instructions.

MOV00K, lines 1-13

	Name	*Intonation Unit*	*Gesture/gaze unit*
1.	Karin	okay.	places RH on M
2.	Kara	okay [( )]	turns TS, BH on TS
3.	Karin	[okay .. so] we start there?	RIF point to North Gate, LH on TS
4.	Kara	do we need to write (that down or anything)?	
5.			takes pen
6.		yeah (I think so),	
7.		do you want (a go with this)?	
8.	Karin	uhm .. go ahead,	
9.	Kara	okay .. what should I do,	
10.		should I write it down?	
11.	Karin	yes I think so yeah,	LH on TS
12.	Kara		turns TS in her direction
13.	Karin	you should write at the (bottom).	point to bottom of TS

This example demonstrates that Kara is actively ratifying her role, looking for a way to be included as a co-participant in the task rather than just being an observer. After her question in line 4, she does, in fact, not wait for a reply but picks up the pen right away. However, there is a reiteration in line 10, this time asking for confirmation. Kara turns the task sheet in her direction to take notes, but as line 13 shows, Karin is still signaling that she is charge by telling Kara where to write "at the bottom" (line 13) accompanied by a nonverbal specification. In terms of joint attention, the distribution of activity roles influences the individuals' attention as well. Karin has to monitor Kara's actions very closely and in fact, one can find longer pauses and more repetition throughout the interaction because Karin has to wait for Kara to finish taking notes, i.e., to accomplish the task of note taking. Similar to a situation Clark and Krych (2004:78) describe, in which one of the directors waits for the builder to reach and show him the Lego block he has mentioned previously, Karin must wait for the right moments to proceed with the planning to ensure that Kara can follow and understand her moves.

In summary, these three examples have demonstrated how the structures of map/task sheet and activity roles are connected and how they influence the individuals' attention states,

especially in the last case when one participant takes notes and the second participant must monitor this action as well. The exemplary analysis of three excerpts provided these results for the distribution of the map, the task sheet, and the resulting activity roles:

1) If the map and the task sheet are in the middle, then both interactants have equal access to them; the participants have the option of taking the role(s) of instruction reader (IR) and route planner (RP).
2) If the map and the task sheet are in the middle, the interactants have equal access to them; however, the roles are distributed, which means that one person fulfills two functions (IR and RP) where the other person only does RP.
3) If the map and the task sheet are distributed, roles must be distributed and hence there is an IR and a RP.

For the first and the second pattern, it was suggested earlier that the verbal mode is indicative of how the activities are opened. These fifteen interactions will now be revisited focusing on the verbalizations at the beginning and bringing them into a relationship with other modes.

4.4 OPENING THE INTERACTION VERBALLY

The three examples presented earlier demonstrate that speech and gesture, vocal and gestural actions, cannot be separated from one another in a face-to-face environment. Interactants monitor their own and the other person's actions (they gaze at each other) and they produce gestures for themselves (locating something on the map) and the other person (to ground their actions). Thus in focusing on the verbal mode to outline differences and similarities between the interactions, it is important to stress that talk alone does not suffice and that multiple semiotic resources must be considered. MOV00N, for example, represents a prefaced beginning, whereas in MOV00M the interactants begin by reading the instructions. Reading the instructions varies along a scale: from participants reading out the complete instructions on the one end to participants who read the first item and then start to plan the route step by step on the other end. The first example shows an "in-between stage" during which there is an interruption in the reading process.

MOV00L, lines 6-15

	Name	*Intonation Unit*	*Gesture/gaze unit*
6.	Larissa	follow the instructions below,	
7.		to plan your day at the zoo.	
8.		start at the South Gate,	
9.		where is that?	gaze shift to M
10.	Laura	(°right here°)	RH point to South Gate
11.	Larissa	South Gate.	gaze shift to TS
12.	Laura		gaze shift to TS
13.		[at 9 a.m.].	
14.	Larissa	[at 9 a.m.].	
15.		you you go to visit the Bear Grottos,	

In line 9, Larissa is posing a question by which she interrupts the reading process. This is accompanied by a shift in focus, from the instructions to the map. Laura, who looked at the map while Larissa was reading, places her right hand on the map to identify the location of the South Gate. Larissa repeats “South Gate” in line 11 as an affirmation. Both girls then redirect their gaze to the instructions and thereby establish a common focus. This is reflected in the verbal overlap in lines 13 and 14. Larissa then resumes reading.

The following example is a step-by-step planning of the activity Ben and Beth look at the instruction sheet while Ben is reading.

MOV00B, lines 1-11

	Name	*Intonation Unit*	*Gesture/gaze unit*
1.	Ben	start at the North Gate at 9 a.m.	
2.			gaze shift to M
3.	Beth		gaze shift to M
4.		°()°	
5.	Ben	°there's the North Gate°.	RH point to North Gate
6.			gaze shift to TS
7.	Beth		gaze shift to TS

8.	Ben	you go . to visit . the [Camels].	gaze shift to M
9.	Beth	[the Camels] right.	gaze shift to M
10.		the Camels are here.	RH point to camels
11.	Ben	yeah.	

Here the interactants start at the North Gate and plan the route to the first destination. Ben takes on the role of the IR as well as the RP, but Beth is equally involved in the planning process. Their gaze patterns align: their gaze shifts back and forth from the map and the task sheet and they take turns in the verbal mode as well, sometimes overlapping, here in lines 8 and 9.

The excerpt from MOV00T represents the third type of an interaction opening, which is similar to MOV00N in so far as there is a preface to the actual start of the planning process. As both examples demonstrate, regardless of whether the interaction is opened with a question, as it is the case here, or a statement, as it is the case in MOV00N, there are no verbalizations of how activity roles are distributed in fourteen of the fifteen interactions. This is opposed to the earlier assumption that interactants could start out by asking "Do you want to read or should I?", for instance. It again emphasizes the close conjunction between verbal and visual resources.

MOV00T, lines 1-15

	Name	*Intonation unit*	*Gesture/gaze unit*
1.	Tamara	have you got a good sense of direction.	gaze at Tom
2.	Tom	(woo:),	
3.		"I'm CHAP,	gaze at Tamara ←
4.		course I have".	
5.		[((laughs))].	
6.		[I'm a () ((laughs))].	gaze shift to TS
7.	Tamara	[((laughs))].	
8.		oh good,	
9.		'cause I'm a woman,	gaze shift to TS
10.		[I don't have it ((laughs))].	
11.	Tom	yeah: no it's not perfect,	
12.		I'll tell ya.	
13.	Tamara	we need to start now.	

14. Tom yeah:.
15. Tamara ((chuckles)) start at the South Gate at 9 a.m.

Goodwin (2007) reports on a problematic situation in the father and daughter homework interaction. They are sitting in bed next to each other, so in a side-by-side constellation. There is one instance in which the father summons the daughter, namely he calls her name to align her to the activity (Goodwin 2007:63). At the beginning of the conversation here, Tamara directs a question to Tom, who is looking at the map already. One could assume that she would address him by name. However, probably due to the face-to-face orientation and not a side-to-side position (as it is the case with father and daughter) the direction of Tamara's gaze directed at Tom suffices to first elicit a verbal response (line 2) and then a state of mutual gaze (line 3). Tom shifts the frame to a playful one referring to clear gender stereotypes when he says "I'm a chap, course I have". The frame-shift is also marked by a shift in Tom's voice, sounding darker than his usual register to underline the fact that "he is a real man". There is mutual laughter and we can assume that Tamara's statement 'I'm a woman, I don't have it" relates to the frame Tom has established. He then provides a more truthful account of his sense of direction in lines 11 and 12 and even though he stays with the current topic, the frame shifts from a non-serious to a serious one. The actual planning process is prompted by Tamara in line 13 when requests that they start now. She begins to read the instructions in line 15 and takes the IR role. The preface from lines 1-12 is related to the task and the content of the activity. It also enables Tamara to express her lack of a general sense of orientation and, at the same time, to find out about her partner's sense of direction. One possible interpretation of this exchange is that Tamara implicitly expresses that she does not want to be the RP, but would rather like to read the instructions. This assumption is reinforced by the fact that she takes on the role of the IR in the next line. However, taking the remaining transcript of the interaction into account, it becomes obvious that this assumption does not hold because Tamara and Tom take turns reading and planning. The preface could thus be reinterpreted as a lead-in into the activity during which the interactants inform each other of their competences important to the fulfillment of the task. Since both of them do not claim to have an excellent sense of direction, they then take a cooperative stance and co-construct the routes through the zoo.

The last example is an exception to the examples discussed so far. It is the only interaction begun by an explicit question of who is going to read the instructions in lines 3 and 4.

MOV00W, lines 1-10

	Name	*Intonation Unit*	*Gesture/gaze unit*
1.	Wendy	should we read the ..	holds onto TS
2.	Wilma	yeah.	
3.	Wendy	do you want me to do it?	
4.		or (do you wanna).	gaze at Wilma
5.	Wilma	YOU.	gaze at Wendy, RH point at Wendy
6.	Wendy	start at the North Gate at 9 a.m.	
7.			gaze shift to M
8.	Wilma		gaze shift to M
9.	Wendy	[so that's-],	
10.	Wilma	[yeah that's there].	RIF point to North Gate

Wendy looks at Wilma when she starts uttering her question in line 1. The question is, in fact, not completed, but still successfully answered as we can see in line 2. Relevant to this incomplete question is the fact that Wendy places her hand on the task sheet and so Wilma does not need to hear "instructions" to know what Wendy means. Due to the close proximity of the participants and the shared workspace, Wilma can monitor Wendy's hands and perceive her gesture. It is also important to note that Wendy first uses the plural pronoun "we" and then in l. 3 and 4 changes it to refer to the participants individually. She offers two options to Wilma and looks at her when she asks whether Wilma wants to read the instructions. Wilma's answer is heavily marked, both by a stress of "YOU" spoken louder in the verbal modality, her gaze and a pointing gesture directed at Wendy. Similar to other conversations presented, Wilma and Wendy engage in a joint action, they coordinate their individual actions, their gaze patterns align (for example in lines 7 and 8) and the mutual orientation is represented in verbal overlap.

4.5 CONCLUSION

Opening an activity has been shown to be a complex process, which is not scripted and hence offers variation in the realization of the interaction openings. The analysis was put forth in three steps: based on the (nonverbal) distribution of the map and the task sheet, the activity roles were demonstrated to differ. The identified roles include the *instruction reader* and the *route planner*,

with a possible extension, the role of *note taker* (one person either takes notes or marks the routes on the map). Interactions were divided into egalitarian and non-egalitarian with the first option being the more common form of interaction. When both participants verbally contribute to the interaction, they have equal standing, opposed to dyads where one is the predominant planner and the other one the note taker.

Participants pursue a common goal, which means that they want to complete the task that they are given, not only in a timely fashion but also by maintaining the interactive frame. However, as the activity roles and the division of labor clarify, participants perform different things to fulfill the requirements of the task in order to reach the end goal. Some interactants needed more time to start the activity proper and this was reflected in the prefaced interactions presented here. Opening the activity can thus not be limited to a two-pair sequence, as it was proposed for telephone conversations (Schegloff 1968). Rather, activity openings display complex and multimodal structures, which can vary depending on the form of organization. Even though there is neither a script nor a formulaic opening, which also accounts for the variation in the openings, there are nevertheless some similarities across the pairs, such as the negotiation of the map and the task sheet as well as the roles that people attribute to themselves. Shifting the focus to the verbalizations that occur at the beginning in the second half of the analysis demonstrated that the dyads engage in multimodal demonstrations in the planning process, displaying a conjunction of different semiotic fields, including body orientation, gaze, and pointing. All these elements contribute to the framework that they establish and maintain. Depending on the distribution of map and task sheet as well as the activity roles, individual and shared orientation as well as monitoring processes, in this case self- and other-monitoring, were shown to differ. This has consequences for the organization and maintenance of a shared focus of attention. In the following chapters, the relationship between the modes of speech, gesture, and gaze will be further investigated and special attention will be paid to reference, for example via a deictic expression or a pointing gesture, as well as gaze patterns, for instance a gaze shift from the shared workspace to the interaction partner.

In sum, this chapter has addressed a traditional topic of Conversation Analysis, namely the opening sequence of conversation and, in this particular setting, of interaction. Through the investigation of both egalitarian and non-egalitarian organizational forms, it has demonstrated that beginnings of interactions consist of multimodal resources, rather than sequential, two-pair structures.

5 VERBAL AND GESTURAL REFERENCE

One way to monitor the progress and state of the interaction is to look at the distribution of the external representations. They provide information about the activity roles, and the attention states of the participants as well as changes within the interaction. Another important indicator is cohesion across the task performance, in particular the relationship between verbal reference, gesture, and gaze. Participants refer to things in their surroundings with words, with their hands, and with their gaze. Verbal reference can be established to a) the map/the task sheet, b) the activity and c) to the (previous) discourse. Pointing is directed to either one or both of the external representations in the interaction space; gaze can be directed to the map and to the task sheet as well as to the other interactant. Based on McNeill et al.'s (2010) study, co-referential chains will be of special interest in this chapter. Referring expressions, deictic gestures, and gaze are used to coordinate individual actions and they allow participants to position themselves in relation to their own previous actions as well as the other interlocutor's actions. Cohesion is therefore not only visible on a textual level (Halliday and Hasan 1976), but also on an interactional level through the references to self- and other- contributions as well as embodied gestures displayed by the dyad.

McNeill et al. (2010) analyze multiparty war gaming sessions, focusing on how gesture and gaze create F-formations (Kendon 1990) and how co-referential chains can build the basis for hyperphrases. McNeill et al. define a *hyperphrase* as "a nexus of converging, interweaving processes that cannot be completely untangled" (2010:144). Within a hyperphrase nexus, multimodal information can be transported. Verbal and non-verbal features can occur within one hyperphrase and they stretch out as *single production pulse* (McNeill et al. 2010:145). For instance, within one topical unit, repairs and hesitations can occur alongside a single, repeated gesture as well as gaze directed at the listener. When two or more interlocutors engage in a joint F-formation, a space to which they have established shared aceess, to talk about a target, the idea of collaboration can be transported in verbal co-constructions as well as embodied gestures. The mechanisms of dialogue, as McNeill et al. argue, need to be viewed from a multimodal perspective and need to be understood as "thought in context" (2010:164). Growth points—theoretical units reflecting upon the speaker's cognitive state as well as the language-imagery dialectic (McNeill et al. 2010:144)—can be co-inhabited, meaning that two or more participants

align and form coalitions in dialogic floor management. With regard to reference, McNeill and his colleagues differentiate three levels across which references thread: 1) the *object level*, 2) the *meta level* and 3) the *para level*. On the object level, cohesion is created through references to the object world. Reference in this sense does not only mean textual reference, as it is defined by Halliday and Hasan (1976), but it includes references to external representations as well. Meta level reference creates cohesion through references to the discourse itself. For example, when one of the participants says "cause that's what the list suggests", she does not only draw a reference to an object, "the list", but in the larger context, she also refers to the content of the activity and the prior discourse. Para level reference includes individual participants; for instance, "this is the busiest zoo I've ever seen" relates to an individual participant and includes personal experience and background knowledge. When one says "I agree with the assumption" (McNeill et al. 2010:155), the personal pronoun "I" functions as a cohesive tie and references an individual person. In sum, these three levels account for co-referential chains in the structure of discourse, providing information about thematic relationships and content (McNeill et al. 2010:156). McNeill et al. further state that co-referential chains "can span different speakers and so can tie together multiparty hyperphrases and shared growth points in dialogues" (2010:155).

In an earlier study, McNeill and Levy (1993) investigate related notions. The researchers establish a relationship between speech and gesture, analyzing cohesion under a speech-gesture synchrony in narrative discourse. The hand with which the gesture is produced, the shape of the hand, and the space where the hand gesture is produced, are all indicative of topical cohesion. Their findings demonstrate that a single gesture can mark the existence of a referent or a character and that the same gesture can later help to maintain reference to this particular referent/ character. The gesture functions as a *place marker* to distinguish discourse themes as well as characters and voices within a narrative (cf. McNeill and Levy 1993:382).

Speakers contribute to a conversation on different levels using multimodal information in order to create F-formations and to signal membership or non-membership in coalitions. Within coalitions a "package of two such [participation] frameworks" is represented, one being the actual task activity in the present context, the other being "the social group joining in –the meta and para levels, respectively" (Mc Neill et al. 2010:159). Dyadic conversations naturally vary from multiparty discourse as two people cannot form a coalition against an opponent. However, individuals can align and disalign with one another to take a stance, as Scheibman (2007)

explains, for instance. She states that speakers "express stance by allying themselves with (or sometimes separating themselves from) one another" (2007:113). The social group mentioned in McNeill et al.'s definition relates to sociocultural and general beliefs, while the individual and interactive features of stance taking contribute to the task frame (Scheibmann 2007:113). Participants in this study position themselves in relation to the activity, in relation to the other participant, and also in relation to their own prior contributions. Thus, the three levels of reference proposed by McNeill and his fellow researchers can be applied to investigate how interlocutors verbalize and visualize these relations, expressing acceptance or rejection of previously proposed or following moves. Keisanen further explains that "[…] any linguistic or paralinguistic feature of language, or a linguistic construction for that matter, can function as a stance marker" (2007:254). In managing the task through individual and collaborative contributions, stance is expressed in verbalizations, (pointing) gestures as well as withdrawal of gaze or establishment of mutual gaze. The multimodal approach proposed by McNeill et al. (2010) allows for an in-depth analysis of different discourse types and is adapted here to investigate the phenomenon of task management in dyadic interactions as it emerges in different modes.

5.1 REFERENCE IN THE TASK BASED SETTING

In the introduction, the relevance of cohesion and reference in the task based context has been outlined, arguing for a speech-gesture synchrony to fully comprehend the meaning of reference. A longer example will now be adduced. This example derives from an interaction between Ben and Beth. They are relatively close to the beginning of the planning and are mapping out the route to the Australia House. The excerpt was selected because it demonstrates the importance of an assumption of a speech-gesture synthesis with regard to cohesion. There are only two clear cases of referential cohesion by a demonstrative pronoun based on textual cohesion. To understand the other forms of references, visual aids and gestural movements are mandatory.

MOV00B, lines 77-114

	Name	*Intonation Unit*	*Gesture/gaze unit*
77.	Beth	well we go straight forward from the Camels,	RH tracing on M
78.			turns M

79.		and then we turn (0.8) <this diagonal left path.>	RH tracing on M
80.	Ben	yeah.	
81.	Beth	and (1.0),	
82.	Ben	into an [area],	
83.	Beth	[into,]	
84.	Ben	which is accessible to both Habitat Africas.	
85.	Beth	oh we turn right,	nods
86.	Ben	yeah.	
87.	Beth	and [then we could,]	Beth: RH tracing on M
88.	Ben	[and then follow] the path.	
89.	Beth	po- follow the path through the Habitat Africa.	turns M in her direction, gaze at M
90.		yeah?	gaze at Ben
91.	Ben	yeah.	
92.	Beth	so where are we then.	LIF point to Habitat Africa
93.		are we (just over here),	
94.	Ben	we just come out of this-,	RH point to Habitat Africa
95.	Beth	here okay	RH inscription with pen
96.	Ben	(yeah [the end again]).	
97.	Beth	[°here .. I have it here°].	
98.		the Australia House.	
99.		where is that?	
100.	Ben	uhm,	
101.	Beth	(THERE it is).	LH point to Australia House
102.	Ben	yeah.	
103.	Beth	should we just retrace our steps?	
104.	Ben	yeah I think we should.	
105.	Beth	so we come out of this,	LH point to Habitat Africa
106.	Ben	(oh [we uh:),]	
107.	Beth	[thing.]	gaze at Ben

108.	Ben	yeah.	
109.	Beth	should we?	gaze at Ben
110.	Ben	is that the quickest way.	
111.	Beth		turns M in Ben's direction
112.	Ben	(4.0)	
113.		oh we're here.	RH tracing on M
114.		yeah I think we should retrace our steps.	RH circle above M

In line 98 Beth asks "Where is that", referring back to the "Australia House" mentioned in the line before. One does not need a gesture to understand that the demonstrative "that" is referring to the Australia House. In fact, Halliday and Hasan explain that the demonstratives *this, these, that, those* and *the* "have the experiental function of Deictic" (1976: 58), which means they point verbally. In line 109 Ben says "is that the quickest way" without a deictic gesture, just as in the example before. The structure here, however, is more complex since "that" is not simply referring to an entity, such as the Australia House, but to a prior event. Ben's statement indirectly refers to Beth's utterance "should we retrace our steps?" in line 102 and as such it is connected to the whole of the preceding route planning. The example further demonstrates what Halliday and Hasan mean by proximity: they say that there is a tendency to associate *that* with a past-time referent (1976:60). In this particular case, one could say that Ben refers to a prior discourse unit that has been completed and this could explain his choice of *that* instead of *this*. Furthermore, an utterance such as the one in line 100 including a 'locational pro-term' "there" is an indexical expression whose meaning largely depends on its context. The referent of such an expression can vary with its content and one can only recognize the referent via a gestural complementation. The pointing gestures facilitate a clear identification and disambiguation. The next utterance is one where the reference does not lie in the pronoun alone, but is also accompanied by a gesture that is clearly needed to identify the referent. In line 79 Beth says "this diagonal left path", tracing the path on the map at the same time. Since there are various paths on the map, Beth verbally specifies the path as being a diagonal one, leading to the left. However, one needs the clarification through the gesture to understand which particular path is meant because there are several diagonal paths. By tracing the route, she enables Ben to follow her planning and his verbal agreement in the following turn (line 80) illustrates that her gesture was perceived

successfully. There is another similar turn exchange in lines 87 and 88. However, this time it is co-constructed by both Beth and Ben. There is overlap in lines 86 and 87 already and while Ben is saying "follow the path", Beth is indicating the path on the map. Again, "the" alone does not suffice as a clarification of the path. Only through the addition of a gesture, the utterance "the path through" (line 88) becomes meaningful. There are two Habitat Africas, the Savannah and the Forest, and there is a path in between these two parts, which Ben and Beth want to follow. In lines 94 and 104, there are two more occurrences of "this" accompanied by a pointing gesture to indicate a concrete entity on the map.

The excerpt also contains a passage with a meta level reference. In line 102, Beth refers to the prior planning by saying "should we just retrace our steps". In the following lines, Beth is beginning to plan while Ben is still processing Beth's suggestion. This becomes apparent in lines 105 and 109, the latter one having been discussed already. Beth turns the map in Ben's direction (line 110), shifting modes of expression. The modal shift is from verbal to non-verbal by turning the map and thereby replying to Ben's question, who, in turn, continues tracing the route on the map while saying "oh we're here" (line 112). This is followed by a repetition of his own wording "yeah I think we should" from line 103 and a recycling of Beth's question (line 102). The two instances of "I think" in the utterances in lines 103 and 113 account for the para level because Ben refers to himself expressing his opinion. The repetition of Beth's initial suggestion by Ben further provides information about the process of planning. However, language alone does not suffice to clarify this process; the "circle gesture" above the map specifies and visualizes the re-tracing of the steps. This example will be revisited in Chapter 6 to demonstrate how verbatim repetition is applied as an interactive device to produce common ground. In sum, the example demonstrated that cohesions can no longer be reduced to the textual level as Halliday and Hasan have proposed. Gesture and gaze patterns are essential in a consideration of reference as entailing multimodal information transported in dyadic task management.

Returning to the study by McNeill et al. (2010) on multiparty floor management, I will now explain co-referential chains in more detail. The example above demonstrated many cases related to the object level and contained features of meta and para level reference as well. I argue for a necessity to establish parameters to successfully account for the variety of references represented in the current data and to place them in relation to grounding and stance taking. All three levels are closely intertwined and in the following, examples where all three levels co-

occur within one intonation unit or across intonation units within one topical unit will be presented. The brief exchange between AJ and Anna demonstrates a co-occurrence of meta and para level features:

AJ: we're lost at the zoo already,

Anna: well that's to be expected.

AJ's utterance can be classified as a meta level comment, referring to the result of the previous planning of the activity. He evaluates the outcome of Anna's and his actions. He also references the two interactants with the plural pronoun "we". There are two possible interpretations to Anna's response: she either wants to express that due to the size of the zoo, it is no surprise that they lost their way, or alternatively, she could mean that they both have such a poor sense of direction that they do not know where they are. In the latter case, Anna's statement would have to be considered a para level reference because she is referring to personal experience and knowledge and she positions herself in relation to AJ's evaluation, providing a second assessment.

I will now move on to a more detailed description of the three levels, starting with the object and the para level. I will then investigate the meta level, the most extensive level, to establish the features that demonstrate the complexity of this particular level.

5.2 OBJECT LEVEL REFERENCE

References on the object level create cohesion through a reference to the object world; in the data presented, such references are either produced to identify an entity on the map or to indicate instructions on the task sheet. Here are some typical utterances referring to the items (in bold) on the map:

*Head down by **Aardvarks*** (MOV00B, line 116)

*Just go around **Australia House*** (MOV00O, line 187)

*You pass **South Gate Snacks*** (MOV00T, line 73)

All these examples refer by naming the entity as labeled on the map. However, there are references such as "turn right at the **toilets**" (MOV00B, line 36), where the word "toilet" is not explicitly mentioned on the map. There is a little icon with a male and a female person and in the key below the map, this icon is specified as "restrooms". The person uttering "the toilets" needs to establish a link between the icon, the key, and the word on the map before he/she can transfer

this knowledge into speech. Reference to the object level is frequently created via a deictic gesture and references to an object on the map are verbalized by (nominal) demonstratives such as “this” and “that” as well as their plural counterparts “these” and “those”, referring “to the location of some thing” (Halliday and Hasan 1976:58). There are also circumstantial demonstratives, “here” and “there”, indicating “the location of a process in space” (Halliday and Hasan 1976:57), as demonstrated in the following two examples.

MOV00O, lines 118-123

	Name	*Intonation Unit*	*Gesture/gaze unit*
118.	Olga	go see the Dolphin Show at Seven Seas.	
119.		Seven [Seas].	
120.	Olivia	[which] is he:re.	RIF point at Seven Seas
121.	Olga	okay.	
122.		and we're here.	LIF point to Australia House
123.	Olivia	right.	

MOV00O, lines 267-268

	Name	*Intonation unit*	*Gesture/gaze unit*
267.	Olga	then after lunch you want to see the Pachyderm House,	LIF on TS
268.		which is back there?	LIF point to Pachyderm House

In both excerpts, Olga reads the instructions (lines 118 and 267). In the first example, Olivia identifies the referent “Seven Seas” by saying “here” and by pointing to it with the right index finger. Olga (line 122) refers to their current location “here”, pointing to the Australia House. Between the arrival at the Australia House and the identification of Seven Seas, there is an insertion sequence where Olga reads the instructions. By pointing to the Australia House, she establishes a cohesive link between the arrival at the Australia House and the new entity, which they will visit next. Together, both particpants parse the map and create a starting and an end point. In the second excerpt, Olga identifies the location of the Pachyderm House herself—in contrast to the first example, where Olivia pointed to it—by moving her left index finger from the task sheet to the map, pointing to the item on the map saying “there”. “There” is further specified by “back”, which is one possible explanation for her choice of *there* rather than *here*,

since the location is farther away from their current location. Gestures frequently occur at points of topic shifts (Levy and McNeill 1992). Gestural handedness can be indicative of the beginning of a new topic, for example to distinguish new from old information. Gestural handedness describes the production of a gesture, which can be formed with one hand or two hands. There can be shifts from one hand to the other or from a single hand to both hands. In the example above, Olga consistently uses her left hand to point—in contrast to a shift in handedness to mark a new topic. Olga establishes and maintains reference to entities on the map/task sheet with one single hand and even the gesture form, a left index finger point, remains identical. Not only does one need the gestural and visual addition to understand the verbal references, but gesture form and handedness provide further information about topical cohesion and cohesion across the planning process.

Object level reference has been shown to take two basic forms: firstly, participants use the proper name of the referent or attribute a name to an icon on the map. In the previous case, a gesture is optional, whereas in the latter form, a gesture remains optional only if the icon is conventionalized, for example in the case of restrooms. Secondly, object level reference can occur in the form of a demonstrative pronoun and in this case, a gesture is necessary to disambiguate. References to objects on the map/task sheet co-occur with the other two levels in the management of both the interaction and the activity.

5.3 PARA LEVEL REFERENCE

On this level, cohesion is achieved by a reference that includes individual participants. To use an example from McNeill et al.'s study, "I agree with the assumption" (2010:156). The speaker not only refers to himself, but also relates to another person's action: the second assessment relates to a previous action and it signals agreement with this action. The social relations are sustained because the first speaker's assumptions are reinforced by the second speaker's agreeing stance. Kärkkäinen (2007) investigates the usage of *I guess*, in particular in American English, as an epistemic stance marker. She classifies *I guess* as "a subjective marker *par excellence"* and she continues by stating that speakers often use such markers to produce "an action such as an assessment, an opinion, or a (strong) claim, that inherently involves taking a stance or a position" (2007:184). In the recording MOV00D, one of the participants says "I guess we need some paper" (line 153). There is personal reference through the usage of "I" and "we". The utterance is

framed with "I guess", by which the speaker expresses his personal opinion and indexes a stance taking activity. As such, the utterance is classified as a para level reference. *I guess* is not very frequent in the present data; however, turns can be introduced and completed with *I think* and *I suppose*, and *I suppose* might even be the British English equivalent to *I guess*. A new parameter, which can be added to McNeill et al.'s (2010) definition of para level reference, can thus be defined as an utterance which is framed by an epistemic stance marker in order to express stance. In maintaining or challenging social relations, para level references refer to prior actions, similar to meta level references. In fact, the two levels frequently co-occur and are closely tied to the task performance of self- and other-activities relevant to the task. Stating the assumption "I guess we need some paper" does not only contain the opinion of one participant, but it also refers to the planning process and provides a suggestion as to how to solve the task more sufficiently. The utterance combines meta and para level features because the speaker places himself in relation to the stance object (see also Haddington 2007) and he identifies himself as an individual participant who positions himself in relation to the second participant. Following this stance taking activity, the second interlocutor does, in fact, express a divergent stance, disagreeing with the first interlocutor's suggestion by saying "oh I don't know we do, they've got cameras" (MOV00D, lines 154, 155).

Stance markers such as *I guess*, *I think*, or *I believe* express a varying degree in how one positions himself/herself in relation to another. The usages and functions of *I think*, for example, have been investigated in recent years: Aijmer argues for a "scale of pragmaticalization" (1997:1) along which an emergent pragmatic construction such as *I think* developed into a discourse marker or a *modal particle* as she terms it. *I think* can express different aspects of knowledge and Aijmer (1997:21-22) differentiates between a *tentative* and a *deliberate* usage of the construction. A tentative *I think* hedges an utterance to soften an assertion, whereas a deliberative use of the modal particle adds reassurance to an assertion. Simon-Vandenbergen (2000:42) builds on these findings and further explains that there are multiple meanings of the expression, varying depending on intonation, syntax, and the surrounding discourse. As part of a class of "parenthetical verbs", including *believe*, *guess*, and *suppose*, the verb *think* can occur initially, in mid-position, and in final position in a clause (Simon-Vandenbergen 2000:42). In the following, there will be examples of *I think* in different positions varying in degree of assertiveness and reliability.

Participants use the plural pronoun “we” by which they do not index one individual any longer, but they use personal reference to speak about both interactants. With an utterance such as “we’re doing this really awkward way” (MOV00R, line 400), the speaker provides an evaluative comment on the task performance and, hence, meta features are displayed alongside the personal reference. Statements such as “basically we’re just lost now” (MOV00L, line 289) and “we’re stuck in a hole .. basically” (MOV00O, line 172) are framed by the marker “basically” to introduce or complete an evaluative comment related to the task performance, similar to the usage of modal particles. The usage of “we” rather than “I” relativizes the evaluation of the previous planning process because of an avoidance of attributing blame to a single participant. In multiparty discourse, there might be a higher necessity to disassociate oneself from other members, and presumably even more so in business meetings and military contexts. In the data investigated in this study, however, the success of the task lies in the assumption of cooperativeness of both participants. Thus, interactants invite each other to contribute to the planning activities or request input. Some participants, for instance, directly address the other person, marked in bold in the following two utterances: “do **you** think we’ve done enough detail?” (MOV00O) and “**you** choose which way should we go.” (MOV00C). In the first case, the speaker Olga asks the listener Olivia a question to elicit information and to inquire about her opinion. Olga’s question is triggered by the instructions stating “Describe in detail how to get there”.

MOV00O, lines 110-116

	Name	*Intonation unit*	*Gesture/gaze unit*
110.	Olga	do you think we've done enough detail?	gaze at Olivia
111.	Olivia	I think .. we might,	
112.			gaze at Olga, nods
113.	Olga	do you think?	
114.			gaze at TS
115.	Olivia	[I (think).]	
116.	Olga	[okay?]	

Olga not only addresses her interlocutor via the personal verbal reference, but she also raises her eyes from the task sheet to look at Olivia. By directing a question at Olivia, Olga invites her to

co-participate and to provide a subsequent opinion. Olivia's reply in line 111 is introduced by the phrase "I think", recycling Olga's verb choice in the prior utterance. The stance marker is repeated in line 115 to assert the previous statement. The nonverbal signals, the gaze shift and the slight nod in line 112, underline the stance taking process and have an assertive function as well. In the second utterance "you choose which way should we go." (MOV00C), one interactant, Clare, gives the floor to the other interactant, Cloe. Clare thereby signals that the other person's input is required to progress with the next move. Clare's invitation to "choose the way" is accompanied by a right hand gesture that "gives the floor" to Cloe. Cloe replies "we go an interesting way", which includes the evaluative, positive adjective "interesting" to show acceptance and to attribute a positive uptake to Clare's suggestion. Cloe also repositions herself, her upper body moves closer to the table and the map, again to demonstrate that she accepts Clare's invitation to take the floor and to continue planning. Both direct addresses presented above contain the plural personal pronoun "we", probably to signal that the invitation to co-participate is part of the "pair activity" and the idea of "being in this together" is transported. As a result so far, personal reference has been shown to be frequently combined with reference to prior discourse and activity. Certain parameters are interwoven, for example when a meta utterance is framed by a para level feature such as a personal stance marker.

In the following, para level and meta level references will be viewed as ranging along a scale displaying features of either one or both levels simultaneously. In the excerpt from MOV00Q, both types of reference occur in quick succession. Queena and Quintina must go from the Habitat Africa to the Australia House. They have planned the route to the Habitat Africa already and have identified the Australia House. Queena then realizes that there are two parts of the Habitat Africa. This is where the transcript begins.

MOV00Q, lines 42-49

	Name	*Intonation unit*	*Gesture/gaze unit*
42.	Queena	oh Habitat Africa is there as well.	RIF point to Habitat Africa
43.		so if we can follow it down this road.	RIF tracing
44.		all the way round.	RIF tracing
45.		and then come back to see this part of Habitat Africa.	RIF tracing
46.	Quintina	(3.5)	LIF point gaze at M, gaze at TS

47.		isn't it Australia?	gaze at Queena
48.	Queena	this bit I was talking about.	LIF point to TS
49.	Quintina	OH sorry [(I missed it)].	

Queena's utterance in line 42 is a meta level utterance. She points to an area on the map where the Habitat Africa is located and ends her route planning as she arrives at the other part of the Habitat Africa (line 45). The tracing movements from lines 43 to 45 are marked by brief stops in accordance with the utterance completion. After the end of the route description by Queena in line 45, Quintina takes the turn and points at the entity on the map where Queena started her description. Quintina's gaze is directed at the map and then shifts to the task sheet. These nonverbal movements are preparatory to the question in line 47. Quintina looks at Queena when she directs the question at her. Quintina expresses her opinion in the form of a question to soften the disagreeing stance. At the same time, the content of the question functions as a clarification to the divergence in focus, which became apparent in the nonverbal mode right before the utterance. Quintina places her left index finger at the starting point of Queena's description to comprehend and to establish a common focus. When the confusion remains, she verbalizes this problem. This is followed by a specification by Queena including a reference to herself and her prior discourse, taking a divergent stance to Quintina's question. The utterance in line 48 accounts for both a para and a meta level reference. As part of the clarification process, Queena points with her left index finger to direct Quintina's attention to the instructions on the task sheet. Queena draws a distinction in handedness and associates the right index finger with the map and the left index finger with the task sheet.

So far, the investigation of co-referential chains demonstrates that multimodal information is transported in the task management. For instance, when a participant disaligns with another participant's action, the disagreement is highly marked in the verbal and the nonverbal modes. Personal reference and reference to the prior discourse are co-expressed in one intonation and/or gesture unit. On the object level, speaking and gesturing occur while performing the task and performing activities relevant to the task. The para level is applicable when participants employ person deixis to position themselves and to ground their actions. However, another pattern occurs in dyadic discourse: one participant can choose to give up the

floor by inviting the other participant to contribute to the interaction. By addressing the other person, input and opinions are requested and this is another new parameter defined here for the para level.

5.4 META LEVEL REFERENCE

Meta level reference will be investigated in the following, differentiating references to the map and the task sheet from references to previous discourse and task planning units. Generally defined, meta level reference creates cohesion through reference to the prior discourse. In this context, the definition will be broadened to include references to the task sheet and the map itself. For instance, the question "do we draw on this" refers to both participants, "we", it refers to the map, "this", and it provides an activity, "drawing", which enables the participants to solve the task. Leaving a visible mark on the map allows for a recapitulation of the principal stages and steps through the zoo at a later point in the activity. Utterances as the one above will be classified as meta level utterances because they are not directly related to the object level. More precisely, there is a shift in frames from the actual activity to discourse about how to perform the activity, providing reference to the representations map and task sheet.

Secondly, there is reference to the task planning process, the actual performance of the activity. This is similar to McNeill et al.'s (2010) definition of meta level reference because most of these references refer to prior discourse contents and events. I provided one such example earlier when Beth was asking whether they should retrace their steps. Here, one needs to know what happened previously in the planning in order to understand her utterance and the reference. She is creating cohesion by linking a current event to a preceding one. Within the interaction between Beth and Ben, there was self- and other-repetition and I will return to this particular aspect of the interactions in the following chapter. In order to fully understand cohesion and to investigate how reference is maintained across longer passages of discourse, repetition and reiterations of discourse topics as well as gesture recurrence will be considered.

Thirdly, there is meta level reference combined with para level features. For instance, when uncertainties are addressed and when confusion by one or both participants becomes apparent, one or both participants shift from the object level to the meta level and, at the same time, they index themselves to express their opinion. This was demonstrated in the interaction between Queena and Quintina, for instance. Generally speaking, participants draw on the para

level to position themselves and to take a stance in the negotiation of the task procedures. Furthermore, references on the meta level are essential to the task management and in the following, different forms of meta references will be investigated.

5.4.1 Meta level reference to external representations

External representations and manifestations of the activity are represented in the map and the task sheet. The participants are instructed on how to perform the task and naturally, as part of the sense making process, references to the map and the task sheet are found repeatedly. These references can have various forms and I will now classify the meta level references to the external representations. There are some indirect speech acts in the form of questions functioning as a suggestion. Cloe, for instance, plans the route from the South Gate to the Bear Grottos. She indicates directions, for example "turn left", and names landmarks, such as "the Hamill Family play zoo".

MOV00C, lines 70-71

	Name	*Intonation unit*	*Gesture/gaze unit*
70.	Cloe	should we- should we make a scale,	gaze at M, point at M
71.		of this map?	

Cloe suggests scaling the map in order to be more precise about distances. While she speaks, her gaze is directed at the map and she uses the pen to point at it. The turn is highly marked because in all three modes, speech, gesture, and gaze, Cloe expresses the same notion, which is reference to the map. A structurally similar utterance is "should we maybe read through the whole thing first?" (MOV00D, line 24). Dan (on the left in the picture) looks at the task sheet and his question is accompanied by a gesture which is directed at the task sheet. He moves the pen down along the list and then back up to point to the first item (the Camels) on the task sheet. Dan's suggestion results in collaborative pointing. David also places the pen on the map, thereby showing that he is also attending to the instructions and supporting Dan's idea to read them. The screenshot below shows the collaborative pointing.

Figure 8 Screenshot of collaborative pointing

Both participants point to the area in the middle of the task sheet and their gaze is directed toward this area. Aspects of the instructions are indicated to each other via the tip of the pen, which means that they visually refer to the object level, but the verbal components refer to next higher level. Two of the three levels conflate: gesture and speech co-occur, but thread across the object and the meta level respectively. In terms of coalitions, it becomes apparent that David follows Dan's invitation to co-participate and he aligns with Dan. The gesture reflects upon the idea of being cooperative. David positions himself by following Dan's pointing movement, taking a positive stance, and as a result, there is mutual pointing. Both participants display high involvement in the task. In both examples, the participants' gaze patterns are directed to the map and the task sheet, even when asking a question. Opposed to research done by Goodwin (1980), for example, who established rules for the establishment of mutual gaze, the gaze behavior of the participants in these two contexts differs from regular conversation. The findings of this study in fact demonstrate that the gaze patterns in task based interactions are of a special nature and vary from everyday conversation.

Another kind of meta reference is found in the direct and indirect references which transport the idea of hetoronomy, for example in the following excerpt.

MOV00A, lines 132-140

	Name	*Intonation unit*	*Gesture/gaze unit*
132.	Anna	but we just already gone past there?	gaze shift M and TS
133.	AJ	(got to) go A:LL the way back.	raised eyebrows
134.		that's alright.	gaze at Anna
135.	Anna	yeah so really then?	gaze at AJ ←

136.		it would be more sensible to go this way?	tracing
137.		and that path round.	tracing
138.	AJ	yeah.	
139.	Anna	well we've got no choice,	LH palm raised
140.		cause that's what the list suggests.	gaze at TS, point at TS

Anna and AJ have reached a completion point in their planning and they look at the instructions to find out where to go to next. The excerpt begins with a complaint by Anna in line 132. The instructions state that the participants should pass the Great Bear Wilderness, which Anna and AJ have done already. Anna's gaze shifts between the map and the task sheet because she is reading the instructions and searching for the entities on the map. AJ's reaction in line 133 is marked by raised eyebrows, an evaluation in itself, which occurs before the actual verbal evaluation in line 134. The evaluative statement "that's alright" occurs in synchrony with a gaze shift from the map to Anna. Mutual gaze is established in line 135 by Anna. When her focus shifts to the activity again, she redirects her gaze to the map and outlines the path above the map. In lines 139 and 140, there is an evaluation and a direct reference to the list, which depicts the participants' awareness of being given instructions to follow. When Anna refers to the list in speech, she also directs her gaze and her pointing gesture at the task sheet, attributing importance to this unit. It also marks the end of the meta discourse, after which Anna and AJ return to the actual planning on the object level.

In the next example, there is a reference to the object level, which is made visible in the pointing and tracing movements (lines 24 and 25). Madeline's left index finger rests on the symbol for the camels (line 24) and her thumb traces the path "round there" in line 25. There is also an indirect meta level reference to the instructions and the setting of the activity.

MOV00M, lines 24-28

	Name	*Intonation unit*	*Gesture/gaze unit*
24.	Madeline	do we have to go all the way,	LIF on M
25.		round there?	tracing with thumb
26.	Madeline	(2.0)	gaze shift to TS

27.	Marina		gaze shift to TS
28.	Marina	uhm I think so?	

The beginning of the utterance in line 24, "do we have to" signals Madeline's awareness of being required to follow the instructions. She includes both herself and Marina by using the plural personal pronoun "we". Both Madeline's and Marina's gaze then shifts to the task sheet and after a verbal pause, Marina replies "uhm I think so?" (line 27). Marina's reply accounts for a para level feature because she refers to herself and expresses her opinion. The modal particle *I think* occurs in a cluster (see Aijmer 1997) and in its parenthetical reading in combination with *so*, it accounts for a weak assertive (see Simon-Vandenbergen 2000).

Quintina and Queena point to the North Gate collaboratively to begin the interaction. Quintina then retracts her right hand and points with her left index finger. This is in anticipation of Queena's tracing movement. In a joint effort, Quintina verbalizes Queena's gesture. She brings the starting point and the end point into a dynamic relationship by tracing the way in between these two. Quintina's and Queena's index fingers meet at the end point of Queena's tracing movement (line 9 in the transcript). The screenshot depicts the successful establishment of a common focus (lines 8 and 9 in the transcript).

Figure 9 Screenshot of establishment of shared focus

MOV00Q, lines 8-14

	Name	*Intonation unit*	*Gesture/gaze unit*
8.	Quintina		LIF point to path below North Gate
9.	Queena	so we go down there.	RIF tracing
10.		(1.0)	
11.		can we go that way?	

12.	Quintina	u:hm,	
13.	Queena	yeah 'cause there's people there.	LIF point to M
14.	Quintina	yeah.	

This example is interesting for various reasons. The meta reference itself is found in line 11 “can we go that way?”, again bearing awareness of being given instructions and of the sense-making process as a whole. The intonation unit contains person deixis “we”, a signal of the joint activity, as well as the referential expression “that way”. It refers to the path that Queena has identified: while she says “so we go down there.” (line 9), she traces the path with her right index finger on the map. The utterance is followed by a pause (line 10), during which Queena does not receive any feedback from Quintina. This is probably the reason why she ask a meta level question, providing an opportunity for Quintina to reply and to continue with the planning. However, Queena’s gaze remains on the map and she does not look at her interlocutor. Quintina’s reply “uhm” in line 12 is insufficient, signaling that she is uncertain and needs time to think. Queena thus finds an answer to the question herself. In line 13, she provides an explanation and points to an area on the map where she can see people. Quintina takes an affirmative stance “yeah” and then they continue with the task.

To summarize thus far, the analysis of task based interactions supports and supplements McNeill et al.’s (2010) findings: dyadic discourse is similar to multiparty discourse in terms of shared co-referential chains. Participants express their cooperation and involvement in verbal co-constructions, mutual pointing, and mutual gaze. The dyadic interactions stand in contrast to multiparty interactions with regard to the dominant participant, however. McNeill et al. (2010:146-148) describe patterns for pointing and gaze direction: generally, the dominant participant is the source of pointing and the target of gaze. In the current study, however, there are differences to these findings. Both participants are the source of pointing and the targets of gaze and deictic gestures are usually the map and the task sheet to which both participants either refer individually or simultaneously. A close analysis of these differences will be presented in Chapter 7 and Chapter 8.

5.4.2 Meta level reference as an evaluation

The forms of evaluation displayed in the progress of the activity can be investigated as another element of meta level references. Evaluation can relate to one's own actions, to the other person's actions, rendering *self*- and *other*-evaluation, and it can also entail comments about some prior action related to the task.

In its function as referencing a prior event or discourse unit, meta reference fulfills the role of expressing an evaluative stance. Thompson and Hunston provide an operationalized account of evaluation in text to talk about "language expressing opinion" (2000:2). They use *evaluation* to include synonymous terms such as *appraisal* and *stance* and they define the term as

> the broad cover term for the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about. That attitude may relate to certainty or obligation or desirability or any of a number of other sets of values (Thompson and Hunston 2000:5).

There are three forms of evaluation which can be identified in this context: first, there is evaluation that relates to the task planning; second, evaluation can take the form of self-evaluation; and third, there is evaluation that relates to someone else's actions. The first form of evaluation refers to something that has been previously proposed or said as in the following passage. It represents discourse about the activity and ends with an evaluative remark by one of the participants.

MOV00R, lines 396-400

	Name	*Intonation unit*	*Gesture/gaze unit*
396.	Rita	hold on no,	
397.		we got all this other [stuff,]	gaze shift to M
398.	Rose	[to go to?]	gaze shift to M
399.		[u:hm]	
400.	Rita	[we're] doing this really awkward way.	

The evaluation "awkward" modified with the intensifier "really" in line 400 is a result of the previous planning process. Rita points out that she and Rose still have a few more things from

the list to see. In lines 397 and 398, there is overlapping speech during which Rose complements Rita's utterance. Both participants shift their gaze to the M and Rita then evaluates their current route description, using the progressive form "we're doing". This assessment launches a new description and a re-planning of the routes.

In its second form, evaluation can relate to one's own utterances and actions in taking a position toward them. Hence, one can establish a close relationship to the para level. Fiona and Flavia have just begun the interaction and now search for the first entity to visit.

MOV00F, lines 3-10

	Name	*Intonation unit*	*Gesture/gaze unit*
3.	Fiona	uh let's see,	
4.		(1.5) I'll look from the center,	
5.		[((laugh))]	
6.	Flavia	[((laugh))]	
7.	Fiona	so I can find it.	
8.		uh ..	
9.	Flavia	here it is.	RIF point to Bear Grottos
10.	Fiona	a:lright okay.	

The two utterances by Fiona in lines 4 and 7 account for the meta level because she explains what she is going to do in order to successfully locate the first animal house on the map. She draws a personal reference to herself "I" and thereby informs Fiona of her individual actions which she contributes to the task management. The statement "so I can find it" functions both as an explanation as well as an evaluation of her own actions. The evaluative tokens "alright" and "okay" in line 10 relate to Flavia's contribution "here it is" and positively assess Flavia's identification of the relevant entity on the map.

The next excerpt contains self-evaluation and the stance marker "I think" in two of its usages. Xandra's and Xenia's next stop is the Great Bear Wilderness. The excerpt begins with Xenia planning the route. Due to the participants' seating position, the referent of each point is not recognizable on the map. Pointing movements, however, are still visible and indicated accordingly in the transcript.

MOV00X, lines 97-115

	Name	*Intonation unit*	*Gesture/gaze unit*
97.	Xenia:	°yeah so we have to go through there,°	RIF
98.		°so we get there,°	RIF
99.		°from there,°	RIF tracing
100.	Xandra:	°so which way would you go then?°	
101.	Xenia:	the same as yours.	RIF
102.		I think. (2.3)	
103.		mh, (5.6)	gaze shift to TS
104.			gaze shift to M
105.		uhm ( )	RIF point
106.		there .. are we,	RIF point
107.	Xandra:	°(yeah.)°	RH point
108.		start from there. (2.9)	
109.	Xenia:	°and then,°	
110.		that's grass.	
111.		isn't it right?	
112.		u:hm,	
113.	Xandra:	I think it is [grass yeah.]	nodding
114.	Xenia:	[(it doesn't really say,]	RIF point
115.		in the key.	RIF point to key

Xenia points twice when she says “there” and ends her description with a right index finger trace (line 99). In line 100, Xandra addresses Xenia to ask which way she would go. Xenia’s reply in lines 101 and 102 falls into the para level category because of the stance marker “I think”. It frames the statement and functions as a relativizer to the alignment expressed in “same as yours”. The particle is placed at the end of phrase and is structurally different from “I think” used to introduce a stance utterance because it weakens the utterance instead of reinforcing it. The particle is followed by two longer (filled) pauses during which Xenia shifts her gaze from map to task sheet and back. Xandra’s gaze is directed at the map the whole time and when Xenia says “there .. are we” accompanied by a right index finger point, Xandra forcefully points to the same

area with her right hand. In this case, the gesture carries more significance than the speech. The utterance, presumably "yeah", is almost inaudible, but the gesture follows immediately after Xenia's utterance and signals strong agreement. Starting in line 110, a question that has been asked earlier is reiterated. There is a space on the map which both participants identify as a patch of grass. However, they are not completely certain about this since there is no explanation in the key. This is seen in Xenia's utterance and her two right index finger pointing gestures (lines 114 and 115). Xandra states her opinion in line 113, introducing her utterance with a deliberate "I think" and ending the intonation unit with a reinforcing "yeah". The issue is repeated later in the interaction when Xandra and Xenia recapitulate the whole planning process. Repetition will be analyzed in-depth in the next chapter. However, the repetition is of interest in the context of meta and para level reference and contains another instance of a stance marker.

MOVOOX, lines 374-378

	Name	*Intonation unit*	*Gesture/gaze unit*
374.	Xandra	okay- uhm: .. we come out from our left,	RIF point
375.		past the patch of grass,	RIF tracing
376.		which I think it is.	raises and lowers head
377.		°patch of grass,°	
378.	Xenia:	((laughs))	

The area where the patch of grass is located is highlighted with a tracing movement on the map (line 375). The para level reference in line 376 is accompanied by a head movement. Xandra slightly raises her head and lowers it again to indicate that she is expressing her opinion. Halliday classifies this usage of *I think* as a "qualificative comment" (1994:131) since the speaker is commenting on a proposition. In combination with the repetition of the phrase "patch of grass" in line 377 and the head movement, "I think" re-affirms Xandra's judgment about the previously identified area. The example did not only present self-evaluation, but also other-evaluation.

The third type of meta reference identified in this study contains an assessment which relates to someone else's actions. The topical unit in the following transcript is concerned with finding the butterflies. Wilma and Wendy have some difficulty locating the butterflies on the map and this is represented in speech and in the distribution of the map. The participants engage

in a joint effort to find the butterflies on the map. They do so by giving up control over the shared interaction space, i.e., one person "owns" the map for a short period while the other person looks from a more distant perspective. Points, traces, and most of the gaze behaviors, are directed at the map and the task sheet and thus relate to the object level; the verbalizations, however, take place above the object level, consisting of meta and para level utterances.

MOV00W, lines 315-335

	Name	*Intonation Unit*	*Gesture/ gaze unit*
315.	Wilma	but there's not even a picture of THEM?	gaze at M
316.	Wendy	oh no I've been looking for pictures.	gaze at M
317.	Wilma	(12.5)	searching for butterflies on M
318.			retrieves RH
319.	Wendy	°let me have a look°.	turns M in her direction
320.		(1.0)	
321.		right ... [(huge animals are-).]	RIF point to top of M
322.	Wilma	[what did that say then?]	RIF point to bottom left corner of M
323.	Wendy	[°La Gran-°].	RIF trace (reading)
324.	Wilma	[oh Mold-A-Rama].	
325.	Wendy	HUH?	gaze at Wilma
326.		Mold-A-Rama.	gaze at M, RIF point
327.		what's Mold-A-Rama.	
328.	Wilma	I thought it said moth.	
329.	Wendy	huh?	
330.	Wilma	thought it said moth.	
331.	Wendy	that's not a butterfly.	
332.		WEll?	
333.	Wilma	(well they might be in the same ( )).	
334.			gaze at Wendy
335.	Wendy	well that's true ((snuffles)) ... H.	

At the beginning of the transcript, the map is turned in Wilma's direction and her right hand is placed on it to signal that she is the "owner" of the map at this moment in the interaction. Her meta level reference in line 315 contains a direct reference to the butterflies "them". In line 316, there is another instance of a combined meta and para level reference because Wendy refers to a previous activity that she herself had carried out. When she had control over the map, she had looked for a picture of the butterfly house already. There is a long verbal pause during which Wilma scans the map before she retrieves her right hand. This signals that she is ready to complete her turn, probably because she was unsuccessful in finding the symbol for the butterflies. Wendy now turns the map and utters "let me have a look", again combining para and meta level features. There are two changes in F-formation, from shared interaction space to individual ownership of the map, first by Wilma and then by Wendy. This is the reason why the participants' individual foci differ (lines 322 and 323) and cooperation fails. The verbal overlap shows that they talk about two different things. Wendy points to an area in the bottom left corner of the map and Wilma traces across the map right above this area. This difference in focus leads to a double clarification on the meta level: Wendy mistakes "Mold-A-Rama" for "moth", which she associates with butterflies. When she says "oh Mold-A-Rama" in line 324, she realizes that she misread the label. Her utterance also prompts Wilma's attention to this particular space on the map. Wendy positions herself in lines 328 and 330, establishing a reference to herself and her thoughts. Aijmer (1997:7) attributes structural and formal flexibility to the construction *I think* to differentiate it, for instance, from a discourse marker such as *I know*. Formal flexibility means that the particle "I think" can vary in tense, modality, or aspect. Wendy uses the marker in the past tense to refer to her previous assumption that "Mold-A-Rama" said "Moth". She omits the personal pronoun "I" in the repeated phrasing in line 330.

In summary, the meta level can be divided into references aimed at the external representations, either directly or indirectly, and into references which fulfill an evaluative function. Attitudes and opinions toward the activity performance are expressed.

5.5 CONCLUSION

Cohesion and the notion of co-referential chains are viewed in relation to stance taking and positioning within and across topical units throughout the progress of the activity performance. Similar to a multiparty setting, the investigation of two-participant interactions emphasizes the

importance of a multimodal approach to language. Taking deictic gestures and gaze into a discourse analytic account, it is illustrated how individual actions are carried out, including evaluation of and alignment with one's own contributions. At the same time, in assuming that two (or more) people form coalitions and express allegiance on multiple levels, stance taking toward someone else's utterances and actions become apparent. All three levels, object, meta, and para, are closely intertwined and they reflect upon the actions of the two participants. These three co-referential chains also illustrate that cohesion must be investigated in interaction and based on the inclusion of multiple modes. Verbal references in the form of naming entities and/or the use of pronouns and demonstratives are oftentimes accompanied by pointing—with the finger, the pen, and also the participants' gaze. Pointing, however, can also occur without speech, highlighting the communicative nature of gesture.

The two interactants create and maintain F-formations through both individual and joint actions as they share control over the space and the external representations in front of them. The power and control over the interaction space and the representations can be distributed to shift from a joint to an individual focus. Changes in the distribution of the map and the task sheet are sometimes verbalized and can contain a para level feature.

The results of the investigation yield two new features which can be added to McNeill et al.'s (2010) understanding of co-referential chains, based on a discourse analytic perspective. Firstly, epistemic stance markers in the form of verbs and modal particles such as *I think*, *I suppose*, and *I guess*, are applied to assert an opinion or to soften an utterance. Secondly, in addition to reference to oneself expressed in the usage of the pronouns "I" or the "me", for instance, there are two other ways in which participants establish reference to a person. Interlocutors directly address the other party to invite their input, suggestions, and feedback. Lastly, they use the plural form "we", which frequently occurs in combination with questions and statements accounting for the meta level. The plural form of the personal pronoun entails a sense of solidarity and "togetherness", illustrating that participants' assumption of jointly fulfilling the task as a pair activity.

Para level reference co-occurs with meta level reference, a level that includes references to prior discourse themes as well as references to the task performance, to the sense making process, and to evaluations. Verbal meta level utterances can be accompanied by a gesture and a gaze pattern directed to the object level, i.e., there is direct reference to an entity on the map/task

sheet. Differentiating the co-referential levels and the modalities of speech, gesture, and gaze thus allows for a detailed analysis of the individual processes to unite them into one complete picture. In terms of establishing and maintaining cooperation, verbal overlap and co-constructions, collaborative pointing, and mutual gaze occur. Gestures are embodied by the co-participant and sometimes they are verbalized by the other party, i.e., one participant traces a route and the other participant describes this path in speech. The dyads draw on multiple modes to mutually achieve the task and their individual actions complement each other.

In contrast to a multiparty setting where there might be a dominant speaker, the forms of asymmetry described by McNeill et al. cannot be maintained for the current purposes. The target and the source of pointing and gaze differ, i.e., that map and task sheet influence the gesture and gaze patterns of the participants, as I will demonstrate in Chapter 8. Yet another feature, which has received attention in traditional discourse analysis, can inform us of cohesion across the task performance: repetition. Verbal repetition and gesture recurrence provide information about co-referential chains. These features will be investigated in Chapter 6 as they allow for a study of cohesion across longer discourse passages.

6 REPETITION AND COMMON GROUND

In the interactive task in this study, the interactants need to ground their verbal and gestural communicative acts to reach the goal of the current purpose, to complete the task they are given. They update their common ground throughout the activity and, indeed, some dyads reinforce their common ground toward the end of the interaction. It is proposed that they do so by a repetition of the activity planning to recontextualize previous contributions. It is also argued that gesture recurrence and mimicry establish a shared focus, help comprehension, and result in common ground. Thus, verbal repetition and gesture recurrence are presented as two means to increment common ground, either co-occurring or substituting one another.

Johnstone writes that repetition "creates a shared universe of discourse" (1987:207), thereby relating to the cohesive nature of repetition. Parallel to Stivers' (2008) finding that there is preference for agreement toward the end of a story, the repetition of descriptions of routes through the zoo suggests that participants want to ensure that they both agree on the steps they are taking right now and on those they have taken so far. Repeated route descriptions can also occur to ensure that they remember proceedings (correctly). Repetition is thus a cognitive device that facilitates memorization and understanding. It can also express stance and can be used as a conversational and/or interactional device. Norrick (1987) relates to Jefferson when he differentiates different kinds of repetition and introduces intention as a key feature of repetition in conversation. He writes:

> So, first, there is repetition of unconnected lexical items, then, repetition of phrases which perform no significant operation on their originals, and, finally, repetition predominantly reflecting the exigencies of face-to-face conversation. All of these attract less cognitive attention, create less cohesion, and contribute less to the progress of a conversation than corresponding phrases or larger chunks repeated intentionally to perform specific operations (Norrick 1987:248).

Repeats in this study are performed by the participants with a specific purpose in mind. The repetition allows them to take a certain stance and to align with another person. Through these actions, they ground their actions and reach common ground.

Bazzanella (2011:246) explains that repetition can either mean the reappearance of the same formal event or a variation in the repetition of the event. Repetition can be further divided into *monological* and *dialogical* repetition, or in Tannen's terms (1989:54) into *self-* and *allo-repetition*. The former means that the speaker repeats himself/herself and the latter that a person repeats something another person has said. Words, phrases, sentences and even whole texts can be repeated, either verbatim or in a paraphrased form. In this sense, repetition connects new utterances to previous utterances and hence is a cohesive device (see also Halliday and Hasan 1976). Bazzanella (2011:249) further explains that the repetition of phrases or sentences can also function as stance-taking, and as such, repetition is polyvalent, which means that it can be used to express either agreement or disagreement with a preceding utterance. In an earlier study, Bazzanella (1993) has shown that repetition occurs along a scale, which represents the double-polarity of repetition and varies from complete agreement to disagreement/opposition. Similar to this, the repetition of planned routes might concur with the first planning or it might undergo changes. Alternative versions to the first planning are sometimes introduced with meta-utterances such as "I suppose there's different things you can-, different ways you can do it," (MOV00D, l. 301-302). Dialogical repetition fulfils an interactional function and creates interpersonal involvement, a characteristic of repetition which Tannen defines as follows: "It [repetition] provides a resource to keep talk going, where talk itself is a show of involvement, of willingness to interact, to serve positive face" (1989:52). In sum, repetition fulfils four purposes, which are *production*, *comprehension*, *connection*, and *interaction* (Tannen 1989:48). The studies presented so far all investigate repetition in terms of verbal repetition, but do not incorporate the repetition of nonverbal features into their study objective. In this study, however, the phenomenon of repetition is investigated based on verbal and nonverbal reiterated occurrences of certain items. The feature that I will focus on here is the repetition of references to items on the map as well as the repetition of route descriptions. The references can take the form of single words and proper names, for example of animal exhibits, as well as whole phrases and larger chunks of discourse. I will also look at the recurrence and repetition of gestures. Repetition is found at all stages of the activity, fairly close to the start, in mid-activity and at the end of the activity to close the planning process.

Howarth and Anderson (2007) study how new objects are introduced in dialogue. They claim that the conversational setting, face-to-face versus video-mediated, as well as the cognitive

load, high or low, influence the way two interlocutors collaborate with each other in The Map Task. Howarth and Anderson indeed show that time pressure and setting of the conversation influences how new entities are introduced into a dialogue. Their findings further show that there is articulatory reduction, i.e., a faster articulation of words, which occurs when words that refer to the same object are repeated (Howarth and Anderson 2007:289). Speakers thus draw a distinction between new and old information and words are articulated less clearly and more quickly when they refer to a previously introduced entity. However, this depends on whether the setting is monologic or dialogic (see Hupet and Chantraine 1992). The cost of the referring behavior is high when speakers address different groups of listeners and also when there is no contribution by the listener. However, if the listener is present and can cooperate in the task, the cost of referring decreases because the listener's contribution is important in producing a successful reference (Hupet and Chantraine 1992:493). In summary, these studies demonstrate that references and referring expressions decrease in length and number of words when repeated. This is what Clark and his colleagues view as the collaborative process of referring.

Reference can also be created with gestures. In analogy to the findings for verbal reference, it is one of the assumptions of this study that deictic gestures referring to entities on the map will decrease or will disappear completely in repeated versions. This assumption is in accordance with McNeill and Levy's (1993:365) proposal of a general principle suggesting that fewer gestures occur when information is already presupposed. This claim also relates to Levy and McNeill's idea that *communicative dynamism (CD)* (1992:279) runs across stretches of discourse, opposed to only the clause level. CD is high when new topics are introduced in narratives and conversations. Speakers increase "the quantity of spoken and gestural output at episode junctures" (McNeill and Levy 1992:299). Conversely, when CD is low, and this is assumed to be the case when phrases and larger chunks are reiterated, verbalizations should be shorter and gesture usage should decrease. I will first present some examples of a repeated description at the beginning of the interaction and then continue with repetitions occurring during the planning as well as repetitions occurring at activity closings.

6.1 REPETITION AT THE BEGINNING OF THE INTERACTION

Ben and Beth are supposed to go from the North Gate to the Camels. Due to some problems in origo and orientation, this first path from A to B is repeated several times, both in speech and in gesture. Both participants hold a pen and use it to point.

MOV00B, lines 8-43

	Name	*Intonation Unit*	*Gesture/gaze unit*
8.	Beth	so we come through this gate,	RH point to North Gate
9.		okay?	pen on M
10.		(do we?)	
11.	Ben	so you go through the [gate,]	RH point to North Gate
12.	Beth	[through the gate]	pen on M
13.	Ben	turn right,	RH tracing
14.	Beth	NO?	turns her head
15.	Ben	yeah.	RH point
16.	Beth	right,	RH tracing
17.		yeah,	pen on M
18.	Ben	past the toilets	RH tracing
19.	Beth	goo:d?	RH tracing
20.	Ben	(move a bit away from it)	
21.		turn [right,]	RH tracing
22.	Beth	[turn right?]	RH tracing
23.	Ben	at the camels,	repeated pointing
24.	Beth	and then go to the camels.	repeated pointing
10 seconds of conversation omitted here			
33.	Ben	walk through (the) North Gate,	
34.	Beth		RH point to North Gate
35.	Ben	turn [right],	
36.	Beth	[turn] right .. at the toilets,	RH point to toilets
37.		[and then],	

38.	Ben	[after the] mh female toilets,	
39.	Beth	I need to look at it this way.	turns M in her direction
40.		(2.0)	
41.		okay ... right.	
42.		and then,	
43.	Ben	and head towards camels.	RH point to camels

The transcript can be investigated under several aspects of repetition. For example, in lines 8, 11 and 12, the reference to the North Gate is reiterated three times. Beth identifies "this gate" with a pointing gesture and Ben slightly rephrases the utterance and says "the gate", but maintains the same pointing gesture to the North Gate. While Ben points, Beth's pen remains on the map, so that as a result, there is collaborative pointing. The overlap in gesture is also represented in speech: Beth echoes Ben's utterance (see Norrick 1987), a process that Tannen calls *shadowing*, defined as "repeating what is being heard with a split-second delay" (1989:88). Shadowing is also seen in lines 21 and 22 as well as 35 and 36. Beth and Ben do not only shadow each other's words, but they also shadow each other's gestures. In both modalities, cooperativeness and involvement are expressed. From lines 18 to 22, Ben outlines the route on the map and Beth follows his steps, so to speak. There is a positive evaluation "good" in line 19, which further stresses the affirmative stance already demonstrated in the nonverbal mode. Shadowing and overlap, as Tannen explains, are rapport-building and express the cooperative stance of both participants. By shadowing, participants can also overcome problems in origo, which can arise due to different perspectives and the differing individual foci. An example of this is seen in line 14, when Beth disagrees with Ben's suggestion to turn right. "No" is uttered louder and with a rising intonation to express her strong disagreement. Ben however insists on the direction and uses the pen to point out where he wants to turn right (line 15). This is how he visualizes his perspective for Beth. After the agreement token "yeah" uttered by Beth (line 17), Ben continues to outline the path which they should take. Once they have reached the camels, there is repeated pointing (lines 23 and 24) to mark the arrival at their destination. Beth also expands on Ben's statement when she says "and then go to the camels" (line 24). After a few seconds during which the two interactants re-read the instructions, they reintroduce the route description. This time, there is no tracing at all and the pointing gestures also decrease. Beth restates one of Ben's

specifications when she says "[turn] right .. at the toilets" to show affirmation, to reinforce her earlier positive evaluation, and to ensure mutual understanding. Ben also recycles his utterance and expands on it when he specifies that the toilets are the "female toilets" (line 38).

In sum, through the reiteration of the previously mentioned items, textual coherence is enhanced. The affirmative repeats of phrases and gestures reassert the rapport building process that takes place at the beginning of the interaction. The joint activities are grounded and re-grounded throughout the development of the exchange above. Based on the common ground that they now have established, the planning can continue. Repeated episodes of the planning process are also found while planning. I will now continue with a longer excerpt from a single interaction, outlining the development of the planning process.

6.2 REPETITION WHILE PLANNING

In MOV00R, Rose and Rita start from the South Gate and both of them are planning the routes. Rita, however, is the predominant pointer. She is the "note taker" because she holds a pen in her hand and uses it throughout the planning process to point, trace, and mark entities on the map. The labels "point" and "tracing" in the transcript mean that Rita points/traces with the pen unless otherwise indicated. Rose points less frequently and if so, without a pen.

MOV00R, lines 10-22

	Name	*Intonation unit*	*Gesture/gaze unit*
10.	Rose	so we enter the South Gate?	
11.	Rita	go straight past The Swamp .. forward.	RH tracing
12.	Rose	ye:s.	
13.	Rita	straight on past the Swamp.	RH tracing
14.	Rose	YE:S.	
15.	Rita	mh,	
16.		just right,	RH point to fountain
17.	Rose	make a right by Roosevelt Fountain?	
18.	Rita	the Bear Grotto.	
19.	Rose	and there are the Bear Grottos?	
20.	Rita	and then to Fragile Kingdom which is,	

21.	Rose		RIF point to Fragile Kingdom
22.	Rita	.. there.	

The first repetition is found in line 13. Rita reiterates her previous utterance (line 11) and re-traces the route past the Swamp. The repetition of words and gesture implies that this is a problematic instance in the planning and it can be assumed that Rita is insecure about her choice. In favor of this assumption speaks Rose's reaction: she replies "yes" twice, but the second time it is spoken louder to reinforce her positive stance toward Rita's suggestion. The repetition also represents the two categories of *production* and *comprehension* (Tannen 1989:48-49). It enables Rita to produce more fluent speech – there is no pause in the repetition – and at the same time she gains time to think about the next step. This relates to the phenomenon that Howarth and Anderson (2007) describe and it supports their argument that the speaker can produce a reference more easily when the listener is cooperative. Rita is reassured because of Rose's verbal support and she can assume that common ground is successfully established. In lines 16 and 17, there is a co-construction of the route. Rita's description is insufficient in the verbal mode and only her pointing gesture clarifies what she means by "just right". Due to the accessibility of the map to Rose, she can see what Rita refers to and elaborates on Rita's utterance. Tannen (1989:65) describes similar features of repetition when she introduces the notion of expansion. In Tannen's example, the speaker herself elaborates on a previous utterance, which in turn leads the addressee to a transformation of the initial question and an answer to it. In lines 17 and 19, Rose elaborates on Rita's contribution and specifies which locations they pass and finally reach. The excerpt closes with another co-construction. Rita plans the next move and verbally refers to the Fragile Kingdom (line 20) and Rose complements this utterance with a point to the Fragile Kingdom in line 21. Rita then refers to the location "there" verbally. In this example, the repetition of entities occurred early in the interaction and it portrays how the interactants make sense of the activity. They use self- and allo-repetition, in particular elaboration strategies, to reach agreement. The example also demonstrates the close cooperation and building of common ground between the two interactants.

Rita and Rose complete the list of activities after about 7min 30sec (235 intonation units). They have now reached a point where they can decide to end the session or to continue their interaction. Both of them re-direct their gaze to the instructions and they look at them silently for

a while. This is another completion point at which one or both of them could establish mutual gaze, for instance, to signal the end of the task. However, Rose reads the beginning of the instructions "you go °visit°" again. Her turn initiates a repetition of the previous route description. Rita and Rose indeed agree that they should repeat all their steps to make sure "they [we] got the routes and stuff right" (line 243). The transcript starts in line 248 when Rita mentions the Fragile Kingdom. This is the second item on the list after the Bear Grottos and the destination they want to reach.

MOV00R, lines 248-276

	Name	*Intonation Unit*	*Gesture/gaze unit*
248.	Rita	from there to Fragile Kingdom.	RH tracing
249.		so back to Roosevelt Fountain.	RH tracing
250.		north .. °right°?	
251.	Rose	NOrth.	
252.	Rita	would you?	
253.	Rose	is that an [entrance?]	RIF point to M
254.	Rita	[here?]	RH point to Pavillions
255.		YEAH would you go THAT way.	marks path on M
256.	Rose	O:H [yes,]	
257.	Rita	[o:r-]	
258.		(3.2)	
259.		or would it be like,	RH tracing
260.		you have to go that way.	RH tracing
261.	Rose	yeah:.	
262.		(1.0)	
263.		oh hold on.	
264.		if we g- go like north,	RIF point to M
265.		into The Pavillions.	
266.	Rita	right and then,	
267.		from The Pavillions into the Fragile Kingdom.	repeated pointing
268.	Rose	yeah:.	

269.	Rita	yeah.
270.		right so,
271.		from the Bear Grottos pa- north past the Roosevelt Fountain,
272.		(t-) and into The Pavillions,
273.		and [then] from there,
274.	Rose	[yeah].
275.	Rita	like east into The Fragile Kingdom yeah?
276.	Rose	yeah .. yeah.

Lines 252 to 269 present a passage where common ground is established and the prior initial planning is re-evaluated. In the first round, Rita and Rose describe a similar path, but they are already uncertain about which of the two paths they should take. Rita, for example, says "depends on which way we go" (line 24). This problem is re-introduced by her in line 252, although she cannot complete her utterance at this point. Rose inserts another question (line 253), which seems to relate to Rita's question as well, because they point to the same area on the map. Rita formulates her question and stresses "THAT way" (line 255) while she highlights the way on the map. The visible trace that Rita leaves on the map together with the stress on the demonstrative pronoun offers a possible solution to the problem of which way to go. Rita describes an alternative way in lines 259/260. Because of the other visible line drawn earlier Rita and Rose can compare the two routes. Rose's uptake on the problematization becomes apparent in line 263 and is followed by a potential solution. The co-construction in lines 265 to 267 illustrates the verbalization of the route, which has previously been marked on the map. They list the individual entities that they pass on their way to the Fragile Kingdom. What follows in lines 270 onward is a compact summary of the path they take and this time, there are no gestures at all. The continuation token "yeah" is uttered repeatedly by Rose throughout the interaction and in this particular case it encourages Rita to progress with her planning. Rita also incorporates Rose's solution to the path problem: Rose suggests to go "into The Pavillons" (line 265), which Rita echoes twice to show acceptance and agreement. This is how they reach a common basis and an agreed-upon solution at the end of unit. In sum, from the first planning to the second planning, the frequency of repairs and hesitations decreased. Even though the overall references to entities on the map decrease as well, they still use the names of the animal exhibits instead of

demonstratives. The problematic issues are reconsidered and this is where specifications are made through the use of gestures. The verbal reference to the path occurs via the usage of the demonstrative “that”, which requires a gestural accompaniment for an unambiguous identification of the referent. In the third repetition, starting in line 270, the usage of gestures decreases to zero and the path is described sequentially giving cardinal directions and naming the locations they must pass.

In summary, a repeated planning occurs to negotiate individual routes or even chunks of routes. When the planning is successful, when it is accepted by both participants, the repeated versions are shorter. This process is called ‘streamlining’ because there are fewer pauses, hesitations, and repairs as well as a decrease in the use of gesture. This development is indicative of the achievement of common ground because when both participants share the same assumptions (and beliefs), then there is no need for an elaboration via speech and gesture.

6.3 REPETITION AS SENSE MAKING AND FAMILIARIZATION PROCESS

Repetition can function as a sense making device. The participants need time to acquaint themselves with the task, with the map, and with the instructions. Naturally, this process will be most noticeable at the start of the interaction, but it can occur throughout the planning process as well. The first example (MOV00B) demonstrates how there can be confusion due to a lack of common focus. Beth names the Australia House as the next place to visit and then identifies it with a pointing gesture on the map. In line 94 in the transcript, she still has her left index finger on the map and gazes at Ben to make a suggestion of how to reach the Australia House.

MOV00B, lines 94-107

	Name	*Intonation Unit*	*Gesture/gaze unit*
94.	Beth	should we just retrace our steps.	gaze at Ben, LIF on M
95.	Ben	yeah I think we should.	
96.	Beth	so we come out of the,	
97.	Ben	oh-uh we uh:,	
98.	Beth	thing.	LIF point to Habitat Africa, gaze at Ben
99.	Ben	yeah.	

100.	Beth	should we?	
101.			[gaze at Ben]
102.	Ben	[is that the quickest way.]	
103.	Beth		turns M in Ben's direction
104.		((laugh))	
105.	Ben	ah we're here.	RH point to Habitat Africa
106.			circling gesture
107.		yeah I think we should really retrace our steps.	circling gesture

Beth parses the visual field via her pointing gesture and one can assume that Ben, who is also looking at the M, successfully identifies the area and the referent of the point. Hence, there is immediate agreement in line 95, which, in turn, prompts Beth to continue planning. While Beth plans the next step (lines 96 and 98), Ben utters some hesitations markers (line 97), which causes Beth to look at him and finally, to give up the floor to Ben. This is marked by words and gesture. Beth initiates a question-answer adjacency pair and thereby invites Ben to take up the turn. In line 102, Ben finally states his concerns due to his uncertainty of whether Beth's suggestion relates to "the quickest way". Beth now literally gives the turn to Ben by moving the map in his direction. In line 105, it becomes apparent that Ben was not aware of their current location when Beth set out to plan the route. He makes a circling gesture above the map whereby he parses the visual field for himself and this gesture is repeated alongside the verbatim repetition of Beth's initial utterance. Grounding takes place on the four levels described by Clark (1996). Beth brings Ben to attend to her words and gestures in order to identify her utterance and gesture. When Ben's confusion becomes apparent, Beth aids Ben in understanding what she means by allowing him to take "her" perspective. He considers Beth's words and gestures and, as a result, at the end of the unit, grounding is successful. This is also demonstrated in the fact that they establish a shared focus and reach agreement, which is expressed in the repetition of the phrase "retrace our steps" and a gesture that carries out the steps they will retrace.

In MOV00D there are three sequential repetitions of the route description. Due to spatial limitations, I will present selected excerpts of the interaction to demonstrate how repetition is used as a sense-making device and how it depicts the process of familiarization. In the second and third repetition, the participants negotiate what they each view as given or debatable. Dan

and David search for the Motor Safari, which they take to go to “La Gran Cocina” to meet their friends for lunch. Both of them hold a pen in their right hand and use it to point. Dan reads out the instructions in lines 47 to 50, which is typical of the first planning phase and is part of the familiarization process.

MOV00D, lines 47-67

	Name	*Intonation Unit*	*Gesture/gaze unit*
47.	Dan	after the show,	
48.		you want to go to: the Motor Safari,	
49.		and meet your friends for lunch,	
50.		... .t at La Gran Co- Cocina.	
51.		(2.8)	
52.		[so the Motor Safari.]	
53.	David	[there's the Gran Co-,]	RH point to La Gran Cocina
54.	Dan	(where's) the Motor Safari.	
55.			pen on La Gran Cocina
56.	David	mh.	
57.		(9.1)	
58.		(H).	
59.		I can't see it.	moves arm up
60.	Dan	°yeah°.	
61.	David	is it- OH?	repeated pointing
62.		that’s it,	
63.		°where is it°,	
64.		that's there,	RH point to Motor Safari
65.		there's one there yeah,	RH point to Motor Safari
66.		that's THAT's the nearest one.	RH point to Motor Safari
67.		yeah °that's the nearest one°.	RH point to Motor Safari

After a pause of 2.8sec where there is not any speech nor is there a gesture, it becomes obvious that Dan is looking for the Motor Safari while David is searching for La Gran Cocina (lines 52

and 53). Prior to their utterances, their individual foci differed, which is reflected in the output in speech. What follows is a close-knit cooperation in the verbal and the gestural mode. In line 53, David identifies the Gran Cocina. Note that word "Cocina" is not even fully uttered, but the pointing gesture clearly shows the referent. Dan repeats his utterance from line 52 and this time he formulates a clear question "Where's the Motor Safari?". In line 55, he places his pen on the symbol for the Gran Cocina in acknowledgement of David's contribution. The pen remains on this item while both of them search for the Motor Safari. After a long pause during which they familiarize themselves with the map, David says "I can't see it", referring to the Motor Safari. He moves his right arm so that the pen points to the North Gate. The pen functions as an extension of his arm and finger and as such it is used as an orientation tool. By redirecting it, he re-orients and starts his search from the top of the map. Because of the movement of the pen, the process of changing the perspective becomes observable and provides information that is not presented in speech. David then points repeatedly to the symbol for the Motor Safari and identifies the individual stops on the map. In line 67 he recycles his previous utterance as a reinforcement and positive evaluation of his direction giving.

Pointing occurs frequently in this first planning and correlates with demonstratives such as "there" and "that". The completion of the first round is marked by the participants' return to the North Gate. David starts anew when he refers to the instructions "okay if you arrive at nine?" (line 88), using the same strategy that Rose used in the second example and hence initiates the second route description. Two things happen now: firstly, the demonstrative "there", which was used exclusively in the first round, changes to "here". This shift signals proximity to the speaker and demonstrates that a familiarization process has taken place. Secondly, the interactants ground their stance by visibly tracing the paths on the map. Instead of pointing, David now connects the individual stops by drawing lines along the way as they go from "here around there, and we could go down here" (lines 90-91). At this stage, there is shared attention as both participants look at the map following David's movement and thus the use of proper names decreases. Nevertheless, tracing occurs throughout.

In the third repetition, the use of demonstratives decreases because Dan and David now use the proper names of the animal exhibits. This is due to the fact that Dan assigns numbers to the houses they have visited. He writes a digit next to each item on the task sheet.

MOV00D, lines 160-190

	Name	*Intonation Unit*	*Gesture/gaze unit*
160.	Dan	so we start at the North Gate,	RH point at TS
161.		first thing we would,	RH point to M
162.		first thing we would see i:s,	RH point to Australia House
163.	David	the Australia House.	RH point
164.	Dan	Australia House,	
165.		that would be number one.	writes number 1 on TS
166.		(2.9)	
167.	David	then we would go:?	
168.		go straight ahead to see the Camels.	RH tracing
169.			gaze at TS
170.	Dan		writes number 2 on TS
171.	David	then go to Habitat Africa,	gaze at TS
172.	Dan		writes number 3 on TS
173.		okay.	
174.	David	then go round past the Aardvarks,	RH tracing
175.		back past the Australia House,	RH tracing
176.		past the Butterflies,	RH tracing
177.		down past the (Pavi)-Pavillions,	RH tracing
178.		go through the .. Fragile Kingdom,	RH point to Fragile Kingdom
179.		to Seven Seas?	RH point to Seven Seas, gaze at M
180.	Dan		gaze at M
181.		( )	
182.	David	but if we just,	
183.	Dan	(),	
184.		(detail)	
185.	David	(well if you're just gonna ( ))	
186.	Dan	so: .. so that's,	
187.		so Seven Seas is gonna be number,	

188.	David	[four.]	
189.	Dan	[four.]	writes number 4 on TS
190.	David	°yeah°.	

Lines 161-164 represent a dense co-construction, both verbally and gesturally. Dan points to the Australia House and David complements this pointing gesture in the verbal mode. Dan recycles his own utterance in reverse order and notes number one on the task sheet. In lines 168 to 173 it becomes apparent how Dan's note taking activities influence the course of the interaction. David looks from the map to the task sheet twice (lines 169, 171) to check whether Dan has understood his description and whether he has completed the note taking process. The agreement token "okay" (line 173) marks the end of the note taking and signals to David that he can continue. David traces the way "round" and "past" and then points to the Fragile Kingdom and the Seven Seas (lines 178, 179). By tracing the route David depicts the movement through the zoo and by pointing he indicates the stops along the way. The following exchange (lines 180-185) is not completely audible, but there must be some kind of disagreement. It is assumed that Dan cannot follow David's description because most of the animal exhibits that David mentions, for example the Aardvarks, the butterflies, and the Pavillions, are not on the list. Dan problematizes this issue and asks for less detail. This is seen in the fact that they reconsider Dan's route description and agree that the Seven Seas will be number four. Embellishments at this point in the planning are not necessary and can even hinder shared understanding. When there is an assumed agreement between two participants, an elaboration by one of them can lead to misunderstandings and confusions. The other person might disalign with the speaker's description and request less detail. In other words, when there is agreement, or when one person believes that the other person shares his/her beliefs, i.e., when these are considered to be common beliefs, then the process of streamlining is the preferred action.

So far it has been shown that repetition can occur mid-activity to ensure mutual agreement and understanding before the planning continues. The participants expand individual routes, either by self- or allo-expansion. However, when agreement is assumed, there is a preference for a non-expanded path description and usually the use of gestures will also decrease. I will now adduce a brief excerpt in which repetition is used to express disagreement before I will continue with repeated route descriptions found toward the end of the activity.

6.4 REPETITION AS DISAGREEMENT

Repetition fulfilling the role of disagreement is not very frequent in the study's dataset. This is not to say that disagreement does not occur. However, if it does, it is mostly in the form of an utterance such as "no" or even a whole phrase as in the example below. Quintina and Queena are close to finishing the activity. Queena suggests to walk "up again" (line 235) to return to the North Gate. Quintina, however, takes a divergent stance in line 238.

MOV00Q, lines 234-244

	Name	*Intonation Unit*	*Gesture/gaze unit*
234.	Quintina	return to the North GATE.	gaze at M
235.	Queena	so it's just up again.	RIF tracing
236.	Quintina		RIF point at M
237.	Queena	and to the exit.	RIF point to North Gate
238.	Quintina	unless we walk round.	RIF tracing
239.	Queena	°we could do,°	
240.	Quintina	'cause the motor coach is in the way.	
241.	Queena	AH yeah.	
242.	Quintina	and then it's just basically round and up.	RIF tracing, point
243.	Queena	>round and up<.	RIF tracing, point
244.		yeah.	gaze at Quintina

Both participants point and trace collaboratively, which allows them to closely follow each other's descriptions. The divergent stance is thus said in a softened form to sustain the interactive frame. Quintina even provides an explanation as to why she disagrees: "'cause the motor coach is in the way" (line 240). What follows is another example of how repetition in quick succession allows for an achievement of common ground. Queena shadows Quintina's words "round and up", spoken quickly, and mimicks her tracing and deictic gesture from the current location to the North Gate. "Yeah" in line 244 is a confirmation token that emphasizes the agreeing stance. Even though this example contains disagreement, it did not occur via repetition. The repetition, and this is the predominant pattern in the data analyzed, functions as agreeing stance taking in order to establish shared understanding.

There is one instance, however, found in MOV00T, where allo-repetition expresses disagreement and in fact occurs twice. Tamara and Tom want to go on the Motor Safari and they are currently searching for the closest stop. Tom uses a pen to point.

MOV00T, lines 455-467

	Name	*Intonation Unit*	*Gesture/gaze unit*
455.	Tamara	it's the stops though,	LIF point to Motor Safari
456.		[isn't it?]	RIF point
457.	Tom	[o:h so: nice],	repeated pointing
458.		ye-yeah so there's one stop,	RH point to Motor Safari
459.		that's the nearest stop innit.	
460.	Tamara	that's not a stop,	RIF point
461.		that's the Information Kiosk.	tracing, RIF point
462.	Tom	NO that is a stop,	RH point to Motor Safari
463.			RH point to Motor Safari
464.	Tamara	OH is it?	
465.	Tom	that is a stop,	RH point to Motor Safari
466.		same as that .. see?	RH point to Motor Safari
467.	Tamara	U:H yeah	nodding

Tamara uses her left index finger and her right index finger to point to several icons representing the Motor Safari (lines 455 and 456). Tom points repeatedly to one particular icon on the map while uttering a positive stance in line 457. He identifies this "stop" as one of many and explains that this is the closest one. Even though we find the question tag "innit" (line 459) at the end of his utterance, he does not intend it as a real question addressed to Tamara. As we see in the following turns, he is convinced that the identified spot is the closest stop to their current location. Tamara recycles Tom's utterance "there's one stop" (line 458) in line 460 when she says "that's not a stop". Through the paraphrased repetition, Tamara disaligns with Tom and states that the place which he has indexed is in fact the Information Kiosk. In line 462, there is the second disagreement, again through a repeat. The rephrasing by Tom is almost verbatim, but instead of using Tamara's negative phrasing "is not", he introduces the disagreeing stance with a loud "no" followed by a positive phrasing of Tamara's utterance. The two pointing gestures in

lines 462 and 463 are of special interest here because the first point in line 462 is carried out with the pen. Tom then takes the cap off the pen and repeats his point as if he wanted to express more detail with the thinner tip of the pen. Tom reinforces his stance a second time by reiterating "that is a stop" (line 465), probably also because he does not receive immediate alignment. Only in line 467 does Tamara agree and nod.

In sum, the feature of repetition occurs as a sense-making device, both with agreeing and disagreeing stance. Disagreement is often softened to maintain the interactive and cooperative frame. Gesture is applied to specify the meaning of verbal utterances and to explain decisions regarding the route planning.

6.5 REPETITION TO CLOSE THE ACTIVITY

I will now continue with examples of repetitions of chunks toward the end of the activity. Endings are a compact summary of the previous planning process. As summaries, the repetitions at the end should be streamlined versions, as opposed to embellishments of the activity planning. The term *ending* refers to a formal concept and is found, as the term implies, at the end of a conversation or a narrative (see Abbott 2008). In narrative research, both the terms *ending* and *closing* are used to refer to the same concept. Norrick (2007:132) explains that storytellers need to mark the end of their story for their listeners, so that the listeners are prepared to respond appropriately. Tellers can use formulaic closings, for example to link the story to the present time and the current conversation topic, or they can also provide a summary of the story told. In Labovian terms, this would be the story *coda*, by which tellers signal that the narrative is finished. Tannen (1989:69) states that episodes in a conversation are often bounded: repetition at the beginning of a conversation sets the theme whereas repetition at the end of a conversation terminates the episode and forms a kind of coda. This, as Tannen argues, is due to the ritualized nature of opening and closing in conversation. In the present context, the design and the structure of the experiment inform the participants of the end point of their activity. Once they have reached the end of the list of instructions, they have completed the task. However, the participants do not end the planning process abruptly. Again, as it has been shown for the openings of the interactions, there are no clear closing formulae available, but by contrast with the openings, participants here draw on formulaic expressions they know from other discourse types. Hence we find endings such as "that's it" or "here we are" to mark the completion of the

activity. These endings are also often marked by a shift in body posture and the establishment of mutual gaze. There are references to the researcher, for example when the participants say that they should inform her (the researcher) that they are finished, turn their heads toward the door to establish eye contact, or rise from their seats to open the door to the other room. Completing the activity is thus different from ending the actual experiment session.

In MOV00O, Olga and Olivia reach the North Gate, from which they exit the zoo. The end of the planning activity leads into a conversation about possible things to do during the next zoo visit.

MOV00O, lines 294-318

	Name	*Intonation Unit*	*Gesture/gaze unit*
294.	Olga	and return to the North Gate at 4 p.m.	gaze at TS
295.	Olivia	so we've gone past the Special Events Arena,	RIF tracing, gaze shift to M
296.	Olga	yeah?	
297.	Olivia	.. to the butterflies,	RIF tracing
298.		and then BACK OUT?	RIF tracing
299.	Olga	so perhaps next time,	LIF circle on M
300.		we could visit that bit.	gaze at Olivia
301.		[((laugh))].	
302.	Olivia	[((laugh))].	gaze at Olga
303.		we were missing out on all the giraffes [and stuff?]	
304.	Olga	[YEAH?]	
305.		we did that stuff,	LIF tracing
306.		didn't [we]?	
307.	Olivia	[o:h?]	
308.	Olga	but never mind.	
309.	Olivia	that's the busy side,	RIF tracing above map
310.	Olga		gaze at Olivia
311.	Olivia	that’s [why:?]	gaze at Olga
312.	Olga	[yeah.]	gaze shift to M

313.	Olivia	busy ((giggle)),	
314.	Olga		gaze at Olivia
315.	Olivia	busy that we can't even get out of?	
316.		(1.0)	
317.		alright should we say?	
318.	Olga	yeah.	

In line 294 Olga reads the last bit of the instructions and Olivia describes the route from their current location back to the exit in lines 295 and 298, tracing the route from the special events arena, passing by the butterflies and reaching the North Gate. When she says "and BACK OUT?" the activity and the planning process are finished. This is marked by the louder speech and the rising intonation at the end of the unit. However, Olga and Olivia do not end their conversation at this point. Olga makes a suggestion for a future visit to the zoo (line 299). This statement refers back to an earlier observation of hers. She says that they are "sort of missing out that bit" (line 291), accompanied by a circling gesture on the map to identify "the bit", which is the part of the zoo that they do not visit. The utterance in line 299 is again accompanied by the same circling gesture that Olga used when she said that they were "missing out". The recurrent gesture thus marks the area that they do not visit and the physical space on the map is associated with the same hand (left hand) and the same gesture space.

Olivia's utterance in line 308 also refers back to Olga's utterance in line 299. With regard to content, she provides a possible explanation as to why they did not visit this particular part of the zoo. With her gesture she traces the busy side of the zoo on the map. Even though the gesture is not identical, i.e., it is not performed in a circle, Olivia still mimics Olga's movement above the map and identifies the same area. McNeill writes, "mimicry is a social interactive response" (2008:8), so when someone repeats the gesture of somebody else, it creates comprehension and provides access to underlying meanings. Gestures in this respect then fulfill the same function that Tannen and others have described for verbal allo-repetition, namely that speech and gesture recurrence enables us to understand the meaning of somebody else's utterance or gesture.

Olivia's statement in line 308 grounds the mutual activities of both participants by a reference to a meta-utterance, an evaluative remark, made by Olga. When Olga read the instructions out loud, she ended by saying "so: it's a pretty full long day" (line 33). Olivia agreed

by saying “busy day”, which in turn was repeated by Olga as a confirmation. The notion of “being busy” is introduced at the beginning of the activity planning and recurs toward the closing of the activity. It is continued in line 314, again presenting a reference to a prior contribution, i.e., a previous problem that occurred because Olga and Olivia were not able to proceed with their route planning. Olga and Olivia lost their way at the zoo at some point and Olivia recontextualizes the current exchange to their prior exchange.

Toward the end of the interaction, there is a brief pause (line 315), which signals the transition to the actual ending. Olivia closes the experiment session by her utterance in line 314, eliciting a positive reply from Olga, and at the same time she stands up and walks toward the door. Olga’s agreement is uttered while Olivia turns around to the door.

Thus far, it has been demonstrated that individual items and events can be repeated toward the end of the activity. The acceptance process (Clark and Wilkes-Gibbs 1986) takes place in several steps and certain notions are reintroduced to reconsider if they are still accepted by both participants. As Fetzer writes, “[...] a reference to context and a reference to meaning signify that there is some controversy about the communicative status of a contribution while at the same time providing a frame of reference for its recontextualization and reevaluation” (2004:2). The notion of visiting a busy zoo, for example, was introduced at the beginning of the interaction and was re-introduced at the end not only to establish common ground, but also to link the beginning and the ending of the activity. There is gesture recurrence, both of self- and other-gestures, in accordance to the topic of “unseen parts of the zoo”. The recurrence of these gestures further expresses the agreeing stance of both participants and helps them to renew the mutual agreement on this topic.

6.6 Conclusion

Repetition as a means to structure discourse has received great attention in interactional sociolinguistics, for example for conversational repetition from Tannen (1989) and Norrick (1987). Stivers (2008) and Bazzanella (2011) present current research on repetition in storytelling and face-to-face interaction. These approaches to repetition place their focus on the functions of verbal repeats, thereby excluding gesture and the importance of gesture recurrence. By contrast with these studies, the current investigation assumed that the phenomenon of repetition is displayed both in speech and in gesture, complementing each other. It was further

assumed that grounding is closely related to repetition in speech and in gesture. Participants can ground their own utterances by repeating them and they can position themselves toward somebody else's utterance. The same holds for gestures. The analysis incorporated self- and other-repetition of verbal and gestural components. Another assumption of this study was that referring expressions and gestures decrease when repeated.

Repetitions, it has been shown, occur across different stages in the interaction. They can appear fairly close to the beginning, in mid-activity and toward the end of the interaction. Depending on when they occur, they fulfill different functions respectively, from sense-making to checking and reinforcement to closing. With regard to gesture, the following pattern is observed: gestures recur as a disambiguation device. Tracing, pointing and even marking paths on the map are all used to clarify, reconsider or reinforce a previous route description, and thus they make the process of grounding visible. Tracing is employed in unison by both participants to show and establish a mutual focus resulting in mutual understanding. The use of gestures only decreases when agreement is reached and often correlates with shorter verbal description as well. When there is a repetition of references of path descriptions at the start of the interaction, they reflect upon the strategies that the participants apply to make sense of the task, individually and mutually. Referring becomes a collaborative process in which verbal references are accompanied by pointing gestures and by tracing, either by oneself or by the other person, thereby facilitating the overt establishment of common ground. The participants demonstrate their active involvement as they signal their willingness to cooperate and to create rapport. Repetition as a sense-making device portrays the process of familiarization: when route descriptions are reiterated at a later point in the interaction, they usually refer back to a problematic issue in the previous planning. Hence, when there is doubt and consequently room for disagreement, individual references or whole chunks of references are reintroduced. Here embellishments in the form of elaborations and expansions of individual routes occur. When there is (assumed) agreement, however, the preferred action is to streamline the repeated chunks. Only when there is agreement do the participants continue with their actions.

Repetition within disagreement does not occur frequently. If it does, however, it is similar to an agreeing stance because speech, gesture, and the recurrence of gesture are all used to clarify one's own viewpoint and to bring the addressee to comprehend the meaning of utterance and gesture. Repetitions, especially in the middle of the planning, are characterized by an ongoing

acceptance process in order to reach mutual agreement. Repetition is used throughout the individual interactions to signal that a common basis is reached and hence, repetition is essential to common ground. When items are reiterated toward the closing of the interaction, they either relate to some previous notions or they function as a memorization device. By assigning numbers to the stops along the way or by reintroducing landmarks and important stops, the two participants ensure memorization of shared and agreed routes. Previous episodes or idea units are bounded and can be considered a coda, a proper closing device to the task. Repetition and gesture recurrence create cohesion across the different stages of the activity and can relate to the object, meta, and para level. Common ground becomes a product of the interaction rather than only being a precondition of the interaction (see also McNeill 2010). In sum, the feature of verbal repetition has been brought into a relationship with the repetition of gesture to demonstrate how these two modes co-occur as means of production of cohesion and the achievement of common ground.

7 DEICTIC GESTURES

In their primary function, the term *deictic gesture* denotes those kinds of gestures which indicate the item or the person to which they point. They are often considered to be visual aids to indexical utterances (see Fludernik 1989/90). Deictic utterances acquire meaning in the context of their usage: demonstrative pronouns can either refer to something in the object world, for example "the car over there", or they can provide deixis by pointing to certain parts in the discourse itself, for example "this paragraph". The index finger is considered the most prototypical "tool" for pointing, although deictic gestures can have various forms and shapes. For the present purpose, pointing is restricted to gestures produced with the hand(s) or the pen, excluding lip or nose points, for instance. This study expands on the traditional view of a deictic gesture and its primary function, arguing for a differentiation within the class of deictics, depending on the functions they fulfill in the process of the task performance. The findings of this study allow for an identification of four major forms of pointing. First, there is the *single pointing gesture*, a movement usually directed at the map and at the task sheet by an individual participant. Second, there is the repeated pointing gesture, which is also directed at the map and the task sheet. Third, there is *tracing*, a special form of pointing which brings two features into a dynamic relationship. Fourth, there is *collaborative pointing*, which can include single and repeated pointing gestures as well as tracing, hereafter called *collaborative tracing*. All deictic gestures can be produced with the right or the left hand and changes in handedness, i.e., changes from the right to the left hand and vice versa, can occur throughout the interaction. Deictic gestures can be carried out on the map/task sheet, sometimes leaving a visible mark. However, they can also be performed in the air above the map/task sheet, which means that there is no contact between the pointing finger and the piece of paper. Some participants use the pen as an extension of their hand or in place of their index finger.

Deictic gestures can replace or they can supplement deictic expressions in speech. In the case of replacement, the deictic gesture alone creates cohesion by referring to an object or a person. In the case of supplementation, a verbal reference will be accompanied by a pointing gesture, which underlines the content and the meaning of the utterance. Such a gesture is called a *co-speech deictic gesture*. Pointing does not simply mean to index something; rather, it is a complex process and in the following it will be argued that pointing is an essential factor of the

organization and achievement of the task at hand. It is assumed that the identified forms of pointing fulfill different communicative and interactional functions and occur at different moments in the course of the interaction.

7.1 SINGLE POINTING GESTURE

The single point is the most basic form of the four types of pointing gestures. It can take the shape of a right index finger point, a left index finger point, and can be extended with a pen. Pointing gestures are also sometimes produced with the thumb or the whole hand. The single point is the most frequent of the four types because it can occur throughout the whole interaction, generally as an indicator of an entity on the map or a phrase on the task sheet. Participants point individually, resulting in either sequential or mutual points. The deictic gesture can be accompanied by speech: an individual participant can be the source of both speech and gesture or these two modes can be distributed across the two participants. Frequently, participants will discuss the shortest and most efficient routes through the zoo. Exemplary for the present data, MOVOOG contains an instance of differentiation of two paths through deictic gestures: one participant presents two possible orders in which they can visit the first four animal houses listed on the task sheet. He points to the first item on the map, then lifts the pen to point to the second one, and so forth. He glosses each of the four entities with a /dən/ sound. He does not provide the names of the lexical items. The gesture and the order in which the points occur are the single source of contrast between order one and order two. The pointing gestures are also the only means of identifying the referent. Due to the visual access to the map and a prior reading of the instructions, the other participant can see and comprehend the first participant's actions, following the two possible paths to take from the starting point to the final destination.
Visual orientation is essential even if pointing and speech co-occur. The next excerpt contains a co-speech deictic gesture in line 27. When Cloe says "there", she points to the location of the Fragile Kingdom and places the pen (held in her right hand) on the map. The gesture builds a reference to the object level while the utterance "there" builds on textual reference and Clare's previous contribution "Fragile Kingdom".

MOVOOC, lines 24-27

	Name	*Intonation unit*	*Gesture unit*
24.	Clare	and then we go to the Fragile Kingdom.	LH on map
25.		(1.0)	
26.		°where's the Fragile Kingdom°.	retrieves LH
27.	Cloe	there.	RH point to Fragile Kingdom

The excerpt does not only present a co-speech gesture performed as a single point. Before Cloe's turn in line 27, Clare's left hand rests on the map, more specifically, her left index finger is placed on the "Roosevelt Fountain". This is the starting point from where they need to go to the "Fragile Kingdom". Thus, in speech Clare refers to a future event in the planning whereas her resting finger indicates a past stop, which she has successfully identified. She removes her left hand when she utters "where's the Fragile Kingdom" (line 26), spoken softly, and this invites Cloe to take the turn.

Thus far, two things have been demonstrated: firstly, deictic gestures require a visual field to which they are directed and all participants of an interaction need access to this area in order to perform and comprehend pointing actions. Secondly, a single deictic gesture is not only used to indicate a new entity, but also functions as a place holder to mark a certain space for oneself and for the other party. This particular space represents old information and it can be connected to a new spatial area. One way of connecting two individual spaces, it is assumed, is to relate them via a tracing movement. This is demonstrated in the next example. As the interaction progresses, Clare and Cloe describe the path from the Roosevelt Fountain to the Fragile Kingdom. Cloe is the active participant as she outlines the route. She moves the pen along the path, above the map, and she pauses briefly after naming the entities they pass along their way.

MOV00C, lines 36-45

	Name	*Intonation unit*	*Gesture unit*
36.	Cloe	go past the Safari Grill Restaurant?	RH tracing
37.		and East Mall.	RH tracing
38.		and then round?	RH tracing

39.		…	RH point
40.		°past the Seven Seas°,	RH tracing
41.	Clare		Nod
42.	Cloe	and down,	RH tracing
43.		.. into the Fragile Kingdom.	RH point to Fragile Kingdom
44.	Clare	yeah.	
45.	Cloe		pen moves down to M

The tracing process correlates with the verbal representations "past" (lines 36 and 40), "round" (line 38), and "down" (line 42). The tracing movements are not a continuous movement. For instance, at the end of the intonation units in lines 36 and 37, there are brief stops, which means that the pen rests above the map before the movement is continued. In line 39, there is a more distinct right hand point, which co-occurs with a verbal pause. There is another right hand point in line 43, indicating the Fragile Kingdom. It marks the arrival at their destination. Clare takes an affirmative stance in line 44 and as a result, Cloe now places the pen on the map. This completion of the right hand point from line 43 firstly illustrates an agreeing second stance by Cloe and secondly, it demonstrates that both participants have successfully established a common focus. This focus is maintained as a visible marker of the destination they have reached and from where they will continue their way.

As a result so far, single points have been shown to function as an indicator of items on the map, either accompanied by a sound or otherwise identified by using the proper name or a demonstrative. Single pointing gestures also co-occur within topical units to mark short pauses and breaks. They can express agreement, for instance at the end of a unit, to visualize the completion of this unit. If the pointing gesture is maintained, it is used as a place marker which aides in connecting the current locations with a new location on the map. In the screenshot, both participants place their pens on the map. In this situation, the pen which is held in an upright position marks one of the stops of the Motor Safari. It is the closest one to the participants' current location. From there, they need to go to the Roosevelt Fountain on the Motor Safari, so they search for the stop closest to the Fountain. As a result, the identification of the closest stop by the participant on the left yields this cross-pattern:

Figure 10 Screenshot of cross pattern (marking start and final destination)

The activity of pointing once has been established as the most frequent and the most basic form of a deictic gesture. There are variations to this form since the gesture can be produced with a tool other than the index finger and it can occur in different locations in the interaction space. For example, gesture space can vary in terms of vertical space, depending on whether the gesture is performed in the air above the map/task sheet or on the map/task sheet. Gesture space can also vary along a horizontal level, drawing a left-right distinction, for instance. Out of the single deictic gesture, at least two other forms of pointing can develop: tracing and collaborative pointing.

7.2 TRACING

I will now continue with the investigation of tracing movements taking single deictic gestures into consideration where applicable. In Goodwin's study (2003), the term *tracing* is used when archaeologists highlight the shape of a feature marked on a map, bringing this feature into a relationship with the actual instantiation of the feature in the soil. Goodwin argues that a single thing, which is treated as one thing in speech, can manifest itself in two different spaces. For example, in the work of archaeologists, one space would be the area in the soil where a certain object is located. The second space would be the map, for instance, on which the object is documented. These two spaces are put into a relationship and are indicated in a single pointing gesture, in tracing (2003:14-15). As part of a "systematic progression within pointing"

(2003:16), Goodwin introduces the concept of 'inscription', which is applicable when the tracing process leaves a visible mark on the map. In the present context, the idea of a tracing gesture will be adapted to relate to a moving finger or pen which indicates paths on the map as the participants move from point A to point B to point C and so on. Moreover, this study extends Goodwin's work as it takes lexical items into account. The claim is that tracing co-occurs with certain lexical items representing directions and movements in speech. The stops and paths can be outlined in the air, they can be outlined with a lowered finger/pen, which touches the map, or they can be performed leaving an actual mark on the map. In the latter case, this enduring mark on the map is considered an inscription. Inscriptions only occur for those pairs of interactants who use the pen and it is then assumed that an inscription is the result of negotiations and agreement across the two participants. The visible mark on the map cannot be erased easily and so an agreeing stance by both interactants needs to precede the inscription.

Earlier an example from interaction MOV00G was adduced. Gabriel was the participant who indicated individual entities with the pen uttering /dən/ simultaneously. He provided two alternative orders and Gavin, the second participant, expressed his preference. The following transcript continues the earlier description and shows how Gavin positions himself in relation to Gabriel's two path suggestions.

MOV00G, lines 68-78

	Name	*Intonation unit*	*Gesture unit*
68.	Gavin	I pro-	RH point
69.		PERsonally I 'd probably do the Camels first.	RH point to Camels
70.	Gabriel	[yeah],	
71.	Gavin	['cause] then you can go rou:nd,	RH tracing
72.		past the Aardvarks.	
73.	Gabriel	.. >that's true,<	
74.		you can just go like in a .. a big C like shape.	RH tracing
75.	Gavin	mhm.	
76.	Gabriel	°and come round (here),°	RH tracing
77.		YEAH,	
78.		do camels first then.	

After a false start in line 68, Gavin takes a clear stance and positions himself in relation to Gabriel's two alternative propositions. He introduces the para level reference with the stance marker "personally" and he complements his utterance with a right hand point to the Camels. Gabriel's agreement in line 70 is in overlap with Gavin's explanation as to why he would like to visit the Camels first. The tracing gesture in line 71 takes the shape of a half circle to visualize the way "round", going past the Aardwarks, which will lead them to the Australia House. Gabriel expresses agreement in line 73 and further elaborates on this agreeing stance in the following intonation unit. He says "you can just go like in a .. a big C like shape". The utterance is accompanied by a tracing gesture in the form of the letter C and is carried out in the air above the map. The whole movement is completed with another right hand trace in line 76, which again has the shape of a half circle. Tracing has occurred three times in this short excerpt and in all three cases, the form of the tracing movement correlated with a verbal description, "round" and "C like shape". The route descriptions are compressed, which means that not every lexical item is named. The information is packaged in the gesture and the visualization of the route(s). The unit is completed with another stance taking utterance, which demonstrates the process of understanding and finally agreeing with the other person.

In order to accomplish this agreement, mutual orientation to a shared interaction space is necessary. Adhering to this agreement, Gabriel and Gavin start anew to describe the route from the North Gate to the Camels. They divide the activity roles: Gavin inscribes the path on the map, i.e., he draws an X on the starting point, the North Gate, followed by a line along the path to the Camels, where he draws another X, and he continues this until they reach the Australia House. Gabriel, who watches Gavin's movement, verbalizes the inscription by naming the entities which they visit. From the North Gate to the Australia House, Gavin produces one single movement consisting of several components: the gesture begins with a single pointing gesture, which is visualized as an X on the map, and the gesture is continued as a tracing movement with one more point as a mid-stop and another point highlighting the arrival at the last stop. The marking of the single points is accompanied by speech; it is a co-construction since the verbal components, i.e., the names of the animal exhibits, are provided by the other participant. A single point, it was argued earlier, functions as a place marker and is endured when there is a visible mark.

Deictic gestures and tracing even fulfill the described functions when there is disagreement, as an analysis of the next excerpt will illustrate. Janet and Jennifer have just begun the planning activity and as part of the sense making process, a disagreeing stance occurs.

MOV00J, lines 7-20

	Name	*Intonation unit*	*Gesture unit*
7.	Janet	there is North Gate,	RH point to North Gate
8.		a:nd,	
9.		[we want to go] to the camels,	
10.	Jennifer	[yeah right,]	
11.		so we are here,	[RH inscription]
12.	Janet		[RH point to North Gate]
13.		yeah.	
14.	Jennifer	right,	
15.		so North Gate,	
16.		go to visit the Camels.	
17.	Janet	so you come in,	RH tracing
18.		…	
19.		a:nd?	RH tracing
20.	Jennifer	you can't just go across there.	RH tracing

Janet identifies the starting point in line 7 and receives uptake on this identification in line 11. Jennifer draws a line on the M to indicate where they are right now "so we are here", which coincides with a right hand point by Janet (line 12), prior to the agreement token "yeah" in line 13. Jennifer's inscription of the starting point on the map is a safe activity in the interaction, which she can perform without any expected negative evaluation by Janet. However, not all route suggestions receive a positive evaluation. In lines 17 and 19, there are two tracing movements, the first of which co-occurs with a verb that expresses movement as well, "come in". There is short pause in line 18 and the next utterance "a:nd" is prolonged. The duration of "and" is long enough for the tracing gesture to the Camels to be completed. However, Janet moves the pen across a grey area on the map, which leads to disagreement expressed by Jennifer (line 20). The tracing movement in line 20 combined with the utterance "across there" mirrors

Janet's movement and the content of her planning activity. The embodied repeated gesture highlights the problematic area for Jennifer to illustrate the reason for disagreement. This critical point in the interaction is resolved in laughter by both participants.

Thus far, three features of tracing gestures can be summarized: through tracing, participants relate two or more points dynamically. Frequently, the gesture contains more information than speech, i.e., the tracing gesture co-occurs with certain lexical items expressing movement, for example *round*, *past*, and *down*. The gesture illustrates the individual items which are passed, even when they are not verbalized in speech. A shared interaction space and mutual orientation underlie these activities. Tracing as a special kind of pointing is part of the gesture's progression and development. Tracing can remain constant regardless of lexical content, i.e., the tracing movement highlights an area in the shared interaction space and this action can both show agreement and disagreement. In such a case, speech needs to be considered to comprehend how common ground is built.

7.3 REPEATED POINTING GESTURE

Repeated pointing is defined as a continuous lowering and raising movement of the finger or the pen directed to the same entity. Some instances of repeated points have been provided in previous sections already. Repeated points can occur throughout the interaction and, like a single point, they mark a place on the map/task sheet. However, it is argued that repeated points differ from single points as they are used as a means of reinforcement. In MOV00G, the participants are asked to see the Dolphin Show at Seven Seas.

MOV00G, lines 56-62

	Name	*Intonation Unit*	*Gesture unit*
56.	Gavin	ten thirty .. at the dolphins.	
57.	Gabriel:	so which one is closest to the dolphins?	
58.		at all then,	
59.	Gavin:	yeah .. uh ( )	
60.		where are the dolphins on this ( ),	
61.	Gabriel:	Seven Seas,	repeated pointing
62.	Gavin:	oh yeah yeah,	RH point to Seven Seas

This brief exchange between Gavin and Gabriel represents a highly problematic instance in the interaction. The participants need to locate the Dolphins on the map in order to plan the route. At the same time, they try to be as efficient as possible. The repeated point (line 61) follows two questions, which need to be answered in order to progress successfully. The repeated point thus highlights the crucial element around which the problematization evolves. The repeated gestural reference to the Seven Seas allows Gavin to readjust his focus and the success is demonstrated in his reaction in line 62. He agrees and points to the Seven Seas as well. Gavin is now able to attend to Gabriel's question from line 57 so that they can continue their activities. Repeated pointing as reinforcement can also coincide with an evaluation. A longer passage of the following example was presented earlier. Tom's reaction in line 455 "o:h so: nice" is a positive evaluation relating to Tamara's prior statement. The stops of the Motor Safari represent a crucial element at the present stage of the interaction. By pointing to one of the stops repeatedly, Tom displays comprehension and agreement toward Tamara's contribution.

MOV00T, lines 453-455

	Name	*Intonation Unit*	*Gesture unit*
453.	Tamara	it's the stops though,	LIF point to Motor Safari
454.		[isn't it?]	RIF point to Motor Safari
455.	Tom	[o:h so: nice],	repeated pointing

Repeated pointing is used as a stance taking device, for example when two (or more) options are available. In the following transcript, the participants' activity roles are partially divided: Susan and Sabrina are both route planners because they take regular turns in mutually constructing the routes; however, only Susan holds a pen in her hand and outlines the routes on the map.

MOV00S, lines 134-146

	Name	*Intonation unit*	*Gesture unit*
134.	Susan	okay so FROM here yeah,	RH point to Australia House
135.		from the Australia House,	RH point to Australia House
136.		you go straight ON,	RH tracing
137.	Sabrina	you go f- you go right from the Australia House.	
138.	Susan	yeah you go right from the Australia House,	RH tracing

139.		then you go:?	
140.		(1.0)	
141.		[down at the-],	RH tracing
142.	Sabrina	[take] the FIRST or the second-,	
143.		either the first or the second road,	
144.	Susan	this one.	repeated pointing
145.	Sabrina	the second road after the Carousel.	
146.	Susan	yeah the second road after the Carousel,	

The two single RH points in lines 134 and 135 co-occur with "here" and "Australia House". The demonstrative pronoun precedes the naming of the actual referent whereas the deictic gesture is prolonged and directed at the same location on the map twice. The tracing gestures in lines 136, 138, and 141 are performed in synchrony with lexical items expressing movement (*straight on, go right from, down at*). In line 142, Sabrina contributes to the interaction after a longer pause and some hesitant speech by Susan. Sabrina provides two alternative choices, either the first or the second road, and she repeats her utterance in line 143. Following this repetition, Susan repeatedly points at one of the two roads which Sabrina had previously mentioned. In pointing repeatedly, she makes a decision and underlines this decision. Sabrina fills the pointing gesture with content because she specifies which road to take and elaborates on where to find this road (line 145). Susan expresses agreement through "yeah" and a verbatim repetition of Sabrina's phrasing (line 146).

In addition to highlighting and reinforcing important elements for the planning process, there is a tendency for repeated pointing gestures to be displayed at the end of a unit, which means that they mark the arrival at a designated stop. Rita and Rose go from the Australia House to the Butterflies; Rita is the more active planner and pointer in this excerpt.

MOVOOR, lines 198-214

	Name	*Intonation Unit*	*Gesture unit*
198.	Rita	and from there we go to the butterflies.	
199.		so from there,	
200.		(1.0)	

201.		west out of it,	
202.	Rose	we have to walk back [out,]	
203.	Rita	[to] Australia House,	
204.	Rose	Australia House.	
205.	Rita	go that way,	RH point to M
206.		would you then,	
207.		turn right,	
208.			RH tracing
209.		right again,	
210.	Rose	yeah pas-	
211.	Rita	right again,	RH tracing
212.	Rose	yeah.	
213.	Rita	and then left into,	RH tracing
214.		... the Butterfly House.	repeated pointing

From lines 198 to 204, Rose tries to contribute to the activity, but starting in line 205, Rita claims the turn for herself and begins to describe the path to the butterflies. There are a few right hand traces outlining the path that leads them right several times (lines 208, 211). There is a tracing gesture which co-occurs with “left into” and the preposition “into” emphasizes the fact that they enter the Butterfly House. When Rita verbalizes “the Butterfly House” in line 214, she marks the arrival at this place with a repeated pointing gesture.

Repeated pointing has been shown to take the role of reinforcing and highlighting elements of the planning process. Repeated pointing also aids re-orientation within the path planning and is applied to mark the arrival and completion of the path planning.

7.4 COLLABORATIVE POINTING

As the participants share an interaction space, one prerequisite for the successful establishment of common ground is the achievement of both shared orientation and focus. Pointing gestures by an individual have already been shown to have manifold functions. By pointing mutually at an entity on the map/task sheet or by tracing collaboratively, the participants display sense making and comprehension processes. According to Goodwin, *collaborative pointing* occurs in

anticipation of somebody else's actions and "constitutes an elegant solution of the problem of how to mutually demonstrate that each participant can independently parse a complex visual field into the specific phenomena that are relevant to the accomplishment of the task at hand" (2003:27). This study supports Goodwin's findings with regard to single points. There are two exemplary instances in the following two excerpts.

MOV00A, lines 13-15

	Name	*Intonation unit*	*Gesture unit*
13.	Anna	there is North,	
14.		so South is [here,]	LIF point to South Gate
15.	AJ	[yeah,]	RIF point to South Gate

MOV00A, lines 159-163

	Name	*Intonation unit*	*Gesture/gaze unit*
159.	AJ	we need to go to the uh Bear Grottos.	circling gesture
160.		(1.0)	circling gesture
161.	Anna	the Bear Grottos	
162.	AJ	which is .. [THERE?]	RIF point to Bear Grottos
163.	Anna	[there] yeah.	RH point to Bear Grottos

In both examples, the single pointing gesture co-occurs with the identification of an entity on the map. In the first case, Anna and AJ search for the South Gate and when Anna points and says "here", AJ agrees and points to the same location. In the second case, the two participants want to visit the Bear Grottos and AJ's circling gesture in the air above the map (lines 159 and 160) visualize the scanning of the map. In lines 162 and 163, AJ and Anna utter "there" in overlap and point at the same time. In both examples, the participants did not only point mutually, but also simultaneously and the collaborative pointing gesture co-occurred with verbal overlap. Their orientation was identical and they agreed with each other in words and gesture.

Goodwin's notion of collaborative pointing, however, must be extended. In this study, it is suggested that collaborative pointing can further occur as an embodied gesture, namely in following and mimicking someone else's gesture. Mc Neill writes,

> Human bodies offer identical possibilities for embodiment of sense and meaning. This is the foundation of mimicry and its role in unraveling the contexts of other speakers. Mimicry is a kind of borrowed embodiment – borrowing significant actions of the other. Gestures are a natural form of such embodiment with language, which makes mimicry a powerful tool for accessing another speaker's meaning" (2008:10).

As such, participants do not only demonstrate that they can parse a field individually, but in fact, it is argued that collaborative pointing occurs to overcome differences in origo and orientation, borrowing the other person's gesture to comprehend their propositions. In other words, by pointing collaboratively, shared understanding and agreement emerge. In addition, *collaborative tracing* here emerges as a special kind of collaborative pointing, which provides information about the state of comprehension and which is found at moments in the interaction where there is a lack of shared orientation and hence, potential for disagreement. In the previous chapter on repetition, collaborative pointing was discussed briefly in the exchange between Ben and Beth. Due to a difference in individual foci, disagreement to a proposed route direction emerged: Ben suggested to turn right, but Beth did not agree with this proposition. However, by following and incorporating Ben's tracing and pointing gestures, a shift in origo occurred and Beth then inhabited Ben's perspective. As a result, the arrival at the Camels was underlined with a mutual and repeated pointing gesture by both participants.

Goodwin considers collaborative pointing a solution to parsing an area individually and independently. However, the process of parsing a visual field can also precede a problematic instance in a task. In other words, collaborative pointing is also the result of a previously completed parsing process which is only visualized when necessary. The next excerpt contains such an example. Wendy uses her left index finger to outline a straight path from the North Gate to the Camels. She receives a positive reply from Wilma in line 59, which occurs in anticipation of the final destination, the Camels (line 60).

MOV00W, lines 58-68

	Name	*Intonation Unit*	*Gesture unit*
58.	Wendy	°we go along [that way,]°	LIF tracing
59.	Wilma	[yeah,]	
60.	Wendy	to the CAMELS?	LIF point to Camels

61.		and then we go to Habitat Africa=,	RIF point to Camels
62.		=so you just continue along that road,	RIF tracing
63.		don't you?	
64.	Wilma	well .. well no,	
65.		cause you gotta go left here?	RIF tracing
66.	Wendy	oh yeah go left there,	RIF tracing
67.	Wilma	here it is.	RIF tracing
68.	Wendy	and go in there ... .h.	RIF tracing

The next stop is the Habitat Africa and at this point, there is a shift in handedness: Wendy's right index finger is put on the map and it touches her left index finger, which was used to trace (line 61). Then there is a shift and she uses her right index finger to trace the path. This gesture accompanies the utterance "so you just continue along that road" (line 62). However, she uses a tag question "don't you", which invites Wilma's feedback and in fact, a hedged disagreeing stance "well no" (line 64) occurs. Wilma must have thought about the route to the Habitat Africa already because the disagreement and alternative route description follow without pauses and hesitations. As a visual aid and an explanation for her disagreement, Wilma traces the path with her right index finger (line 66) and Wendy follows her movement immediately. Wendy has now taken Wilma's perspective, which is also shown in the verbalization of the embodied gesture, demonstrating agreement (see line 66). Collaborative tracing is thus used a correction tool for an earlier planned path and at the same time, it allows for a readjustment of focus and the establishment of a shared focus.

The right index finger trace in line 67 is different from the earlier tracing gestures: Wilma identifies the Habitat Africa in speech "here it is" and usually, one would expect a static indexical gesture directed at this location. However, Wilma outlines the shape of the Habitat Africa, which means that she moves her right index finger along the grey area representing the Habitat Africa on the map. Her right index finger touches the map as she traces the form of the Habitat Africa and this gesture is considered an inscription, produced to emphasize the reason of her previous disagreement. In line 68, Wendy displays an agreeing stance, marking the arrival at the animal exhibit in her utterance and mimicking Wilma's tracing gesture. Agreement and common ground were ratified in speech and in gesture and this is the basis for a continuation of

the task. An embodied gesture as part of the collaborative pointing process can co-occur with verbal elaboration as well.

In the following example, Fiona embodies Flavia's tracing gesture, which means that she borrows the gesture to make meaning of it (see Pereira 2011 for a detailed discussion of this example). However—and this is in contrast to McNeill's statement—as part of the process of understanding the meaning of the other person's gesture, there is verbal elaboration.

MOVOOF, lines 14-40

	Name	*Intonation unit*	*Gesture unit*
14.	Flavia	straight on,	RIF tracing
15.		past the zoo,	RIF tracing
16.		and then go North?	RIF tracing
17.	Fiona		LIF point to M
18.		you'd go through those?	
19.		I won't.	
20.		(°look at that line°.)	
21.		don't you have to go round?	
22.			RIF point to M
23.		uhm,	
24.		((laughter))	
25.	Flavia	uh don't (mind),	
26.		you can arrive through the North Gate as well.	RIF point to North Gate
27.		so that must just be like [the entrance °bit°,]	RIF tracing
28.	Fiona	[oh alright] okay.	
29.		yeah.	
30.	Flavia	you go like that,	RIF tracing
31.		and then like that.	RIF tracing
32.	Fiona	so straight up,	RIF tracing
33.		past the Trading Post,	
34.		past- past the Snacks,	
35.		past the Trading Post,	

36.		... so a bit on the right,	RIF point to M
37.		past the Zoo?	RIF tracing
38.	Flavia	yeah and then [west.]	RIF tracing
39.	Fiona	[to the] fountain,	RIF tracing
40.		and right at the fountain.	

In lines 14 to 16 and then again in lines 30 and 31, Flavia describes the route from the South Gate to the Bear Grottos. The gestures are identical in the first and in the second description, outlining the path up and to the right with the right index finger. A comparison of the intonation units, however, demonstrates that the verbal descriptions in lines 30 and 31 are less precise and less detailed than they are from lines 14 to 16. The gesture form is maintained, even identical, while speech is less specific. In the repeated route description, the visualization is thus more important in order for the observer to understand and to take Flavia's perspective. The repetition occurs due to a problem in parsing the map and a lack of an established shared focus. Fiona's left index finger is on the map while Flavia speaks (lines 14 to 16). In line 17, preceding speech, Fiona points to the South Gate area and verbalizes where she sees a problem (lines 18 to 21). Instead of going through the gate, she suggests to take the path around it and identifies the location with a right index finger point in line 22. Flavia provides a solution by comparison: she identifies the North Gate and traces what she considers to be "the entrance °bit°," (line 27) and these actions receive a positive feedback from Fiona. After the repetition, which was already addressed, Fiona embodies Flavia's tracing movement. The form of the movement is identical; however, it is carried out slower due to the fact that speech is more elaborate. There is a self-correction in lines 34 and 35 and the gesture is paused. In lines 38 and 39, both participants trace the route collaboratively and there is overlap in speech. Thus, a shared focus was successfully established and mutual gestures are now possible.

7.5 Conclusion

In the previous chapters, it has been shown that referential expressions and gestures are used repeatedly throughout the activity solving process. Generally speaking, each type of gesture is attributed a primary function. For example, the primary function of iconics lies in the representation of a concrete object and beats put emphasis on the speech they accompany. In this

view, deictic gestures are defined as indexicals, establishing a relationship to a referent in the surrounding environment. This chapter, however, argued for a multitude of functions of deictic gestures. First of all, a differentiation of various kinds of deictic gestures was proposed: the single deictic gesture, the repeated deictic gesture, and the tracing gesture all fall under the superordinate term *deictic gesture*. These different forms of pointing can be used collaboratively, which means by both participants. The term *collaborative pointing* was borrowed from Goodwin (2003) to include single and repeated deictic gestures. In analogy to this, the term *collaborative tracing* was introduced to differentiate collaborative pointing from tracing, allowing for a separate layer of analysis, and distinguishing a static from a dynamic gesture. Secondly, it has been shown that the different types of deictic gestures occur at different stages in the task planning, thus fulfilling various functions. The investigation yielded the following results: all forms of deictic gestures can occur alone or they can co-occur with speech. There are variations to these patterns: one interlocutor might speak and point at the same time; speech and gesture can also be distributed, namely a gesture can precede or follow one's own speech or a gesture can accompany someone else's speech.

Single pointing gestures occur throughout the planning process, at the beginning of a new unit, in the middle as well as at the end of a unit. The analysis demonstrated that they correlate with names of entities on the map or demonstrative pronouns such as "here" and "there". In this respect, they incorporate the function of indication of an entity. In contrast to this, tracing frequently co-occurs with lexical items and verbs expressing movement, for example *go round, go past, come down*, and so forth. Due to the limited dataset, a quantification would not yield significant results. Future research which is based on a larger amount of data, however, could quantify the results of the present study. Nevertheless, tracing entails the meaning of movement, and as such, it is a dynamic gesture. A single point, however, lacks this aspect of movement, representing a static gesture, especially when a single point is maintained.

Tracing gestures are carried out in the air above the map as well as with the pointing finger touching the map; as a variation to this, participants use the pen leaving visible marks on the map. This is considered an inscription. An inscription has the meaning of a lasting component as it is visible throughout the activity. It is the result of a grounding process, a result of a negotiation that led to agreement. Thus, deictic gestures can be differentiated for gesture

space, depending on the distance or closeness of the pointing finger and the map/task sheet. Tracing gestures can also take the form of semi- and complete circling gestures.

Repeated pointing is a means of reinforcement and emphasis. It occurs as a result of a search process, namely to highlight the end of the search. In addition to this, repeated pointing is found at the end of a unit, for example to mark the arrival at the final destination. Hence, repeated pointing correlates with the completion of a process, especially when this process needs special emphasis.

In the previous chapters, different means of how the two interlocutors solve the task have been addressed. In the present chapter, pointing and tracing in collaboration adds another layer to this crucial aspect of the task management. Not only do the two participants demonstrate that they can perform the task independently, but more importantly, they engage in a social encounter and their actions underline how they coordinate the task. In collaborative pointing and tracing, we can find the visualizations of a sense-making and interactive process, as a mutual focus is established in a shared interaction space. Using the available resources, different patterns emerge as part of the grounding and stance-taking process. Gesture recurrence and mimicry establish a shared focus, facilitating comprehension and resulting in common ground.

8 EYE GAZE

Commenting on the scholarly discourse on gaze, Kendon writes that "most of the work on gaze in interaction has concentrated upon measuring the amounts of looking one interactant addresses to another and how this is correlated with various psychological and situational variables" (1990:89). Recent studies continue to be located in the field of psychology and much research is conducted in the computer sciences, placing a focus on human-robot interaction and using eye-tracking experiments, for instance. This study and this chapter in particular, address the phenomenon of eye gaze from a linguistic perspective. A relationship between gaze and other phenomena, especially speech and gesture, is established, presenting gaze as another meaningful resource for the organization of joint action. Gaze is present everywhere and we all look at things all the time. In order to be able to investigate eye gaze, the term *gaze* must first be conceptualized.

Cook defines gaze as "looking at another person in or between the eyes, or, more generally, in the upper half of the face" (1977:328). Since an eye-tracking device was not used, this study cannot account for the exact location of the gaze direction, whether it is in the eyes or in between the eyes of someone else. In the following, eye gaze at someone is only taken into consideration when there is a visible movement of the eyes directed at the upper half of the body and the face of the other person. Sometimes gaze is also accompanied by head movement. Due to the seated position of the participants, their looks are directed at the upper half of the body, usually at the face, of the other participant. By contrast with Cook's understanding of gaze, it is proposed here that the target of gaze is not necessarily a person. People look at objects in their surrounding and often they attend to multiple visual fields in their environment. In the present context, *gaze direction* is differentiated for three targets: 1) gaze can be directed at the other participant; 2) gaze can be directed at the map; 3) gaze can be directed at the task sheet. In addition to *gaze direction*, there is *gaze shift*, which includes shifts from the map to the task sheet and vice versa, or gaze shifts between one of the external representations and the other person. Finally, there is *mutual gaze*, which is generally defined as establishing eye-contact with another person (Cook 1977). In addition, there is gaze, which is directed at a shared interaction space and an object in this space. In the present context, the map and the task sheet are such objects, lying in the shared interaction space in which gesture and gaze are performed. This type of gaze will

be labeled *shared gaze focus* and it indicates that the two participants jointly attend to the map or the task sheet. In order to disambiguate between two items, speech, gesture, or both resources together are applied as a means of specification.

Kendon (1990) explains that gaze aversion and ending of mutual gaze regulates the emotional arousal in social encounters. *Cut-off gaze*, it has been argued, is due to embarrassment, shame, and aggression (see for example Argyle and Cook 1976). Gaze avoidance is present in an example from MOV00B. The instructions state that the participants would like to see the "Dolphin Show". Ben and Beth must find the icon and the name "Dolphin Show" on the map. However, they have difficulty finding it and Beth says "I don't think it's on here" (MOV00B, line 151). Ben immediately questions Beth's statement by saying "cause it wouldn't be" (line 152). After Ben has found the icon for Dolphin Show on the map, he recycles Beth's utterance, saying "it's not on here" and looks at Beth. Beth, however, does not reciprocate his gaze presumably because she perceives Ben's utterance as a face-threat.

There are certain patterns of when to look, for how long, and also reasons why people look at each other. Goffman (1963) attributes a crucial role to gaze direction as a means to initiate and maintain social encounters. Goodwin (1980) and Bavelas et al. (2002a) argue for asymmetrical gaze patterns, which means that speakers, on the one hand, look frequently, but for short periods. Listeners, on the other hand, look relatively long at the speaker (see also Kendon 1967). Bavelas et al. (2002a) record unacquainted people who are asked to tell close-call stories. As a result of the analysis, they state that speakers usually seek listeners' attention. Once they have received the listeners' feedback, the telling of the stories is performed more fluently. There is a "gaze window" (Bavelas et al. 2002a:577), in which the individual roles remain unchallenged. By deploying gaze as a response, for instance, listeners signal attention without competing for the turn. In natural speech, Goodwin (1980) argues, speakers produce restarts and pauses especially at turn-beginning. Looking at the gaze behavior of participants in these conversations, Goodwin states that there is an increase in coherent sentences when speakers receive listeners' attention. He establishes rules for different gaze patterns, maintaining the asymmetrical nature of gaze behavior. According to him, there is a preference for speakers to obtain listeners' gaze when they are looking at them. Listeners should look at speakers when they are gazing at the listeners. In a more recent study, Haddington (2006) presents gaze as a central element of communication. Based on conversations among friends, he attributes three

functions to gaze: first, gaze and verbal assessments can be a resource for stance taking. Interactants co-participate and elaborate on talk by looking at a present visible assessable, for instance. Second, when interlocutors produce an agreeing second stance, they engage in mutual gaze. Haddington explains that these agreeing actions are often realized by recycling linguistic material and structures from the previous speaker, such as "It's beautiful" followed by an agreement "Yeah, it's really… It's very pretty" (2006:300-303). Third, gaze shifts or cut-off gaze, can express a divergent stance, such as looking away to signal disagreement. Research on gaze in relation to other behavior is not frequently reported on and much research has been based on conversational or narrative data. However, much of the talk people produce at the workplace, as a member of a club, or even at home, centers around an activity. Attention is frequently directed at the activity, which is being performed, for example cooking, driving, or looking at a computer screen. People perform these activities and focus on them even while talking to other people. Gerhardt (2007), for instance, studies gaze behavior of football fans who are engaged in the activity of watching a football game on television. The neutral gaze target in this setting is the television because subjects direct their gaze at this object to follow the game. One trigger of gaze shifts, according to Gerhardt (2007), is humor. By looking at the other person, the football viewers establish and retain a joking frame. In her study of gaze organization, Tiittula (2007) focuses on a single interaction between a sales representative and two customers. The interactants are involved in a business transaction, looking through a catalogue to find articles to buy. The triad stands in a side-by-side position. There is a hierarchical structure because the expert, the sales person, presents relevant articles to the customers. As in the present study, there is an object, the catalogue, which becomes central to the interaction. The participants' roles, however, are not pre-defined by the event, in contrast to the sales event described by Tiittula. The roles in the present study, it has been shown, emerge in the course of the interaction. The participants' body postures can shift from face-to-face to side-by-side and vice versa.

This study presents an approach to gaze work in relation to speech and to deictic gestures from a linguistic perspective. It is proposed that participants' gaze patterns differ from the asymmetrical pattern described for conversations and story-telling because much attention is directed at the task, the task performance, and the achievement of the task at hand. Gaze direction, gaze shifts, and mutual gaze, it is argued, enable interactants to position themselves in a joint activity and they can demonstrate their attention states via their gaze work. It will be

demonstrated how gaze patterns are organized, how and when they are employed with other verbal and non-verbal features to co-ordinate the task and the interaction. In the previous chapters, it has been shown that gesture and gaze often co-occur to create F-formations; when participants point to an area collaboratively, they also direct their gaze to this area. The map and the task sheet provide a natural site for gesture and gaze behavior and in the present study, much of the focus lies on these external representations. In Chapter 4, different activity roles have been established as they emerge at the beginning of the interaction. It has been indicated that gaze and attention states differ depending on the distribution of the roles. In the following, gaze patterns will be differentiated for the different settings, with distributed roles, on the one hand, and maintained equal roles, on the other hand. Some general findings about gaze will be presented before gaze patterns in relation to activity roles will be investigated.

8.1 SOME GENERAL FINDINGS ON GAZE

In studying gaze in task-based interactions, one of the first questions one must ask is whether gaze work is similar to or different from conversation, narrative, or experimental, computer-aided communication. If gaze patterns vary from other discourse genres, the second question to answer is how these patterns vary. Generally speaking, the participants address most of their attention to the map and the task sheet and, as such, the pattern differs from face-to-face oriented encounters. At the beginning of the sessions, participants usually look at each other. To begin the task, some dyads address questions to one another, such as "Should we read the instructions?", while other dyads start nonverbally. In all cases, there is a shift from a face-to-face orientation toward the task sheet and later to the map. The beginning of the activity is thus accompanied by a re-orientation and a gaze shift to the interaction space. The importance of the participants' visual orientation and the visual access to the reference space has also been demonstrated in Chapter 7 with respect to deictic gestures. Especially when speech is not present, a joint focus to the map is essential in order to see what is being pointed at and in order to comprehend what these gestures mean in relation to the task performance as a whole.

As a phenomenon of human-human interaction, people can perceive things in their surrounding without directly looking at them. Both participants are in close proximity of one another so that they can perceive behaviors, for example shifts in body posture or gestures directed at them, without looking up. For instance, in MOV00C, there is a para level reference

"you choose which way should we go", which is accompanied by a gesture, arm and hand outstretched, pointing at the other person (see also Chapter 5, p. 75). However, this gesture is not accompanied by a look directed at the other person, and the addressee does not look up either. The utterance and the gesture suffice to identify the referent of the pronoun "you" and to mark the action as relevant to the task. The para level reference is followed by a path description (see also Chapter 7, pp. 116-117), which is carried out by Cloe. Both participants look at the map while Cloe outlines the path. Since there are no objections or interruptions coming from Clare—she offered the floor to Cloe—Cloe continues the path planning and the shared focus is maintained. Only at the end of the unit, when the Fragile Kingdom is reached, is there a gaze shift by both participants, directing their attention to the task sheet. Again, there is no gaze re-direction of one or both participants at each other. This indicates that when the activities are performed smoothly and in agreement, namely when common ground is assumed, gaze and mutual gaze are not a necessary requirement for the successful completion of the task.

Gaze direction toward one of the external representations is frequent. Gaze shifts from the map to the task sheet regularly occur at the end of a unit, generally by both participants, but sometimes also just by one, for example when roles are distributed. Gaze shifts also occur in mid-activity, mainly to re-read parts of the instructions. Gaze shifts fulfill a control function to ensure that the activities are performed according to the instructions. In Chapter 5 (pp. 79-80), an example was adduced where Quintina shifts her gaze between the map and the task sheet before she looks at Queena (MOV00Q, lines 46 and 47). There is a difference in individual orientation and the two participants focus on two different entities on the map. Quintina's gaze behavior demonstrates that she is comparing the instructions to their current location on the map. She monitors Queena's path description and, thus, her gaze work has a control function. These patterns are in part influenced by the nature of the experiment. However, Gerhard, for instance, presents similar findings for a media reception situation, during which there is "a lack of gaze over long stretches of talk" (2007:98). In the activity-based setting, the asymmetrical gaze pattern, which has been described for speaker-listener interactions, is frequently suspended. Gaze shift occurs between the map and the task sheet; gaze redirection toward the other interactant, however, is lower in frequency. Differentiating the two forms of organization, egalitarian and non-egalitarian, will further supplement this claim.

8.2 GAZE PATTERNS IN INTERACTIONS WITH DISTRIBUTED ROLES

Interactions with distributed roles are those forms of interactions where one person is the route planner and the other person is the instruction reader and sometimes even the note taker. With regard to natural conversation, Goodwin's analysis (1980) shows that the speaker wants to obtain the listener's gaze in order to continue with his or her turn. If one applies this rule to the current setting and to interactions where there is a route planner and a note taker, the gaze pattern should be the following: the route planner will shift his or her gaze toward the note taker to wait for a signal that he or she can continue with the planning. This signal could be a verbal response token, for example "okay", or a nonverbal signal, such as a gaze shift up toward the speaker. However, taking an example from MOVOOK, Goodwin's rule does not apply to the scenario. In the transcript, gaze shifts are indicated, which means that one looks away from the map and the task sheet respectively to look at the other sheet. In the whole passage, there is not a single visible gaze shift which is directed at the other participant. The targets of both participants' gaze are the map and the task sheet. Kara is the note taker and Karin the route planner. They go from the North Gate to the Camels and continue to go to the Habitat Africa.

MOV00K, lines 25-51

	Name	*Intonation Unit*	*Gesture/gaze unit*
25.	Kara	the [camels .. first.]	
26.	Karin	[the camels first] yeah,	gaze shift to TS
27.		so as you go in,	gaze shift to M, RIF point
28.		turn right,	RIF tracing
29.		(1.0)	gaze shift to TS
30.	Kara	°okay°,	gaze shift to TS
31.			gaze shift to M
32.			takes notes
33.	Karin	what was it there?	
34.		Habitat Africa,	gaze shift to M
35.	Kara		takes notes for 7 sec.
36.	Karin	°just go in,°	RIF point to North Gate

37.		°turn right?°	RIF point
38.	Kara		gaze shift at M
39.	Karin	uhm,	
40.	Kara	Habitat?	LIF point to M
41.	Karin	it's down there,	RIF point to Habitat Africa
42.		isn't it?	
43.	Kara	yeah,	
44.	Karin	u:h carry on walking,	RIF tracing
45.		then turn left?	
46.	Kara		gaze shift to TS; takes notes for 5 sec.
47.	Karin	.. and then the Habitat,	gaze shift to TS
48.		is on the right hand side.	
49.			gaze shift to M
50.	Kara	is it turn LEFT?	
51.	Karin	YEAH YEAH.	gaze at M

The first gaze shift from the map to the task sheet occurs in line 26, when Karin takes an agreeing stance with Kara's utterance in line 25. In order to agree with her, Karin has to look at the instructions as well. As she begins to plan the route to the Camels, her gaze shifts back to the map and she outlines the path with her right index finger. In line 29, there is a verbal pause during which Karin looks at the task sheet and as one can see in lines 32 and 33, this is to find out about the next stop. This is necessary because Kara, after a gaze shift between task sheet and the map (lines 30 and 31) is now taking notes, which means that she cannot read the next line on the instruction sheet to Karin. While Karin is a step ahead in the interaction, looking for the next entity (lines 32 and 33), Kara is noting down what has previously been planned. At this point, there is an asymmetry in the interaction and the individual visual attention of each participant differs as well. This is a result of the distributional organization, which emerged at the beginning of the interaction and which is now maintained. After a pause of seven seconds, during which Kara is still taking notes, there is a verbal repetition of the initial route planning by Karin. She points to the map twice (see lines 35 and 36). Kara has now completed the task of taking notes and thus redirects her focus to the map. The difference in the participants' attention is now

resolved. Kara's left index finger points to the map in the air while she says "Habitat?" with a rising intonation (line 39), demonstrating that she is attending to the next step in the activity. Her gesture triggers an identification of the entity by Karin, who has had the time to look for the item on the map while Kara was still taking notes. While Karin plans the route in lines 43 and 44, Kara also looks at the map. Here, the listener's attention is focused at the route planner's actions, allowing for comprehension and the opportunity to assist in the planning or to correct, if necessary. Once Karin has finished, Kara returns to taking notes. After about five seconds, Karin completes the description to the Habitat Africa and when she says "Habitat" in line 46, her gaze shifts to the task sheet to monitor Kara's actions. After the completion of her utterance in line 47, she shifts her gaze back to the map and it remains focused to the map. In lines 49 and 50, there are two opportunities for both participants to look at each other. However, Kara's gaze is directed at the task sheet, she focuses on the note taking activity, and Karin's gaze remains on the map. Even though Kara addresses a question at Karin, she does not look at her. And Karin does not look up to receive the question, nor does she shift her gaze when she replies "yeah" twice, putting emphasis on both tokens.

Goodwin's (1980:275) rule, which he formulates for natural speech, namely that the speaker wants to obtain the listener's gaze during his/her turn, is not fulfilled in the present interaction. Rossano et al. (2009) report on different cultural practices in the utilization of gaze in questions sequences. Gaze is both established and sustained when a speaker addresses a question at a recipient (Rossano et al. 2009:193). The results of their study indicate that there are cross-cultural differences, especially with regard to signaling recipiency of questions. Frequently, listeners do not gaze back at the speaker, even though the speaker is looking at them (Rossano et al. 2009:213). The present example demonstrates that the orientation toward the map and the task sheet is more important than the orientation toward the other person. Gaze is sustained by both interactants. The close proximity of both interactants might be one reason; another reason is the distribution of the external representations because there is an unspoken agreement of who has control over which one of the two sheets. As a result, there is a distribution in activity roles and these roles are maintained, which means that participants do not compete for each other's roles or speaking turns.

Goodwin's second rule, which states that "[a] recipient should be gazing at the speaker when the speaker is gazing at the hearer" (1980:287), is not applicable to a setting like the one

above either, where the roles are distributed during the activity. The speaker does not try to establish eye contact with the listener because the listener's actions are visible through other resources. By looking at the task sheet, the route planner is informed of the present state of the note taking activity, she monitors the process, and she can perceive when to wait and when to continue. This is seen in the two longer pauses: there is no speech, but the task is still in progress. The distributional setting is successful and acceptable for both participants. The speaker can monitor whether her actions are effective without receiving immediate feedback in the form of a gaze response from the note taker.

For Goodwin's second rule, egalitarian interactions will be considered since the assumption is that with equal roles, there will be more opportunities to step out of the planning process to look at the other person. If this is the case, then gaze redirections should occur with frame shifts, namely they should be directed at the other interlocutor whenever there is "off-task" commentary. So far, the data suggest that gaze patterns vary depending on the interactional organization of the activity.

8.3 GAZE PATTERNS IN INTERACTIONS WITH EQUAL ROLES

In the following, I will focus on egalitarian interactions to first account for gaze re-directions addressed at the other participant and second for mutual gaze, namely both participants looking up to look at each other rather than maintaining a shared focus to the interaction space with the map and the task sheet. Gaze patterns will be differentiated for agreement and disagreement.

8.3.1 Gaze and agreement

The next example was presented in Chapter 6 and it will be revisited in this chapter because it reinforces the claim of a close-knit relation between speech, gesture, and gaze. Not only does the excerpt contain repetitive structures in both speech and gesture, but there are also meta references which are accompanied by gaze shifts. It was stated earlier that certain issues and important topics are re-introduced in the interaction as part of the acceptance-process and as a means to ensure mutual agreement in order to reach common ground. By looking at each other, the participants further demonstrate their cooperativeness. The achievement of mutual gaze is indicated with the symbol "←" in the transcript.

MOV00O, lines 294-318

	Name	*Intonation Unit*	*Gesture/gaze unit*
294.	Olga	and return to the North Gate at 4 p.m.	gaze at TS
295.	Olivia	so we've gone past the Special Events Arena,	RIF tracing, gaze shift to M
296.	Olga	yeah?	
297.	Olivia	.. to the butterflies,	RIF tracing
298.		and then BACK OUT?	RIF tracing
299.	Olga	so perhaps next time,	LIF circle on M
300.		we could visit that bit.	gaze at Olivia
301.		[((laugh))].	
302.	Olivia	[((laugh))].	gaze at Olga ←
303.		we were missing out on all the giraffes [and stuff?]	
304.	Olga	[YEAH?]	
305.		we did that stuff,	LIF tracing
306.		didn't [we]?	
307.	Olivia	[o:h?]	
308.	Olga	but never mind.	
309.	Olivia	that's the busy side,	RIF tracing above map
310.	Olga		gaze at Olivia
311.	Olivia	that's [why:?]	gaze at Olga ←
312.	Olga	[yeah.]	gaze shift to M
313.	Olivia	busy ((giggle)),	
314.	Olga		gaze at Olivia
315.	Olivia	busy that we can't even get out of?	
316.		(1.0)	
317.		alright should we say?	
318.	Olga	yeah.	

At the beginning of the transcript, there are several gaze shifts from the map to the task sheet. Olga reads the instructions (line 294) and starting in line 295, Olivia plans the path back to the

North Gate to complete the day. What follows the completion point is a meta-reference by Olga (lines 299 and 300). When Olga says "that bit", she shifts her gaze from the map to Olivia and her suggestion for the next visit is framed with laughter. Olivia reciprocates this playful frame in three modalities: she laughs and she establishes mutual gaze (line 302). Then, in line 303, she elaborates on Olga's statement, expressing a positive uptake of Olga's utterance. Mutual gaze is again established in lines 310 and 311, initiated by the listener Olga and thus following the rule stating that the listener should look at the speaker shortly before the speaker looks at the listener. In contrast to the claim made by Bavelas et al. (2002a) that a listener's gaze response terminates the speaker's gaze, however, the listener Olga shifts her gaze back at the map while she takes an agreeing stance "yeah" before Olivia looks away. In line 313, Olga looks at Olivia one more time, but does not receive a gaze response from her interlocutor, whose gaze is wandering around the room. Eye contact was initiated and reciprocated twice; in both cases, mutual gaze was a result of a meta-referential utterance and it co-occurred with gestures outlining an area and a path (lines 299-300 and 309). Thus, is both cases, the units are highly marked, drawing all three modalities simultaneously.

Speakers establish gaze with the listener to invite a reaction and this is often the case when there is an "off-task" comment, for example in the form of a suggestion that is not presented in the instructions. "Off-task" commentary can include, but is not limited to, meta and para level references. Any kind language which is above the level of the direct task-related language can account for off-task comments, for instance when attitudes and values are expressed. Stance-taking activities in the form of evaluations and assessments have been investigated earlier. Such evaluative comments have been shown to relate to previous statements, taking a stance to an utterance, and to the task performance itself. Such contexts, it is argued, trigger gaze; in other words, gaze should accompany evaluative, personal, or humorous comments, marking that something special is happening. In the following excerpt, Anna initiates such a playful frame and adds events to the day at the zoo, which are not provided in the activity description.

MOV00A, lines 35-45

	Name	*Intonation unit*	*Gesture/gaze unit*
35.	Anna	should we go for a drink in the pub first?	

36.		((laughter))	gaze at AJ
37.	AJ	((smiling)) yeah,	
38.		[I think we should.]	
39.			gaze at Anna
40.	Anna	[((laughing)) and then have something to eat?]	gaze at AJ
41.	AJ	[yeah yeah]	
42.		relax [a bit.]	gaze at M
43.	Anna	[go for] a drink in the pub.	gaze at M
44.	AJ	yeah .. we want a good day out.	gaze at Anna
45.	Anna	yeah?	gaze at M

In line 35, Anna asks a question which, in fact, deviates from the instructions. It can be considered an off-task comment. She makes a suggestion which relates to a real-life event, something one might do during a zoo visit. The question functions as a suggestion and it is framed by following laughter. Anna now looks up from the map to look at AJ and to obtain a response during a potential moment for disagreement. Since her question is not related the task performance per se, AJ could, potentially, take a disagreeing stance. However, he provides a positive facial expression, he is smiling, and he agrees verbally with "yeah". In line 38, AJ elaborates on the agreeing stance and expresses his opinion, framed by the epistemic stance marker "I think". This utterance is followed by a gaze shift, which results in mutual gaze. Both Anna and AJ are now looking at each other, while their speech overlaps. This signals a high involvement and affiliation by both participants. AJ adds another aspect "relax a bit" to the conversation in line 42, before Anna repeats her initial question in line 43. This time, she formulates a statement "go for a drink in the pub". Since AJ has positively evaluated her suggestion, she can now make this statement based on common ground. Thus, it is not necessary to establish mutual gaze at one another again at this point in the interaction. In fact, both participants direct their attention to the map. Toward the completion of the unit, AJ provides another meta-reference, including both Anna and himself via the personal pronoun "we". He gazes at Anna to signal further alignment. Anna agrees verbally, but does not reciprocate AJ's gaze. This is probably due to the fact that they have reached common ground earlier, on the one hand, and that they are about the continue with the task, on the other hand.

In sum, gaze was initiated by the speaker at a point when she deviated from the instructions and this required the listener's feedback. AJ, the listener, understood and considered Anna's suggestion and together with the initiation of eye contact, mutual gaze was established. The overlapping speech, moreover, expressed agreement and a convergent stance. Building on this, common ground was reassured, for example via verbal elaborations, and maintained until the end of the unit.

8.3.2 Gaze and disagreement

The first two examples of egalitarian interactions present instances where the dyads maintain a friendly working atmosphere, their meta utterances are accompanied by gaze, and their actions are based on agreement. The next two examples present instances in the interactions where there is disagreement and the participants must disalign with one another. In the first excerpt, gaze is initiated by speaker. A hedged form of an agreeing stance precedes the actual disagreement. Laura and Larissa plan the path from the Bear Grottos to the Fragile Kingdom. After some confusion about entrance and exit of the Bear Grottos, Laura repeats the previous route description from lines 63 to 66.

MOV00L, lines 63-75

	Name	*Intonation unit*	*Gesture/gaze unit*
63.	Laura	yeah you come out from here,	
64.		then you have to turn left,	
65.		back to the fountain,	
66.		(1.0) <and then take a right?>	gaze at Larissa
67.	Larissa	yeah you could do that,	gaze at Laura
68.	Laura		gaze at Larissa
69.	Larissa	or you could go uhm,	
70.	Laura		gaze at Larissa
71.	Larissa	and hav- come out of the bear's grotto,	
72.		and have uhm [lu- lunch at Safari Grill Restaurant],	gaze at Laura ←
73.	Laura	[((giggle))].	

74.	Larissa	which I would do ((giggle)).
75.		and THEN go to ... the Fragile Kingdom.

In line 66, Laura closes the route description with a glance at Larissa. Her utterance is spoken more slowly than the preceding utterances and is ended with a rising intonation. Previous to this, there was confusion and discussion about the route, and this uncertainty is transported to Laura's utterance in line 66. She looks up to receive feedback from Larissa, who reciprocates with a brief glance at Laura when she says "yeah" (line 67). Then she shifts her focus back to the map immediately. Taking speech into account, one can see that she is about to present two alternatives. She uses the phrasing "you could" to imply that Laura's proposition is only one possibility. She does not express complete agreement in her utterance in line 67. The differing stance is expressed from lines 69 onward and the alternative is introduced with "or". There is another gaze shift at Laura in line 72, which accompanies a meta level expression, suggesting to have lunch. Laura's reaction, a giggle, is a positive evaluation of the suggestion and, thus, Larissa reinforces her opinion with a para level reference "which I would do" in line 74. The unit is completed with Larissa's utterance in line 75, marking the arrival at the final destination. Mutual gaze is established once in the interaction and the duration is very brief. Once Larissa has taken the role of route planner, Laura displays the "typical" listener behavior according to Goodwin's (1980) rule, looking twice at Larissa. The interaction between Laura and Larissa demonstrated a case where the disagreeing stance was presented in a softened form, acknowledging Laura's contribution, but providing an alternative, personal preference.

In the following excerpt, Tamara and Tom disagree about a path description which Tamara proposes in lines 439 to 441. She traces the path twice, first up and then back down. In line 444, Tom makes an alternative suggestion and outlines the route on the map as well. Starting in line 445, there is a shift in frames, from the actual activity to a playful key.

MOV00T, lines 438-452

	Name	*Intonation unit*	*Gesture/gaze unit*
438.	Tamara	so we gotta find the quickest route.	
439.		so it'd be this,	LIF tracing
440.		°wouldn't it?°	LIF tracing
441.		past .. the Rice Centre.	LIF point to Rice Center

442.	Tom	would you?	
443.	Tamara	I think so.	
444.	Tom	wouldn't it be quicker .. going that way.	RH tracing
445.		are we gonna have our first row.	
446.	Tamara	[((giggles))].	
447.	Tom	[((giggles))] are we married by the way?	RH repeated point at Tamara
448.		[((laughs))].	
449.	Tamara	[((laughing))] probably.	gaze at Tom
450.	Tom	((laughing)) we definitely gonna have a row then.	
451.	Tamara	you can walk?	RIF point at Tom
452.		I'll get the train.	

Tom's utterance "are we gonna have our first row" (line 445) relates to the preceding discourse in which both participants disagree with one another. Tom adopts a playful and humorous key as a form of appeasement. The frame shift is also marked by giggles and laughter. Tamara's reaction in line 446 is accompanied by a shift in posture; she leans back, but keeps the visual focus on the map. When Tom says "are we married by the way" (line 447), his gaze remains directed at the map as well. Contrary to the expectation that this meta-task utterance should be accompanied by a gaze redirection, the focus of both participants' attention is on the area where the map lies. However, Tom resorts to a different modality to receive feedback from Tamara. His meta-utterance in line 447 is accompanied by a repeated pointing gesture with the pen. Since his right hand and Tamara's left hand are in close proximity, he moves the pen back and forth three times to touch Tamara's hand. The movement and the physical contact trigger a gaze shift and Tamara now looks at Tom. This is perceived as a re-affirmation, demonstrated in the partial repetition and the emphasis "definitely" in Tom's utterance in line 450. In lines 451 and 452, Tamara further elaborates on the idea of a married couple having a row. She looks at the map, and parallel to Tom's gesture, she uses her right index finger to point at Tom when she says "you can walk" (line 451). The exchange is completed after Tamara's utterance in line 452 and both participants now shift back into the task performance proper.

There are two important implications resulting from the analysis of egalitarian interactions. Firstly, both gaze redirection and establishment of mutual gaze co-occur with comments above the task-level. Off-task commentary prompts frame shifts and functions as an invitation to leave the state of joint attention, which is directed to the task. However, verbal prompts can occur without gaze shifts. In these cases, the focus remains on one of the external representations, whereas speech relates to the level above the task proper. In general, off-task commentary provides an opportunity for gaze redirections, there is an opportunity to establish mutual gaze. However, it is not a necessity for gaze redirections to happen. One possible explanation lies in the availability of other resources, namely speech and gesture, to mark frame shifts. Pointing gestures are used to identify a referent, for example the other person, and due to the participants' position and their close seating arrangements, eye gaze is not mandatory. In meta and para settings, speech and gesture take a primary role, whereas gaze fulfills a secondary role. Gaze shifts might occur, but the focus can also remain addressed to the task level. Secondly, a comparison of interactional forms with and without mutual gaze demonstrates that there is variation in the levels of cooperativeness and positioning. When there is awareness of a shift to the meta-level and when a positive stance is taken toward this, then gaze at one another is mutually achieved. However, the stance-taking activity sometimes includes disagreement, for example when participants discuss certain steps in the activity. At these moments in the interaction, there is no shared common ground and this is expressed in eye gaze. Participants not only look away, as Haddington (2006) reports, but in fact, they do not look at all. In order to collectively accomplish the task, individual activities are finely tuned and different resources are available to shape the coordination of the task. There are phases within the interaction, from the beginning of the task to end, which are marked by certain behaviors as well. At the start of the interaction, the distribution of the map and the task sheet plays an important role. Toward the end of the interaction, a summary of route descriptions is a means of highlighting the end of the activity.

8.4 GAZE, REPETITION, AND ENDINGS

The closings of the activity were studied earlier in relationship to repetition and meta level references. In repeated, summarized verbal route descriptions, there is often a decrease in gesture. Eye gaze will now be taken into account as another feature in the investigation of

repetition and interaction closings. Gaze has been studied at the beginning of a conversation (Goodwin 1980) as well as over stretches of discourse (Bavelas et al. 2002a). However, dyadic gaze patterns with regard to repetition within one interaction as well as gaze behavior while closing the activity has not been investigated in earlier research. I will adduce an exemplary interaction, portraying gaze patterns with repetition as well as endings of interactions. Anna and AJ are in the middle of the task planning and they are on their way to the Pachyderm House and pass other animal exhibits along the way.

MOV00A, lines 159-170

	Name	*Intonation unit*	*Gesture/gaze unit*
159.	Anna	UP we go past Feather and Scales?	
160.		.. up past Tropic World?	
161.		and Primates Journey?	
162.		I think we'd- we'd probably get a bit lost there,	gaze at AJ
163.		cause we get mesmerized by the u:h monkeys,	
164.			gaze at M
165.	AJ	definitely .. yeah °I was- I was a little-	
166.		[I wanna go to Baboon-]	gaze at Anna
167.	Anna	specially [BABOON Island]	gaze at AJ ←
168.		((laughter))	
169.	AJ	yeah I always wanted to go there.	gaze at Anna
170.	Anna		gaze at AJ

Anna plans the path and inserts comments above the task level in lines 162 and 163. The frame shift is indicated by a shift in body posture and gaze directed at AJ. She looks at him for the whole duration of her utterances in lines 162 and 163 before she looks at the map again. AJ does not reciprocate Anna's gaze yet, but still orients toward the map. However, he agrees verbally in line 165 and gazes at Anna when he expresses what he would like to do (line 166). In overlap with AJ's utterance, Anna contributes to the interaction and establishes mutual gaze. Both participants take a convergent stance and AJ further elaborates on the idea of going to Baboon Island. He inserts a para level reference and provides information about his personal background "I always wanted to go there". AJ and Anna now look at each other, they have stepped out of the

activity proper and they face each other. The return to the activity and to the task-level is signaled by a re-positioning—both of them lean in—as well as a resulting joint re-orientation to the map. Toward the end of the interaction, Anna describes the path to the South Gate. As part of the description, the participants pass places which they have visited before, including the Primates Journey.

MOV00A, lines 230-237

	Name	*Intonation unit*	*Gesture/gaze unit*
230.	Anna	we come BAck up towards the butterflies,	
231.		and return to the South Gate?	
232.	AJ		gaze shift to TS
233.	Anna	all the way back down?	
234.		ROund?	RH tracing
235.		through the Tropic World and Primates Journey?	RH tracing
236.	AJ	we like the monkeys?	gaze at Anna
237.	Anna	RIGHT through the SWAMP?	RH tracing

When Anna says "through the Tropic World and Primates Journey" (line 235), she repeats two entities, which they have visited earlier (see lines 160-161). In their previous planning, they paid special attention to the Primates Journey and the Baboon Island. There are monkeys in both animal houses, which are situated across from each other. AJ refers back to the previous event in which they agreed that they would spend some time at Baboon Island. He says "we like the monkeys", establishing a cohesive link to a prior discourse unit. He includes Anna and himself in the utterance and this is based on the common ground they have reached earlier. AJ's utterance is accompanied by a gaze shift at Anna, who does not reciprocate his look at this time in the interaction. Presumably, after agreement has been reached, there is no necessity to look at the other person again. Tiittula (2007) further explains that continuous eye contact is not always required as it sometimes delays key activities relevant to a task. Anna's main goal is to describe the path back to the South Gate. She terminates the activity, the planning itself is completed with the utterance "that's it" (line 246), but the interaction continues after the task completion.

MOV00A, lines 244-256

	Name	*Intonation unit*	*Gesture/gaze unit*
244.	Anna	back to South Gate.	
245.	AJ	hurray.	gaze at Anna
246.	Anna	°that's it.°	
247.	AJ	what a day we just had.	gaze at Anna
248.	Anna	((laugh))	
249.	AJ	((laugh))	
250.			turns around to door
251.	Anna	[too much walking?]	
252.	AJ	yeah .. well	gaze at Anna
253.	Anna	not enough .. bloody riding?	
254.	AJ		gaze at M
255.		we just should have stayed at the pub all day.	gaze at Anna
256.	Anna	yes we should have.	gaze at AJ ←

The end of the activity is expressed twice. In line 244, Anna says “back to the South Gate”, which is the last item on the map, and she puts emphasis on the ending by saying “that’s it” in line 246. She terminates the activity proper. AJ’s contributions “hurray” (line 245) and “what a day we just had” (line 247) evaluate the whole of the activity and they function as a coda to the actual end of the experiment session. AJ’s talk about the task opens a possibility to continue the interaction. Both evaluations co-occur with a gaze shift toward Anna. Anna laughs and turns to her left to reach for her bag. She does not look at AJ and thus he perceives that the interaction is finished. When he turns around toward the door, there is a delayed response to his utterance. Anna takes an evaluative stance as well, stating her feelings about the day at the zoo (line 251). Her reply triggers a gaze response by AJ, who returns to the interaction. His body orients back to the interaction space and his gaze even returns to the map. This co-occurs with Anna’s utterance “bloody riding” (line 253). In lines 255 and 256, there is mutual gaze. AJ recycles one of Anna’s early suggestions about going to the pub. He establishes a connection between two phases of the activity, beginning and ending. The synchrony with which the participants look at each other at the end of the interaction demonstrates their mutual awareness and understanding that the session

is now completed. Generally speaking, there should be mutual gaze at the end of the task. A lack of mutual gaze at the end of the activity signals that the activity is not completed yet and this can re-initiate a repetition of the route planning (see for example the interaction between Rita and Rose in Chapter 6, p. 100). A visual investigation of the video recordings of the interactions demonstrates that two things happen to close the activity: at the end of the object-focused activity, a repositioning takes place and participants return to a face-to-face position. It is at this point that they look at each other to signal their agreement on finishing the experiment.

8.5 CONCLUSION

The investigation of eye gaze in relation to speech and gesture in an activity-based context demonstrates deviation from everyday conversation and storytelling events. There is a terminological problem with the traditional understanding of gaze meaning to look at someone. Based on this definition, eye gaze is frequently suspended in the task-based setting. There is no eye contact with the other interactant over long stretches in the interaction. The beginning and the ending of the task performance is framed with an orientation toward and away from the interaction space. However, if the term *gaze* is inclusive of eye direction toward an activity, then the phenomena of shared gaze focus and mutual gaze can be investigated in the task-based context. Eye gaze plays an important, yet different role from the rules and findings established for face-to-face dialogue. The establishment of eye contact fulfills the function of initiating and maintaining social encounters. However, it is not a prerequisite when other resources are available. The interaction is maintained even though much attention, individual and joint, is concentrated on the shared interaction space and the two pieces of paper. They are the targets of the gaze work, the map being the more frequent gaze target, followed by the task sheet. Some of the traditional gaze patterns (Goodwin 1980, Bavelas et al. 2002a) are also present in the current context, for example when the "listener" looks at the "speaker". However, the present study argues that gaze work, especially in a task-based situation, is much more complex and more detailed than the traditional rules for gaze suggest.

Gaze fulfills a control function. In all types of interactions, a shift between the map and the task sheet helps to ascertain that the instructions are being followed. Over some stretches of discourse, there is a difference in the participants' individual orientations, which means that they attend to different things in their shared reference space. In interactions with distributed roles, the

participant who plans the routes can monitor and control the note taker's actions. When there is a difference in orientation, gaze shifts can help to comprehend what the other person is saying and to ensure that this concurs with the instructions. In interactions with equal roles, pointing gestures can be applied to direct the other interactant's gaze to a certain space and entity within this space.

Gaze re-directions at the other interactant co-occur with meta commentary. The verbal and the visual prompts signal the shift from the task level to the meta level. Looking at someone invites this person to leave the state of joint attention addressed to the task to engage in the meta level. Mutual gaze is triggered by meta commentary and is a result of such a frame shift. In particular, mutual gaze is achieved when the participants take an agreeing stance. In contrast, when there is disagreement, mutual gaze is often avoided. Mutual gaze, i.e., looking at each other, is also suspended when the interactants act on common ground. The investigation of activity closings demonstrated that gaze is indicative of whether the activity is finished or not. When gaze by either one of the two participants is not reciprocated at the end of the unit, the activity is not closed as the gaze initiator does not receive affirmation. However, once mutual gaze is established at the end of the activity, both participants signal each other that there are no unanswered questions or concerns about the activity. Mutual gaze demonstrates agreement about the ending of the activity and leads to a closing of the session.

Gaze cannot be studied in isolation from other nonverbal phenomena and speech. As this investigation demonstrates, organizational and interactional patterns are expressed through a range of resources. Body posture and orientation as well as gesture help to coordinate actions and to position oneself in relation to both the task and the other person.

9 CONCLUSIONS

This study presented an investigation of task-based discourse as a special type of discourse, establishing a relationship between classic topics in discourse analysis and interactional sociolinguistics with a gesture based approach to language. To conclude this work, I will first summarize the most important findings and theoretical implications of this study. I will end with an outlook for future research and possibilities for expanding the present research.

The results of this study derive from the examination of audio-visual data and yield practical implications for the way human communication is organized and shaped. The findings demonstrate the need for adapted frameworks in linguistic studies of communication and language. Linguistic approaches to language and communication focus on spoken aspects of language, to the exclusion of nonverbal features and phenomena. Gesticulation and eye gaze are often treated as extraneous to speech and considered paralinguistic phenomena. The present study, however, views gesture and gaze as integral components of language and of human communication, arguing for a change in linguistic approaches to 'language'.

This study focused on a particular type of discourse concerned with a goal-oriented task in order to demonstrate that the language used in such settings consists both of speech and gesture, making it central to communicative encounters. Dyads of interactants create and maintain joint action, engage in F-formations (Kendon 1990, McNeill et al. 2010) and negotiate steps within the interaction space, contributing individually and mutually to the activity. Within these interactions, certain structures emerge and unfold as features of these interactions. Individual aspects of the dyadic interactions have been treated in five analytic chapters, highlighting certain features in each chapter respectively. However, these phenomena are not to be understood as separate or unrelated. Rather, they are presented as contributing to the interaction as a whole, constituting each other, and creating and shaping joint actions.

The participants were given a map of Brookfield Zoo in Chicago and a task sheet with different activities. Similar methods are utilized in the Map Task (Howarth und Anderson 2007), the Traveling Salesman Problem (Wiener and Tenbrink 2008) or in direction giving and route planning activities (Klein 1982, Kita 2003b). The participants worked together in pairs and were instructed to plan their day at the zoo according to the instructions on the task sheet. The interactions were filmed with two video cameras. Transcriptions were produced based on the

conventions in the Appendix for the verbal part of the recordings and extended to include nonverbal phenomena in a separate column, labeled "gesture/gaze unit" analogous to the inclusion of intonation units in transcriptions. Gestures were transcribed for handedness (right hand or left hand as well as right/left index finger) as well as type of gesture (single and repeated pointing gesture, tracing). Gaze was indicated for gaze direction (map, task sheet, interlocutor), gaze shifts, and mutual gaze ("→").

This study has re-examined classic linguistic research topics, in particular openings, cohesion and reference, as well as repetition, drawing on previous research and contributing new findings and implications for further research based on the inclusion of gesture and gaze. Pointing gestures and eye gaze were investigated individually, establishing relationships to the previously examined phenomena from a nonverbal communication based perspective. The analysis of gesture and gaze allowed for a consideration of different modes, portraying the complexity of the interactions (see also Norris 2011).

Chapter 4 presented some general remarks about the interaction as a whole and the beginnings of interactions, in particular, proposing a relationship between the activity governance and the beginning of the activity. In terms of interaction openings, the present data contrasts with two assumptions made in linguistics: first, sequential research done in Conversation Analysis (CA), in particular the sequencing structure at the beginning of a conversation, for example shown in Schegloff's (1968) work on telephone conversation openings; and second, the traditional notions of speaker and hearer as the prototypical roles of participants in conversations. The activity at hand is not a conventionalized event, such as a telephone conversation. As such, there are no predetermined or ritualized formulas, which the participants can presuppose. Rather, certain structures emerge as the interaction progresses and the participants draw on multiple resources to ground their actions. They must signal their availability to cooperate and they must ratify their individual as well as their activity-related roles within the framework of the interaction. In contrast with the static roles of speaker and hearer, the notion of interactants and participants as well as the notion of activity roles represent the dynamic aspect of the participation of people in interactions. Chapter 4 established three activity roles established as a product of the performance of the activity: 1) the route planner, 2) the instruction reader, and 3) the note taker. Whereas the first two activity roles are present in all interactions under investigation, the third role is only realized in some of the interactions. This

reflects the fact that the event is largely unscripted and that participants negotiate the procedures relevant to the organization and management of the task as it unfolds and progresses. It also reflects the construction of the participants' own identities in the activity in which they engage. Future research could relate these findings to identity work in general to investigate how membership and involvement in different communities of practice and across different discourse genres is expressed in language, including speech, gesture, and other nonverbal communicative phenomena.

From a CA perspective, conversational openings display a two-part structure, they are realized in adjacency pairs such as question-answer or summons-answer. In the present context, however, there is sometimes a complete lack of speech at the beginning of the activity or, if speech is present, it occurs in the form of reading out the instructions. Transferring the instructions onto the map, the first route description can then also follow without an insertion of an adjacency pair. Reasons for this deviation from a sequential opening in conversation may be found in the nature of the task, providing material which can be employed to demonstrate engagement and readiness to begin. On close inspection of the interactions, taking visual phenomena into account, it becomes apparent that modes such as body orientation and shifts in body posture frame the beginning of the activity proper. For instance, a shift from face-to-face orientation to an orientation toward the objects on the table demonstrates a willingness to become involved in the activity. Together with the shift in body posture, eye gaze is also directed to the interaction space to mark the readiness to start the activity. All of these cues account for the beginning of the interaction and demonstrate that openings are more complex than proposed for conversation. Frobenius (2011:825) reportes similar findings for the openings of vlogs, explaining that some vloggers do not use conventionalized formulas and openings. Presumably, in different discourse genres, participants utilize different strategies to signal the beginning of the discourse and the activity and to show involvement. Interactions begin before the first utterances are produced and more research needs to be done in this area to make general statements about available and applied strategies in interactions.

A shared goal of the participants in the present study is the completion of the task. Two different patterns of how to achieve this goal emerged at the beginning of the activity: either the activity is opened nonverbally or there is a combination of speech and gesture. Nonverbal openings were shown to be realized through the distribution of the map and the task sheet, which

resulted in different organizational patterns and distributed control over the objects (map and task sheet) in the interaction space. This division of labor did not only influence the activity roles, but also changed pointing and gaze patterns of the dyads as orientation and monitoring activities differed.

In order to coordinate the task, the participants must maintain coherence with their own actions as well as someone else's actions. Various aspects play a role in this respect. Common ground is the basis of joint action since common ground is partially assumed based on cultural background, age, ethnicity, and other variables. However, as part of the joint action, common ground must also be negotiated and established in relation to the activity. There is an increase in common ground as the participants interact with one another. The establishment of common ground is further demonstrated in different stance taking activities. Phenomena such as reference and repetition were reanalyzed in this study as means of positioning. All of these features are meaningful constituents of the coordinative and interactive work between the dyads.

Chapter 5 addressed cohesion and reference from a speech-gesture synchrony perspective, arguing that the linguistic perception of cohesion as a textual phenomenon is insufficient. Extending McNeill et al.'s (2010) study of multiparty discourse, the present study investigated co-referential chains in dyadic discourse to supplement the few studies which have been carried out on cohesion and gesture. Certain parameters were established for the object, meta, and para level respectively. From a linguistic perspective, the para level is not only marked by reference to the speaker, as proposed by McNeill et al. (2010), but more specifically, it is frequently framed by an epistemic stance marker, such as *I think, I believe,* or *I suppose*, which expresses attitudes and opinions. Participants also use the plural noun "we" on this level to express the notion of solidarity and mutual involvement in the activity. Based on these findings, the connection of the para level to the meta level was further substantiated. Meta level references relate to previous discourse, more specifically they either refer to the task-performance and the external representations or they function as evaluations. Evaluation is a component of stance taking, as it is presented in DuBois' (2007) stance triangle. Evaluation is predominantly expressed in speech, but can also be signaled in gaze aversion to display a disagreeing stance, thereby evaluating a situation. In addition, while verbal evaluations often occur as meta level utterances, gaze remains directed to the reference space, thereby linking the object and the meta level through two different modes. The visual focus remains on the reference space while speech

addresses issues about the activity. This particular aspect of discourse related to a task has also been confirmed by the investigation of gaze behavior in Chapter 8. Similar to speech and gaze, speech and gesture can be co-expressive as well. For example, a reference can be directed to an entity on the map through a pointing gesture, while talk about this entity connects the current event to a previous event in the planning process. Through the resources of speech, gesture, and gaze, which can all be directed to the object level, but can also be distributed across the levels to express individual aspects of the planning process, participants create F-formations and align with one another. Cohesion and reference, it has been found, are also displayed in gesture recurrence and embodiment of gesture. This study contributed to the understanding of how gesture recurrence, similar to verbal repetition, creates connections across larger discourse chunks and events within the interaction.

Chapter 6 established different functions of repetition, from sense making to recapitulation toward the end of the activity. The recurrence of gesture and the borrowing of another's gesture links old and new information and it highlights and reiterates certain components of the route planning. In repeating a gesture, and this has been demonstrated for repeated pointing in particular, activities can be marked as important and they can be emphasized. Mimicking someone else's words and gesture aids comprehension, signals uptake, and indicates agreement. Gesture is therefore employed as a meaningful tool in the interaction and especially the cases where speech is absent demonstrate that gesture is communicative.

Repetitive gesture is similar to repetitive speech in another respect. For referring expressions, it was found that they decrease in length and number when repeated (see Hupet and Chantraine 1992; Levy and McNeill 1992). In analogy to these findings, one of my hypotheses has received support, namely that there is also a decrease in gesture when repeated. In particular when agreement has been reached previously in the interaction and when participants act on this agreement, the use of gesture becomes less frequent and the verbal repetition takes the form of a summary of the previous route planning. Summaries of route planning activities are usually found to conclude a unit or toward the end of the activity to close it.

Repetition and repeated pointing are sometimes used as signals of disagreement to reinforce opinions and to negotiate common ground. More often, however, repetition is a form of agreeing stance. Repeating and incorporating the other participant's pointing or tracing movement is a form of co-construction, signaling alignment. The recurrence of a topic is not

only marked in speech, but also in handedness. For instance, gesture space can be divided into the space where the map and the task sheet lie. The right and the left hand, respectively, can then be associated with one of the external representations and this distribution is maintained or reiterated when a topic is re-introduced in the interaction. This demonstrates once again how meaninful gesture is, since it provides information about the organization of the activity.

The following figure illustrates the interplay of multiple modes as they are utilized in the organization of the activity. It visualizes the connection between object, meta, and para level to demonstrate how they create cohesion. It also summarizes the available resources for the modes of speech, gesture, and gaze, as presented in this study. These micro-level phenomena create common ground and joint action and are relevant in the organization of the interaction as a whole.

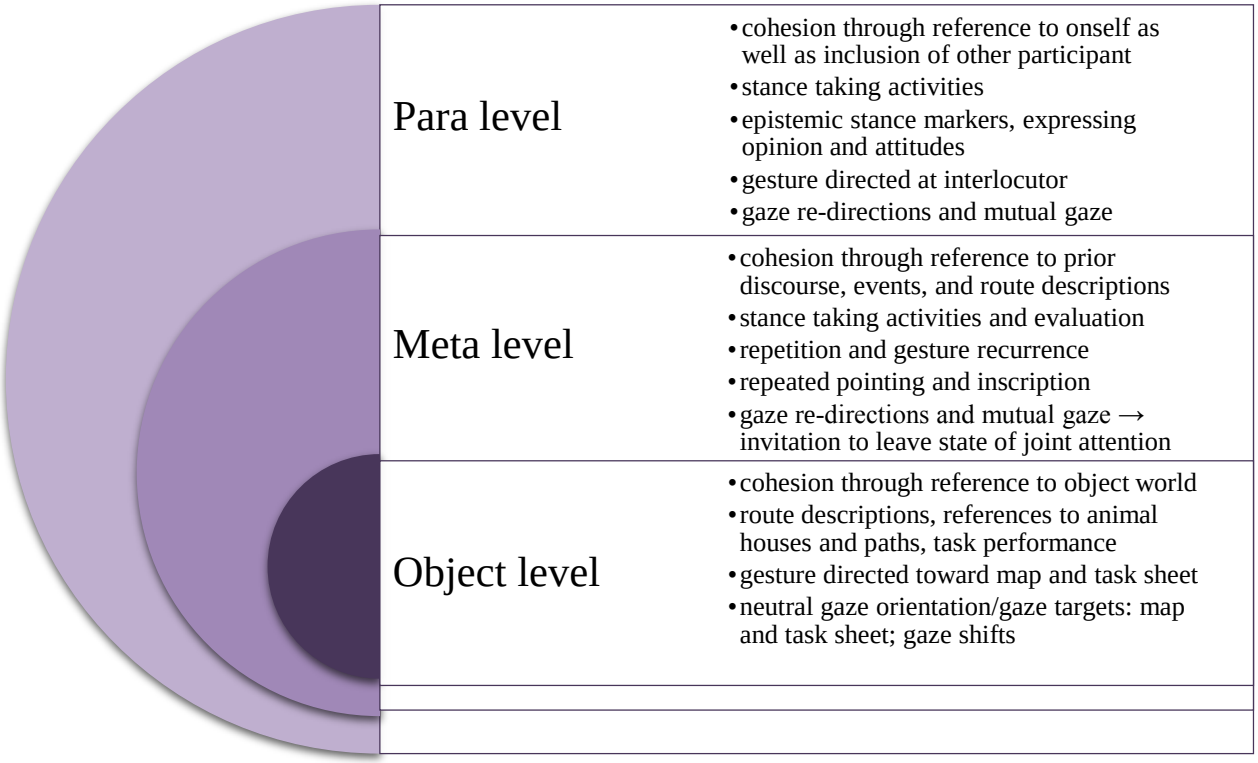

Figure 11 Organization of interaction based on different phenomena

The data derived from the second part of the research project, which were excluded from analysis in this study, offer an opportunity for future research on cohesion and handedness. Pereira (2013) investigates these data to show how handedness and gesture space are employed in storytelling to create different perspectives and identities. Future research could profitably analyze how cohesion is created across the two sessions. Other questions could include, "Are certain aspects of the planning phase reiterated in the telling phase" or "Is the differentiation of gesture space and handedness which was established in the planning maintained in the telling phase?".

Chapter 7 identified four types of deictic gesture, arguing that the terms *gesture* and *deictic gesture*, in particular, are not fixed categories, attributing the function of indexicality to deictic gestures. Rather, deictic gestures are dynamic entities, taking different forms and occurring in different spaces. Each type of deictic gesture fulfills a different function in the planning process. Single pointing gestures are the most frequent, functioning as indexing and referencing items on the map. They conflate with the object level and occur throughout the activity. Repeated pointing gestures have a highlighting function and either emphasize steps within the activity or occur at the end of a unit to mark the arrival at the final destination. Hence, repeated pointing is structurally different from single pointing because repeated pointing only occurs at certain points in the activity. Repeated pointing also has a close connection to the meta level. When it co-occurs with repetition and meta level references, it reinforces propositions and route suggestions. Tracing entails the notion of movement and frequently co-occurs with words and phrases expressing movement, for example *go up/down*, *go past*, *go round*. The last type of deictic gesture identified here is a subgroup containing all three types of deictic gestures described so far. *Collaborative pointing*, a term borrowed from Goodwin (2003), and *collaborative tracing*, labeled as such in analogy to Goodwin's terminology, are those gestures which are carried out by both participants. Collaborative pointing, it is was found in the present study, is the equivalent of verbal co-constructions and overlap. These results for tracing and the collaborative gestures show that the same notion is expressed in different modes. These notions can relate to movement, but also to involvement and cooperation. Future research based on a larger amount of data could establish correlations between certain lexical items and tracing to quantify their co-occurrence.

Chapter 8 demonstrated that gaze patterns in task-based interactions vary from the turn taking organization and the patterns described for conversation (Goodwin 1980, Bavelas et al. 2002a). Here, gaze is not an indicator of turn-taking activities, but rather an indicator of interactants' attention states. More specifically, while speech can express notions about the activity, gaze remains directed to the task level itself, demonstrating that people attend to various things at the same time. The rules for gaze, which were established in earlier studies, must be revised, based on larger amounts of data as well as different discourse genres, investigating similarities and differences across types of activities and conversations.

The results of this study show that gaze fulfills a control function. Each participant can monitor another's actions and moves in the route planning due to the shared workspace. Individual and shared orientation are made clear through speech and deictic gesture. The gaze target of both participants is the map, followed by the task sheet. Frequently, the participants' gaze patterns align, which means that they first share the same focus to the map and then shift their eye gaze from the map to the task sheet simultaneously to maintain the shared focus. Mutual gaze is often suspended, though. Sometimes, it is established when participants take an agreeing stance. However, in the present setting, mutual gaze is not always established with agreement. It was found that participants maintain gaze direction to the material objects of map and task sheet when they act on common ground, i.e., when they share the same assumptions. When there is disagreement, mutual gaze is usually avoided, also because participants can draw on other modes to express their opinion or to receive the interlocutor's feedback. There is a need for quantification and maybe even the employment of eye-tracking technology to fully account for the role of eye gaze in this particular setting and in different settings in general.

To conclude, language, communication, and interaction are central to discourse analysis and thus, any work in this area must acknowledge the significance of speech, gesture, gaze and other nonverbal phenomena to describe them in different settings and for different communities of practice. This study has investigated task-based interactions from a multimodal perspective, yielding important implications for our understanding of language and providing ideas for future explorations in the study of language, communication, and interaction.

10 Deutsche Zusammenfassung

Die vorliegende Studie zieht audio-visuelle Daten heran, um sich mit den Phänomenen der gesprochenen Sprache, Gestik und Blick basierend auf aufgabenorientierten dyadischen Interaktionen im Englischen zu beschäftigen.

Diese Arbeit basiert vor allem auf zwei großen theoretischen Bereichen, zum einen der Diskursanalyse und hier im speziellen der interaktionalen Soziolinguistik, und zum anderen der Gestikforschung, wie sie zum Beispiel von McNeill (1992, 2005, 2012) und Kendon (1980, 2004) verstanden wird. Weiterhin zieht die vorliegende Studie Arbeiten aus der multimodalen Forschung (Schmitt 2007, Norris 2004, 2011) und aus der Forschung zu Blick (Kendon 1967, Goodwin 1980, Bavelas et al. 2002a) heran, um aufzuzeigen, wie verbale und nonverbale Phänomene zu Interaktion und Koordination von Aktivitäten beitragen.

Kommunikation und Sprache müssen als Kombination von multimodalen Ebenen verstanden werden, das heißt, Sprache ist nicht nur ein geschriebenes und gesprochenes System, sondern konstituiert sich aus Worten und Gesten zugleich. Linguistische und diskursanalytische Studien müssen ihren Blickwinkel auf Sprache und Kommunikation folglich ändern. Nicht nur verbale Aspekte sind in diesen Analysen zu berücksichtigen, sondern auch nonverbale Modi müssen inkorporiert werden, um ein Gesamtbild von Sprache und Kommunikation zeichnen zu können.

Im Folgenden stelle ich die Daten, die die Grundlage dieser Studie bilden vor, präsentiere die zentralen Fragen und Hypothesen dieser Studie, fasse die wichtigsten Ergebnisse zusammen und gebe einen Ausblick auf künftige Forschungsfragen und Forschungsschwerpunkte.

Den Untersuchungsgegenstand dieser Arbeit bilden einundzwanzig anwendungsbasierte Interaktionen, die 2009 in Birmingham an der Birmingham City University in England erhoben wurden. Alle Teilnehmer der Studie sind Muttersprachler des Englischen und sind zum Zeitpunkt der Aufnahmen als Studenten an der Birmingham City University eingeschrieben. Die Interaktionen sind mit zwei Videokameras gefilmt worden, wobei jeweils eine Kamera auf einen Teilnehmer gerichtet ist, um visuelle und nonverbale Phänomene festhalten zu können. Die Teilnehmer haben ihr schriftliches Einverständnis zur Verwendung von Audio-, Video- und Bildmaterial gegeben. Alle Namen der Teilnehmer wurden anonymisiert.

Der vorliegende Datensatz bildet nur einen Teil einer größeren Sammlung von Daten. Anschließend an den anwendungsbasierten Teil der Studie, auf die sich die vorliegende Arbeit konzentriert, wurden weitere Daten zum sogenannten „storytelling“, zum Erzählen erhoben, die sich inhaltlich auf den ersten Teil der Studie beziehen. Im anwendungsbasierten Teil der Studie arbeiten jeweils zwei Interaktionspartner zusammen. Den Teilnehmern wird eine Aufgabe gestellt, die sie gemeinschaftlich lösen sollen. Als Teil der Aufgabe werden Materialien gestellt, und zwar sind dies eine Karte des Zoos „Brookfield Zoo“ in Chicago sowie ein Aufgabenblatt, auf dem verschiedene Aktivitäten und Tierhäuser genannt sind (siehe Kapitel 3 für die Aufgabenblätter). Die Teilnehmer sind aufgefordert die Instruktionen zu lesen und danach die einzelnen Entitäten auf der Karte zu lokalisieren, um zu beschreiben, wie sie zu den jeweiligen Orten gelangen. Ähnliche Methoden zur Datenerhebung kennt man aus Studien in denen die sogenannte „Map Task“ eingesetzt wird (Anderson und Boyle 1994; Howarth und Anderson 2007) oder von dem „Traveling Salesman Problem“ (Wiener und Tenbrink 2008). Cohen und Harrison (1973), Klein (1982), Wunderlich und Reinelt (1982) sowie Kita (2003b) haben Daten aufgrund von Wegbeschreibungen und Routenplanungen erhoben und analysiert. Goodwin (2003) untersucht ähnliche Phänomene in der Interaktion zwischen Archäologen, die relevante Objekte auf einer Karte festhalten und auf diese Karte referieren, wenn sie über die Objekte sprechen.

Der Analyseteil dieser Studie setzt sich aus fünf Kapiteln zusammen, die inhaltlich aufeinander referieren, aber als eigenständige Einheiten gelesen werden können. Die ersten drei Kapitel beschäftigen sich mit klassisch diskursanalytischen Themen, und zwar der Gesprächseröffnung und im konkreten Fall der Eröffnung der Interaktion (Kapitel 4). In Kapitel 5 werden Kohäsion, im Speziellen Referenz behandelt. Die Wiederholung von Wörtern, Phrasen und ganzen Intonationseinheiten werden in Kapitel 6 untersucht. In diesen ersten drei analytischen Kapiteln werden etablierte Forschungsgegenstände in Verbindung mit neuen Aspekten, nämlich Gesten und Blick, gebracht. In den letzten beiden Analysekapiteln werden Zeigegesten und Blick eingehender untersucht, um die vorausgehenden Analyseergebnisse durch einen veränderten Fokus zu beleuchten und zu verstärken.

Die Untersuchung der Eröffnung der Interaktionen in Kapitel 3 zeigt, dass die sequentielle Gesprächseröffnung, wie sie zum Beispiel von Schegloff (1968) für Telefongespräche beschrieben wurde und wie sie in der Konversationsanalyse etabliert ist, zu

eng gefasst ist. Die Eröffnung einer Interaktion zeichnet sich nicht ausschließlich durch zwei aufeinanderfolgende Redebeiträge aus, wie zum Beispiel Frage-Antwort oder Gruß-Gegengruß, aus. Vielmehr zeigen die Ergebnisse dieser Untersuchung, dass der Beginn einer neuen Aktivität durch multimodale Signale aufgezeigt wird. Hierzu zählen die Körperhaltung und –orientierung, die Blickrichtung und auch die Organisation der Objekte im Interaktionsraum. Durch eine Loslösung von einer Gesicht-zu-Gesicht-Orientierung („face-to-face orientation“) hin zu einer Orientierung in Richtung des Interaktionsraumes demonstrieren die Teilnehmer ihr Interesse an der Aufgabe und ihre Bereitschaft diese zu beginnen. Mit der körperlichen Repositionierung geht auch eine Neuorientierung der Blickrichtung einher. Interaktionen können nonverbal oder aber durch eine Kombination von Sprache und nonverbalen Phänomenen eröffnet werden. Die Verteilung der Karte und des Aufgabenblattes spielen hierbei eine wichtige Rolle. Durch die Verteilung dieser Objekte signalisieren die Teilnehmer ihre Einbindung in die Aufgabe und schreiben sich selbst und dem anderen Teilnehmer Aktivitätsrollen („activity roles“) zu. Zwei Hauptarten der Organisationsform und drei Arten von Aktivitätsrollen resultieren daraus: Erstens, egalitäre Organisationsformen und zweitens Organisationformen mit verteilten Rollen. Egalitäre Interaktionen zeichnen sich dadurch aus, dass die Karte und das Aufgabenblatt in der Mitte liegen bleiben, was bedeutet, dass beide Teilnehmer gleichen Zugang zu diesen Objekten haben. Beide Teilnehmer können daher auch die Rolle des Planers/des Wegbeschreibers einnehmen und zugleich der Leser der Instruktionen sein. Darüber hinaus können diese Rollen am Anfang oder über den Verlauf der Aktivität verteilt werden, so dass zum Beispiel eine Person die Instruktionen laut vorliest, während der andere diese bereits auf die Karte transferiert. Diese Aktivitäten zeigen auf, dass die Interaktionseröffnung keine sequentielle Struktur aufweist, sondern vielmehr durch das Heranziehen von verschiedenen Modi auf diversen Ebenen eingeleitet wird.

In der zweiten Organisationsform mit verteilten Rollen, werden auch die Karte und das Aufgabenblatt verteilt. Eine Person ist dann der Wegbeschreiber während die andere Person vorliest und in manchen Fällen sogar Notizen macht, also als „note taker“ fungiert. In diesem Falle unterscheiden sich beispielsweise Monitoring-Aktivitäten und Blickverhalten von den egalitären Interaktionen, da der Blick häufiger zwischen Karte und Aufgabenblatt wandert. Diese Aktivitäten wurden in Kapiteln 7 und 8 basierend auf Gestik- und Blickverhalten genauer untersucht.

Zusammenfassend lässt sich bisher sagen, dass die Organisationsformen und Aktivitätsrollen Auskunft darüber geben, wie die einzelnen Teilnehmer sich in die Aufgabe einbringen, wie sich selbst wahrnehmen und wie sie sich zueinander positionieren. Diese Ergebnisse könnten beispielsweise durch künftige Arbeiten zu Identität und zur Schaffung von Gruppenzugehörigkeit in verschiedenen „communities of practice“ und Diskursarten eingehender untersucht werden, um das Zusammenspiel von Sprache und visuellen Phänomenen auch in anderen Diskursarten und Kontexten aufzuzeigen.

Die Untersuchung von Kohäsion und Referenz in Kapitel 4 stützt sich auf eine Studie von McNeill et al. (2010) und analysiert koreferenzielle Ketten („co-referential chains“), die sich durch den Diskurs ziehen. Die vorliegende Studie argumentiert hier, dass Kohäsion und Referenz nicht nur textuelle Phänomene sind, die Textpassagen und einzelne Elemente miteinander verbinden, wie zum Beispiel Halliday und Hasan (1976) dies etabliert haben. Vielmehr werden Kohäsion und Referenz als Phänomene verstanden, die durch Sprache und Gesten zugleich kenntlich gemacht werden und die sich auch von einem Interaktionspartner auf den anderen übertragen können. Damit sind zum Beispiel das Spiegeln und das Einverleiben („mimicking“ und „embodiment“) von Gesten gemeint.

McNeill et al. (2010) unterscheiden drei Ebenen, auf denen referiert werden kann. Dies sind die *Objektebene*, die *Metaebene* und die *Paraebene*. In dieser Studie wurden diese drei Ebenen aus einer linguistischen Perspektive beleuchtet, Parameter wurden aufgestellt, und die Verbindung von Sprache zu Gestik und Blick auf den einzelnen Ebenen sowie ebenenübergreifend herausgearbeitet. Die Ergebnisse zeigen, dass die Paraebene oftmals durch sprachliche Ausdrücke, die den Standpunkt verdeutlichen („stance markers“), markiert wird. Zeigegesten und Blickwechsel kommen auf dieser Ebene seltener als auf der Objekt- und der Metaebene vor. Oftmals ist die Paraebene aber eng mit der Metaebene verwoben. Standpunktmarker leiten die Intonationseinheit ein oder folgen dieser und rahmen somit die Metaaussage. Auf der Metaebene finden sich Referenzen, die sich auf den Aktivitätsverlauf beziehen oder diesen evaluieren. Sie sind deshalb nicht direkt an die Objektebene, die Ebene der konkreten Planung, gebunden, sondern beziehen sich auf Vorausgegangenes im Diskurs und greifen Ereignisse wieder auf. Die Metaebene und die Paraebene stellen die Ebenen dar, auf denen deutlich wird, wie die Teilnehmer sich zueinander und auch zu ihren eigenen und gegenseitigen Aktivitäten positionieren. Positionierung und Evaluation sind zwei Komponenten

des sogenannten „stance taking“, wie DuBois (2007) es beispielsweise konzipiert. Durch Blickrichtung und Zeigegestik können die Teilnehmer sich aneinander orientieren und anpassen (im Englischen wird dies mit dem Ausdruck „alignment“ bezeichnet) und schaffen so einen gemeinsamen Grundstein („common ground“; siehe Stalnaker 2002; Clark et al. 1983) von dem aus sie agieren. Auf der Metaebene treten auch Phänomene wie „mutual gaze“, die Blickorientierung zueinander auf, um den Interaktionspartner einzuladen, sich von der Objektebene zu lösen und sich auf die Metaebene zu begeben, beispielsweise um sich zum vorher Gesagten zu positionieren.

Durch die Analyse der drei verbundenen Ebenen, hat die vorliegende Arbeit auch gezeigt, dass „common ground“ kein Konstrukt ist, das in seiner Ganzheit vorausgesetzt werden kann. Teilweise beruht „common ground“ auf sozioökomischen und kulturellen Gemeinsamkeiten, also den Dingen, die wir aufgrund unserer Erfahrungen und unseres Hintergrundes als gegeben ansehen. „Common ground“ wird aber auch geschaffen und verhandelt und folglich vergrößern sich im Verlauf der Interaktion die Gemeinsamkeiten, auf denen Aktivitäten stattfinden. „Common ground“ spiegelt sich in Wiederholung, sowohl sprachlicher als auch gestischer, wider.

In Kapitel 6 wird das Phänomen der sprachlichen Wiederholung analysiert. Diese dient zum einen der Evaluierung, aber auch der Rückversicherung, dass der andere Teilnehmer den Wegbeschreibungen zustimmt. Durch kollaboratives Zeigen auf eine Entität wird ebenfalls Zustimmung ausgedrückt. Das wiederholte Zeigen auf eine Entität sowie das wiederholte Aufzeigen von Wegen durch den Zoo dient als Verstärkung der eigenen Beiträge und verdeutlicht diese Beiträge durch die Visualisierung für den Interaktionspartner. Wiederholungen treten sowohl im Verlauf der Interaktion, aber auch am Ende der Interaktion als Form der Interaktionsbeendigung auf. Im Verlauf der Aktivitätsplanung und –umsetzung dient die Wiederholung vor allem dem Verständnis und der Verarbeitung der Aufgabe. Das Ende der Interaktion bzw. der Aktivität ist, wie die Eröffnung auch, kein sequentielles Phänomen. Entscheidend für die Beendigung der Aktivität sind neben verbalen Phrasen wie „that’s it“ oder „back at the North Gate“ vor allem die Zuwendung zueinander und die Etablierung von beiderseitigem Blick zueinander („mutual gaze“).

Für sprachliche Referenz haben Studien von Hupet und Chantraine (1992) oder Howarth und Anderson (2007) gezeigt, dass sprachliche Ausdrücke in der Wiederholung kürzer werden

und schneller gesprochen werden. Ähnliches lässt sich durch die Ergebnisse dieser Studie für Gesten in wiederholten Einheiten sagen. Auch sie nehmen in den wiederholten Passagen ab. Die Wiederholung von Gesten lässt darüber hinaus Aussagen über Händigkeit („handedness“) treffen. Händigkeit meint die jeweilige Hand mit der eine Geste produziert wird. Im Verlauf der Aktivität kann eine Unterscheidung von rechter und linker Hand getroffen werden, um eine Zuweisung der jeweiligen Hand zu einem bestimmten Thema oder einem bestimmten Bereich im Interaktionsraum über längere Passagen vorzunehmen. Wenn diese Strukturen rekurrent sind, erteilen sie auch Auskunft über thematische oder personenspezifische Zusammenhänge und schaffen somit Kohärenz (siehe auch Levy und McNeill 1993). Zukünftige Forschung könnte beispielsweise die Phänomene der Kohäsion, Rekurrenz und Händigkeit im zweiten Teil, dem narrativen Teil, der Daten untersuchen, um zu analysieren, ob diese Phänomene aus dem aufgabenbasierten Teil in den narrativen Teil übertragen werden.

In Kapitel 7 hat diese Studie vier Arten von Zeigegesten etabliert. Die Zeigegeste ist nicht nur ein Mittel, um Indexikalität auszudrücken, und kennt nicht nur eine Form, nämlich die des Indexfingers, sondern kann verschiedene Formen aufweisen und diverse Funktionen erfüllen. Zeigegesten können beispielsweise mit dem Daumen oder durch eine Verlängerung, wie etwa ein Stift oder eine laparoskopische Kamera (Koschmann et al. 2010) ausgeführt werden. Enfield (2001) beschreibt das Phänomen des Zeigens mit dem Mund und den Lippen („lip-pointing“), wie es in Laos zu beobachten ist. Die Zeigegeste erfüllt weiterhin diverse Funktionen. Eine einfache Zeigegeste richtet sich zumeist auf eine Entität und spezifiziert diese, vor allem wenn die verbale Referenz alleine ungenügend ist. Dies ist zum Beispiel der Fall, wenn nur durch Begriffe wie „hier“ oder „dort“ referiert wird. Ohne den visuellen Zugang zur Karte und die begleitende Zeigegeste, wären solche Referenzen für den Interaktionspartner im vorliegenden Untersuchungsgegenstand nicht eindeutig zu identifizieren, vor allem wenn zuvor der Referent nicht namentlich genannt wurde. Wiederholtes Zeigen auf eine Entität signalisiert zum Einen eine Verstärkung des Gesagten, markiert auf struktureller Ebene aber auch den Abschluss von Einheiten und das Erreichen des Ziels. Wiederholte Zeigegesten drücken zudem Übereinkunft und Zustimmung aus. Sie sind also ein Mittel um „common ground“ zu schaffen. Neben der einfachen und der wiederholten Zeigegeste, ist auch die sogenannte „tracing“-Geste analysiert worden. „Tracing“, ein Begriff von Goodwin (2003), beschreibt einen Aspekt des Zeigens, bei dem zwei Entitäten in eine dynamische Verbindung gebracht werden. Im aktuellen Datensatz

wäre dies zum Beispiel eine Geste, die zuerst auf einen Startpunkt gerichtet wird, und die dann fortläuft, entlang des Weges sozusagen, und beim Erreichen des Endpunktes aufhört. Wenn diese Geste mit Hilfe des Stiftes ausgeführt wird, dann kann sie auch eine Markierung auf der Karte hinterlassen. Diese spezielle Form nennt sich dann „inscription". Alle drei bisher beschriebenen Formen können auch von beiden Teilnehmern gemeinschaftlich durchgeführt werden, was sich dann „collaborative pointing" (Goodwin 2003) und, in Anlehnung an Goodwins Begriff, „collaborative tracing" nennt. Gemeinschaftliches Zeigen ist, laut Goodwin (2003), eine Möglichkeit, ein visuelles Feld individuell einzuteilen und dies oftmals in Antizipation des Beitrages des Interaktionspartners. In der vorliegenden Arbeit wurde allerdings auch gezeigt, dass gemeinschaftliches Zeigen dann auftritt, wenn es zu Problemen und Unklarheiten in der Interaktion kommt. Hier dient wiederholtes Zeigen der Verstärkung des eigenen Standpunktes. Durch das Aufnehmen und Widerspiegeln der Geste des Anderen, prozessiert man die Wegbeschreibung und schafft Verständnis zugleich. Als Resultat kann die Aktivität auf gemeinsam etablierten Grund, auf „common ground", fortgesetzt werden.

Im letzten analytischen Kapitel (Kapitel 8) wurden Blick und Blickrichtung der beiden Teilnehmer untersucht. Goodwin (1980) und Bavelas et al. (2002a) stellen Regeln für das Blickverhalten des Sprechers und des Zuhörers in Konversationen auf. Laut diesen Studien, unterscheiden sich das Blickverhalten des Zuhörers und des Sprechers, sie verlaufen entlang eines asymmetrischen Blickmusters. Bavelas et al. sprechen von einem „gaze window" (2002a:577), einem Fenster bzw. einem Zeitraum, in dem der Zuhörer den Sprecher anschauen kann, ohne dessen Rolle als Sprecher übernehmen zu wollen. Goodwins (1980) Regeln besagen, dass der Zuhörer den Sprecher häufiger und länger anschaut als der Sprecher den Zuhörer anschaut. Der Zuhörer soll weiterhin den Sprecher anschauen, wenn der Sprecher Blickkontakt aufnehmen will. In dem vorliegenden Datensatz lassen sich diese Regeln sowie das asymmetrische Blickmuster nicht bestätigen. Vielmehr hat diese Studie herausgefunden, dass die Blickorganisation sich verändert sobald zwei oder mehr Personen mit einer Aktivität beschäftigt sind (siehe hierzu auch Gerhardt 2007, Tiittula 2007). Die Grundorientierung richtet sich dann nicht zum Interaktionspartner, sondern liegt viel mehr auf der Aktivität und den Objekten, über die gesprochen wird. Die Nähe der beiden Interaktionspartner zueinander spielt ebenfalls eine wichtige Rolle, da Änderungen in der Körperhaltung und der Positionierung wahrgenommen werden können, ohne aufzublicken. Das Ziel des Blickes („gaze target") sind also meist die

Karte gefolgt von dem Arbeitsblatt. Blickwechsel („gaze shift“) finden zumeist zwischen der Karte und dem Arbeitsblatt statt. Dies ist zum Einen wichtig, da die Instruktionen während der Ausführung der Aufgabe immer wieder gelesen werden müssen. Zum Anderen erfüllt der Blickwechsel eine Kontrollfunktion, da dadurch immer wieder sichergestellt werden kann, dass die Aufgabe erfüllt wird. In Interaktionen mit verteilten Rollen, erlaubt der Kontrollblick dem Wegbeschreiber herauszufinden, ob der Teilnehmer, der die Wegbeschreibung schriftlich niederhält, den Ausführungen folgen kann. Der Blick zu den Aktivitäten des Schreibers erfüllt somit eine Monitoringfunktion, da der Planer sehen kann, wann er mit der Planung des Weges fortfahren kann und wann er warten muss. Der Schreiber, entgegen der Regeln, die für Zuhörerverhalten aufgestellt wurden, muss hier nicht aufblicken, um dem Planer zu signalisieren, dass er fertig ist. Dies zeigen die Schreibaktivitäten und auch der Blickwechsel hin zur Karte auf, sobald der Schreibprozess beendet ist. In diesen Fällen unterscheiden sich zuweilen die „attention states“, die individuellen Orientierungen der Partner. Der Wegbeschreiber fokusiert sich auf die Karte und die Planung der Route. Der Schreiber konzentriert seinen Fokus auf das Schreiben, wodurch zwei unterschiedliche Orientierungen entstehen. Erst wenn der Schreiber wieder zur Karte blickt, ist ein Zustand von „joint attention“, von gemeinschaftlicher Aufmerksamkeit, hergestellt.

Die Interaktionspartner schauen sich selten gegenseitig an, da der gemeinsame Fokus auf der Aktivität liegt. Gegenseitiger Blick zueinander wird also über lange Passagen des Diskurses unterdrückt, findet sich aber auf der Metaebene, zum Beispiel wenn ein Teilnehmer einen Vorschlag unterbreitet, der von der Aufgabenstellung abweicht. Der Blick zu dem anderen Teilnehmer ist also eine Aufforderung, Zustimmung zu erhalten. Wird der Blick entgegnet, dann wird die Metaaktivität fortgesetzt. Metaaktivitäten können aber auch ohne gegenseitigen Blick zueinander stattfinden. Dies ist aufgrund der Nähe, aber auch durch das Vorhandensein anderer Ressourcen möglich. So kann beispielsweise eine Zeigegeste, mit der Teilnehmer A auf Teilnehmer B referiert, wahrgenommen werden, ohne den Blick aus der Grundorientierung zu lösen. Auch die verbalen Kontributionen auf der Metaebene erlauben eine Fokussierung zur Karte oder zum Arbeitsblatt. Die Aufforderung sich von der Objektebene zu lösen und zur Metaebene zu folgen, kann also von einem Blickwechsel zueinander begleitet werden, dies muss aber nicht der Fall sein. Wenn beide Interaktionsteilnehmer zuvor schon „common ground“ etabliert haben, dann können sie auf diesem agieren und die Objektebene verlassen,

ohne einander anzuschauen. Gegenseitiger Blick wird auch oft unterdrückt, wenn die Teilnehmer nicht mit dem Beitrag des Anderen übereinstimmen („gaze aversion“). Hier werden dann wiederholte Zeigegesten verwendet, um den eigenen Standpunkt zu verdeutlichen. Auch die teilweise Wiederholung von Phrasen tritt auf, um die Aussage des Anderen zu hinterfragen. Das Ende der Aufgabe wird durch mehrere Modi markiert und Blick spielt hier eine wichtige Rolle. Wird am Ende kein „mutual gaze“ etabliert, dann signalisieren die Teilnehmer einander, dass sie mit der Aufgabe noch nicht fertig sind. Oftmals werden dann einzelne Teile der Interaktion wieder aufgegriffen und zur Sicherstellung der Richtigkeit der Planung wiederholt. Erst wenn beide Teilnehmer einander ansehen, findet auch eine körperliche Repositionierung statt und auf verbaler Ebene wird ausgesprochen, dass man die Aufgabe und die Interaktion nun beendet.

Durch die Erhebung größerer Datenmengen könnten zukünftig die Ergebnisse für Zeigegesten und Blick quantifiziert werden. Auch könnte der Einsatz von „eye tracking“ Technologie zur Unterstützung der Analyse der Blickrichtung und des Blickwechsels eingesetzt werden. Gestik und Blick müssen weiterhin in verschiedenen Typen des Diskurses und Formen der Interaktion systematisch untersucht werden.

Zusammenfassend lässt sich sagen, dass die vorliegende Studie fünf zentrale Aspekte der Interaktion zwischen zwei Partnern, die eine Aufgabe gemeinschaftlich lösen, herausgearbeitet hat. Die Eröffnung der Aktivität, Kohäsion und Referenz, verbale und gestische Wiederholungen sowie Zeigegesten und Blick wurden eingehend analysiert. Hierbei ist hervorzuheben, dass diese Untersuchungsgegenstände zwar einzeln analysiert wurden, aber als ein zusammenhängendes System zu betrachten sind. All diese Phänomene wurden von der Mikroebene aus analysiert und tragen als Einheit gesehen zu einem neuen Verständnis von Sprache und Kommunikation bei.

Die vorliegende Arbeit hat Terminologie und Ansätze aus den Bereichen der Diskursanalyse, der Multimodalität und der Gestikstudien herangezogen, um den Gegenstand der Sprache allgemein und der aktivitätsbezogenen Interaktion im Speziellen zu beleuchten und ihre Merkmale herauszuarbeiten. Regeln, die beispielsweise für Gespräche und Erzählungen aus diskursanalytischer Perspektive aufgestellt wurden, müssen zukünftig revidiert werden. Sprache und Kommunikation bestehen nicht nur aus verbalen Phänomenen, sondern beinhalten auch nonverbale Modi des Ausdrucks. Linguistische Ansätze können deshalb durch die Forschung im Bereich der Gestikstudien angereichert werden. Die Gestikstudien hingegen können von

Arbeiten aus der Diskursanalyse profitieren. Die vorliegende Arbeit hat einen Beitrag zu beiden Bereichen geleistet und anwendungsbasierten Diskurs aus multimodaler Perspektive analysiert, woraus wichtige Implikationen für das Verständnis und die Analyse von Sprache, Kommunikation und Interaktion resultieren.

11 APPENDIX

Transcription Conventions

she's out.	Period shows falling tone in the preceding element.
oh yeah?	Question mark shows rising tone in the preceding element.
nine, ten	Comma indicates a level, continuing intonation.
DAMN	Capitals show heavy stress or indicate that speech is louder than the surrounding discourse.
°dearest°	Utterances spoken more softly than the surrounding discourse are framed by degree signs.
(2.0)	Numbers in parentheses indicate timed pauses.

If the duration of the pauses is not crucial and not timed:

..	a truncated ellipsis is used to indicate pauses of one-half second or less.
...	An ellipsis is used to indicate a pause of more than a half- second.
ha:rd	The colon indicates the prolonging of the prior sound or syllable.
bu- but	A single dash indicates a cut-off with a glottal stop.
[and so-]	Square brackets on successive lines mark
[why] her?	beginning and end of overlapping talk.
and=	Equals signs on successive lines show latching
=then	between turns.
H	Clearly audible breath sounds are indicated with a capital H.
.h	Inhalations are denoted with a period, followed by a small h. Longer inhalations are depicted with multiple hs as in .hhhh
h	Exhalations are denoted with a small h (without a preceding period). A longer exhalation is denoted by multiple hs.
.t	Alveolar suction click
( )	In the case that utterances cannot be transcribed with certainty empty parentheses are employed
(hard work)	If there is a likely interpretation, the questionable words appear within the parentheses.
/ /	slashes are used for phonetic transcriptions
((laugh))	Aspects of the utterance, such as whispers, coughing, and laughter, are indicated with double parentheses.
Numbering conventions:	Number each intonation unit consecutively (e.g. from 1 to n).

List of figures

12 REFERENCES

Abbé de l'Epée, Charles-Michel. 1776. *Institutions des sourds-et-muets par la voie des signes méthodiques. Ouvrage qui contient le Projet d' une Langue Universelle, par l'entremise des Signes naturelles, assujettis à une Méthode*. Paris: Nyon.

Abbott, H. Porter. 2008. *The Cambridge Introduction to Narrative, 2nd edition*. Cambridge: Cambridge University Press.

Aijmer, Karin. 1997. *I think* – and English modal particle. *Modality in Germanic languages. Historical and comparative perspectives*, ed. by Toril Swan; and Olaf J. Westwik, 1-47. Berlin: Mouton de Gruyter.

Allan, Keith. 2013. *The Oxford handbook of the history of linguistics*. Oxford: Oxford University Press.

Anderson, Anne H.; and Elizabeth A. Boyle. 1994. Forms of introduction in dialogues: their discourse contexts and communicative consequences. *Language and Cognitive Processes* 9 (1). 101-122.

Argyle, Michael. 1975. *Bodily Communication*. London: Methuen.

Argyle, Michael; and Mark Cook. 1976. *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Bakeman, Roger; and Lauren B. Adamson. 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development* 55. 1278-1289.

Bamberg, Michael. 2006. Stories: Big or small? Why do we care? *Narrative Inquiry* 16. 147-155.

Baron-Cohen, Simon. 1995. The Eye Direction Detector (EDD) and the Shared Attention Mechanism (SAM): two cases for evolutionary psychology. *Joint attention: its origins and role in development*, ed. by Chris Moore; and Philip J. Dunham, 41-59. Hillsdale: Erlbaum.

Bateson, Gregory. 1972. *Steps to an ecology of mind*. Chicago: Chicago University Press.

Bavelas, Janet Beavin; et al. 1992. Interactive Gestures. *Discourse Processes* 15. 469-489.

Bavelas, Janet Beavin; Linda Coates; and Trudy Johnson. 2002a. Listener response as a collaborative process: The role of gaze. *Journal of Communication* 52. 566-580.

Bavelas, Janet Beavin et al. 2002b. An experimental study of when and how speakers use gestures to communicate. *Gesture* 2 (1). 1-17.

Bavelas, Janet Beavin; and Jennifer Gerwing. 2007. Conversational hand gestures and facial displays in face-to-face dialogue. *Social Communication*, ed. by Klaus Fiedler, 283-308. New York: Psychology Press.

Bayliss, Andrew P. et al. 2013. "Gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General* 142 (1). 76-92.

Baynham, Mike. 2011. Stance, positioning, and alignment in narratives of professional experience. *Language in Society* 40. 63-74.

Bazzanella, Carla.1993. Dialogic repetition. *Dialoganalyse IV*, ed. by Heinrich Löffler, 285-294. Tübingen: Niemeyer.

----. 2011. Redundancy, repetition, and intensity in discourse. *Language Sciences* 33. 243 254.

Beukeboom, Camiel J. 2009. When words feel right: How affective expressions of listeners change a speaker's language use. *European Journal of Psychology* 39. 747-756.

Birdwhistell, Ray L. 1970. *Kinesics and context: essays on body motion communication.* Philadelphia: University of Pennsylvania Press.

Buck, Ross. 1984. *The communication of emotion*. New York: Guilford Press.

Cienki, Alan. 2009. Conceptual Metaphor Theory in light of research on speakers' gestures. *Cognitive Semiotics* 5 (1-2). 349-366.

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

----. 2003. Pointing and placing. *Pointing: Where language, culture, and cognition meet*, ed. by Sotaro Kita, 243-268. New York: Psychology Press.

Clark, Herbert H.; Robert Schreuder; and Samuel Buttrick. 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior* 22. 245-258.

Clark, Herbert H.; and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22. 1-39.

Clark, Herbert H.; and Meredyth A. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50. 62-81.

Clinton, Hillary. Video Message to the Fulbright Scholars. www.youtube.com/watch?v=MsWwPo47Y0U (20.09.2013)

Cohen, Akiba A.; and Randall P. Harrison. 1973. "Intentionality in the use of hand illustrators in face-to-face communication situations". *Journal of Personality and Social Psychology* 28 (2). 276-279.

Cook, Mark. 1977. Gaze and mutual gaze in social encounters. *American Scientist* 3. 328-333.

Davies, Bronwyn; and Rom Harré. 2007. Positioning: the discursive production of selves. *Journal for the Theory of Social Behavior* 20 (1). 43-63. http://www.massey.ac.nz/~alock/position/position.htm (26.04.2014)

Davis, Jeffrey E. 2010. *Hand talk. Sign language among American Indian nations.* Cambridge: Cambridge University Press.

De Jorio, Andrea. 2000. *Gesture in Naples and gesture in classical antiquity*. Bloomington: Indiana University Press. (originally printed in 1832)

Dooley, Robert A.; and Stephen H. Levinsohn. 2001. *Analyzing discourse. A manual of basic concepts.* Dallas: SIL International.

Du Bois, John W. 2007. The stance triangle. *Stancetaking in discourse: subjectivity, evaluation, interaction*, ed. by Robert Englebretson, 139-182. Amsterdam: Benjamins.

Edwards, Jane A. 2001. The transcription of discourse. *The handbook of Discourse Analysis*, ed by. Deborah Schiffrin; Deborah Tannen; and Heidi E. Hamilton, 321-348. Oxford: Blackwell.

Efron, David. 1941. *Gesture and Environment*. New York: King Crown Press.

Ekman, Paul; and Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1 (1). 49-98.

Ekman, Paul. 1997. Should we call it expression or communication? *Innovation* 10 (4). 333-344.

Emery, Nathan J. 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews* 24. 581-604.

Enfield, Nick J. 2001. 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture* 1 (2). 185-211.

----. 2009. *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.

Fetzer, Anita. 2004. *Recontextualizing context: grammaticality meets appropriateness*. Amsterdam: Benjamins.

Fludernik, Monika. 1989/90. Jespersen's shifters: Reflections on deixis and subjectivity in language. *Klagenfurter Beiträge zur Sprachwissenschaft* 15/16. 97-116.

Frobenius, Maximiliane. 2011. Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics* 43 (3). 814-827.

Gerhardt, Cornelia. 2007. Watching television: The dilemma of gaze. *Toegepaste Taalwetenschap in Artikelen* 77. 91-101.

Gibbon, Dafydd. 2009. Gesture theory is linguistics: On modelling multimodality as prosody. *Proceedings of the 23rd Pacific Asia Conference on Language, Information, and Computation*, ed. by Olivia Kwong, 9-18. Hong Kong: City University of Hong Kong Press.

Goffman, Erving. 1963. *Behavior in public places: Notes on the social organization of gatherings*. New York: The Free Press.

----. 1974. *Frame Analysis: An essay on the organization of experience.* New York: Harper and Row.

----. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.

----. 1982. *The interaction order*. http://www2.asanet.org/governance/PresidentialAddress1982.pdf (29.08.2013)

Goldin-Meadow, Susan. 2003a. *Hearing gesture: How our hands help us think*. Cambridge MA: Belknap Press.

----. 2003b. *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press.

Goodwin, Charles. 1980. Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological Inquiry* 50 (3-4). 272-302.

----. 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32. 1489-1522.

----. 2003. Pointing as situated practice. *Pointing: Where language, culture, and cognition meet*, ed. by Sotaro Kita, 217-241. New York: Psychology Press.

----. 2007. Participation, stance and affect in the organization of activities. *Discourse and Society* 18 (1). 53-73.

Goodwin, Charles; and Marjorie Harness Goodwin. 1987. Concurrent operations on talk: Notes on the interactive organization of assessments. *IPRA Papers in Pragmatics* 1 (1). 1-54.

Grice, Herbert P. 1989. *Studies in the way of words*. Cambridge: Harvard University Press. (originally printed in 1957)

Gritten, Anthony; and Elaine King (eds.) 2006. *Music and gesture*. Hampshire: Ashgate.

----. 2011. *New perspectives on music and gesture*. Hampshire: Ashgate.

Gullberg, Marianne; and Steven G. McCafferty. 2008. Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition* 30. 133-146. http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=1611433&fileOId=3912709 (20.04.2014)

Gumperz, John J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.

Haddington, Pentti. 2006. The organization of gaze and assessments as resources for stance taking. *Text & Talk* 26 (3). 281-328.

----. 2007. Positioning and alignment as activities of stancetaking in news interviews. *Stancetaking in discourse: subjectivity, evaluation, interaction*, ed. by Robert Englebretson, 283-317. Amsterdam: Benjamins.

Halliday, M.A.K. 1994. *An introduction to functional grammar*. London: Hodder Arnold.

Halliday, M.A.K.; and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman: London.

Haviland, John B. 2011. Nouns, Verbs, and Constituents in an Emerging 'Tzotzil' Sign Language. *Representing Language: Essays in Honor of Judith Aissen*, ed. by Rodrigo Gutiérrez-Bravo, Line Mikkelsen and Eric Potsdam, 157-171. California Digital Library eScholarship Repository. Linguistic Research Center,University of California, Santa Cruz. http://escholarship.org/uc/item/0vf4s9tk#page-1 (17.11.2013)

Holenstein, Elmar. 2008. Semiotics as a cognitive science. *Cognitive Semiotics* 3. 6-19.

Howarth, Barbara; and Anne H. Anderson. 2007. Introducing objects in spoken dialogue: The influence of conversational setting and cognitive load on the articulation and use of referring expressions. *Language and Cognitive Processes* 22 (2). 272-296.

Huang, Yan. 2011. Types of inference: entailment, presupposition, and implicature. *Foundations of pragmatics*, ed. by Wolfram Bublitz; and Neal R. Norrick, 397-421. Berlin: de Gruyter Mouton.

Hupet, Michel; and Yves Chantraine. 1992. Changes in repeated references: collaboration or repetition effects? *Journal of Psycholinguistic Research* 21 (6). 485-496.

International Society for Gesture Studies (ISGS). http://www.gesturestudies.com/ (06.07.2013)

Izard, Carroll E. 1977. *Human emotions*. New York: Plenum Press.

Johnstone, Barbara. 1987. Perspectives on repetition. *Text* 7 (3). 205-214.

Kanda, Toshiyuki; Ishiguro, Horoshi; and Toru Ishida. 2001. Psychological analysis of human-robot interaction. *Proceedings of IEEE International Conference on Robotics and Automation* 4. 4166-4173.

Kärkkäinen, Elise. 2007. The role of *I guess* in conversational stancetaking. *Stancetaking in discourse: subjectivity, evaluation, interaction*, ed. by Robert Englebretson, 183-219. Amsterdam: Benjamins.

Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26. 22-63.

----. 1980. Gesticulation and speech: two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, ed. by Mary Ritchie Key, 207-227. The Hague: Mouton.

----. 1990. *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

----. 2004. *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Keisanen, Tiina. 2007. Stancetaking as interactional activity. Challenging the prior speaker. *Stancetaking in discourse: subjectivity, evaluation, interaction*, ed. by Robert Englebretson, 253-281. Amsterdam: Benjamins.

Kita, Sotaro (ed.) 2003a. *Pointing: where language, culture, and cognition meet.* Mahwah: Erlbaum.

----. 2003b. Interplay of gaze, hand, torso orientation, and language in pointing. *Pointing: Where Language, Culture, and Cognition Meet*, ed. by Sotaro Kita, 307-328. New York: Psychology Press.

Klein, Wolfgang. 1982. Local deixis in route directions. Speech, place, and action. *Studies in deixis and related topics*, ed. by Robert. J. Jarvella; and Wolfgang Klein, 161-182. Norwich: Wiley.

Koschmann, Timothy; Le Baron, Curtis; Goodwin, Charles; and Paul Feltovich. 2010. “Can you see the cystic artery yet?” A simple matter of trust. *Journal of Pragmatics* 43 (2). 521-541.

Krauss, Robert M; Chen, Yihsiu; and Purnima Chawla. 1996. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology* 28. 389-450.

Krauss, Robert M.; Chen, Yihsiu; and Gottesmann, Rebecca F. 2000. Lexical gestures and lexical access: a process model. *Language and Gesture*, ed. by David McNeill, 261-283. Cambridge: Cambridge University Press.

Kreuz, Roger J.; and Monica A. Riordan. 2011. The transcription of face-to-face interaction. *Foundations of pragmatics*, ed. by Wolfram Bublitz; and Neal R. Norrick, 657-679. Berlin: De Gruyter Mouton.

Labov, William. 1972. *Language in the Inner City: Studies in the Black English Vernacular*. Oxford: Blackwell.

Lascarides, Alex; and Matthew Stone. 2009. Discourse coherence and gesture interpretation. *Gesture* 9 (2). 147-180.

Levy, Elena T.; and David McNeill. 1992. Speech, gesture, and discourse. *Discourse Processes* 15. 277-301.

Map of Brookfield Zoo. http://www.czs.org/CZS/Brookfield/Zoo-Map/Brookfield-Zoo-map-2010 (01.11.2010)

McCawley, James D. 1999. Participant roles, frames, and speech acts. *Linguistics and Philosophy* 22 (6). 595-619.

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

----. 2005. *Gesture and thought*. Chicago: University of Chicago Press.

----. 2008. Gestures of Power and the Power of Gestures. *Proceedings of the Berlin Ritual Conference 2008*, ed. by Erika Fischer-Lichte; and Christoph Wulf, 1-13. http://mcneilllab.uchicago.edu/pdfs/gsts_of_pwr_and_pwr_of_gests.pdf (29.12.2013)

----. 2012. How *language began. Gesture and speech in human evolution*. Cambridge: Cambridge University Press.

McNeill, David; and Elena T. Levy. 1993. Cohesion and gesture. *Discourse Processes* 16 (4). 363-386.

McNeill, David et al. 2010. Mind Merging. *Expressing oneself/expressing one's self: communication, cognition, language, and identity*, ed. by Ezequiel Morsella, 143-164. New York: Psychology Press.

Misak, Cheryl (ed.) 2004. *The Cambridge companion to Peirce*. Cambridge: Cambridge University Press.

Mondada, Lorenza. 2007. Interaktionsraum und Koordinierung. *Koordination. Analysen zur multimodalen Interaktion*, ed. by Reinhold Schmitt, 55-93. Tübingen: Narr.

Moore, Chris; and Philip J. Dunham (eds.) 1995. *Joint Attention: Its Origins and Role in Development*. Hillsdale: Erlbaum.

Morris, Charles W. 1938. *Foundations of the theory of signs*. Chicago: Chicago University Press.

Morris, Desmond. *Desmond Morris talks about "Owl"*. http://www.desmond-morris.com/index.php (24.01.2014)

Mortensen, Kristian. 2012. Conversation Analysis and multimodality. *Conversation Analysis and Applied Linguistics. The encyclopedia of Applied Linguistics*, ed. by Johannes Wagner; and Kristian Mortensen, 1061-1068. Oxford: Wiley-Blackwell.

Murray, Stephen O. 1998. *American Sociolinguistics. Theorists and theory groups*. Amsterdam: Benjamins.

Norrick, Neal R. 1987. Functions of repetition on conversation. *Text* 7 (3). 245-264.

----. 2007. Conversational storytelling. *The Cambridge companion to narrative*, ed. by David Herman, 127-141. Cambridge: Cambridge University Press.

Norris, Sigrid. 2004. *Analyzing multimodal interaction: A methodological framework*. London: Routledge.

----. 2011. *Identity in (inter)action. Introducing multimodal (inter)action analysis*. Berlin: De Gruyter Mouton.

Nöth, Winfried. 2011. Semiotic foundations of pragmatics. *Foundations of pragmatics*, ed. by Wolfram Bublitz; and Neal R. Norrick, 167-202. Berlin: de Gruyter.

Ochs, Elinor. 1979. Transcription as theory. *Developmental Pragmatics*, ed. by Elinor Ochs; and Bambi Schieffelin, 43-72. New York: Academic Press. http://www.sscnet.ucla.edu/anthro/faculty/ochs/articles/ochs1979.pdf (16.04.2014)

Pape, Helmut. 2004. *Charles S. Peirce zur Einführung*. Hamburg: Junius.

Peirce, Charles Sanders. 1981. *Writings of Charles S. Peirce. A chronological edition, vol. 1*. Bloomington: Indiana University Press.

Penz, Hermine. 2007. Building common ground through metapragmatic comments in international project work. *Metapragmatics in Use*, ed. by Wolfram Bublitz; and Axel Hübler, 263-291. Amsterdam: Benjamins.

Peräkyla, Anssi; and Marja-Leena Sorjonen. 2012. *Emotion in interaction*. Oxford: Oxford University Press.

Pereira, Gerardine M. 2011. "Turn right at the camels": The relevance of deictic gestures in task based discourse. *Toegepaste Taalwetenschap in Artikelen* 85 (1). 123-133.

----. 2013. Creating and maintaining perspectives and identities in narratives with gesture space and gestural handedness. *Saarland Working Papers in Linguistics* 4. 41-56.

Pomerantz, Anita. 1984. Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. *Structures of social action: Studies in Conversation Analysis*, ed. by J. Maxwell Atkinson; and John Heritage, 57-101. Cambridge: Cambridge University Press.

Raymond, Geoffrey; and John Heritage. 2006. The epistemics of social relations: Owning grandchildren. *Language in Society* 35 (5). 677-705.

Rossano, Federico; Brown, Penelope; and Stephen C. Levinson. 2009. Gaze, questioning, and culture. *Conversation Analysis: comparative perspectives*, ed. by Jack Sidnell, 187-249. Cambridge: Cambridge University Press.

Rossini, Nicla. 2012. *Reinterpreting gesture as language*. Amsterdam: IOS Press.

Saeed, John I. 2003. *Semantics*, 2nd ed. Oxford: Blackwell.

Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70 (6). 1075-1095.

----. 1972. Notes on a conversational practice: Formulating place. *Studies in social interaction*, ed. by David Sudnow, 75-119. New York: MacMillan.

Scheibman, Joanne. 2007. Subjective and intersubjective uses of generalizations in English conversations. *Stancetaking in discourse: subjectivity, evaluation, interaction*, ed. by Robert Englebretson, 111-138. Amsterdam: Benjamins.

Schiffrin, Deborah. 1994. *Approaches to Discourse*. Oxford: Blackwell.

Schmitt, Reinhold (ed.) 2007. *Koordination. Analysen zur multimodalen Interaktion*. Tübingen: Narr.

Schmitt, Reinhold; and Arnulf Deppermann. 2007. Monitoring und Koordination als Voraussetzung der multimodalen Konstitution von Interaktionsräumen. *Koordination. Analysen zur multimodalen Interaktion*, ed. by Reinhold Schmitt, 95-128. Tübingen: Narr.

Short, T. L. 2004. The development of Peirce's theory of signs. *The Cambridge Companion to Peirce*, ed. by Cheryl Misak, 214-240. Cambridge: Cambridge University Press.

Sidner, Candace L. et al. 2004. *Explorations in engagement for humans and robots*. http://www.merl.com/papers/docs/TR2004-048.pdf (02.09.2013)

Simon-Vandenbergen, Anne-Marie. 2000. The functions of *I think* in political discourse. *International Journal of Applied Linguistics* 10 (1). 41-63.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25. 701-721. http://semantics.uchicago.edu/kennedy/classes/f07/pragmatics/stalnaker02.pdf (12.09.2013)

Staudte, Maria. 2010. *Joint attention in spoken human-robot interaction*. Universität des Saarlandes Dissertation. http://scidok.sulb.uni-saarland.de/volltexte/2010/3242/ (26.08.2013)

Stivers, Tanya. 2008. Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction* 41 (1). 31-57.

Streeck, Jürgen. 1996. How to do things with things: Objets trouvés and symbolization. *Human Studies* 19 (4). 365-384.

----. 2009. *Gesturecraft. The manu-facture of meaning.* Amsterdam: Benjamins.

Streeck, Jürgen; Goodwin, Charles; and Curtis LeBaron. 2011. *Embodied interaction. Language and body in the material world*. Cambridge: Cambridge University Press

Supalla, Ted. 2002. Making historical sign language materials accessible: A prototype database of early ASL. *Sign Language and Linguistics* 4. 285-297.

----. 2003. Revisiting visual analogy in ASL classifier predicates. *Perspectives on classifier constructions in sign language*, ed. by Karen Emmorey, 249-258. Mahwah: Erlbaum.

Tannen, Deborah. 1979. What's in a Frame? Surface Evidence for Underlying Expectations. *New Directions in Discourse Processing*, ed. by Roy Freedle, 137-181. Norwood: Ablex.

----. 1986. *That's Not What I Meant. How Conversational Style Makes or Breaks Your Relations with Others.* New York: Morrow.

----. 1989. *Talking voices. Repetition, dialogue, and imagery in conversational discourse.* Cambridge: Cambridge University Press.

Thompson, Geoff; and Susan Hunston. 2000. Evaluation: An introduction. *Evaluation in text. Authorial stance and the construction of discourse*, ed. by Susan Hunston; and Geoff Thompson, 1-26. Oxford: Oxford University Press.

Tiittula, Liisa. 2007. Blickorganisation in der side-by-side-Positionierung am Beispiel eines Geschäftsgesprächs. *Koordination. Analysen zur multimodalen Interaktion*, ed. by Reinhold Schmitt, 225-261. Tübingen: Narr.

Tomasello, Michael. 1995. Joint attention as social cognition. *Joint attention: its origins and role in development*, ed. by Chris Moore; and Philip J. Dunham, 103-130. Hillsdale: Erlbaum.

Tylor, Edward Burnet. 1865. *Researches into the early history of mankind and the development of civilization*. London: Murray.

Wiener, Jan M.; and Thora Tenbrink. 2008. *Traveling salesman problem: The human case*. http://www.jan-wiener.de/publications/TSP-the_human_case.pdf (25.06.2012)

Wunderlich, Dieter; and Rudolph Reinelt. 1982. How to Get There From Here. *Speech, Place, and Action*, ed. by Robert J. Jarvella; and Wolfgang Klein, 183-201. Norwich: Wiley.