From Communicative Strategies to Cognitive Modelling

Kerstin Fischer, Reinhard Moratz
University of Bremen, Bremen, Germany
kerstinf@uni-bremen.de, moratz@tzi.de

Abstract

The question addressed in this paper is what we can learn from natural language interaction between humans and robots for modelling the robot's cognitive structure. Experiments were carried out which show which communicative strategies human users employ in the interaction with an artificial communication partner. The communicative strategies found in the interaction show a complex mental model of the domain under consideration, that is, spatial instruction. The findings provide insights for future modelling of the robot's cognitive abilities.

1. Introduction and Motivation

What should a robot be able to do and which cognitive capabilities should it have? This question will be addressed in this paper for a particular domain which is central for the interaction between humans and robots: spatial instruction. Modelling the robot's cognitive structure for spatial instruction involves many critical choices; researchers have a number of criteria at which they can orient. For instance, from an AI perspective, the cognitive modelling of the robot should rely on the criterion of cognitive adequacy. The evidence for cognitively adequate modelling of the robot's behaviour (the weak AI perspective) may be experimental data on human behaviour. This was also our starting point in the study reported on in this paper. However, it turned out that human users often display a very peculiar behaviour in human-robot interaction, for instance, regarding the perspective taken, which differs from the interaction among humans (Moratz and Fischer, 2000) . The assumption underlying the current investigation is therefore that for efficient human-robot communication, the modelling of the robot has to rest upon the users' mental model of the robot and of the domain, and that these two models become apparent by analysing which strategies human users employ in the communication with a robot.

The methodology used here is to have human users interact with a robot which was modelled on the basis of what is known about spatial reference in human-to-human communication. Our procedure was consequntly firstly to design a robot. Secondly, an experimental setup was developed in order to identify the strategies that speakers employ in the interaction with the robot. Thirdly, experiments were then carried out in a joint attention scenario, that is, participants and the robot were attending to the same set of objects distributed in the room. The participants' task was to get the robot to move to some goal objects pointed at by the leader of the experiment by typing natural language sentences into a computer. If the instruction was processed successfully, the robot rolled to the goal objects indicated, if something went wrong, the robot just printed *error* on the screen and the participant reformulated her instruction. Finally, the participants utterances were analysed linguistically in order to identify the strategies employed.

The procedure rests on findings regarding adaptation processes in human-computer interaction that show that users and the system interactively achieve a common mode of communication. For instance, Amalberti et al. (1993) argue that while speakers initially approach human and artificial communication partners very differently, the two types of linguistic behaviour become more and more similar over time, if the users are confronted with exactly the same linguistic output by their communication partners. In these experiments, the output was manipulated by a human 'wizard' who did not know whether the participants were instructed to be talking to a machine or to a human communication partner. Amalberti et al.'s results indicate that humans design their utterances differently for human and artificial communication partners, but that because of adaptation processes these differences may disappear. Thus, while the users' linguistic behaviour is on the one hand determined by their conceptualization of the recipient for whom the utterances are designed (Schegloff, 1972; Sacks et al., 1974; Roche, 1989), the recipient design can change during the interaction.

In error resolution contexts, speakers could be shown to adapt to the linguistic properties of their artificial communication partners' utterances in order to increase understandability (Oviatt, MacEachern,

Levow, 1998; Oviatt, Bernard, Levow, 1998; Levow, 1998). Finally, speakers have been found to be extremely patient in what they endure with malfunctioning artificial communicators (Fischer, 1999). First results in human-robot interaction have shown that also in this kind of setting speakers adapt to their artificial communication partner on the basis of its linguistic and behavioural output (Moratz and Fischer, 2000).

Since human users thus adapt to their communication partners, it was expected that if attempts to instruct the robot turned out to be unsuccessful, users would change their strategy and try another one, for instance, a different type of spatial reference, a different perspective or different lexical material, so that the experiments would provide us with a rich overview of the strategies speakers preferably use in the interaction with a robot. These strategies, and, as we shall see, their order of employment, point to the users' mental models of spatial instruction and of robots. The mental models elicited may not only influence the interaction, the findings can also be used in the modelling of the robot's cognitive structure.

2. The Robot

Four components interact in the robot used, whose architecture has been outlined in more detail in Habel et al. (1999): the language interpretation component, the spatial reference component, the sensor component and the acting component.

The **language interpretation component** is based on Combinatory Categorial Grammar (CCG) (Steedman, 1996). A domain dependent version of the lexical categorial system was developed and the grammatical rules were adapted to fit German word order. Specific categorial rules were derived from the original CCG rules in order to make incremental and efficient processing of natural language instructions like *go to the left block* or *move to the front cube* possible (Hildebrandt and Eikmeyer, 1999). The version of the system used for the experiments does not foresee linguistic output by the system but only consists of a language *understanding* component: The only feedback the users get is the system utterance *error* or some action carried out by the robot. Limiting the system's output garantees that the users' behaviour is not 'shaped' (Zoltan-Ford, 1991) by the robot. In contrast, initiating new communicative strategies on the basis of previous error messages reveals the users' implicit theories about what may have gone wrong and how they make sense out of the robot's behaviour. That is, by depriving users of detailed linguistic feedback by the system, their reactions may reveal their mental model of the robot.

The **spatial reference component** implements the computational model of projective relations described in section 3. It maps the spatial reference expressions of the given command to the relational description delivered from the sensor component. For interpreting commands it uses the computational model of reference systems for projective relations described in the next section.

The **sensor component** uses a video camera. Our orientation to cognitive adequacy in the design of the communicative behavior of the robot influenced our decision to use a sensory equipment resembling human perceptual capabilities (Moratz, 1997). The camera is fixed on top of a pole with a wide angle lens looking below to the close area in front of the robot. That is, on top of the robot, a wooden construction was applied which holds the camera. So the robot has an overview of the floor. The robot thus appears to have a long neck – which is responsible for its nick name *giraffe* (see figure 1). Images are processed with region-based object recognition (Moratz, 1997). The spatial arrangement of these regions is delivered to the spatial reference component as a qualitative relational description.

Figure 1: The robot *Giraffe*

The **acting component** manages the control of the mobile robot (Pioneer 1). The motoric actions the robot can perform are turns and straight movements (Röhrig, 1998). The actions can be carried out by passing a control sequence to the motors. The component can carry out simple obstacle avoidance and path-planning (Habel et al., 1999).

The interaction between the components consists of a superior instruction-reaction cycle between the lan-

guage interpretation component and the spatial reference component; subordinate to this cycle is a perception-action cycle started by the acting component, which assumes the planning function and which controls both the sensor component and the acting component.

3. The Computational Model of Reference Systems for Projective Relations

For the implementation of the robot's verbal strategies of spatial reference, results from psychology and psycholinguistics on spatial expressions in human-to-human communication (Levinson, 1996; Levelt, 1996) were used. In a joint attention scenario, like the one we used, spatial instruction is usually achieved by goal descriptions like *go to the right cube*. The verbal expressions employed typically contain specifications of projective relations (e.g. "left") that are dependent on a specific perspective or point of view (Herrmann and Grabowski, 1994). Projective relations use a *reference object*, a *reference direction* and qualitative *angular sectors* as the directional component to specify regions in which the object referred to, the *target object*, lies. Reference objects can be the speaker, the listener, or other, explicitly referred to, salient objects (e.g. *from my point of view, the coin is to the right of the ball*). In the communication between humans, speakers typically use their own direction of view as reference direction; only in some situations, for instance, in the communication with children (Long, 1982), speakers use the listener's reference system in order to simplify reference resolution for the listener (Herrmann and Grabowski, 1994).

To model robot-centered reference systems, all objects are arranged in a bird's-eye view. This amounts to a projection of the objects onto the plane $\mathcal{D}$ on which the robot can move. The projection of an object O onto the plane $\mathcal{D}$ is called $p_{\mathcal{D}}(O)$. The center μ of this area can be used as point-like representation O' of the object O: $O' = \mu(p_{\mathcal{D}}(O))$.

For a reference system, a reference object RO' and a reference axis $\vec{r}$ are required. This reference axis is a directed line from the robot through the point RO'. These geometric elements partition the plane into a left and a right half-plane.

However, interpretable instructions may differ from instructions one would want to generate; for instance, one may want to be able to interpret a reference to an object as *the front object*, even if would be most relevantly be characterized as being left. We therefore use two different partitions of the plane, one for interpretation (acception of spatial expressions as reference to a specific object) and one for generation. This is motivated by the desire that the acceptor model for the direction instructions be more tolerant than the generating model. The generating model is a complete, disjunct, partition of the visible area. The acceptor model is a superset of every direction instruction. Therefore, the acceptance areas overlap. However, in the current system we only interpret (and do not yet generate) spatial expressions and therefore make no use of generating model so far.

The partitioning into a left and a right half-plane constitutes an acceptor model for the directions "left of" and "right of" relative to the reference object. The dichotomy front/back is modelled similarly by using another axis orthogonal to the reference axis. With some reference frames, however, front and back are exchanged (see below). We therefore conceptualize it as a qualitative distinction, as suggested, for instance, by Freksa (cf. Zimmermann and Freksa, 1996).

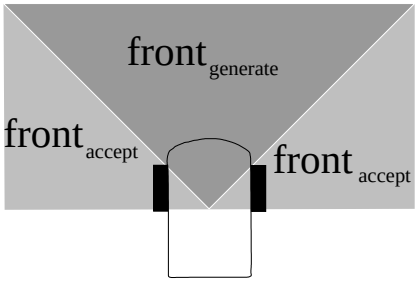

Figure 2: The robot as reference object

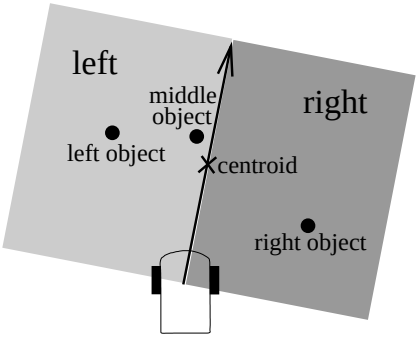

Figure 3: Group based reference

We need to model spatial references with three kinds of reference objects, namely, the robot, another salient object or a group of objects. If the robot is chosen as reference object, the reference direction is naturally given by its view direction. The view direction of the robot is its symmetry axis and therefore a salient structure to be observed by the instructor. Then, the acceptance area and the generating area for "front" are those depicited in figure 2.

If the localisation object is closer to another salient object than to the robot, this object is a convenient reference object. In this case, there are two ways of deriving a reference direction. One is given by the directed straight line from the robot to the reference object (for instance, through the centers of their projections). This is again adapted to the robot's view. In this variant of three-point localisation, the "in front of" sector is directed towards the robot. The front/back-dichotomy is inverted, relative to the reference direction (Herrmann and Grabowski, 1994).

In cases with a group of similar objects, human instructors may use references with respect to the whole group, for example, *go to the left block*. Then the centroid of the group can be treated as the reference object. Analogous to the three point model, the reference direction is given by the directed straight line from the robot center to the group centroid. This virtual reference object is the origin of acceptance areas and generation areas for relations similar to three-point localisation. The object closest to the group centroid can be understood as the central object, and objects to the left or right of the centroid can be referred to as "left" and "right" object respectively, as in the instruction *go to the left object* (see figure 3). The robot developed was tested in the following in the interaction with naive human users.

4. Communicative Strategies of Spatial Instruction

The experimental procedure taken was to employ the robot in interaction with human users in order to see which strategies users employ initially and how their strategies are adapted during the interaction with the system.

Experimental Design

An experimental setting was developed in which the users' task was to make the robot move to particular locations pointed at by the conductor of the experiment; pointing was used in order to avoid verbal expression or pictures of the scene which would impose a particular perspective, for example, the bird's-eye view. Users were instructed to use natural language sentences typed into a computer to move the robot; they were seated in front of a computer into which they could type their instructions. When they turned around, they perceived a scene in which, for instance, a number of cubes were placed on the floor together with the robot (see figure 4).

15 different participants carried out about 40 attempts to move the robot within about 30 minutes time each. About half of the participants were computer scientists working on different aspects of artificial intelligence, the other half were students of other subjects and secretaries. Their sentences were protocolled; altogether 603 instructions were elicited. Additionally, the participants' verbal behaviour during the experiments was recorded in order to capture their self-talk; usually participants announce their strategies or their ideas about what is going wrong even if they do not get any feedback from the experimenter. After the experiments, participants were asked as to what they believe the robot could and could not understand, which strategies they believed had not been successful, and whether their beliefs about the robot have changed during the interaction.

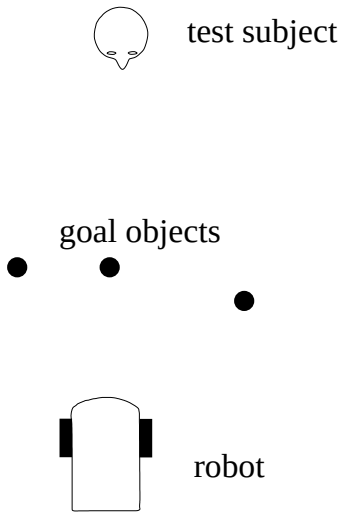

Figure 4: The Experimental Setup

The question addressed in the human-robot interaction experiments is which communicative strategies users employ and how do they change over time.

Experimental Results

The communicative strategies the users were found to employ reveal an order in the type of instructions. Our hypothesis is that this order mirrors the users' mental model (Gentner and Stevens, 1983) of spatial instruction. The model concerns the supposed difficulty and basicness of each type of spatial instruction.

In particular, only half of the participants consistently used the goal-naming strategy observable in joint attention scenarios in natural conversation, that is, to name the reference object itself. These participants were mostly computer scientists familiar with work and goals in artificial intelligence.[1] Examples of this strategy are *fahr bis zum rechten Würfel [***drive up to the right cube***], fahr zu dem Klotz, der vor Dir liegt*

[1] They were, of course, unfamiliar with the robot and the aims of the experiments.

*[***drive to the block which is in front of you***], geh zu dem vorderen Würfel [***walk to the front cube***]*. This strategiy was the one expected and implemented on the basis of findings from human-to-human communication, so that these instructions were usually successful, unless there were orthographic, lexical, or syntactic problems. In the case of failure, these participants used path-naming strategies, if successful, they sticked to the goal-naming strategy.

The other half of the participants, mostly the computer-naive users, first tried out another strategy, in particular, giving path descriptions, decomposing the main action in more primitive actions, such as *move forward*, *go backwards*, or *turn left*. This strategy seemed very natural to the participants, and they were on the whole quite exasperated to find that this strategy did not work with the robot, which, as explained above, was designed on the basis of findings from human-to-human communication and thus could understand only goal-decriptions. Example sentences typed by the participants are the following: *fahr 1 Meter geradeaus [***drive 1 meter ahead***], rolle ein wenig nach vorn [***roll a bit forward***], fahre nach Norden von Dir aus gesehen [***drive north from your point of view***], links [***left***], los Du lahme Kiste vorwärts nach rechts [***come on you lame box ahead to the right***], bewege Dich Richtung Schrank [***move in the direction of the wardrobe***]*.

If the path descriptions did not work, the participants did not try out a description of the goal object, which the robot would have understood. Instead, they used descriptions of movements, for instance *fahre [***drive***], bewege Dich mit einer positiven Geschwindigkeit in irgendeine Richtung [***move with positive speed in some direction***], sitz [***sit***], spring [***jump***], Drehung! [***turn!***]*. Some participants who had used this strategy, employed afterwards a fourth one, namely to specify the instrumental actions necessary for such movement, for example: *drehe Deine hinteren Rollen [***turn your rear wheels***]* or *Motor an [***engine on***]*.

Thus, a fixed order of instructional strategies becomes apparent: If goal descriptions were unsuccessful (for other, possibly lexical or orthographic reasons), users tried out path descriptions, or they started off with path descriptions immediately. If path descriptions turned out unsuccessful, participants employed descriptions of movements. If these revealed insufficient, users attempted to instruct the robot by describing actions instrumental to movement in general.[2] Thus, the order of instructions shows the following hierarchy of difficulty and basicness:

goal description
> path description
> movement description
> description of actions instrumental to movement

An example which shows the development from path descriptions to the description of actions instrumental to movement and the users' attention to orthography is the following:

Command:	*Fahre nach rechts vorn.*
	[drive straight ahead to the right]
Command:	*Drehe dich um 45 Grad nach rechts.*
	[turn 45 degrees to the right]
Command:	*Drehe dich nach rechts.*
	[turn to the right]
Command:	*Fahre 10 cm nach vorn.*
	[drive 10 cm ahead]
Command:	*Los*
	[come on]
Command:	*Fare los.*
	[start driving]
Command:	*Fahre los.*
	[start driving]
Command:	*Motor an.*
	[engine on]

Because the experimental situation for those subjects who started off with the 'wrong' strategy from the outset was so depressing, the conductor of the experiments sometimes attempted to make the subjects change their strategy (these attempts were, of course, recorded and documented). However, participants turned out to be very hesitant about changing their ways of instructing the robot, that is, they did only reluctantly change their instructional strategy if that meant breaking the order described.

5. Mental Models of Human-Robot Spatial Instruction

Our hypothesis is that the fixed order of instructional strategies reflects the participants' mental model (Gentner and Stevens, 1983; Allen, 1997) of the domain of spatial instruction: namely that they regard knowledge about how to move along a path instrumental to moving towards a goal object, that they regard knowing how to move at all instrumental to moving along a path, and that they consider knowing about how to use one's facilities for moving instrumental for moving. Moreover, participants appear to be unable to imagine that a robot could know how to move to a goal object without being able to *understand* path instructions.

Here a further interesting aspect emerges: While

[2] There was only one participant who did not employ the strategies in this order but who explicitly raised the question of whether goal or path descriptions would be appropriate during self-talk. After having started with two unsuccessful path descriptions, he switched to goal descriptions, which happened not to be understandable to the robot either. His later attempts exhibit the same order as found for the other participants.

knowledge about how to move along a path may indeed be a necessary prerequisite to moving towards a goal object, the problem in this case was not that the robot would not have known how to *move* along a path, it just could not *understand communication* about it. The users' conceptualizations of the robot and its navigational capabilities thus exhibit the following properties: Robots may need more basic strategies than goal description, which are used in natural conversations among humans. This is supported by the fact that participants, often even explicitly, assumed the robot's point of view. That is, participants were consistently found to use the communication partner's perspective in deciding, for instance, what is left or right or what is front and back, a strategy taken in human-to-human conversation only with somehow restricted communication partners, such as children.[3] The second part of the mental model is that human users seem to assume that the robot's capabilities are directly reflected in its communicative abilities; if it does not *understand* a path description, it is assumed not to *know* how to move along a path. For a robot, its cognitive and communicative abilities may be completely distinct; they may even be, as in our case, implemented by different system developers. This does not seem to be what humans expect about their communication partner. Besides the mental model of the robot, the users' communicative strategies reveal furthermore a mental model of spatial instruction. This model implies a fixed hierarchy of difficulty and basicness of spatial concepts, which may be due to the users' own embodied experience with spatial navigation.

For the cognitive modelling of robotic systems these findings may have direct consequences; in particular, for half of the participants the experiments were less an effective dialogue with the robot than a depressing or amusing experience of communicative failure. In order to create efficient human-robot communication, system designers may therefore have to consider the mental models users employ.

6. Conclusion

The idea underlying the current paper is that the communicative strategies humans employ in human-robot communication point to their mental models guiding their instruction of robots for moving in space. Experiments in human-robot interaction were carried out which show that the participants' communicative strategies to approach an artificial communication partner differ very much from those found regarding spatial reference among humans. In particular, robots were treated as a communication partner whose point of view needs to be taken, like with children, and who needs more basic instructions than a human interlocutor. Moreover, participants found it impossible to imagine that a robot could do something that it could not communicate about, contrary to the real capabilities of the robot under consideration and possibly of artificial systems in general. These aspects of the conceptualization of the communication partner should be considered in modelling the cognitive structure of artificial communicators (see also Fischer (2000)).

Furthermore, a mental model of spatial instruction was identified which ranks the different types of spatial instruction according to their basicness and difficulty. The robot's inability to respond to instructions assumed to be more basic led to communication breakdown. While the work on adaptation processes in human-computer interaction and also the adaptation processes recorded in this study (Moratz and Fischer, 2000; Fischer, 2001) point to the fact that users will adapt to a number of linguistic and behavioural peculiarities of robots with ease, the mental model of basicness and difficulty elicited in the experiments described is obviously very difficult to overcome, as is shown by the participants' reluctance to change their strategies. This means that we have to account for the mental model of spatial navigation users assume in the interaction with the robot if we want to create successful human-robot communication. In this respect it may also be telling that most of the participants who used the type of instruction that the robot was implemented for were computer scientists; it seems that especially for interfaces with computer naive users their mental models have to be considered. Thus, future design of a robot for spatial instruction has to account for the hierarchy of difficult and of basic mechanisms for spatial navigation *plus* means to communicate about them, as the users of such robots expect. Analysing the communication between humans and robots seems to be a suitable way to elicit such mental models.

[3] This finding may, unfortunately, also be an artifact of our scenario design in which the participants had to turn between looking at the scene and typing into the computer.

References

Allen, R. (1997). Mental models and user models. In M. Helander, T. Landauer, and P. Prablu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed.). Amsterdam: Elsevier.

Amalberti, R., N. Carbonell, and P. Falzon (1993). User representations of computer systems in human–computer speech interaction. *International Journal of Man–Machine Studies 38*, 547–566.

Fischer, K. (1999). Repeats, reformulations, and emotional speech: Evidence for the design of human-computer speech interfaces. In H.-J. Bullinger and

J. Ziegler (Eds.), *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany*, pp. 560–565. Lawrence Erlbaum Ass., London.

Fischer, K. (2000). What is a situation? *Gothenburg Papers in Computational Linguistics 00-5*, 85–92.

Fischer, K. (2001). How much common ground do we need for speaking? In P. Kühnlein, H. Rieser, and H. Zeevat (Eds.), *Proceedings of the 5th Workshop on Formal Semantics and Pragmatics of Dialogue, Bi-Dialog 2001, Bielefeld, June 14-16th, 2001*, pp. 313–320.

Gentner, D. and A. Stevens (Eds.) (1983). *Mental Models*. Hillsdale: Erlbaum.

Habel, C., B. Hildebrandt, and R. Moratz (1999). Interactive robot navigation based on qualitative spatial representations. In I. Wachsmuth and B. Jung (Eds.), *Proceedings Kogwis99*, St. Augustin, pp. 219–225. infix.

Herrmann, T. and J. Grabowski (1994). *Sprechen: Psychologie der Sprachproduktion*. Heidelberg: Spektrum Verlag.

Hildebrandt, B. and H.-J. Eikmeyer (1999). *"Sprachverarbeitung mit Combinatory Categorial Grammar: Inkrementalität & Effizienz"*. Bielefeld: SFB 360: Situierte Künstliche Kommunikatoren, Report 99/05.

Levelt, W. J. M. (1996). Perspective Taking and Ellipsis in Spatial Descriptions. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett (Eds.), *Language and Space*, pp. 77–109. Cambridge, MA: MIT Press.

Levinson, S. C. (1996). Frames of Reference and Molyneux's Question: Crosslinguistic Evidence. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett (Eds.), *Language and Space*, pp. 109–169. Cambridge, MA: MIT Press.

Levow, G.-A. (1998). Characterizing and recognizing spoken corrections in human-com puter dialogue. In *Proceedings of Coling/ACL '98*.

Long, M. H. (1982). Adaption an den Lerner. Die Aushandlung verstehbarer Eingabe in Gesprächen zwischen muttersprachlichen Sprechern und Lernern. *Zeitschrift für Literaturwissenschaft und Linguistik 12*, 100–119.

Moratz, R. (1997). *Visuelle Objekterkennung als kognitive Simulation*. Diski 174. Sankt Augustin: Infix.

Moratz, R. and K. Fischer (2000). A cognitive model of spatial reference for human-robot communication. In *Proceedings of the 12th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2000, November 13-15 2000, Vancouver, British Columbia, Canada*, pp. 222–228.

Oviatt, S., J. Bernard, and G.-A. Levow (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech 41*(3-4), 419–442.

Oviatt, S., M. MacEachern, and G.-A. Levow (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication 24*, 87–110.

Roche, J. (1989). *Xenolekte. Struktur und Variation im Deutsch gegenüber Ausländern*. Berlin, New York: de Gruyter.

Röhrig, R. (1998). *Repräsentation und Verarbeitung von qualitativem Orientierungswissen*. Hamburg: Universität Hamburg, Dissertation.

Sacks, H., E. A. Schegloff, and G. Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language 50*(4), 696–735.

Schegloff, E. A. (1972). Notes on a conversational practise: Formulating place. In D. Sudnow (Ed.), *Studies in Social Interaction*, pp. 75–119. New York: Free Press.

Steedman, M. (1996). *Surface Structure and Interpretation*. Cambridge, MA: MIT Press.

Zimmermann, K. and C. Freksa (1996). Qualitative spatial reasoning using orientation, distance, and path knowledge. *Applied Intelligence 6*, 49–58.

Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *International journal of Man-Machine Studies* (34), 527–647.