# Communicative Aspects of Human-Robot Interaction

Thora Tenbrink
Department for Informatics
Spatial Cognition Priority Program
Vogt-Kölln-Str. 30
D-22527 Hamburg, Germany

tenbrink@informatik.uni-hamburg.de

## 1. Introduction

How do we communicate with robots? People are today increasingly used to interacting linguistically with computer systems, such as telephone answering machines that are able to ask about the user's aims in order to inform them, for instance, about train connections. In contrast to such systems made purely for dialogue or others which may be used to solve specific computational tasks, most robots have to deal with their spatial surroundings, as they are built to move and navigate in the real world. Human-computer interaction in general is a field that, although it has been adressed in several investigations (e.g., Amalberti et al., 1993; Zoltan-Ford, 1991), still leaves many questions unanswered. However, there has been even less research on how humans communicate with robots as compared to human-computer interaction in general.

Moreover, as has become clear from decades of intensive research on spatial cognition (e.g., Levinson 1996), it is by no means intuitively obvious how humans usually communicate about space, while space and spatial communication is central to dealing with robots. Much variation, even among humans, is possible concerning, for instance, reference systems (Eschenbach et al. 1998), levels of granularity in referring to objects situated in space (Habel 1991), and levels of specification regarding the task to be fulfilled (as observed by Moratz & Fischer 2000 in an experiment on human-robot communication). As communicating with robots navigating in space necessarily involves communicating about space, it is necessary to determine how humans would naturally refer to their spatial surroundings when interacting with a robot.

One of the major decisive aspects in the design of computer systems that are built for interacting with humans is that, by far, not every aspect known about human-computer interaction, about space, or regarding robot functionality can be implemented. In the worst case, the selection comes about more or less accidentally, based on the designer's predispositions, on the system's computing limitations, and on the available algorithms. Ideally, a careful concept of what to consider and which aspects to leave out is developed on the basis of the intended functions. Regarding a dialogue module, in the latter case the designer needs information about which aspects of linguistic communication will have to be implemented for successful and effective human-computer interaction.

This paper will outline some basic aspects of what characterizes human-robot interaction in contrast to other kinds of interaction, such as communication with children or foreigners. Here, the robot's looks – humanoid or not – will not play a major role. The question at hand is rather whether or to what degree humans expect a robot to behave linguistically like a human being.

The paper is structured as follows. I will start with some general remarks about modern robots' strengths and weaknesses which may or may not be known to human users, including some specific aspects concerning dialogue systems. Next, I will turn to natural human-human interaction to point to a selection of relevant linguistic insights regarding how humans normally talk to each other. Then, I will contrast these points with robot interaction partners, and point out why communication may fail in that specific kind of communication. The paper will conclude with some suggestions concerning how to account for such communication failures by designing intelligent output that anticipates human users' linguistic actions and reactions.

## *2. Robot Functionalities*

Modern computer systems, especially robots, are built for interacting with human users to the purpose of fulfilling more or less specific tasks. Although industry or service robots sometimes create the impression of acting fairly intelligently, they can do only what their designer has created them for. They are equipped with specific perception, language understanding, computing, and task fulfilling abilities which enable them to function according to the user's expectations. Thus, robots may be built for entertainment such as Sony's "dog" that basically behaves like a dog, but that is also being trained to learn language (Kaplan 2000). Other robots may be designed for service, for industry, medical help, or for the assistance of handicapped people (Röfer & Lankenau, 2000).

For spatially situated robots, there are limitations regarding either computing power or computing time (or both). Robots need to know and be able to deal with different aspects of the world than other computers do, such as information about their spatial surroundings, perception, navigation facilities, etc. All robots built today are limited to their own, highly specified functions which need to be carefully designed in order not to overload the system, or collapse for other reasons. For instance, the robot "Kismet" (Breazeal & Scassellati, 1999) consists of a stereo active vision system augmented with facial features for emotive expression. This robot, which is able to react to visual input by appropriate facial expressions in a rather convincing way, has no further abilities such as moving around in space, or talking. Similarly, other robots might be built solely to walk up stairs, or, involving very different kinds of problems, to participate in a robot soccer game (e.g., Weigel et al. 2000).

Moreover, there are limitations regarding the kinds of features, or abilities, to be implemented. Generally, the perceptual abilities of robots are on a strictly metric (or quantitative) level. Standardly designed robots are capable of measuring exact distances, and they can distinguish formally defined colors. Anything that is abstractly definable and formally translatable into code can be implemented fairly easily. Qualitative information such as "to the left of", however, has to be translated into formal calculi in order to enable artificial systems to use it in a way that resembles human usage. Although humans generally know what "to the left of" means, this kind of reference is always dependent on the reference system that is used, yielding complications in generating a formal definition. Thus, "to the left of" can be intended to mean to the left of another object from the speaker's, the listener's, or a third entity's point of view (Tenbrink & Moratz, forthc.). Humans usually decide about the intended reference system by means of extralinguistic cues such as pointing, a shared focus of attention, or a walking direction. However, in cases of doubt, they will normally ask appropriate questions to confirm their assumption.

With regard to colors, the formal specification (a quantitative kind of information) reflects only part of what humans use to categorize their color perception (Belpaeme et al., 1998). With changes of light, the same color may be perceived as decisively different by both human and robot perception systems. Humans, however, would nevertheless tend to assign the same category to the color because they – perhaps unknowingly – take into account the light's influence. Thus, their categorization systems are to be regarded as both quantitative and qualitative. In order to enable a robot to do likewise, such human knowledge – which humans possess and deal with without necessarily being aware of it – needs to be implemented.

Much effort is today invested into enabling robots to act and communicate more like humans. For example, research in the area of spatial cognition carried out in the Spatial Cognition Priority Program (SPP-RK; supported by the DFG, German Science Foundation) aims at specifying human knowledge about space. How do we acquire knowledge about spatial surroundings, how do we use it in order to fulfill spatial tasks, how do we reason about spatial relations? Answers to these questions are modelled and implemented into computer systems. Thus, SPP researchers (Moratz et al., 2000) have enabled a robot to distinguish between human categories such as *left* and *right*. The remarkable point is that this robot does not need nor use any information about metrically specified distances and angles. It is built to act as a human might act, based on cognitively adequate modelling of spatial relations.

Enabling a robot to distinguish left and right is not quite the same as making a robot understand natural language instructions like "go to the left (or right)". Robots consist of a variety of modules each of which fulfills its own function. The integration of the modules then leads to the robots' capability of fulfilling specific tasks. Thus, if equipped with a language module, robots can communicate linguistically with humans; and, furthermore equipped with the necessary task fulfilling modules, the robot may also act according to the human's instructions. Moratz' robot, after having been equipped with the module that enables it to act like a human, was then connected to a language understanding system (Hildebrandt & Eikmeyer 1999) that contained instructions like "go to the left". Accordingly, this robot could both parse the meaning of this natural language instruction, and act appropriately because it also possessed the necessary equipment for acting.

Like the other modules, language understanding systems are also subject to strong limitations caused by the need of computing power, the size of the linguistic database, etc. One problem that applies for dialogue systems perhaps more than for most other modules is the impact of computing time. If the system needs several minutes to compute an answer to a given question, or to react to a linguistic instruction, communication will almost certainly fail. Johnstone et al. (1994) describe how natural turn-taking is hampered by the system's unnaturally long pauses, yielding overlapping speech and misunderstandings of pragmatic intentions. Such problems might lead system designers to decide that parsing must necessarily happen on a shallow level in order to facilitate fluent dialogues. The drawback of this is that there will, in all likelihood, be many cases of misinterpretation, causing severe communication failures.

A further problem is that humans addressing robots may not use whatever linguistic level they choose. To the contrary, robots can understand and produce linguistically exactly what their language module allows them to. Thus, each linguistically active robot needs its own domain-dependent language module which is appropriate for the needs of the specific interaction situation that the robot is intended for.

From the user's perspective, there is a high potential for communication problems. At least, some information about the robot's linguistic equipment is necessary. In the worst case, the users need to be explicitly informed about the exact commands that their artificial interaction partner can deal with. Thus, they are confronted with an interaction situation which is artificial and uncomfortable because of its unfamiliarity compared to other kinds of linguistic interaction.

## ***3. Varieties of linguistic interaction***

Humans interacting with humans usually do not have to consider (consciously) the limitations of a specific interaction situation. They are not used to talking to an interaction partner that does not know anything but a single domain. Generally, humans rely strongly on shared world knowledge, and they refer to extralinguistic (mostly, visual) aspects of their surroundings perceived in much the same way by both interaction partners. To a high degree shared by both is also what they know about conventions of communication, such as those specified by Grice (1975) and in the field of discourse analysis (Quasthoff 1995; Brown & Yule 1983).

However, register theory reveals that humans adapt considerably to what they believe about their communication partner. For example, humans talking to people who look like foreigners are likely to switch to a register known as "foreigner talk", which is characterized by short sentences and words, and at times even a reduced grammar. Likewise, children are usually addressed by their parents (and other adults) in what has become known as "motherese" or "babytalk". Adults take into account that children have little linguistic expertise and therefore use simplistic expressions which are sometimes transformed to diminutives (which some people like to use in repeated formats, such as *sleepy-sleepy* or the like). Moreover, they tend to use common, preferably short words and not too specified vocabulary, as well as shorter sentences with fewer subordinations than towards other adults. Newport et al. (1977) point out that "features of maternal speech related to conversational meaning correlate with a wide variety of measures of the child listeners, including age, vocabulary size and syntactic sophistication." Thus, it is well-known that humans are quite good at adapting to interaction partners who have different cognitive abilities or a different level of linguistic sophistication than they have themselves.

A closer look at the interaction situations just outlined reveals several patterns that are not exactly reflected in human-robot interaction. Foreigners, on the one hand, may be conceptualized as having much knowledge in many domains in the world, conceivably on about the same level as their native interaction partner. All they lack is sophistication in the language that is required in that specific situation. Children, on the other hand, have only little knowledge of the world in addition to a low level of linguistic expertise. However, they do possess a small but growing repertory of concepts which are best characterized as qualitative. This repertory is not confined to any specific domain, but rather limited to the children's experiences in their everyday life.

These patterns, which can reasonably be assumed to be part of every human being's world (and discourse) knowledge – again, without a need of awareness of such knowledge during a given interaction situation – shed light on some major differences to human-robot communication.

Like in "foreigner talk" or "motherese", human users are required to adapt to the conceptual and linguistic limitations of their interaction partner. However, they are generally only poorly informed about the robot's linguistic (and functional) abilities. Moreover, they lack experience with robot interaction partners, as many users, especially participants in human-robot interaction experiments, have never before dealt with a robot, which is in clear contrast to their experience with foreigners and children. Worse still, robots may differ considerably in their abilities as they are built specifically for different functions.

In general, it can be assumed that robots have little or no human-like general knowledge of the world, but highly specified knowledge concerning one specific domain, which is overwhelmingly of a quantitative kind, and some robots also have linguistic knowledge in that specific domain. That this is the case, however, is not so obviously apparent that every human user can be assumed to be aware of these facts about robots. However, they can safely be assumed to be relevant to effective human-robot interaction.

## *4. Adaptation to robots*

Little is known so far about what human users actually tend to believe about their robot interaction partners. What linguistic expertise do human users expect of the artificial tool they are interacting with? What are their – conscious or unconscious – hypotheses about a robot’s perceptual equipment? How does their conceptualization of the robot’s abilities influence their communication strategies?

Experimental work (Moratz & Fischer 2000) reveals that even those users who knew only very little about robots, interacted with the robot in a way that reflected their beliefs about robots, of which they were – according to their verbal comments during the experiment, and the questionnaires they filled in afterwards – only partly aware. The users' linguistic behavior reflected underlying hypotheses concerning, for instance, the robot's linguistic expertise ("does it speak German or English?" "what kind of vocabulary do I have to use?"), its perceptual equipment ("what does the robot see? where is its front?"), and its knowledge of the world. The users varied their utterances according to their beliefs about what might have been the cause for occurring communication problems. It should be added that they did not get any help by the system at all, because all that the users got for an answer, when communication failed, was "error".

Many of the human users did not expect their artificial interaction partner to understand qualitative instructions like "turn to the left". This is reasonable because robots are, up to now, rarely equipped with knowledge about qualitative human concepts. Thus, the system users considered that the robot might need quantitative information about distances, angles, etc.; and tried to provide such kinds of information accordingly.

Moreover, some people felt inclined to instruct the robot not on the basis of human conceptualizations such as specifications of goals ("go to the left box"), but rather in terms of smaller units such as path descriptions ("go to the left"), descriptions of movements ("move forward") or subsidiary actions like turning particular wheels. After having tried, and failed with, an instruction containing a goal or path, users turn to more basic conceptions such as specific descriptions of movements to make themselves understood. However, almost none of the human subjects in the experiment used a goal instruction after having failed with an instruction involving subsidiary actions.

Conceivably, the motivation behind this is that goals and paths are more complex to their interaction partner than minor actions specified in more detail. After all, the instruction "go to the left box" requires the robot, for instance, to locate the intended object by perceiving and identifying it correctly, to match this identification to the utterance, to move its wheels in the correct direction, and to stop in appropriate distance with regard to the box. The users in the experiment seemed to think that if they explain to the robot in detail what it needs to do, they make it easier for the robot. The more complex the action is considered, the more difficult it will be for the robot to understand the corresponding instruction, and act accordingly. Thus, easier or less complex instructions entailing only few subsidiary actions will be understood more easily.

However, this requires the robot to understand as much as it can do. Although people, as described above, seemed to know that robots do not know much language and that they need to adapt their vocabulary and linguistic style to the robot's abilities, they seemed to assume unconsciously that the robot needs to understand at least as much on a linguistic level as it can do functionally.

Of course, robots function in a completely different way. A robot's language understanding system does not necessarily correspond exactly to its task fulfilling abilities. Thus, the language module may entail instructions like "Turn to the left"; other modules may enable the robot to fulfill all of the necessary actions like moving the wheels; yet the language module does not need to be prepared for instructions like "Move your wheels".

That actions and the language used to describe them may not coincide is not altogether new to humans. Young children are much more likely to react to utterances like "Go to the door" than to

"Put one feet before the other, using your muscles, until you reach the rectangular opening". Parents usually exhibit no difficulties in using the kind of language that their children can understand, and they know that the child will use her legs to reach the door without explicitly having been told so. Thus, they naturally use a goal description (at least in cases where the goal is visible) without speculating that the task might be too complex.

However, parents share a lot of abilities, knowledge, and common ground with their children, and they can also remember their own childhood. This makes it a lot easier to find the kind of language that the child can understand. Humans interacting with robots lack that knowledge (see Fischer, forthc.), so they have to rely on their hypotheses.

Obviously, communication failures between human users and robots are not due to the humans' general inability to adapt to interaction partners with different cognitive abilities than they possess themselves. If they do not know the exact abilities of their interaction partner, e.g. in talking to a foreigner, they use clues of the ongoing dialogue to find out, and they adapt rather fast to what they have found.

In human-robot interaction, however, inadequate conceptualizations of the robot's functionality and lack of knowledge concerning its linguistic and technical features may influence the success of the dialogue to a high degree. Thus, if there is an unconscious underlying assumption that robots need to understand as much as they can do, more than the usual discourse adaptation processes might be needed to rule out such an assumption. In order to communicate, human users need information concerning some general features of robots, and, more specifically, the linguistic and functional abilities of the robot they are dealing with. If the ongoing discourse does not provide any specific and well-tuned clues concerning these facts, the human users might have to try out many different kinds of variation regarding differing levels of linguistic interaction before they find out how to communicate with their artificial interaction partner.

## *5. Intelligently designed dialogue systems*

The insight that communication between humans and robots is complicated by lack of knowledge does not necessarily lead to the conclusion that human users need to be instructed specifically before they are in a position to approach a robot linguistically. Instead, it should be possible to enable people to learn and adapt while using and talking to the robot. In the experiment described above, the robot offered next to no feedback at all. Thus, the human users had almost no chance to find out what to do to get through to their interaction partner, except if they happened to use the right kind of instruction by chance. In a different scenario, however, the robot, instead of waiting passively for the user to give instructions, might initially ask a question such as: "Which of the three boxes shall I go to?". By way of this one short question, users can extract information about several useful issues at once: First, they can conclude that the robot already perceives a group of objects such that they may instruct it as to which one of them it should approach. Second, they do not have to worry about goal instructions being too complex for the robot to fulfill, because the robot already asked about the goal itself. Third, the kind of language to be used does not need to be wondered about, as the question is stated in English, and the object is identified by an intuitively suitable label, namely, "box".

However, not all communication problems are solved this easily. In more complex robot instruction scenarios, language understanding systems are needed which react to the users' linguistic input so that the users are, where necessary, informed about the specific features of the system they are dealing with. This enables them to address the robot in an adequate manner. Thus, a robot that is not equipped to understand comparisons might answer to an instruction like "Go to the larger box" by: "Sorry, I didn't understand. Shall I go to the leftmost box from my point of view?", thereby telling the user unobtrusively that it can understand qualitative directions but not comparisons. Regarding the ambiguity of spatial reference systems, a robot needs to be able to determine which point of view human users are likely to use, even if they do not explicitly state it. Thus, "go to the left" is ambiguous in itself; however, if human users are predisposed to use the robot's point of view, the ambiguity can be accounted for by using the robot's perspective per default.

Experiments like the above, and experiments that specifically address various issues such as naturally used reference systems, are necessary to work out human users' underlying assumptions about (linguistically equipped) robots, and their natural predispositions, in order to build dialogue systems in a way that accounts for them. "Intelligent" systems, in the end, will be those that seem to anticipate what their users believe and expect, and behave accordingly, both functionally and linguistically.

Human users react strongly to the linguistic output of artificial systems (Fischer 2000), adapting to what the robot's feedback suggests to them about its functionality. Because the users usually have a strong motivation to be understood by their interaction partner, they acknowledge feedback very fast and use it to adapt to the system. Thus, if the system’s output provides human users with useful information concerning the robot’s perceptual and linguistic abilities, this enables them to update their knowledge about robots and dialogue systems – especially, the robot and dialogue system they are actually dealing with –, and to adapt their communicative strategies to the situation-specific requirements. The humans' conceptualization of their interaction partner changes through the influence of the system's output, yielding considerable differences in the users' linguistic behavior. For example, although Amalberti et al. (1993) report that human users tend to regard an artificial dialogue system as a tool rather than a participant in collective problem solving, later research (Fischer & Batliner 2000) reveals that people react to apologies from the system by calming down and getting more cooperative, even when they were just getting angry because of system malfunctions. Such a reaction can be regarded as similar to interaction with a human being. Thus, an intelligently designed dialogue system should be able to trigger human users' linguistic reactions in a communicatively effective way.

Effective communication is dependent on mutual understanding. By themselves, robots cannot understand; implying that their "understanding" needs to be implemented in as much detail as possible. Conversely, what human users do not understand at first, they must be told by the system. More experiments on human conceptualizations are needed to work out in more detail what kind of information or feedback would enhance effective communication. Ultimately, only an adaptive language module that is not only able to anticipate human user's reactions, but also react appropriately to utterances that reveal misconceptions, can solve and prevent severe communication problems between human and artificial interaction partners.

## References

Amalberti, R., N. Carbonell, & P. Falzon. 1993. User Representations of Computer Systems in Human-Computer speech interaction. *International Journal of Man-Machine Studies*, 38:547-566.

Belpaeme, T., Steels, L. & Van Looveren, J. 1998. The Construction and Acquisition of Visual Categories, in Birk, A. and Demiris, J. (eds.), *Proceedings of the 6th European Workshop on Learning Robots*, Lecture Notes on Artificial Intelligence, Springer.

Breazeal, C. & Scassellati, B. 1999. A context-dependent attention system for a social robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence* (IJCAI99). Stockholm, Sweden. 1146—1151.

Brown, G. & Yule, G. 1983. *Discourse analysis*. Cambridge: Cambridge University Press.

Eschenbach, C., C. Habel & A. Leßmöllmann 1998. Multiple frames of reference in interpreting complex projective terms. In Patrick Olivier (ed.), *Spatial Language: Cognitive and Computational Aspects*. Dordrecht: Kluwer.

Fischer, K. (forthc.) How much common ground do we need for speaking? In *Proceedings of Bidialog 2001, Fifth Workshop on the Semantics and Pragmatics of Dialogue*, June $14^{th}$ - $16^{th}$, 2001 at ZiF, Bielefeld.

Fischer, K. 2000. What Is a Situation? *Proceedings of Götalog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue*, Göteborg University, 15-17 June 2000. Gothenburg Papers in Computational Linguistics 00-5 (pp. 85-92).

Fischer, K. & Batliner, A. 2000: What Makes Speakers Angry in Human-Computer Conversation. In *Proceedings of the Third Workshop on Human-Computer Conversation*, Bellagio, Italy, 3-5 July 2000.

Grice, H. P. 1975. Logic and conversation. In: Cole, Peter; Morgan, Jerry (Eds.), *Syntax and semantics*. New York, San Francisco, London (Vol 3: 41-58).

Habel, C. 1991. Hierarchical Representations of Spatial Knowledge: Aspects of Embedding and Granularity. *Second International Colloquium on Cognitive Science (San Sebastian, 7-11. May 1991).*

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

Johnstone, A., U. Berry, & T. Nguyen. 1994. There was a long pause: influencing turn-taking behaviour in human-human and human-computer spoken dialogues. *Int. J. Human-Computer Studies 41*, 383-411.

Kaplan, F. 2000. Talking AIBO: First Experimentation of Verbal Interactions with an Autonomous Four-legged Robot. In Nijholt, A., Heylen, D. & Jokinen, K. (eds.), *Learning to Behave: Interacting agents*. cele-twente Workshop on Language Technology, pp. 57-63.

Levinson, S.C. 1996. Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M.A. Peterson, L. Nadel & M.F. Garrett (eds.), *Language and Space* (pp. 109-169). Cambridge, MA: MIT Press.

Moratz, R. & K. Fischer. 2000. Cognitive Adequate Modelling of Spatial Reference in Human-Robot Interaction. *International Conference on Tools with AI*, Vancouver 2000.

Moratz, R., Renz, J., & Wolter, D. 2000. Qualitative Spatial Reasoning about Line Segments. In W. Horn (Ed.), *Proceedings of the 14th European Conference on Artificial Intelligence* (ECAI'00). Amsterdam: IOS Press.

Quasthoff, U.M. (ed.). 1995. *Aspects of oral communication*. Mouton: de Gruyter.

Röfer, T., & Lankenau, A. 2000. Architecture and Applications of the Bremen Autonomous Wheelchair. In P. Wang (Ed.), *Information Sciences 126*:1-4. Elsevier Science BV (pp. 1-20).

Tenbrink, T. & Moratz, R. (forthc.). Group-based spatial reference in human-robot interaction. *5. Fachtagung der Gesellschaft für Kognitionswissenschaft*, 25.-28.9.2001, Leipzig.

Weigel, T., Auerbach, W., Dietl, M., Dümler, B., Gutmann, J.-S., Marko, K., Müller, K., Nebel, B., Szerbakowski, B., & Thiel, M. 2000. CS Freiburg: Doing the right thing in a group. In P. Stone, G. Kraetzschmar, & T. Balch (Ed.), *RoboCup-2000: Robot Soccer World Cup IV*. Berlin, Heidelberg, New York: Springer.

Zoltan-Ford, E. 1991. How to Get People to Say and Type what Computers Can Understand. *International Journal of Man-Machine Studies* 34: 527-547.