KERSTIN FISCHER (BREMEN)

Linguistic Methods for Investigating Concepts in Use

Abstract

In this paper, we present a method for investigating concepts in use. Besides providing an inventory of mental representations as they can be inferred from the speakers' linguistic behaviour, also their conditions of usage are being analysed. The methodology proposed rests on the assumption that common sense reasoning is observable in the speakers' displays of their interpretations of the preceding utterances, which in the human-robot communication data investigated here are reformulations of the previous utterance. The linguistic choices speakers make and the order in which they appear allow us to infer the speakers' underlying hypotheses about language, the communication partner, and the world.

1 Studying Concepts in Use

The focus of this paper is on how our mental representations guide our linguistic output. Thus, we are interested in how humans structure their experience and represent information about the world, for instance, about the orientation of objects in a room or an agent's movement through space, and the linguistic consequences of such representations. For example, we can assume that the way spatial configurations, such as rooms with objects in them, are mentally represented may influence the way speakers talk about them. The most natural source to look at when we want to find out about what goes on in people's heads may be psychology or, with respect to language, psycholinguistics. Thus, we may want to look at psychological literature to understand speakers' mental representations of spatial configurations.

Experimental psychology provides us with a rich body of data on how humans may store spatial knowledge. For instance, a large body of research shows that one such representation of spatial knowledge may be as a mental model (e.g. JOHNSON-LAIRD 1989). Mental models are analogous to imagined diagrams of states of affairs, they are calculable, finite, economical (such that there may be a few and only a few alternative representations), and their inner structure corresponds to the situation represented (HABEL 1995). However, spatial reasoning does not seem to rely exclusively on mental models; according to ROBERTS & NEWTON (2001), people may also employ strategies that point to very different underlying mental representations. For instance, participants in experiments may be shown to be field dependent or field independent, assimilators or explorers, adaptors or innovators, visualizers or verbalizers, holists or serialists, or convergent or divergent thinkers (2001: 142). Thus, there may be competing strategies based on different underlying representations and various ways of using them among different speakers. According to ROBERTS & NEWTON (2001: 140-141), experimental psychology so far does not account for the variability observable in strategy selection. The reason may be that psychological experiments, forced by the methodological constraints of the discipline, may have to be so controlled that they do not allow sufficiently detailed inferences about the speakers' real representations in actual situations and how they make use of them. Thus, it is unclear how people behave in unrestricted settings in which they have several possible strategies available and can choose freely among them. The conditions for building on one representation rather than another, the motivations for strategy selection, are thus not clear, and a different set of methods is necessary for identifying representations in use.

In contrast, variability in linguistic strategies has been investigated in great detail in register theory, in which it is held that the situation has a great impact on the speakers' linguistic choices (e.g. HALLIDAY 1985). However, what we can observe when we look at situated talk is that the speakers' concepts of the situation may be very different, and that these differences are strongly reflected in the participants' linguistic behaviour. What a situation consists in is therefore also determined by the speakers' mental representation of it (FISCHER 2000). Thus, the problem is not only that we find a diversity of possible representations, but there may also be a variety of strategies of using them, depending on the speakers' conceptualization of what the situation consists in. We therefore need to look at actual situations and investigate the concepts in use. The question then is how far linguistic methods can support answering these questions.

2 Using Human-Robot Communication as a Methodological Resource

The method proposed relies on perspectives developed in the framework of ethnomethodological conversation analysis (CA), which aims at describing "how common-sense reasoning and social action work" (HERITAGE 1988: 127). While the problems dealt with in conversation analysis traditionally center around the sequential organisation of conversations, the analysis of practises in dealing with aspects of the context, such as the recipient (e.g. SCHEGLOFF

1972), categories like gender (e.g. SCHEGLOFF 1997), or the impact of institutional settings (e.g. HERITAGE & SORJONEN 1994), have continuously been subject to empirical analysis as well.

One of the main assumptions of CA is that a description of interaction should only rest upon those descriptive categories of which it can be shown that the speakers themselves attend to them, i.e. that they are not carried into the analyses as explanatory background by the analyst. This methodological precaution guarantees the idenfication of only those conditioning factors to which speakers are oriented. Conversation analysis has therefore concentrated on the observable, on aspects that leave their traces in the sequential organisation of talk. Thus, SACKS (1985: 20) holds that for members, that is, the participants in a speech community, "activities are observables. They see activities." This is the case because "persons go about seeing activities (...) by reference to some procedures that they take it properly occur as the activities occur" (ibid.). As such, "observability is not specific to each activity but is learned as a general phenomenon" (SACKS 1985: 20). For instance, it is very natural for parents to see that their child has brushed her teeth by seeing the traces of tooth paste on her cheeks. Similarly, speakers in a conversation make inferences about the activities their communication partners are engaged in by using a huge repertoire of methods (hence: ethno*method*ology) and thus arrive at interpretations of each others' utterances.

What we can observe in interactions, then, are the traces of a number of methods the speakers themselves make use of. "To recognize *what* is said *means* to recognize *how a person is speaking"* (GARFINKEL 1972: 320, emphasis original). Common understanding therefore "has necessarily an operational structure" (GARFINKEL 1972: 321). The displays of the procedures involved, however, may occur only for "all practical purposes" (NOTHDURFT 1984). It is therefore proposed here to focus on interactions in which "all practical purposes" make it necessary that much attention is devoted to the structures we are interested in. That is, we should elicit the data in situations in which speakers are "encouraged" to display their mental representations more overtly than it is necessary in smooth natural conversation among equal human interlocutors. In particular, it is proposed to exploit the conversation analytic notion of "deviant case analysis". Deviant cases are a useful resource for the analyst since in their interactive treatment speakers display their understanding of the regular phenomenon as normatively required and violations against it to be accountable (HERITAGE 1988: 135). If "someone displays in their conduct that they are 'noticing' the absence of a certain type of turn from a coparticipant, then that demonstrates their own orientation to the relevance of the sequence that the analyst is aiming to describe" (HUTCHBY & WOOFFITT 1998: 98).

The use of deviant case analysis we are making here however differs from the use made of it generally in CA. In CA, deviant cases are employed to show the speakers' normative orientation to the expected pattern, which is then of course more frequent than the deviant one (HERITAGE 1988: 135). The proposal made here, in contrast, is not to analyse basically smooth and unproblematic human-to-human conversation at all, but to focus instead on problematic human-to-robot (mostly mis-)communication. In particular, by being forced to produce several attempts as solutions to the same task, the types of information to which speakers attend become evident. The deviance of human-robot interaction compared to natural conversation therefore has the advantage of providing the analyst with a big number of strategies which display the speakers' orientation to particular categories. Human-robot communication is particularly suited since, in contrast to human-computer interaction in which similar problems occur, robots furthermore interact with their environments. The sequential positions of the utterances produced as solutions to the same problem reveal the speakers' hypotheses about what causes the communicative problems because they constitute reformulations of the previous utterances. They thus provide not only interpretative cues as displays of the speakers' own interpretation of their previous utterances, but also display the speakers' concepts of the determining factors of the current situation. These involve at least the speakers' concepts about language, the robot as a commuication partner, and the subject matter, in this case a spatial task.

An example from the data elicited in human-robot communication can show how speakers' linguistic behaviour can reveal their hypotheses about the reasons for the situation of miscommunication between speaker and robot. For instance, the following example from the data described below shows that the participant pays attention to the words she chooses in the interaction with the robot:

(1) VP12: steuere den rechten der drei knöpfe an *[aim at the the rightmost of the three buttons]*

Robot: error

VP12: fahre auf den rechten der roten punkte zu *[drive towards the rightmost of the red points]*

Robot: error

VP12: fahre zu dem rechten roten knopf *[drive to the right red button]*

Robot: error

VP12: steuere auf den rechten der drei roten würfel zu *[aim in the direction of the rightmost of the three red cubes]*

Robot: error

VP12: bewege dich zu dem rechten roten würfel *[move towards the right red cube]*

What can be seen here is that in a situation of miscommunication, the speaker proposes a number of solutions to achieve her task (to make the robot move to one of the objects in the room). Now, her linguistic choices as well as their variability are not arbitrary – with conversation analysis we hold that no details of production should be discarded prior to analysis (SACKS 1984) and that *how* something is said contributes to *what is being said* (GARFINKEL 1972). Thus, we can assume a certain orderliness of the speaker's utterances, and the fact that her utterances all constitute reformulations of each other reveals her hypotheses about what could improve the communication. The utterances get their relevance as indicators of such underlying concepts on the basis of their sequential placement: as solutions to the same problem. What counts as a problem as well as a possible solution is proposed by the speaker herself. In example (1), the speaker's attention to finding the appropriate word can be identified by the traces the hypothesis leaves in the linguistic choices she takes within the sequence: systematically trying out different nouns *(buttons, points, cubes)* and verbs *(aim, drive, move)*. We can therefore assume that this speaker believes the choice of vocabulary to be a potentially problematic aspect of human-robot interaction and that she can solve the communication problems at hand by finding those words the robot can understand. At the same time she displays her concept about the robot as a communication partner to whom not the same range of vocabulary is available and for whom words have to be chosen with particular attention. Thus, her attention also shows that lexical choice is part of her recipient design (SCHEGLOFF 1972).

Returning to the corpus, we can find other instances from other speakers that show attention to vocabulary. This aspect of recipient design is thus a more general phenomenon:

(2) VP15: fahre schräg nach rechts oben *[drive diagonally up to the right]*

Robot: error

VP15: bewege dich nach rechts oben *[move up to the right]*

The fact that this type of solution to the communication problem is not only taken in one instance but is attended to in several other situations of communicative failure with the same speaker as well as with several other speakers shows that the hypothesis that the choice of words constitutes a potential problem in human-robot communication is a systematic aspect attended to in the interaction between humans and robots (cf. FISCHER 2001).

Both human-computer and human-robot interaction furthermore provide us with additional data not usually available in human-to-human communication. Thus, users often produce self-talk in which they give accounts of their doings and in which they reveal their interpretations and explanations about what is going on. This behaviour can be explained on the basis of GOFFMAN's (1978) analysis of response cries in that speakers give accounts to overhearers, such as the person running the experiments, of their partly strange linguistic behaviour triggered by the system's malfunctions. The following two examples from human-to-computer appointment scheduling dialogues (e.g. FISCHER 1999) show how speakers, partly in self-talk, partly directed towards the system, announce their discourse strategies, here repetition. The first speaker does so by means of *well, again*, and the second more explicitly by announcing her strategy with *I repeat*:

(3)[1] e0375204: wie wäre es am Dienstag, dem neunzehnten ersten, *von zwö*lf bis vieRzehn Uhr? *[how about Tuesday, 19th of January, from 12 to 2pm?]*

s0375205: Mittwoch, der sechste erste, von acht bis zehn Uhr ist schon belegt. *[Wednesday 6th of January is already occupied.]*

e0375205: hm? tja. noch mal. wie wäre es mit Dienstag, dem NEUNzehnten ersten, von ZWÖLF bis vieRzehn Uhr? *[hm? well, again. how about Tuesday, 19th of January, from 12 to 2pm?]*

(4) e0194301: Die < L > nstag, zwö < L > lfter e < L > rster, < P > zwa < L > nzig bis zwei < L > undzwanzig U < L > hr. *[Tuesday, 12th of January, 8 to 10pm.]*

s0194302: ich habe Sie nicht verstanden. *[I did not understand.]*

e0194302: ich wiederhole. Dienstag, den zwölften ersten, zwischen zwanzig und zweiundzwanzig Uhr. *[I repeat. Tuesday, 12th of January, 8 to 10pm.]*

---

[1] Transcription Conventions: <L> : syllable lengthening; syllables in capital letters: strong stress; single letters in capital letters: hyperarticulated sounds.

Besides self-talk, the use of human-robot communication allows us to draw on further resources than are usually available to conversation analysts. Since both human-computer and human-robot interaction still often occur in laboratory settings because of the prototype nature of the systems involved, it is quite natural to have speakers fill out a questionnaire after the conversation with the system. These data can give important information as to which kinds of information speakers attend to and which they use as explanations for the communicative problems arising themselves (cf. FISCHER 2001). Regarding the above example (1), the speaker noted in the questionnaire after the recording that she believes that she did not always manage to find words the robot could understand. Thus, the speaker reveals that she attends to the choice of words as a conversational resource that is relevant in the interaction with this particular communication partner. Consequently the data support the previous analyses of her participant categories by means of evidence from those further data sources.

As another resource, besides the naturally occuring self-talk, also the “thinking aloud” technique can be used. This constitutes however a very unnatural situation and should therefore not be used. In contrast, a method called “constructive interaction” has already been employed successfully in system diagnosis (SUCHMAN 1987, DOUGLAS 1995). In this method, usually two speakers interact both with the system and with each other, and they may decide jointly about how to approach the system. The interactions therefore exhibit the participants’ hypotheses about the functioning of the system. For the current experiment, these kinds of data were not elicited, but use of the constructive interaction technique is being planned for future research.

Using human-robot interaction furthermore allows us to control a number of external factors without rendering the experimental situation implausible. In contrast to natural conversation the situation in which such dialogues are carried out are already very unnatural – even more so since robots are usually prototypes only available in research laboratories. However, participants usually accept the deviant situation, and display an amazing degree of patience with the shortcomings of such systems (BATLINER et al. forthcoming). The interaction with artificial systems can therefore be used for controlling individual factors of the communicative situation. Moreover, human-computer and human-robot interaction involve artificial communication partners which need not be “given” in the same way in which humans are “given”. Instead, artificial communication partners are being designed on the basis of computer scientists’ ideas about communication. That is, one of the speakers is made both transparent (regarding its functioning, motivations, intentions etc.) and manipulable to the analyst. Similarly, crucial aspects of the communicative situation, such as the processing time which may affect the turn-taking system, the definition of the situation as serious or informal, or the appearance of the robot, can be manipulated in the study without necessarily increasing the unnaturalness of the already unnatural situation. This allows the controlled investigation of a number of variables that crucially influence communicative processes both in the interaction with artificial systems and in human-to-human communication (such as recipient design, alignment, interactive negotiation, or the role of linguistic feedback), but that could hardly be controlled in naturally occurring settings.

To look at the speakers' attention to particular linguistic choices that may reveal their concepts about aspects of the situation is however not the only method linguistics provides for the analysis of general conceptual structures. In cognitive linguistics, the relationship between linguistic structure and general cognitive structures also constitutes the focus of linguistic investigation. Concepts proposed here are, for instance, idealized cognitive models (e.g. LAKOFF 1987a), cognitive metaphors (e.g. LAKOFF & JOHNSON 1980), and frames and scripts (e.g. FILLMORE & ATKINS 1992) that are proposed to represent organisations of our experience of, and interaction with, the world. The success of this research direction shows that linguistic methods can be very well suited for investigating cognitive structures. Cognitive linguistics however usually does not consider the conditions of usage of the cognitive structures as they are encoded in linguistic structures, and may, for instance, only provide lists of possible cognitive metaphors (e.g. LAKOFF & JOHNSON 1980, 1999, LAKOFF & NUÑEZ 2001, LAKOFF 1997b, 2002). For example, metaphorical concepts of the same phenomenon may sometimes have opposite orientations, for instance, we can both *fill in* and *fill out* a form (LINDNER 1982). Cognitive linguistic analysis reveals the different underlying concepts, but usually does not show the conditions of variability for the different orientations. In contrast, combining the method proposed with a metaphor analysis of the data elicited guarantees that the cognitive structures identifiable in the interactions provide us additionally with information about how speakers make use of them in real situations and what the conditions are that determine the decision for one rather than for another by analysing at the same time to which aspects of the situation the speakers attend. Additional methods of linguistic analysis, for instance also a functional linguistic analysis (e.g. MARTIN 1992), can be employed as long as it can be shown that the representations elicited are attended to concepts in action.

3 Example Analysis

In the following, an example analysis of spatial instructions in human-robot interaction is presented which can indicate how far the method proposed may take us.

3.1 Data

The focus of this study is the speakers' mental representations as they become apparent in interaction with language processing robots. That is, our interest lies in the resources that participants make use of in the interaction. The robot used for data elicitation is based on the commercially available Pioneer 1 with an additional elevated camera, which allows the recognition of objects in the scene. The robot is equiped with a natural language understanding system and can thus process instructions like *move to the red block* typed into a keyboard. It is furthermore equipped with facilities for path planning so that the robot can indeed move to the red block indicated.

In the interaction the robot itself did not generate natural language utterances. Having the robot produce linguistic output would have prompted the participants with particular linguistic structures, and as ZOLTAN-FORD (1991) has shown, speakers are likely to pick up the linguistic material presented to them. The robot designed thus only processed natural language without being able to generate utterances itself. The only output was either *error* printed on the screen, or successful action. Thus, human users who were not successful from the start were challenged to try out many different kinds of spatial instruction on the basis of their own ideas about what may have caused the communicative failure. This allows us to identify a whole range of different strategies employed by the speakers on the basis of their own hypotheses about spatial instruction, the robot and its capabilities, without the speakers being prompted to a particular solution by the robot (i.e. its designers). For data elicitation, it was therefore necessary that speakers interact with a robot that itself proposes as little linguistic material as possible.

The users' task was to make the robot move to particular locations pointed at by the experimentator; pointing was used in order to avoid verbal expression or pictures of the scene which would impose a particular perspective on the scene, for example, the view from above. Users were instructed to use natural language sentences typed into a computer to move the robot; they were seated in front of a computer in which they typed 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, which was placed either at a 90 degree angle or opposite the participant, as shown in figure (1). The arrangements of the cubes were varied; moreover, in one quarter of the settings, a cardboard box was added to the setting in order to trigger instructions referring to the box as a salient object.

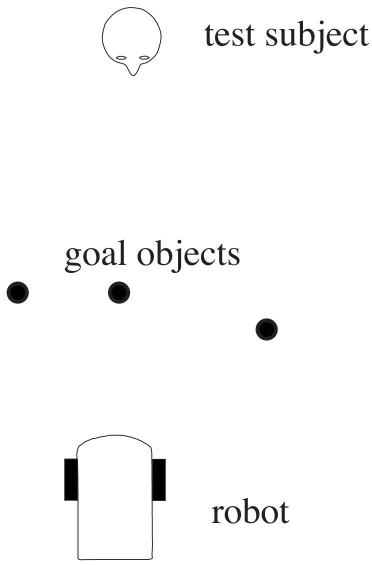

Figure 1: The setting of the experiment

Fifteen different participants carried out an average of 30 attempts to move the robot each within about 30 minutes. Altogether 476 instructions were elicited. The instructions were protocolled, and the speakers' verbal behaviour during the experiments was recorded in order to capture self-talk in which they announced their strategies or their ideas about what was going wrong. After the experiments, participants were asked to fill out a questionnaire as to what they believed the robot could and could not understand, which strategies they believed to be unsuccessful, and whether their beliefs about the robot had changed during the interaction.

3.2 Results

The data, in spite of their unnaturalness and in spite of the relatively small number of utterances available, show a surprising range of phenomena, some of which have been described elsewhere. For instance, the fact that speakers employ a particular referential strategy that relies on reference to the objects as a group (MORATZ, FISCHER & TENBRINK 2001, TENBRINK, FISCHER & MORATZ to appear) had hardly been discussed elsewhere before. Furthermore, it has been shown that the data at hand reveal what speakers attend to as common ground (FISCHER 2001). Here we shall concentrate on three areas: the participants' hypotheses about language, their hypotheses about the communication partner, and their hypotheses about spatial instruction.

3.2.1 Concepts of Language

To start with a trivial observation, the following examples[2] illustrate the speakers' attention to orthography:

(5) VP3: Fahre nach norden von die aus gesehen *[Drive to the north from the (sic!) point of view]*

VP3: Fahre nach Norden von Dir aus gesehen *[Drive to the North from your point of view]*

(6) VP11: 10cm

VP11: 10 cm

(7) VP3: Fahre nach vorn *[Drive ahead]*

VP3: fahre nach vorn *[drive ahead]*

(8) VP5: dreh dich nach rechts *[turn to the right]*

VP5: drehe dich nach rechts *[turn to the right]*

Underlying such variation is the hypothesis that using a particular spelling variant can make understanding more difficult. The different linguistic strategies thus show both hypotheses about the language – one orthographic variant may be more useful than another – and about the communication partner since the robot is displayed as a communication partner who may get confused by different spellings.

But not only spelling is attended to as potentially problematic. Also morphological reduction is apparent from the following examples:

(9) VP3: drehe dich *[turn yourself]*

VP3: Drehen *[turning]*

VP3: Drehen um 180° *[turning by 180°]*[3]

---

[2] In order not to increase the length of the paper unnecessarily, the robot's standard utterance *error* is being left out in the examples following.

(10) VP8: fahre geradeaus etwas rechts *[drive ahead a bit to the right]*

VP8: geradeaus etwas rechts fahren *[driving ahead a bit to the right]*

In the following example, besides morphological reduction, we find additional specification regarding the material of the object:

(11) VP3: Fahre geradeaus zum roten Würfel *[Drive ahead to the red cube]*

VP3: Fahre zum vorderen Holzwürfel *[Drive ahead to the wooden cube]*

VP3: Gehe zu Würfel *[Go to cube]*

Thus, speakers do not only leave out information for the robot to understand the utterance better, but sometimes they regard additional information necessary (see section 3.2.2. for further discussion).

Sometimes the speakers shorten their instructions by leaving out the verb:

(12) VP4: gehe nach links *[go to the left]*

VP4: links *[left]*

(13) VP5: fahr zu dem klotz vor dir *[drive to the block in front of you]*

VP5: klotz *[block]*

There are furthermore a number of examples that show that speakers attend to syntactic constructions like relative clauses:

(14) VP2: geh zu dem Würfel, der rechts des Kartons liegt *[go to the cube which is right of the card board box]*

VP2: geh zum Würfel rechts des Kartons *[go to the cube right of the card board box]*

(15) VP5: fahr zu dem klotz der hinter dem kasten liegt *[drive to the block that is lying behind the box]*

VP5: fahr zu dem klotz hinter dem kasten *[drive to the block behind the box]*

Like the previous examples, the following instance shows that the speakers attend to syntax. However, the example also indicates that the participants may have very unnatural ideas about the natural language that natural language processing systems can process:

---

[3] The example furthermore illustrates nicely how speakers attempt to be more precise for the robot, see also below section 3.2.2.

(16) VP5: fahre zum klotz rechts von dir *[drive to the block to your right]*

VP5: fahre zum rechts von dir klotz *[drive to the right of you block]*

That speakers attend to the choice of words is illustrated, in addition to those presented in the previous section, in the following examples:

(17) VP10: Fahre zum rechten Klotz *[Drive to the right block]*

VP10: Fahre zum Baustein rechts vorne. *[Drive to the constituent part right in the front.]*

(18) VP9: Fahre um das Hindernis *[Drive around the obstacle]*

VP9: Fahre um den Karton *[Drive around the card board box]*

Speakers may even attend to the choice of language itself:

(19) VP11: fahre vorwaers *[drive ahead]*

VP11: go

VP11: move forward

VP11: Fahre *[Drive]*

What the analyses so far have shown is that speakers attend to orthography, syntax, morphology, vocabulary, and the language itself as resources that they can use to adapt to the requirements of the situations. This shows on the one hand what the situation that "requires" particular choices is held to be, and how the linguistic choices themselves are conceptualized. When we now look at the whole spectrum of choices made, supported by the findings regarding the communication partner and the spatial domain to be discussed below, we can see that the linguistic variation observable exhibits a particular direction, namely most of the choices exhibit a reduction of complexity. Thus, the speakers display to us not only their attention to these aspects as potentially relevant in the situation, but also their hypotheses about what makes language easier to process for a robot. Thus a relative clause is obviously conceptualized as more difficult to understand than a prepositional phrase, and an inflected verb seems to be held to be less understandable than an uninflected verb. Such observations can be supported by findings in other domains in which linguistic difficulty is an issue, for instance, in foreigner talk (ROCHE 1989). The data thus exhibit the participants' folk theories about language.

3.2.2 Concepts of the Communication Partner

As we have seen in the previous section, participants attended to their linguistic choices such that they constantly made their utterances easier to understand for the robot. Furthermore, speakers wondered, both during the experiments and in the questionnaires, about the linguistic capabilities of the robot, asking whether it would understand particular words or syntactic constructions, such as relative clauses. Thus, they attended to the fact that the robot could have limited linguistic capabilities and therefore displayed their concept of the robot as a communication partner that is restricted in its capabilities. It may be interesting to note here that although the robot indeed *appears* to be restricted in its linguistic capabilities, this is in most cases not the reason for the communicative failure. Its parsing capabilities, its ability to cope with spelling variants and the comprehensiveness of its lexicon are actually not too bad. The communicative failure was really produced by the robot's inability to process anything but goal instructions (see next section). The speakers' attention to the linguistic properties of their utterances stems thus from their concept about what robots could be good at and what they could have problems with, as well as their theories about what makes language and spatial instructions easy or difficult to process. The linguistically restricted communication partner is thus the speakers' own construct.

Now, as we have already seen in the previous section, it is not always the case that speakers attempt to produce utterances that are less complex for their communication partner, the robot. For instance, the following examples show constructions which would, for example, not make an utterance easier to process for a human interlocutor:

(20) VP15: fahre nach rechts *[drive to the right]*

VP15: fahre kurs 010 *[drive direction 010]*

(21) VP8: vorwärts *[ahead]*

VP8: geradeaus ca. 10 cm *[ahead about 10 cm]*

VP8: bewege dich vorwärts um 10 cm *[move ahead 10 cm]*

(22) VP8: geradeaus etwa 2 Uhr *[ahead about 2 o'clock]*

VP8: los du lahme kiste vorwärts etwas nach rechts *[come on you lame box ahead a bit to the right]*

VP8: vorwärts fahren und etwas nach rechts *[driving ahead a bit to the right]*

VP8: vorwärts etwa 45 Grad nach rechts halten *[ahead about 45 degrees to the right]*

These linguistic choices show that the speakers consider a more formal way of specifying the instructions helpful. This finding is supported, for instance, by an observation on human-computer interaction made by HITZENBERGER & WOMSER-HACKER (1995), namely that in communication with artificial systems, speakers may use more formal language.

Furthermore, in their attention to the choice of words, speakers also tried out more abstract categories, thus displaying their hypothesis that a robot can reason about objects or obstacles but not about cubes or card board boxes (see example 17).

Additionally, unlike in communication among humans, the speakers in our experiment consistently took the robot's perspective. That they nevertheless attended to the point of view as relevant information is apparent from an incident in which the user had firstly taken the robot's point of view as origin, but due to some other mistake, the instruction was not carried out successfully. The user then announced that she had found out that the robot was using her perspective "after all", and she tried out the next instruction accordingly. In another experiment, the user's first question was where the front of the robot was. Thus, human users attend to the point of view as an informational resource, while at the same time they consistently take the robot's perspective, as long as they do not have evidence that this could not be the right strategy. This linguistic behaviour has otherwise been found, for instance, in communication with children and thus indicates that speakers regard the robot as a communication partner who is not capable of taking the speaker's perspective.

That artificial communication partners are conceptualized differently from human communication partners and that this may influence the linguistic choices speakers make has also been found in other studies (AMALBERTI et al 1993, JOHNSTONE et al. 1994, FISCHER 2000, 2001, FISCHER & MORATZ 2001). The concept of the robot evolving here is thus one of a communication partner who is linguistically highly restricted, who needs precise, formal, definitions, who may find abstract categories easier to process than concrete ones, and who cannot take its communication partner's point of view.

3.2.3 Concepts of Spatial Instruction

The communicative strategies the users were found to employ reveal an order in the type of instructions. In particular, only half of the participants started out by using 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.[4] Examples of this strategy are *fahr bis zum rechten Würfel* ***[drive up to the right cube],*** *fahr zu dem Klotz, der vor Dir liegt* **[drive to the block which is in front of you],** *geh zu dem vorderen Würfel* ***[walk to the front cube].*** This strategy was the one expected and implemented, so that these instructions were usually successful, unless there were orthographic, lexical, or syntactic problems. In such cases, these participants used path-naming strategies; if successful, they continued to use the goal-naming strategy. The following example shows how a speaker switches from a previously successful goal-instruction to a path-instruction because of a spelling mistake:

(23) VP7: Fahre zum rechten Würfel *[Drive to the right cube]*

Robot: (successful action)

VP7: Fahre zum linken Würfel *[Drive to the left cube]*

Robot: (successful action)

VP7: Faahre zum rechten Würfel *[Drive to the right cube]*

Robot: error

VP7: Fahre nach rechts *[Drive to the right]*

Robot: error

Similarly, in the following example, the initial goal-instruction is unsuccessful because of a wrong word. Thus the speaker tries path-instructions and then sticks to them:

(24) VP13: fahre zum rechten klötzchen *[drive to the right little-block]*

VP13: fahre vorwärts *[drive straight]*

VP13: Fahre vorwärts! *[Drive straight!]*

VP13: 50cm nach vorne *[50cm ahead]*

VP13: vorwärts *[straight]*

The other half of the participants, mostly the computer-naive users, already started out by giving path descriptions, decomposing the main action into more primitive actions, such as *move forward*, *go backwards*, or *turn left*. Half of the participants thus initially used path specifications to instruct the robot. This strategy seemed very natural to the participants, and they were on the whole quite perplexed to find that this strategy did not work with the robot.

---

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

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 afterwards employed 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]***. The examples display this order of types of instructions:

(25) VP11: Fahre zum Schrank *[Drive to the wardrobe]*

VP11: Roll nach vorne *[Roll ahead]*

VP11: Dreh dein rechtes Rad *[Turn your right wheel]*

(26) VP14: Fahre 10 cm nach vorn. *[Drive 10 cm ahead.]*

VP14: Fare los. *[Start driving.]*

VP14: Fahre los. *[Start driving.]*

VP14: 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 themselves to be insufficient, users attempted to instruct the robot by describing actions instrumental to movement in general.[5] From this order of instructions we can infer, supported by the findings on linguistic choices and on the conception of the communication partner, the robot, that what is at issue here is again a hierarchy of difficulty and basicness.

---

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

Additional evidence comes from the fact that because the experimental situation for those subjects who started off with the “wrong” strategy from the outset was so depressing, the leader 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.

The fixed order of instructional strategies can be inferred to reflect the participants’ mental representation 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.

4 Conclusion and Prospects

In this paper, a method for investigating spatial representations in use has been proposed. An example analysis was carried out on human-robot interaction to show the possibilities and limitations of the method for determining the speakers’ mental concepts on the basis of an investigation of their linguistic strategies in an actual communicative situation.

We have argued that our search for cognitive representations should not be restricted to showing which cognitive structures are available – identifying the resources speakers may make use of constitutes only the first step in the investigation. Additionally we have to investigate which representations speakers make use of in real situations and what influences their conditions of use.

We have then shown how such an analysis can be carried out in a particular situation: quite deviant, compared to natural conversation, human-robot interaction. For this situation it could be shown to which aspects speakers attend when making linguistic choices. In particular, the analysis yielded results with respect to three areas: concepts about what makes language easier or more difficult to process, concepts of the communication partner, the robot, and concepts of spatial instruction. With respect to the latter, the users’ strategies differed in that half of the participants referred to the intended object itself while the other half tried specifying the steps involved to reach the goal instead. Our findings regarding the initial strategy, to name the goal object versus to present a path description, presenting two different concepts of spatial instruction, could be explained by regarding their conditions of production in a situation in which basicness and difficulty is at issue. Thus, the fact that speakers were split in two equally large groups was not explained away by just referring to different user types; ROBERTS & NEWTON (2001: 143) argue explictly against an analysis using "cognitive styles" since this would just mean renaming the phenomenon observed. Instead, the speakers' choices were embedded in a system of choices they were attending to with respect to their understanding of the whole situation. In the case of communicative failure, all participants attended to a particular order of instructional strategies. This variation in strategies points to a mental representation of spatial instruction that is determined by a supposed hierarchy of basicness and difficulty of spatial instruction.

That basicness and difficulty are indeed at issue is supported by the findings on speakers' strategies regarding language and the communication partner: While in some cases the strategies were directed at making the utterance more precise, most concerned making things easier. The aspects attended to were identified as speaker categories and related to the speakers' conceptualization of the robot as a restricted communication partner (like children).

Consequently we could demonstrate some of the underlying mental constructs on the basis of which the speakers make their linguistic choices, as they were attended to in this particular situation, and some of their determining factors. Now, having argued that cognition is situationally bound and that we have to consider the situations in which the concepts attended to are used, is there any use of our findings besides as account of what may happen in human-robot interaction? We have seen a model of spatial instruction that is organized on the basis of basicness and difficulty. We can thus conclude for one thing that speakers attend to the level of difficulty of their instructions. We can furthermore assume that in situations in which difficulty is at issue, the hierarchy will again be of use, which, however, constitutes an empirical question.

We have also identified a number of general strategies employed by the participants, for instance, to adapt both language and spatial instruction to their conceptions of their communication partner. This finding is also likely to be more generally valid. And finally, we have proposed a procedure by means of which the participants' concepts in use can be investigated. Pushing conversation analytic concepts in this somewhat unusual direction opens up a promising new research direction.

For future research, we want to elicit data of human-robot interaction that are complex enough to allow conclusions regarding more complex mental representations, in particular, about spatial configurations. The goal is to identify the linguistic traces of such cognitive representations, such as mental models, and to shed light on their conditions of employment, such as aspects of the spatial setting, or the speakers', possibly differing, conceptualizations of the situation, their commmunication partner, the language, and space. At the same time we will examine the strengths and limitations of the method proposed and contribute to an interdisciplinary discussion on the nature of (particularly spatial) representations.

References

AMALBERTI, R. & CARBONELL, N. & FALZON, P. (1993): User Representations of Computer Systems in Human-Computer Speech Interaction, in: *International Journal of Man-Machine Studies* 38, 547-566.

BATLINER, A. & HUBER, R. & NIEMANN, H. & NÖTH, E. & SPILKER, J. & FISCHER, K. (2000): The Recognition of Emotion, in: WAHLSTER, W. (ed.): *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin etc.: Springer.

DOUGLAS, SARAH A. (1995): Conversation Analysis and HCI-Design, in: *Cambridge Series on Human-Computer Interaction* 10, 184-203.

FILLMORE, CHARLES & ATKINS, BERYL T. (1992): Toward a Frame-Based Lexicon: The Semantics of *risk* and its Neighbors, in: LEHRER, ADRIENNE & KITTAY, EVA F. (eds.): *Frames, Fields and Contrasts*. Hillsdale N.J.: Lawrence Erlbaum, 75-102.

FISCHER, KERSTIN (1999): Repeats, Reformulations, and Emotional Speech: Evidence for the Design of Human-Computer Speech Interfaces, in: BULLINGER, HANS-JÖRG & ZIEGLER, JÜRGEN (eds.): *Human-Computer Interaction: Ergonomics and User Interfaces,* Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany. Lawrence Erlbaum Ass., London, 560-565.

FISCHER, KERSTIN (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, 85-92.

FISCHER, KERSTIN (2001): How much Common Ground Do we Need for Speaking?, in: KÜHNLEIN, PETER & RIESER, HANNES & ZEEVAT, HENK (eds.): *Proceedings of the 5th Workshop on Formal Semantics and Pragmatics of Dialogue,* Bi-Dialog 2001, Bielefeld, June 14-16th, 2001, 313-320.

GARFINKEL, HAROLD (1972): Remarks on Ethnomethodology. Directions in Socolinguistics, in: GUMPERZ, JOHN J. & HYMES, DELL (eds.): *The Ethnography of Communication*. New York etc.: Holt, Rinehart and Winston, 301-324.

GOFFMAN, ERVING (1978): Response Cries, in: *Language* 54, 787-815.

HABEL, CHRISTOPHER (1995): Mentale Modelle, in: *Künstliche Intelligenz* 6, 42.

HALLIDAY, M.A.K. (1985): *An Introduction to Functional Grammar*. London: Arnold.

HERITAGE, JOHN (1988): Explanations as Accounts: A Conversation Analytic Perspective, in: ANTAKI, CHARLES (ed.): *Analysing Everyday Explanation. A Casebook of Methods*. London etc.: Sage, 127-144.

HERITAGE, JOHN & SORJONEN, MARJA-LEENA (1994): Constituting and Maintaining Activities across Sequences: And-Prefacing as a Feature of Question Design, in: *Language in Society* 23, 1-29.

HITZENBERGER, L. & WOMSER-HACKER, C. (1995): Experimentelle Untersuchungen zu multimodalen natürlichsprachigen Dialogen in der Mensch-Computer-Interaktion, in: *Sprache und Datenverarbeitung* 19,1: 51-61.

HUTCHBY, IAN & WOOFFITT, ROBIN (1998): *Conversation Analysis*. Cambridge: Polity.

JOHNSON-LAIRD, P.N. (1989): Mental Models, in: POSNER, MICHAEL I. (ed.): *Foundations of Cognitive Science*. Cambridge M.A.: MIT Press, 469-499.

JOHNSTONE, A. & BERRY, U. & NGYUEN, T. & ASPER, A. (1994): There Was a Long Pause: Influencing Turn-Taking Behaviour in Human-Human and Human-Computer Spoken Dialogues, in: *International Journal of Human-Computer Studies* 41, 383-411.

LAKOFF, GEORGE (1987a): *Women, Fire, and Dangerous Things*. University of Chicago Press.

LAKOFF, GEORGE (1987b): *Moral Politics: What Conservatives Know that Liberals Don't*. University of Chicago Press.

LAKOFF, GEORGE (2002): *Moral Politics: How Liberals and Conservatives Think*. University of Chicago Press.

LAKOFF, GEORGE & JOHNSON, MARK (1980): *Metaphors we Live by*. The University of Chicago Press.

LAKOFF, GEORGE & JOHNSON, MARK (1999): *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. HarperCollins Publishers.

LAKOFF, GEORGE & NUÑEZ, RAFAEL E. (2001): *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics into Being*, Basic Books.

LINDNER, SUE (1982): What Goes up Doesn't necessarily Come down: The Ins and Outs of Opposites, in: *Proceedings of the 18th Meeting of the Chicago Linguistic Society*, 305-323.

MARTIN, JAMES R. (1992): *English Text: System and Structure*. Philadelphia etc.: Benjamins.

MORATZ, REINHARD & FISCHER, KERSTIN & TENBRINK, THORA (2001): Cognitive Modelling of Spatial Reference for Human-Robot Interaction, in: *International Journal on Artificial Intelligence Tools* 10,4: 589-611.

NOTHDURFT, WERNER (1984): *„...äh folgendes Problem äh...". Die interaktive Ausarbeitung „des Problems" in Beratungsgesprächen,* Tübingen: Narr.

ROBERTS, MAXWELL J. & NEWTON, ELIZABETH J. (2001): Understanding Strategy Selection, in: *International Journal of Human-Computer Studies* 54, 137-154.

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

SACKS, HARVEY (1984): Notes on Methodology, in: ATKINSON, J. AND HERITAGE, JOHN (eds.): *Structure of Social Action: Studies in Conversation Analysis*. Cambridge University Press.

SACKS, HARVEY (1985): The Inference-Making Machine: Notes on Observability, in: VAN DIJK, TEUN A. (ed.): *Handbook on Discourse Analysis.* Vol. III: Discourse and Dialogue. London etc.: Academic Press, 13-23.

SCHEGLOFF, EMANUEL A. (1972): Notes on a Conversational Practise: Formulating Place, in: SUDNOW, DAVID (ed.): *Studies in Social Interaction*. New York: Free Press, 1972, 75-119.

SCHEGLOFF, EMANUEL A. (1997): Whose Text, Whose Context? *Discourse & Society* 8,2: 165-187.

SUCHMAN, L. (1987): *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.

TENBRINK, THORA & FISCHER, KERSTIN & MORATZ, REINHARD (to appear): Human Strategies in Linguistic Human-Robot Communication, in: FREKSA, CHRISTIAN (ed.): *Spatial Cognition*.

ZOLTAN-FORD, ELIZABETH (1991): How to Get People to Say and Type what Computers Can Understand, in: *International Journal of Man-Machine Studies* 34, 527-647.

KERSTIN FISCHER
Universität Bremen
Fachbereich 10: Sprach- und Literaturwissenschaften
PF 33 04 40
D-28334 Bremen
E-Mail: kerstinf@uni-bremen.de