Table of Contents

Paper 1

Notes on Analysing Context

Kerstin Fischer
kerstinf@uni-bremen.de
http://www.fb10.uni-bremen.de/anglistik/homepages/fischer.htm

Abstract

The focus of this paper is on the relationship between language and context. For verbal interaction, two types of context can be distinguished, the context of text, comprising what, in computational linguistics, is called a dialogue history, and the context of situation, meaning the surroundings in which utterances occur. While much research has been carried out regarding the mechanisms by means of which what has been said may constitute a shared basis, the description of the context of situation remains systematically unexplored. A method is proposed, relying on the analysis of how speakers make use of contextual factors in human-robot communication with conversation analytic methods. The procedure proposed is exemplified in a study on the speakers' attention to the types of common ground proposed in Clark (1996).

1 Introduction

The problem addressed in this paper concerns the relationship between language and context with the aim of identifying the potential influences of the situation on the linguistic properties of utterances. The questions we have to deal with are thus what context is and how aspects of context determine language usage.

These two questions are of both theoretical and practical importance. On the one hand, it is interesting for our understanding of what communication is and how it works to find out how speakers interact with each other and the context. Thus, Gumperz (1992), for instance, regards the way speakers invoke contextual knowledge as part of communicative competence, that is, the knowledge and abilities speakers employ in verbal interaction. That is, if we do not consider linguistic competence as a static system of grammatical rules but instead attempt to account for the negotiated, dynamic, nature of language in use, it becomes clear that we have to include considerations on the interaction between verbal behaviour and the context.

On the other hand, there are a number of practical reasons to study the relationship between language and context. For instance, the initial motivation for developing register theory were grounded in considerations of language teaching: Students of a second language have to know which linguistic properties to use in which situations. Furthermore, in automatic speech processing, register-based analyses are of great importance. For instance, knowledge about situationally appropriate language use can improve system performance by limiting the resources to those linguistic properties that may actually occur (Krause, 1992b), by providing hints for appropriate choices in natural language generation (Paris and Bateman, 1989), or by supporting automatic disambiguation. For example, the linguistic choices users make in human-robot interaction can be used to restrict or at least weigh the different possible interpretations of a ambiguous utterances.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

kerstinf@uni-bremen.de
http://www.fb10.uni-bremen.de/anglistik/homepages/fischer.htm.

Determining the relationship between context and the speakers' linguistic behaviour is thus of both theoretical and application-oriented relevance.

2 Problems in Analysing Context

Two types of context can be distinguished, *context of situation* and *context of text* (e.g. O'Donnell, 1999). Thus, linguistic choices at a given moment may be influenced by the larger surroundings on the one hand, and on the preceding interaction, the *discourse history*, on the other.

2.1 Context of Text

In dialogical interaction, every utterance constrains the next one, thus opening up new possibilities and restricting choices at the same time. For instance, a question usually makes some kind of answer conditionally relevant. Discourse may thus be represented as a tree or a network with multiple decision points, each point opening up and constraining a number of possible next moves (Berry, 1981; Halliday, 1984; O'Donnell, 1999). However, as work by Clark and collaborateurs (Clark and Marshall, 1981; Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989a,b; Clark and Brennan, 1991), summarized in Clark (1996), shows, it is not sufficient to look at what has been said. Participants have been shown to attend to the *grounding* of information, that is, they ascertain that "the speaker and addressees mutually believe that the addressees have understood what the speaker meant to a criterion sufficient for current purposes" (Clark and Schaefer, 1989b, p. 124-125). Consequently, what constitutes the context of text for the speakers is not what has been only mentioned, but what has been sufficiently grounded, and what can therefore be assumed to be shared by the participants. Much research on context has been carried out on what is shared between the conversational participants on the basis of the discourse record of the current situation. That is, much work has addressed aspects of grounding, the process by which individuals add information to the common ground (Traum, 1994; Ginzburg, 1998; O'Donnell, 1999). Thus, research has concentrated on the augmentation of the propositions representing the assumed shared knowledge on the basis of what is said in a discourse situation (Larrson et al., 2000). The context of text, the history of grounded discourse, is yet only one part of the context relevant for the linguistic choices to be made.

2.2 Context of Situation

The context of situation concerns the surroundings in which utterances occur. However, what is unclear is what such surroundings include. Self-evidently we do not want to describe every aspect of the surrounding world in arbitrary detail, attempting to provide a full description of the situation. That is, we want to limit our description of context to those aspects that condition the linguistic usage. But how can that be done? Consider the following example (Capra, 1996, p. 41-42):

> "a simple experiment often performed in introductory physics courses. The professor drops an object from a certain height and shows her students with a simple formula from Newtonian physics how to calculate the time it takes for the object to reach the ground. As with most Newtonian physics, this calculation will neglect the resistance of the air and will therefore not be completely accurate. Indeed, if the object to be dropped were a feather, the experiment would not work at all.
>
> The professor may be satisified with this "first approximation," or she may want to go a step further and take the air resistance into account by adding a simple term to the formula. The result – second approximation – will be more accurate but still not completely so, because air resistance depends on the temperature and pressure of the air. If the professor is very ambitious, she may derive a much more complicated formula as a third approximation, which would take these variables into account.

> However, the air resistance depends not only on the temperature and air pressure, but also on the air convection – that is, on the large-scale circulation of air particles through the room. The students may observe that this air convection is caused, in addition to an open window, by their breathing patterns; and at this point the professor will probably stop the process of improving the approximation in successive steps.
>
> This simple example shows that the fall of an object is connected in multiple ways to its environment – and, ultimately, to the rest of the universe. No matter how many connections we take into account in our scientific description of the phenomenon, we will always be forced to leave others out. Therefore, scientists can never deal with truth, in the sense of a precise correspondence between the description and the described phenomenon. In science we always deal with limited and approximate descriptions of reality."

The example shows that even in the natural sciences it may be difficult to account for all those factors that may influence the outcome of an experiment. It is likely to be equally difficult to decide what constitutes the context of a speech situation, such that the factors that determine the speakers' linguistic behaviour are accounted for.

A similar example from Edmonds (1999) shows that, besides the fact that everything may be in some relation to everything else, it is the language users' *understanding* of the situation that may matter. This may make things both easier and more difficult: On the one hand, the participants' understanding may be a useful guide to what constitutes the relevant context of situation. On the other hand, facts about mental representations such as the understanding of the situation may be as difficult to obtain as matter-of-fact descriptions of the external environment. Edmonds' considerations furthermore rest on the perception of exceptionality of situations (Edmonds, 1999, p.2):

> "A man is distracted and falls off a small ledge onto a pavement. When he lands his leg breaks. What caused his leg to break? It could be attributed to many things: The hardness of the pavement; the weakness of his femur; the way he landed on the leg; the height of the ledge; the distraction; or even the man's distractability. There seems to be no end to the number of factors one *could* include as causes of the fracture. Whether one *does* count each of these as causes is arbitrary from an absolute external viewpoint. It can depend on the extent to which we judge each of them as *unusual*."

While Edmonds' proposal may help us to deal with extraordinary aspects of context, research in sociolinguistics has long shown to us that there are aspects of the context that are omnipresent and thus not necessarily perceived as extraordinary or unusual and that nevertheless may influence the linguistic properties of utterances very much. Such categories, such as gender, social class or ethnic background, have been shown to correlate with properties of language usage (cf., for instance Fasold, 1990). This relationship has often been analyzed by comparing linguistic utterances in situations that differ preferably only with respect to a single variable, for instance, in Labov's study of the social stratification of /r/ (Labov, 1966). The methodology used is the quantitative and qualitative comparison of two or more linguistic varieties.

Problems related to this approach are that the variety of situations situations can usually not be controlled so that they really differ with respect to a single variable – alternatively the situations people are 'placed' in are quite unnatural. Furthermore, variationist approaches have often been criticized for quantitatively comparing linguistic properties that may actually be functionally distinct and thus cannot be compared on a solid methodological basis (Schiffrin, 1994; Heritage, 1995). Another, more fundamental, problem is that external variables are often not what matters; thus, Fischer (2000) shows that differences in the linguistic behaviour of female and male speakers are not significant in human-computer interaction while differences in the conceptualization of the computer as a communication partner lead to significantly different linguistic behaviour. Similarly, for human-computer interaction Amalberti et al. (1993) show that while initially speakers' verbal behaviour may differ very much depending on whether they believe to be talking to a human being or to a computer, these difference disappear in the course of the interaction. Thus, speakers adapt their linguistic behaviour to their communication partners' linguistic output (which was identical in this study), and they revise their conceptualizations of the communication partner on the basis of the interaction history.

Register theory has provided us with a number of possible factors that may influence the choice of linguistic properties (e.g. Halliday, 1964, 1970). However, while certainly many of these aspects of context may influence the speakers' linguistic behaviour *in general*, we do not know how a particular context influences linguistic choices. For instance, Halliday (1964) proposes *field, mode* and *style* (later *tenor*) as relevant dimensions. Typical examples of registers, such as motherese, foreigner talk or computer talk (e.g. Krause, 1992a), do not include considerations of the field, for instance. Those three alleged[1] registers are mainly determined by the participants. Thus, even if some factors may turn out to influence speech in some situations, it may not be the case that they influence language use in a particular way in others, at least some factors may be perceived as less *extraordinary*, such as the topics of foreigner talk.

It seems thus that both internal and external aspects of context are being selected by the participants depending on their relevance for the participants. If however context is not objective but dependent on the interactants' selections, the contextual categories selected may differ even between the participants themselves. Consequently, there is some need for the interactants to signal to each other in which way they make use of the context in a particular speech situation. Thus, if aspects of context are relevant for the interaction, we can assume that the speakers will display this to their communication partners. Gumperz (1992), in his article on what communicative competence, our ability not only to produce sentences correctly, but to produce them in an interactionally appropriate way, actually consists in, argues that "conversing, as we all know, is a cooperative activity that involves active participation and coordination of moves by two or more participants in the joint production of talk. There are good reasons to believe that such coordination presupposes verbal activities and types of knowledge that are not yet fully understood." Such types of knowledge, which are, in his view, accounted for by neither grammatical, sociolinguistic or other rules nor schemas or scripts or other types of static knowledge, are employed flexibly and dynamically based on the situational requirements – and are indicated as such by the participants themselves during the interaction. Thus, "what we must explain are on-line processing strategies" (Gumperz, 1992, p.278). The strategies by means of which participants indicate context to each other are what Gumperz (1982) introduced as *contextualization cues*: By indicating aspects of context by means of linguistic features such as prosody, mimic, or gesture, speakers provide hints relevant to the interpration of their utterances. The following example (Gumperz, 1982, p. 147) illustrates Gumperz' use of *contextualization cue*:

Teacher:	James, what does this word say?
James:	I don't know?
Teacher:	Well, if you don't want to try someone else will. Freddy?
Freddy:	Is that a 'p' or a 'b'?
Teacher:	(encouragingly) It's a 'p.'
Freddy:	Pen.

Gumperz' analysis treats this example as an instance of intercultural miscommunication since the teacher misses the pupil's (James') contextualization cue that he would like to be encouraged. The background is that in the African-American community rising intonation means a request for encouragement. The methodology used by Gumperz is to have informants interpret utterances, and in the current example, the informants from the African-American community all interpreted James' utterance as such a request. A general methodological problem then arises: How does the analyst know that this type of context may be relevant? How did he get to choose African-American informants in the first place?

[1]But see, for instance, Roche (1989).

Regarding the methodology for *investigating* context, rather than describing the linguistic means by means of which context is activated, once the link has been discovered, Gumperz' approach leaves a number of questions. A more constrained approach is presented by conversation analysts. Schegloff (1997), for instance, proposes using conversation analytic methodology in order to restrict the analysis only to those aspects of context to which speakers can be shown to attend.

To illustrate this methodology, consider the following example from Harvey Sacks. The 'context' is a group therapy meeting in which previously, in a side conversation, Ken has scheduled with Roger when to pick up his radio, which Ken had repaired for Roger (Sacks, 1992, p. 595):

Dan:	Ken, why don't you make these arrangements out of here.
Ken:	((very quiet)) oh.
Al:	Yes, Teacher.
Ken:	((cough))
Dan:	Is that what I sounded like for a moment?
Al:	Yes, Teacher.
Dan:	What'd it make you feel like?
Roger:	Like it reminded us that we're in group therapy.
Al:	I felt like back in school.

Thus, the displays of the interactants to each other show that the group therapy meeting constitutes a relevant context for the participants. However, the example also indicates that participants may invoke contexts that are not part of the current situation, such as the school context. The conversation analytic methodology can thus provide us with what we need: The participants display to each other (and thus to the analyst) what they consider to be the relevant context. The method therefore accounts for both the extraordinary aspects revelant to Edmonds' example, and the omnipresent aspects of the situation as far as participants orient to them themselves.

However, conversation analytic approaches to context have been criticized, for instance, with respect to the following problems: For one thing, the analyses are heavily based on what is being said by the participants. Other aspects of speaker behaviour are only beginning to be taken into account in CA analyses (e.g. Duranti and Goodwin, 1992). Thus, aspects obviously apparent in the situation perceived, such as the participation framework (Goffman, 1979), are only taken into account by the analyst if they leave traces in what is being said. Related to this is the general methodological criticism that displays of attention to particular aspects of the situation are given only 'for all practical purposes' (Nothdurft, 1984, 124-125). Thus, particularly the unproblematic, self-evident knowledge of the context, which we are interested in, is unlikely to be displayed. The third point is that conversation analysis is held to be naive in the belief that analysts should be able to look at the data unbiasedly since they are involved in particular cultural contexts themselves (Deppermann, 1999). In section 3, we shall see how these methodological problems can be accounted for.

2.3 Summary

Summing up the discussion so far, there can be no external view on context: first, external categories may or may not be relevant, and there are no clear limits to what constitutes an external context and what does not, since everything may be related. An approach to context should therefore focus on what is relevant for the participants. If we share a particular cultural context, we may share judgements about the extraordinary and the relevant with the participants, but we may have fewer intuitions about the role of omnipresent factors such as gender, ethnicity or what it means that a conversation takes place in the framework of a group therapy meeting. We therefore have to rely on to what the speakers as participants attend since only to what speakers attend can influence their linguistic behaviour. So using the conversation analytic perspective provides us with a selection of those aspects of context that may shape language usage. However, in normal conversations, particularly the obvious, uncontroversial aspects of the context may not be displayed by the participants to each other as being attended to. It may thus be useful to investigate talk in situations in which aspects of the context are at stake. In the following, such an approach is developed and illustrated.

3 An Approach to Analyzing Context

The approach taken here is analyzing human-robot interaction by means of methods form conversation analysis (CA). While it is certainly useful to base one's investigations on natural conversations since they constitute the most basic type of communication from many points of view (Fillmore, 1981; Diewald, 1991), studying the communication between human speakers and communication partners about whom they do not know much may be particularly suited for showing what aspects of context they consider to be both relevant and at stake. For instance, common ground cannot be presupposed in the interaction between human and artificial communication partners to the same extent as in the communication among humans. Human-robot interaction is furthermore particularly suited since not only the artificial communication partner itself is problemtic for the participants, but also how the robot interacts with the environment, what it perceives, or how it moves. In this particular type of interaction, almost all aspects of common ground may have to be negotiated. Dialogues with so little shared knowledge between the communication partners may thus reveal which types of contextual information are necessary for communication to work because in these cases, context is attended to as potentially problematic. The particular features of the method proposed are outlined in more detail in the following.

3.1 Displays of Understanding

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. In CA, the procedure for identifying speaker categories is the close analysis of the sequential organisation of talk-in-interaction. The idea is that with every utterance, speakers display their understanding of the previous utterance to their communication partner (and therefore to the analyst), thus providing us with a proof criterion for our analyses (Sacks et al., 1974; Clark and Schaefer, 1989b; Heritage, 1995). In the following turn, the first speaker has the opportunity to ratify or reject the interpretation displayed. Also in the communication with (simulated) artificial communication partners this tripartite structure can be identified, for example:

(1) e0376102: blablabla. ähm wie wäre es mit Samstag, <Laugh> dem dreiundzwanzigsten ersten? [*blablabla. uh how about Saturday, the 23rd of January?*]

s0376103: bitte machen Sie einen Vorschlag. [*please propose a date*]

e0376103: mach' ich doch. wie wär's mit Samstag, dem DREIundzwanzigsten ersten, für einen zwei-stündigen Termin, ab <P> zehn Uhr morgens? [*that's what I'm doing. how about Saturday, the 23rd of January, for an appointment of two hours, beginning at 10am?*]

In the first turn of the excerpt, the speaker presents an utterance to her partner. In the following turn, the system asks the speaker to make a proposal, thus displaying its (mis-)understanding of the previous turn. In the third turn, the speaker corrects the system's interpretation by saying that the previous turn had already been a proposal, by repeating and further specifying the previous proposal and by prosodically marking the date for the appointment proposed. That the first turn was meant to be a proposal is thus an interpretation displayed by the participants themselves. Consequently, also in situations with communicative failure the sequential structure of the interaction constitutes a useful resource for the interpretation by the analyst on the basis of speaker categories (Hutchby, 2001).

However, misinterpretations by automatic speech processing systems are frequent, and users do not always take the opportunity to intervene and to reformulate their utterances. The more often speakers are misinterpreted by the system, the less likely they are to reformulate their utterances for their communication partner Fischer (1999). Thus, the system often provides only partial displays of interpretation or even insufficient ones, and users can be found to give up using the third position to correct the misinterpretation after some time. This may be particularly true of the communication with communication partners who are not expected to understand, for instance, metalanguaging, such as shown above in example (1). Yet in the communication with artificial systems, speakers may still display their understanding of the situation, see section 3.4. We can therefore trust in the fact that speakers will display their understanding of the situation to each other or a bystander, and, thus, to the analyst interested in aspects of context.

The next-turn-proof-procedure is also the methodological principle underlying the study of grounding. By displaying understanding to each other, the participants provide acceptance of the information provided by the communication partner, and their ratified information can be regarded as grounded (Clark and Schaefer, 1989a). Presenting information and signalling acceptance together constitute contributions to conversations, which are thus collaboratively achieved. The background assumption is that speakers need to assure themselves "that the others are at that moment attending to, hearing, and trying to understand what they are saying" (Clark and Schaefer, 1989a, p.123). The authors continue: "Speakers need their listeners' help for both these assurrances, so conversation take everyone's coordination. Together, the participants bear the mutual responsibility of assuring that waht is said has been heard and understood before the conversation goes on" (Clark and Schaefer, 1989a, p.123-4). By signalling to the communication partner what has been understood, the participants continuously build up a pool of information available to all participants: " the speaker and her addresses all focus on grounding that content, on placing it among their shared beliefs – technically heir mutual beliefs" (Clark and Schaefer, 1989a, p.124). While Clark and Schaefer (1989a) exemplify the construction of presentation and acceptance on telephone directory inquiries, they argue that it is a "systematic feature of everyday conversations" (Clark and Schaefer, 1989a, p.149). Attention to the grounding of information thus not only constitutes the methodological basis of conversation analysis, but it also allows the analysis of the participants' attention to the establishing of a ratified and shared context of text.

3.2 Deviant Case Analysis

Another way in which the methodology draws on research in conversation analysis is to exploit the notion of 'deviant case analysis'. In CA, deviant cases play an important role since in their interactive treatment, speakers display their attention to the regularity proposed overtly by displaying their expectations (Sacks et al., 1974; Hutchby and Wooffitt, 1998). The methodology is thus based on the idea that deviant cases are not only orderly in themselves, but, as Hutchby and Wooffitt (1998, p. 95-98) argue, 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 and Wooffitt, 1998, p. 98). Thus, by not analysing basically smooth and unproblematic human-to-human conversation, but by focussing instead on problematic human-to-robot (mostly mis-)communication, even omnipresent, self-evident, obvious aspects of context to which speakers attend may become evident.

As we shall see, the speakers' ways of dealing with the deviant situation furthermore reveals a large amount of their background knowledge when faced with communicative problems, since speakers may employ a huge number of strategies which display their orientation to particular categories, which include aspects of the relevant context of situation. The deviance of human-robot communication therefore allows the analysis of all of those aspects of communication speakers build on in natural conversations and which they miss, or are uncertain about, in the communication with artificial systems.

3.3 Controlling the Situation

Using human-robot interaction furthermore allows to controll a number of external factors. In contrast to natural conversation, the situations in which such dialogues are carried out are already very unnatural – even more so since automatic speech processing robots are usually prototypes only available in research laboratories. Introducing some further constraints on the situation does therefore not necessarily result in extreme differences in naturalness. Moreover, human-robot interaction involves an artificial communication partner which need not be 'given' in the same way in which humans are 'given'. Instead, the artificial communication partners are being designed on the basis of computer scientists' ideas about communication. That is, one of the speakers may be 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 particular aspects of the context of situation, and, by manipulating the system's output, the context of text.

3.4 Availability of Further Resources

Human-robot communication provides us with a number of additional data not usually available in natural conversation. Thus, in the communication with artificial communication partners 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 analysis of response cries (Goffman, 1978) in that speakers give accounts to bystanders, like the leader of the experiments, of their partly strange linguistic behaviour triggered by the system's malfunctions. The following two examples show how speakers, partly in self-talk, partly directed towards the system, announce their discourse strategies, here repetition:

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

s0375205: Mittwoch, der sechste erste, von acht bis zehn Uhr ist schon belegt. [*Wednesday, the sixth of January, from eight to ten am is already occupied.*]

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

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

Besides the naturally occuring self-talk, also the 'thinking aloud' technique can be used. Another method that has been already used successfully in system diagnosis, is called 'constructive interaction' (Suchman, 1987; Douglas, 1995) and involves the joint use of an artificial system by two participants.

Furthermore, since human-robot communication still often occurs in laboratory settings because of the prototype nature of the robots involved, it is quite natural to have speakers fill out a questionnaire after the conversation. These data can give important information regarding to which kinds of information speakers really attend and which they use as explanations for themselves. Thus, speakers may reveal their speaker categories to the analyst in those further data sources.

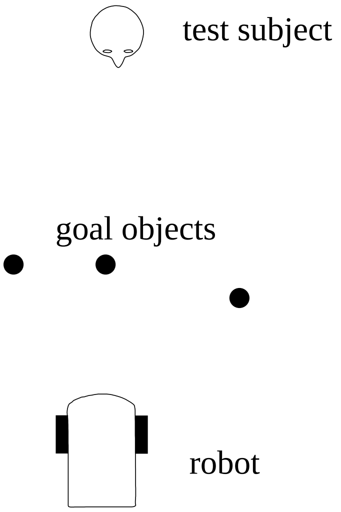

Figure 1.1: The Experimental Setup

4 Example Analysis: Instantiating Clark's Typology of Common Ground

In order to illustrate the method for analyzing aspects of context here proposed, in this paper it is determined to which types of common ground speakers attend as proposed in Clark (1996). The procedure is to investigate human-robot communication as a particular kind of interaction in which common ground is at stake, i.e. in which speakers are uncertain about the common ground they can assume. In particular, analysing what kinds of information about the robot and the context they request during the experiments for producing utterances for their communication partner reveals to which types of common ground they orient.

Clark (1996, p. 92-121) provides us with a list of knowledge types speakers may use as possible resources to establish common ground, besides the discourse record. Such resources include knowledge about the human nature, a common lexicon, knowledge about scripts, and knowing how. Which of these resources speakers really make use of in discourse has rarely been studied, and if so, only in natural conversations (Kreckel, 1981; Clark, 1996).

The results will not only be theoretically interesting because they provide evidence for a typology of common ground, they will also have consequences regarding the methodology proposed.

4.1 Data

The data were elicited in an experimental setting (see Figure 1.1) for which initially, a robot (see Figure 1.2) was designed on the basis of what is known about spatial reference among humans (Moratz and Fischer, 2000). Then, a test scenario was developed in which the users' task was to make the robot move to particular locations pointed at by the leader of the experiment; pointing was used in order to avoid the prompting of verbal expressions and the use of pictures of the scene which would impose a particular perspective. Users were instructed to use natural language sentences typed into a computer to move the robot through a setting in which, for instance, a number of cubes were placed on the floor together with the robot.

Since the main aim of the experiments was to determine how participants naively approach the robot as a communication partner, the only system output was action or an unspecified error message. This avoids the 'shaping' of the users' language by the system's output (Zoltan-Ford, 1991). By being constantly forced to try out new strategies to increase the understandability of their utterances, users furthermore reveal their hypotheses about how the robot may work. The

Figure 1.2: The Robot

users' beliefs about the nature of the robot, about what it can perceive and do, are part of the common ground users expect, that is, to which they attend.

Fifteen different participants carried out about 40 attempts to move the robot within about 30 minutes time each. Altogether 473 instructions were elicited. The sentences were protocolled, and the users' verbal behaviour during the experiments was recorded in order to capture self-talk in which speakers anounce their strategies or their ideas about what is going wrong. After the experiments, participants were asked to put down in a questionnaire what they believed the robot could and could not understand.

4.2 Types of Common Ground Attended to in the Data

Clark (1996) presents a typology of common ground, proposing a number of fine-grained distinctions. In his account, 'total common ground' comprises a 'discourse representation', which consists of a textual and a situational description, as well as the discourse record, and communal and personal common ground (Clark, 1996, p. 54). Common ground therefore constitutes only one part of the context of language use, though an essential part.

Clark (1996, p. 100-120) distinguishes two types of common ground, communal and personal. Both of them can be distinguished into a number of subtypes:

1. communal common ground
    - human nature
    - communal lexicons
    - cultural facts, norms, procedures
    - ineffable background
    - our feeling of other's knowing

2. personal common ground
    - perceptual basis
    - actional basis
    - personal diaries
    - acquaintedness (friends and strangers)
    - personal lexicons

We can now use the evidence from our corpus of human-to-robot communication to show to which of these aspects of common ground from Clark's typology the users really attend in which ways. The different types of common ground users can be found to orient to in the dialogues with the robot will be investigated on the basis of the transcripts of the human-robot interaction itself, the self-talk elicited during the experiments, and the answers participants gave in the questionnaire at the end of each dialogue. By means of this procedure, we can show which the types of common ground are of which users display that they are noticing their absence.[2]

Communal Common Ground

Human Nature The data do not show that speakers attend to their knowledge about the human nature, but since the robot is not a human communication partner, this is not surprising. However, for human-computer conversation in another scenario, participants could be shown to treat the computer like a human being (Fischer, 2000). Thus, in previous analyses human speakers have been found to transfer human attributes to their artificial communication partners. In contrast, in the present human-robot dialogues, the users' self-talk consists of several questions regarding the nature of the robot, for instance, regarding its orientation. In order to be able to employ an intrinsic reference system, the users requested to know where the 'front' of the robot is and what it can perceive. We can count this information need as evidence that participants orient to the robot's nature while formulating spatial instructions.

Communal Lexicons The data show that the human users worry very much about which are the appropriate words to use, i.e. which is the common lexicon between them and the robot. Thus, during the experiments they asked questions like whether one word or another is understandable for the robot. In the questionnaire, six out of the fifteen participants hypothesized that a possible source for their miscommunication with the robot was that they could not find the right words, that they did not know the 'appropriate' lexicon.

However, participants did not only have problems regarding the communal lexicon; the recordings of their utterances during the experiments as well as the questionnaire results reveal that they regarded the following linguistic aspects as problematic: orthography, formality (in particular the formal or informal way to spell imperative verbs in German), but also syntax, for instance, whether relative clauses are allowed, the length and complexity of sentences, the granularity level (especially the question whether they should use natural language or metrical expressions), and, most generally, whether the language of instruction should be German, the native language of the participants, or English, 'the language computers speak.' The data thus support Clark's category, yet it has to be extended to all linguistic levels involved, not just the lexicon.

Cultural Facts, Norms, Procedures The setting in which the experiments were carried out did not require participants to attend to cultural facts, norms, or procedures, such as scripts. However, participants are found to request one property in their communication partner that can be considered to constitute 'normal' or even 'normative' behaviour among human beings such that it can be requested of human speakers, and its lack is accountable in dialogues, namely consistency (Goffman, 1978). When the users found a hypothesis about the behaviour of the robot untrue,

[2]We will focus only on the human users and their beliefs about the robot, and thus spare out the perspective of what the robot may be implemented to believe about what the human users may know.

they complained about its lack of consistency, both verbally and in the questionnaire. Thus participants displayed attention to an instance of cultural norms, even in this restricted setting.

Ineffable Background The example by means of which Clark (1996, p. 110) illustrates the notion of ineffable background is that a person living in San Fransisco is expected to know what Golden Gate Bridge or Coit Tower, for instance, look like. In our data, one particular problem turned up which relates best to the notion of ineffable background, but which differs from Clark's example in a particular way. This problem concerns the way people believe aspects of spatial reasoning to be related. In particular, in the experiments about half of the participants instructed the robot with a strategy which is also most common among humans, namely to name the goal object to which the robot was supposed to move. The other half, however, started off with another type of instruction, namely to describe a path along which the robot was supposed to move. Since the robot was not implemented for this kind of instruction, the system's feedback was only "error". The users' strategy now consisted in proceeding to more and more elementary strategies, up to sentences like 'move your rear wheels.' Similarly, those participants who had initially attempted a goal description but had failed because of some linguistic problem tried path descriptions later. Remarkable is that none but a single participant, who openly wondered about whether path or goal descriptions were more appropriate, returned from path descriptions or more elementary strategies to a goal description, which the robot would have understood. Even if prompted to do so, users were extremely reluctant to change their strategy. Thus, for the participants there was apparently a fixed order of simplicity vs. complexity regarding spatial instruction, which was unrelated to the robot's real implemented capabilities. For them, therefore, knowing how to move along a path constitutes the ineffable background for moving towards a goal object. To return to Clark's example, the participants behaved as if it was impossible to know Coit Tower without knowing that it is in San Francisco. We can thus conclude that participants orient to ineffable background as a source of common ground in the dialogues.

Grading of Information By grading of information, Clark (1996, p. 110-112) understands our knowledge of other people's knowing. He quotes results from experiments which show that we usually have a good idea of what our (human) communication partners know and what they are not likely to know; that is, in general we have a good judgement of the mutuality of information. What our results show is that this is not the case with robots. The participants are uncertain about what language the robot understands, which words, syntactic structures, formality and granularity levels are understandable to it, what it perceives (see below for a discussion of these aspects), and how it interacts with the world. However, the participants' displays of their noticing the absence of this kind of knowledge supports the category as such.

Personal Common Ground

Perceptual Basis A joint perceptual basis constitutes the prototype for personal common ground (Clark, 1996, p. 112). In the dialogues between the human speakers and the robot investigated, the conditions for a joint perceptual basis are not given; the situation is not equally accessible to both participants, that is, the robot's perceptual capabilities are much more restricted than those of a human being. Thus, a robot may not have the information that something is the case, although for the human speaker it is 'obvious'. Accordingly, speakers were found to be much aware of the fact that their perception may differ from the robot's perception, i.e. while a fact perceived indicates to them that something holds, it may not indicate the same fact to their communication partner. Thus, the participants were uncertain about whether the scene perceived by them constitutes the same situation to the robot. The questions participants asked during the experiments were thus: 'what does it see?', 'where is its front?' and even 'does it see anything at all?'. Thus, they displayed attention to determining a common perceptual basis.

Actional Basis The actional basis between the participants is constituted, according to Clark (1996, p. 114), by means of joint action, the prototype being talk. This includes the successful presentation, acceptance and acknowledgement of utterances (Clark and Schaefer, 1989b). When the conversational participants in our experiment were successful in giving an instruction, the robot's resulting action can be seen as an appropriate acceptance and the user's proceeding to the next task as a verfication of this interpretation of the instruction. Users, however, were also found to change their linguistic behaviour on the basis of failed joint action, i.e. when the system answered "error" only. Usually it took the participants several attempts before they succeeded; some participants did not achieve a single joint action at all. However, once they had discovered a way to make themselves understood, they sticked to it; that is, they adapted their linguistic behaviour according to their hypotheses about common ground, accumulated on the basis of previous joint action. Thus, users were found to attend to both successful and unsuccessful joint actions carried out in the interaction with the robot.

Personal Diaries By personal diaries, Clark (1996, p. 114) understands the previous joint actions carried out by the participants. Here it is not entirely clear in which way the personal diary differs from previous joint actions.

Acquaintedness (Friends and Strangers) Because of the limited interaction with the robot, participants hardly acquired acquaintedness with it. However, what the data show is that participants were constantly attempting to increase acquaintance with the robot in order to reduce their uncertainty. As results by Amalberti et al. (1993) show, users indeed adapt to machines in a way that can be described as increasing acquaintance. Thus, after three times (with breaks of at least a week in between) 60 minutes of interaction with the simulated system, participants believing to talk to a computer behaved similarly to those who had been told that they were talking to a human 'wizard'. How far acquaintedness with a robot can go, whether a private language may evolve (see also the problem of acquiring personal lexicons below), cannot be predicted on the basis of the current experiments. What the data do show, however, is that users try to increase the acquaintedness with the robot, that is, that they attend to it.

Personal Lexicons Because of the limited interaction with the robot, participants can not be said to have acquired a personal lexicon with it, though there are interpersonal differences in their linguistic strategies (for instance, in the choice of goal- versus path-based instructions), and thus idiosyncratic communicative means may have developed. In any case, participants gave up using particular words after some time of interaction if they suspected them to be problematic, so that speakers can be argued to attend at least 'negatively' to a common personal lexicon.

5 Conclusions and Prospects

Using the methodology developed in this paper, the example analysis of how speakers attend to the aspects of context proposed in Clark's typology of common ground could be used to support the categories proposed. In particular, the problems users have in their formulating of utterances for the robot as a communication partner point to the fact that we normally know very much about our co-participants by drawing at least on those resources mentioned in Clark's typology. The results of this study show that in the communication with an unfamiliar communication partner users indeed attend to these resources. At the same time, the results indirectly support Clark's hypothesis that we build on all of those above mentioned types of information for our joint actions in human-to-human communication.

Regarding specific categories, it could be shown that the categories related to the linguistic resources have to be extended; all linguistic levels may be part of the negotiation of common ground, not just the lexicon. Furthermore, the common ground also consists of basic theories about how the world works, in this case, that moving towards a goal presupposes knowing how to move along a path and how to use the respective devices for moving (engines, wheels). Knowledge as basic as how to navigate in space is therefore also part of the category **ineffable background**.

Finally, the distinction between actional basis and personal diaries was not found to be useful since the common diaries are built up on the basis of previous joint action.

The method proposed thus turns out to be very useful for analyzing aspects of context. While the analysis so far only relied on a very restricted set of data, mostly the self-talk and the questionnaire data gathered in a very limited interaction with a robot in a very simple scenario of spatial reference, it is obvious how the methodology developed can be extended to investigate in a controlled way particular aspects of the context of situation and the ways in which speakers make use of them.

Bibliography

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

Berry, M. (1981). Systemic linguistics & discourse analysis. In Coulthard, M. and Montgomery, M., editors, *Studies in Discourse Analysis*, pages 120–145. London: Routledge.

Capra, F. (1996). *The Web of Life*. Doubleday, New York: Anchor Books.

Clark, H. and Schaefer, E. F. (1989a). Collaborating on contributions to conversations. In Dietrich, R. and Graumann, C., editors, *Language Processing in Social Context*, pages 123–152. North-Holland: Elsevier.

Clark, H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22:1–39.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In Resnik, L., Levine, J., and Teasley, S., editors, *Perspectives on Socially Shared Cognition*. Academic Press.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I., editors, *Elements of Discourse Understanding*. Cambridge University Press.

Clark, H. H. and Schaefer, E. F. (1989b). Contributing to discourse. *Cognitive Science*, 13:259–294.

Deppermann, A. (1999). *Gespräche Analysieren*. Opladen: Leske & Buderich.

Diewald, G. (1991). *Deixis und Textsorten im Deutschen*. Number 118 in Reihe Germanistische Linguistik. Tübingen: Niemeyer.

Douglas, S. A. (1995). Conversation analysis and hci-design. *Cambridge Series on Human-Computer Interation*, pages 184–203.

Duranti, A. and Goodwin, C., editors (1992). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge University Press.

Edmonds, B. (1999). The pragmatic roots of context. Technical Report CPM-99-52, Manchester Metropolitan University.

Fasold, R. (1990). *The Sociolinguistics of Language. Introduction to Sociolinguistics*, volume II of *Language in Society*. Cambridge: Blackwell.

Fillmore, C. J. (1981). Pragmatics and the description of discourse. In Cole, P., editor, *Radical Pragmatics*, pages 143–166. New York etc.: Academic Press.

Fischer, K. (1999). Repeats, reformulations, and emotional speech: Evidence for the design of human-computer speech interfaces. In Bullinger, H.-J. and Ziegler, J., editors, *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany*, pages 560–565. Lawrence Erlbaum Ass., London.

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(05):85–92.

Ginzburg, J. (1998). Shifting sharing and access to facts about utterances. In Heydrich, W. and Rieser, H., editors, *Proceedings of the 10th European Summer School in Logic, Language and Information – Workshop on "Mutual Knowledge, Common Ground and Public Information"*, pages 30–35.

Goffman, E. (1978). Response cries. *Language*, 54:787–815.

Goffman, E. (1979). Footing. *Semiotica*, 25(1/2):1–29.

Gumperz, J. (1982). *Discourse Strategies.* Number 1 in Studies in Interactional Sociolinguistics. Cambridge University Press.

Gumperz, J. (1992). Communicative competence revisited. In Duranti, A. and Goodwin, C., editors, *Rethinking Context: Language as an Interactive Phenomenon.* Cambridge University Press.

Halliday, M. (1964). The linguistic study of literary text. In Lunt, H., editor, *Proceedings of the Ninth International Congress of Linguists*, pages 302–307. London, The Hague, Paris.

Halliday, M. (1970). Functional diversity in language as seen from a consideration of modality and mood in english. *Foundations of Language*, pages 322–361.

Halliday, M. (1984). Language as code and language as behaviour: A systemic functional interpretation of the nature and ontogenesis of dialogue. In Fawcett, R., Halliday, M., Lamb, S., and Makkai, A., editors, *The Semiotics of Culture and Language*, volume 2: Language and Other Semiotic Systems of Culture, pages 3–35. London: Pinter.

Heritage, J. (1995). Conversation analysis: Methodological aspects. In Quasthoff, U., editor, *Aspects of Oral Communication.*, number 21 in Research in Text Theory, pages 391–418. Berlin/New York: de Gruyter.

Hutchby, I. (2001). *Conversation and Technology: From the Telephone to the Internet.* Cambridge: Polity Press.

Hutchby, I. and Wooffitt, R. (1998). *Conversation Analysis.* Cambridge: Polity Press.

Krause, J. (1992a). Fazit und Ausblick: Registermodell versus metaphorischer Gebrauch von Sprache in der Mensch-Computer- Interaktion. In Krause, J. and Hitzenberger, L., editors, *Computertalk*, number 12 in Sprache und Computer, pages 157–170. Hildesheim: Olms.

Krause, J. (1992b). Natürlichsprachliche Mensch-Computer-Interaktion als technisierte Kommunikation: Die computer talk-Hypothese. In Krause, J. and Hitzenberger, L., editors, *Computertalk*, number 12 in Sprache und Computer, pages 157–170. Hildesheim: Olms.

Kreckel, M. (1981). *Communicative Acts and Shared Knowledge in Natural Discourse.* London etc.: Academic Press.

Labov, W. (1966). The social stratification of english in new york city. Center for Applied Linguistics, Washington.

Larrson, S., Cooper, R., and Engdahl, E. (2000). Question accomodation and information states in dialogues. In *Proceedings of the Third Workshop on Human–Computer–Conversation*, pages 93–98, Bellagio, Italy.

Moratz, R. and Fischer, K. (2000). Cognitively adequate modelling of spatial cognition in human-robot interaction. In *Proceedings of the 12th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2000*, pages 222–228, Vancouver, British Columbia, Canada.

Nothdurft, W. (1984). *"...äh folgendes Problem äh...". Die interaktive Ausarbeitung "des Problems" in Beratungsgesprächen.* Number 57 in Forschungsberichte des Instituts für Deutsche Sprache Mannheim. Tübingen: Narr.

O'Donnell, M. (1999). Context in dynamic modelling. In Ghadessy, M., editor, *Text and Context in Functional Linguistics*. Amsterdam: John Benjamins.

Paris, C. L. and Bateman, J. A. (1989). User modelling and register theory: A congruence of concerns. Technical Report 1990, USC Marina des Rey, CA.

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

Sacks, H. (1992). *Lectures on Conversation*, volume I and II. Oxford, Cambridge: Blackwell. ed. by Gail Jefferson.

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

Schegloff, E. A. (1997). Whose text, whose context? *Discourse & Society*, 8(2):165–187.

Schiffrin, D. (1994). *Approaches to Discourse.* Oxford, Cambridge: Blackwell.

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

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversations.* PhD thesis, University of Rochester.

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.