Paper 30

How much Common Ground Do we Need for Speaking?

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

1 Introduction

The question addressed in this paper is to which types of common ground speakers attend in dialogical interactions. The procedure is to investigate 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. Analysing what they request for producing utterances for their communication partner reveals to which types of common ground they orient.

Most research on common ground has been carried out on what is shared between the conversational participants on the basis of the discourse record of the current situation. Building on work by Clark and collaborateurs (Clark and Marshall, 1981; Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989; Clark and Brennan, 1991), summarized in Clark (1996), much work has addressed aspects of 'grounding', the process by which individuals add information to the common ground (Traum, 1994; Ginzburg, 1998). 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).

However, as Clark (1996, p. 92-121) points out, the shared basis for joint action speakers draw upon consists in a number of further aspects besides the discourse record, and he provides us with a list of knowledge types speakers may use as possible resources to establish common ground. Such resources include knowledge about the human nature, a common lexicon, knowledge about scripts, and knowing how. The problem addressed in this paper, what speakers really draw upon in discourse, has rarely been studied, and if so, only in natural conversations (Kreckel, 1981; Clark, 1996). 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 they consider the necessary common ground for their producing of utterances. 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. 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 how much, and which types of, information is necessary for communication to work because in these cases, common ground is attended to as potentially problematic. The methodology underlying this study relies thus on the conversation analytic principle of 'deviant case analysis', based on the idea that deviant cases are not only orderly in themselves, but, as Hutchby and Wooffitt (1998, p. 95-98) argue for the analysis of sequences, if "someone displays in their conduct that they are 'noticing' the

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).
Paper 30,
kerstinf@uni-bremen.de
http://www.fb10.uni-bremen.de/anglistik/homepages/fischer.htm.

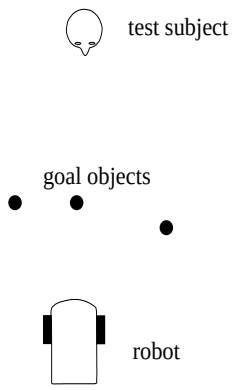

Figure 30.1: The Experimental Setup

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, the types of common ground to which speakers attend may become evident. The methodology is thus to analyse the participants' displays of their noticing the absence of aspects of common ground in human-robot interaction.

The results will not only be theoretically interesting because they provide evidence for a typology of common ground, they will also have important practical consequences regarding the modelling of cognitive agents, such as robots.

2 Data

The data were elicited in an experimental setting (see Figure 30.1) for which initially, a robot (see Figure 30.2) 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 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 603 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.

Figure 30.2: The Robot

3 Types of Common Ground Attended to in the Data

According to Kreckel (1981, p. 29), the background knowledge that plays a role in interactions can be distinguished into three kinds:

1. "what remains idiosyncratic and, thus, has to be labelled experience;
2. what is based on mutual acquaintance oo knowledge about and, thus, can be considered as implicitly or explicitly shared;
3. what is based on separate acquaintance or knowledge about and, thus, may or may not be in common between communicants."

The criterion on the basis of which Kreckel distinguishes what the speakers can build upon as common is thus whether the respective knowledge is acquired jointly or separately by the interactants. Clark's (1996) analysis is more sophisticated in showing how these different types of knowledge relate to each other and by providing further, more fine-grained, distinctions (Clark, 1996, p. 54). Thus 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 (Kreckel's common knowledge) and personal (Kreckel's shared knowledge) common ground.

Clark (1996, p. 92-96) argues that common ground is best be seen as a shared basis between the participants. This means that

> both participants have the information that b holds;
>
> b indicates to both that both have the information that b holds;
>
> b indicates to both that p.

This may give rise to reflexive knowledge about common ground:

(i) both have the information that p and that (i).

Clark (1996, p. 100-120) distinguishes two types of shared bases, 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 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. This may also suggest which kinds of background they preferred to have for formulating their utterances for their communication partner.[1]

3.1 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. 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.

[1] 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.

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 write 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, 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. Participants thus 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 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.

3.2 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?'.

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, 1989). 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. 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 'negatively' to a common personal lexicon.

4 Conclusions and Prospects

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. Thus, 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.

What practical consequences do our results have? Clark (1996, p. 116-120) has argued that conversational participants have techniques for building up common ground, for instance, by deliberately displaying community affiliations. This may point to a way how future systems can be significantly improved: Strategies have to be found by which the artificial system can signal to its interactants what its abilities and strengths are, and thus to inform the human interlocutor about its 'nature', its linguistic and perceptual capabilities, and even its ineffable background. This has to be done subtly and implicitly since too much in-advance instruction has turned out to be unpleasant, at best (Ogden and Bernick, 1996; Fischer and Batliner, 2000). However, the results of this investigation have shown how great the need for a common ground between system and user is and how much users invest to build up hypotheses about how the robot works. These conceptualization and adaptation processes could be exploited for the improvement of future human-robot interaction.

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.

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. (1989). Contributing to discourse. *Cognitive Science*, 13:259–294.

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

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. (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.

Fischer, K. and Batliner, A. (2000). What makes speakers angry in human-computer conversation. In *Proc. of the Third Workshop on Human–Computer–Conversation*, pages 62–67, Bellagio, Italien.

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.

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

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

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.

Ogden, W. and Bernick, P. (1996). Using natural language interfaces. In Helander, M., editor, *Handbook of Human-Computer Interaction.* Elsevier Science Publishers, North Holland.

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.