Why a static interpretation is not sufficient in spatial communication[*]

John A. Bateman
FB10: Faculty of Linguistics and Literary Sciences
University of Bremen
Bremen, Germany
bateman@uni-bremen.de

Kerstin Fischer
FB10: Faculty of Linguistics and Literary Sciences
University of Bremen
Bremen, Germany
kerstinf@uni-bremen.de

Thora Tenbrink
FB10: Faculty of Linguistics and Literary Sciences
University of Bremen
Bremen, Germany
tenbrink@informatik.uni-bremen.de

Abstract

This paper proposes a research methodology for attacking the problem of providing fluent and natural discourse about space and spatially situated tasks between naïve users and robots. We suggest flexible and adaptive ontology mediation, parameterized according to empirically determined discourse and contextual factors, as a suitable architecture with clear applications for the treatment of natural human-human dialog also.

1 Introduction

Linguistic communication about space in a situated human-robot interaction scenario involves more complex and flexible interpretation mechanisms than might be expected. In addition to the inherent complexity of interpreting situationally dependent – and interactively negotiated – spatial language, users unfamiliar with their artificial interlocutor address the system according to assumptions about its abilities that are often inaccurate. This paper aims at identifying central problems involved in spatial communication between humans and robots, and presents our approach to solving them.

Natural language is an essential mode of interaction between users and sophisticated spatially-aware systems such as mobile assistance robots. Providing suitably sophisticated and flexible natural language capabilities for ever more complex interaction scenarios is a major problem. Such interaction needs to be as natural and non-intrusive as possible in order to support the widest possible range of potential users. However, most of the features that make interaction 'natural' still present substantial challenges for dialog systems.

The consideration of robots and humans in real communicative situations presents a valuable research environment for investigating dialog systems further. The abilities, knowledge and linguistic responses of the robots can be varied experimentally in a manner not possible in human-human interaction. The combination of the necessary modules in a complete dialog system is, however, a major research challenge in its own right—one which again has similarities with the development of accounts for human dialog alone. In particular, the lack of appropriate modularizations of the technical components involved in complete systems that combine spatial and linguistic capabilities is a significant bottleneck.

One particularly effective strategy for modularization is the adoption of *linguistically motivated ontologies* that mediate between domain or application knowledge and Human Language Technology (HLT) components. It might therefore be expected that some kind of ontology mediation could be usefully applicable to the modeling of dialog capabilities.

Problematic, however, is that many results concerning natural dialog between humans have

[*] The authors are listed in alphabetical order.

now established that the form of linguistic expressions employed during dialog depends crucially on the interaction situation. This effect is especially strong when dealing with settings which involve discourse participants with differing perceptual and linguistic abilities and preferences or when those settings are themselves highly complex and change while the discourse develops as the participants move around. Such settings represent the standard case in mobile robot interaction scenarios. Current strategies for ontology mediation exhibit a rigidity that is inappropriate for the relationships observed in real interactions. Here, a *negotiation* of mediation within 'conversational' interaction appears crucial both for ontology design and for achieving inter-operability.

This combines several directions of research in a new way. Our purpose in this paper is to briefly motivate the research domain of human-robot interaction as a methodology for investigating dialog and to set out our own investigative methods and modeling goals as projected in our newly established research group[†]. Our particular focus is on the specific effect of a robot interaction partner on the linguistic and spatial choices of a human speaker, and on the negotiative processes involved in achieving common ground. These factors are to be investigated as situational parameters determining the linguistic properties of the language employed. Inter-ontology mediation will then be used to parameterize mappings between situational factors and linguistic processing and generation.

2 Spatial Communication in a Robotic Scenario

The empirical questions we address are located in two problem areas: first, the complexity of the interpretation of spatial expressions based on the considerable variability of implicitly underlying reference systems and the associated negotiation processes between the interactants; second, the peculiarities pertaining to the choice of linguistic expressions in an unfamiliar interaction situation involving an artificial interlocutor. Both problem areas combine in linguistic interaction scenarios in which users are required to communicate with an unfamiliar robot about spatial surroundings.

[†] The research project I1-OntoSpace, directed by John Bateman, Kerstin Fischer, and Reinhard Moratz, is part of the newly established Collaborative Research Center (SFB) "Spatial Cognition" located at the universities of Bremen and Freiburg.

2.1 Interpreting Spatial Expressions

In the communication of spatial information, work has addressed the formal analysis of the meaning of spatial expressions on the one hand, and their choice under certain specifiable circumstances on the other. Lexical semantic approaches discuss the possibility of a core semantic meaning for spatial expressions from which possible deviations can be derived (e.g., Eschenbach 1999); psychologically inspired work has revealed the importance of functional categorizations (Garrod and Sanford 1988; Coventry, 1998); while work from robotics has introduced the notion of a field potential for describing degrees of likelihood of positioning (cf. Stopp *et al.,* 1994).

Psycholinguistic experimental studies on spatial situations focus on different kinds of mental representations that are reflected verbally in the speakers' utterances. A central concept underlying much spatial expression work is that of *reference systems* (Hermann and Grabowski, 1994). Such systems may be *intrinsic, relative*, and *absolute (*Levinson, 1996), all of which may – depending on various factors of the actual situation – be employed from either of three different perspectives: speaker-centered, listener-centered, or third-party point of view. In intrinsic reference systems, objects are located by referring to the intrinsic properties of another entity, such as the speaker's front in "The ball is in front of me". Relative reference systems depend on the presence of a further entity (the so-called *relatum*), as in "The ball is in front of the table". If (at least) one similar entity is present rather than a *different* relatum, speakers employ *group-based* reference (Moratz et al., 2001). Absolute reference systems depend on the earth's cardinal directions, such as *north* or *south*.

In tasks involving route descriptions rather than the localisation of objects, further kinds of perspectives, and combinations of perspectives, are available to the speaker: for instance, one can assume the perspective of an "imaginary wanderer" (an imagined person that walks along the route described). At the same time speakers may refer to landmarks available in the scenery (Hermann and Grabowski, 1994), and they may adapt their linguistic choices of spatial expressions flexibly according to the changing visual perception.

The complexity of the repertory of available spatial reference systems requires a high degree of flexibility and awareness of possible misunderstandings in natural interaction, all the more so as the intended kind of reference is seldom explicitly specified and subsequently retained throughout a communication situation. Rather, it is negotiated and changed during the interaction (Tversky et al., 1999). Especially in cases of qualitative or vague spatial instructions, there is often a need for interactive negotiation about the exact semantics of the instruction (Wachsmuth and Cao, 1995). Conversational participants often agree on situationally dependent representations, such as metonymies, in order to achieve smooth and effective communication (e.g., Rieser, 1996). Reference resolution is achieved in dependence on the visual as well as the linguistic context; influenced, for instance, by the current focus of attention (Kessler et al., 1999). This process may reach a high level of complexity if the discourse situation offers a wide range of conceptual representations.

Speakers also react to their interaction partner's contributions, and appear to attune their linguistic choices to what they believe to be suitable for their partner in the situation at hand. For example, Schegloff (1972) shows how "formulating place" depends on the recipient for whom the description is designed. Schober (1993) found that speakers attend to their hearers' clues as to whether they have understood the instruction in the sense that the references have been *grounded* (Clark and Wilkes-Gibbs, 1986). Garrod and Anderson (1987) found that communication partners interactively developed distinct but consistent description schemes, which reflected different kinds of underlying mental representations. These were dependent on the interaction itself as well as on the given task.

2.2 Communicating with Artificial Interlocutors

In the same way as the use of spatial expressions depends on the recipient and on the negotiation processes between the interlocutors, linguistic choices in general are influenced by the speakers' understanding of their communication partner and the interactive negotiation of this conceptualization during the interaction (cf. Sacks, 1992). This is most obvious in the communication with artificial interlocutors; human-computer interaction has therefore been suggested to constitute a special linguistic register (Krause & Hitzenberger,1992).

Several studies on human-computer interaction have shown that users may differ considerably in the language they direct to a system and thus that the communication with artificial systems is not homogeneous. Instead, the linguistic properties of the utterances depend very much on the way speakers conceptualize the system (cf. Fischer 2000) and on the system's output (cf. Fischer & Batliner, 2000). Since speakers design their speech for their recipients, the way they think about their communication partner may have a strong impact on the form of their utterances. Particularly the conceptualization of the robot as a tool versus as human-like may lead to significantly different linguistic behavior (Fischer 2000).

This insight corresponds to previous findings of psycholinguistic studies that show that the negotiation of linguistic behavior, or *alignment*, takes place on all linguistic levels (Garrod, 1999; Clark, 1996), a result supported and suggested by findings in human-computer interaction (Amalberti et al. 1993; Fischer 1999, 2000) and now again forcefully argued in Pickering and Garrod (in press).

Thus, human-computer conversation cannot be regarded as a single rigidly defined linguistic variety, but instead is constantly negotiated and adapted, influenced both by the system design and by the users' conceptualizations of the system. The question remains as to whether it is possible to specify limits and constraints on this variability. It is precisely this flexibility that makes this kind of interaction particularly relevant for research into the mechanisms of dialogue and interaction in general.

3 Our approach

Our approach has two interacting components. On the one hand, we undertake the controlled elicitation of data in experimental settings, varying particular situational parameters and relating the linguistic choices speakers make in the interaction with the robot to those parameters. On the other hand, we use the data to motivate particular models of the dialogue mechanisms involved, which will then be used for developing a dialogue system enabling more natural communication. The improved system will then be employed for further rounds of experiments.

Experiments, outlined in more detail below, will give insights about the effect different behaviors, appearances, and linguistic output of the robot may have on the users' ways of thinking about the robot and how these different conceptualizations influence the users' linguistic behavior. The linguistic properties affected by different conceptualizations of the robots are expected to be found on all linguistic levels, including lexical choices (e.g. basic level vs. abstract categories), syntactic constructions (e.g. the use of imperative), morphological reduction, and phonological/ prosodic features (e.g., syllable lengthening, speech rate, and other error resolution strategies).

An important aspect of the model to be constructed is the requirement that very different kinds of representations are integrated within a single functioning system. The robots, and their controlling software, typically incorporate spatial representations that may be arbitrarily far removed from those assumed by the human interactants and expressed in their linguistic utterances. The flexibility required during negotiation of shared discourse strategies within any interaction argues strongly against the 'prewired' solutions common in such work.

Here our approach draws on a standard technique for achieving modularity that has been pursued within several branches of NLP. This involves employing, either explicitly or implicitly, ontologies that represent information at various levels of abstraction within a system. One of the first systems to employ this technique for re-use was the Penman text generation system (Mann and Matthiessen, 1985), within which the Penman Upper Model was developed (Bateman et al., 1990). Re-use for natural language generation within this framework was described in detail by, e.g., Bateman (1990), and follows the strategy of subordinating domain model concepts to upper model concepts so that domain concepts inherit the linguistic possibilities for expression available to their superordinate upper model concepts. For example, if a domain object—such as some entity recognized by the robot—is subordinated to the upper model concept `Nondecomposable-Object`, then the generation component knows which kinds of linguistic constructions may be used for describing this concept and which not. This approach has since been widely employed.

Although an obvious benefit of this approach is that it enables semantic specifications to make direct use of domain model concepts, it also has a striking deficiency. This was revealed most clearly in research efforts aimed at *multiple register* generation—that is, the generation by single generic text generation systems of texts belonging to diverse text types and for varying levels of reader/user expertise (cf. Bateman and Paris, 1989). Different registers can easily require single domain concepts to be expressed so divergently that any single upper model concept assignment is invalidated. What one type of text may consider nondecomposable might in another be treated as non-atomic and decomposable into parts.

This necessitates the maintenance of at least two distinct levels of ontological information—one purely linguistic semantic, that of the upper model, and one for the domain model—with flexible mapping relationships between them. In the context of our experiments here, the domain model will focus particularly on spatial relationships and object attributes shown to be relevant for forming spatial expressions.

The preservation of two distinct levels of information is echoed in a range of very diverse approaches (cf., e.g., the distinction between LF and QLF in Alshawi, 1992; in temporal semantics in Herweg, 1991; and ontology-based utterance analysis in Lang, 1991), the extent and range of the flexibility required in mappings between levels has not been mapped out satisfactorily. It has also not been anchored as firmly as is necessary in the details of linguistic negotiation within dialog.

Our experiments will seek to provide crucial data concerning the range and flexibility of the mappings required, focusing on the area of spatial relations and their linguistic expression. The allocation of spatial configurations to particular linguistically expressed orientations is known to vary according to the interactional context and the kinds of spatial configurations encountered. As a very simple illustration, for example, whether a speaker uses 'on' or 'in' a location reflects as much a negotiated reference system within an interaction as it does a statement about *a priori* dimensionality of the location referred to. Such allocations, expressed in terms of links between selections from the spatial and linguistic ontologies, are to be parameterized according to aspects of the negotiated communication revealed by the experiments.

4 Empirically-derived Linguistic Parameterization

We address the parameterization of ontology mappings related to spatial configurations by making the relationship between situational variables and linguistic properties transparent. For instance, the users' choices of spatial reference systems and of strategies for referring to landmarks are central parameters of linguistic variability in human-robot interaction. Users may (justifiably!) be uncertain about what the robot can perceive, and so the lack of mutual and reflexive common ground for the interactants regarding the spatial situation can lead to insecurity about which objects may serve as landmarks and how they can be referred to. This variability can be experimentally controlled. Closely related to this factor are the participants' linguistic and spatial choices concerning group-based reference – a kind of reference system often neglected in the literature – using further similar objects instead of a different object as a relatum to specify the target object's position (Moratz et al., 2001). Such issues are addressed by confronting users with tasks involving different configurations of diverse (similar and differing) objects, the robot, and the user.

4.1 An Example: Perspective Taking

In previous work we have established that controlled experimental settings yield significant information concerning modeling requirements. In Moratz *et al.* (2001) it was shown that all users consistently – and often inexplicitly – took the robot's perspective in their linguistic expressions. Subsequently, in order to see whether the previous finding was an artifact of the experimental setting, we carried out further studies involving different conditions designed for prompting variation in perspective taking. The experimental situation was designed to reveal whether goal instructions would be made relative to the listener (the robot or another human) or the speaker, whether extrinsic reference systems would be employed (e.g. 'drive north'), and whether a group of similar objects would be used for object localization.

We varied a number of different scenarios, experimenting with different positions of human instructor and robot, and with different addressees (a human and a robot instructee). In all settings, the human instructors were told to instruct their (human or robot) communication partner to localize objects on the basis of their spatial position with respect to similar objects by means of written instructions. In one scenario, instructors were seated in front of a computer in which they should type their instructions. The robot's task was to measure the distance between the objects indicated. The position of the robot (a Pioneer 2) with respect to the objects and the instructor was varied (see Figure 1 where the robot is located at Pos. C) in order to identify factors influencing perspective taking. This scenario was repeated with 21 instructors.

In another scenario the same task was carried out by 17 instructors who communicated with a human communication partner by means of written instructions. Here the task for the human instructee (replacing the robot in the previous experiment) was simply to point to the object described (instead of measuring distances).

A preliminary analysis revealed that the data in the human-to-human experiment basically replicate the results obtained by Schober (1998): instructors varied in taking either their own or the other's perspective, marking their choice usually only in the case of problems after a negotiation process. In contrast, in the human-robot scenario, irrespective of the position of the robot with respect to the instructor, instructors only took the robot's perspective. This strategy was often explicitly marked. If the task was to measure the distance between an object and the robot itself, the robot was often not addressed itself but referred to as *robot* or *robby*.

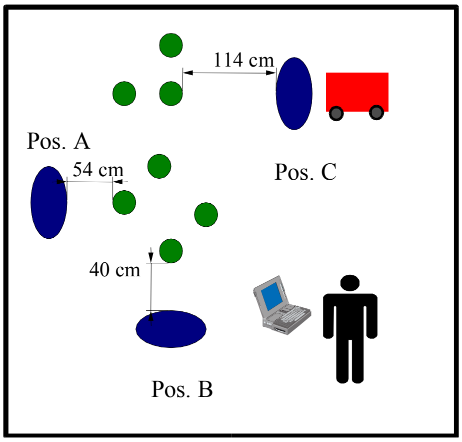

Figure 1. Experimental Scenario.

4.2 Determinants of Linguistic Choices

The results of the previous experiments suggest that in human-robot interaction, irrespective of the particular spatial configuration, human instructors reliably take the robot's point of view. This behavior differs from that found in human-human interaction and so needs to be considered in designing dialog interpretation strategies for such systems.

From this starting point it is necessary to investigate more closely the precise conditions which trigger the difference in observed dialog behavior. So far, we have used only one type of robot in the experiments, and the linguistic output by the robot was tightly constrained for methodological reasons (Fischer, in press). Further plans now are to carry out experiments with a different type of robot (the Aibo from Sony, a robot resembling a small dog) and employing more advanced linguistic capabilities on the part of the robot. This allows us to vary the communicative situation between the original simple robot scenario and human-human communication in a precisely controlled fashion. This should reveal the determining factors for instructor's choices with respect to perspective taking.

Currently we hypothesize that this behavior is conditioned by the level of sophistication of the robot's output on the one hand, and the complexity of the spatial setting on the other. Thus, more challenging scenarios that we now plan to explore will require the human users to employ localization *sequences* in order to achieve the communicative goal, which may consist in instructing the robot to move along a certain path towards a specific location. Along that path diverse kinds of objects can be placed that may or may not be used by the test subjects as landmarks for the route instruction. Here, a specific research issue is that of consistency in the instruction chains used in route instructions. For this purpose, experimental scenarios will be used which at certain places offer themselves for diverse kinds of reference systems. The precise linguistic behavior of the robot during the experiment will also be varied. Our expectation is that moving through an environment with several differing elements will lead the user to employ varying kinds of reference systems in single localization sequences.

4.3 Parameterized Ontology Mapping

We are planning to employ inter-ontology mappings in order to link abstract spatial representations with their possible linguistic expressions in a flexible manner. The users' choices in different experimental scenarios and in different dialog situations allow us to explore the limits and requirements of these variable mappings.

We can illustrate the potential role of parameterization of this kind of mediation by considering again the reference systems introduced above. For each reference system, various objects located spatially need to be assigned to specific roles within the reference system—e.g., as landmarks—and to receive an orientation. If the speaker is facing the robot, and the robot is *assumed* by the speaker to be facing him or her, then a choice of reference system relative to the speaker will allocate 'left' and 'right' in one orientation whereas a choice relative to the robot will allocate the same terms to the directly opposite orientation. This means that the assignment between concrete objects and relations in the spatial domain will be mapped to quite distinct objects and directions in the linguistic domain. This relationship can be captured as a relation between the two ontology levels. Which assignment is made may be left implicit in the utterances of the speaker, may be indicated explicitly or, the case of particular interest to us, may be determined by other factors in the communicative situation—including both how the robot appears to the human and the linguistic behavior of the robot.

The flexibility of mediation is to be modeled by the parameterization of the particular inter-ontology mapping constructed, while the kinds and limits of parameterization needed will be explored experimentally. This relationship is summarized graphically in Figure 2.

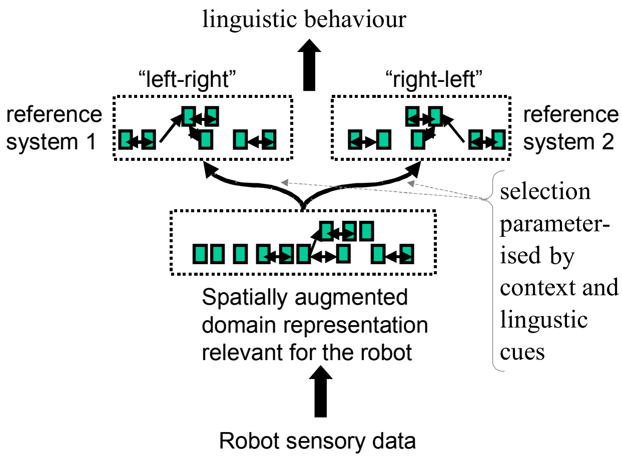

Figure 2. Parameterized Inter-Ontology Mediation.

The references systems are represented by particular configurations of concepts and relations in the linguistic ontology. These are then mapped to the spatial representations. The parameterization of this mapping falls into two main categories: situational and interactional. The situational parameters include the appearance of the robot, the language complexity, spatial complexity and task complexity; the interactional parameters include the particular linguistic forms used in the interaction and the dialog strategies employed.

For the perspective-taking problem illustrated here, for instance, the users' view of the robot seems to entail that their instructional choice will generally be to take the robot's perspective. This is then accounted for by specifying appropriate constraints on the particular subordination employed between spatial and linguistic ontology. More complex considerations will involve the mapping of spatial relations in the spatial ontology to appropriate configurations in the linguistic ontology: this again depends on both situational and interactional context.

5 Conclusion and Outlook

The main feature of the research method we propose is a combination of experimental control and considerable flexibility in the environment in which the human and artificial agents are interacting. This flexibility of the environment is directly mirrored in the language found in the various situations: the linguistic properties of the users' utterances, and the corresponding requirements for natural language analysis and generation in the dialog system, will depend at least on the perspective of the user, the arrangement of objects and landmarks, the user's conception of their artificial communication partner, the interactional history, and whether user and system share the same spatial position. This variability demands not particular solutions to selected problems, but generic and re-usable techniques.

Such techniques may be supported by parameterized ontology mediation of the kind suggested. Moreover, there is a direct link between this kind of approach and earlier work on multiple register generation and analysis. The situational parameters correspond to traditional notions of register variation—i.e., depending on the situational features holding, differing linguistic features occur with higher or lower probability (cf. Biber, 1993). Another way of viewing this process is to say that the domain ontology configurations are associated with differing 'views' of the linguistic system. As the register changes, so does the linguistic behavior observed.

We now take this further by adding in the interactional parameters. This means that the precise correspondences between linguistic and domain configurations, as well as the 'view' of the linguistic system, can vary within an interaction; this has been termed *microregister* variation in Bateman (1986). Microregisters follow closely the spirit of Garrod and Sanford (1988) and the use of dialog interaction to construct 'ontologies' of the domain of discourse on the fly, while more explicitly considering shared (or aligned) linguistic representations as shown to be necessary in Pickering and Garrod (in press).

Acknowledgements

This work is part of the project I1-OntoSpace of the newly established SFB/TR8 "Spatial Cognition" coordinated by Christian Freksa. We are grateful for the financial support of the DFG and to Reinhard Moratz for useful discussions and cooperation in the robotics area.

References

Alshawi, H. (ed.)(1992) *The Core Language Engine,* Cambridge, Massachusetts: MIT Press.

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

Bateman, J.A. (1986) *Utterances in context: towards a systemic theory of the intersubjective achievement of discourse,* PhD dissertation. University of Edinburgh, Scotland.

Bateman, J.A. (1990) Upper Modeling: Organizing Knowledge for Natural Language Processing. In: *Proceedings of the Fifth International Natural Language Generation Workshop,* pp. 54--60. Dawson, PA .

Bateman, J.A. and Paris, C.L. (1989) Phrasing a Text in Terms the User Can Understand. In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence,* pp. 1511--1517. Detroit, Michigan.

Bateman, J.A., Kasper, R.T., Moore, J.D. and Whitney, R.A. (1990) *A general organization of Knowledge for natural language processing: the PENMAN Upper Model,* Marina del Rey, California: USC/Information Sciences Institute .

Biber, D. (1993) Using register-diversified corpora for general language studies. *Computational Linguistics* **19,** 219--242.

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

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

Coventry, K.R. (1998) Spatial Prepositions, Functional Relations, and Lexical Specification. In Olivier, Patrick and Klaus-Peter Gapp (eds.) *Representation and Processing of Spatial Expressions*. Mahwah, New Jersey: Lawrence Erlbaum, pp. 247-262

Eschenbach, C. (1999) Geometric structures of frames of reference and natural language semantics. *Spatial Cognition and Computation*, 1 (4), 329-348.

Fischer, K. (1999) Repeats, Reformulations, and Emotional Speech: Evidence for the Design of Human-Computer Speech Interfaces. In: Bullinger, H. & Ziegler, J. (eds.): *Human-Computer Interaction: Ergonomics and User Interfaces*. Lawrence Erlbaum Ass., London, pp. 560-565.

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

Fischer (in press). Linguistic Methods for Investigating Concepts in Use. In: Stolz, Th. and Kolbe, K. (eds.): Methodologie in der Linguistik. Frankfurt a.M.: Lang.

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

Garrod, S. (1999): The Challenge of Dialogue for Theories of Language Processing. In: Garrod, S. and Pickering, M.J. (eds.): *Language Processing*. Hove: Psychology Press.

Garrod, S. and Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition* 27: 181-218.

Garrod, S.C., and A.J. Sanford (1988) Discourse models as interfaces between language and the spatial world. *Journal of Semantics*, 6:147-160.

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

Herweg, M. (1991) Aspectual requirements of temporal connectives: evidence for a two-level approach to semantics. In: Pustejovsky, J. and Bergler, S., (eds.) *Proceedings of the 1991 ACL Workshop on Lexical Semantics and Knowledge Representation,* pp. 152--164. Berkeley, CA.

Kessler, K., Duwe, I. and Strohner, H. (1999) Grounding Mental Models: Subconceptual Dynamics in the Resolution of Reference in Discourse. In: Gert Rickheit and Christopher Habel (eds.), *Mental Models in Discourse Processing and Reasoning*. Amsterdam: Elsevier.

Krause, J. and Hitzenberger, L (1992). *Computertalk*. Hildesheim: Olms.

Lang, E. (1991) The LILOG ontology from a linguistic point of view. In: Herzog, O. and Rollinger, C. (eds.) *Text understanding in LILOG: integrating computational linguistics and artificial intelligence, Final report on the IBM Germany LILOG-Project,* pp. 464--481. Berlin : Springer.

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

Mann, W.C. and Matthiessen, C.M.I.M. (1985) Demonstration of the Nigel Text Generation Computer Program. In: Benson, J.D. and Greaves, W.S., (eds.) *Systemic Perspectives on Discourse, Volume 1,* pp. 50--83. Norwood, New Jersey : Ablex.

Moratz, R., K. Fischer, and T. Tenbrink. (2001) Cognitive Modelling of Spatial Reference for Human-Robot Interaction. *International Journal On Artificial Intelligence Tools*, 10:4, World Scientific Publishing, Singapur.

Pickering, M.J and Garrod. S. (in press) Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences.*

Rieser, H. (1996) Repräsentations-Metonymie, Perspektive und Koordination in aufgabenorientierten Dialogen. In: C. Umbach, M. Grabski and R. Hörnig (Hrsg.). *Perspektive in Sprache und Raum*. Wiesbaden: Deutscher Universitäts-Verlag.

Sacks, H. (1992): *Lectures on Conversation*. Oxford: Blackwell.

Schegloff, E. A. (1972) Notes on a conversational practice: formulating place. In Giglioli, Per Paolo (ed.): *Language and social context.* Harmondsworth: Penguin.

Schober, M.F. (1993). Spatial Perspective-Taking in Conversation, *Cognition* 47: 1-24.

Schober, M.F. (1998). How addressees affect spatial perspective choice in dialogue. In P.L. Olivier & K.-P. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 231-245). Mahwah, NJ: Lawrence Erlbaum.

Stopp, E., Gapp K.-P., Herzog, G. Laengle, T. and Leuth, T. (1994) Utilizing spatial relations for natural language access to an autonomous mobile robot. *Proceedings of KI-94,* pp39-50.

Tversky, B., P. Lee, and S. Mainwaring. (1999) Why do speakers mix perspectives? *Spatial Cognition and Computation*, 1:399-412.

Wachsmuth, I. and Y. Cao (1995) Interactive Graphics Design with Situated Agents. In W. Strasser and F. Wahl (eds.): *Graphics and Robotics* (pp. 73- 85), Berlin Heidelberg New York: Springer.