Keeping the initiative: an empirically-motivated approach to predicting user-initiated dialogue contributions in HCI

XXXXXXXXXXX

Abstract

In this paper, we address the problem of reducing the unpredictability of user-initiated dialogue contributions in human-computer interaction without explicitly restricting the user's interactive possibilities. On the basis of an empirical analysis of user behaviour in comparable tasks across human-human and human-computer experimental situations, we demonstrate that it is possible to identify conditions under which particular classes of user-initiated contributions will occur. We present the empirical basis of these findings and discuss their consequences for dialogue system design.

1 Introduction

It is increasingly recognised that human-computer dialogue situations can benefit considerably from mixed-initiative interaction (Allen, 1999). Interaction where there is, or appears to be, little control of just when and how the user makes a dialogue contribution and increases the perceived naturalness of an interaction, itself a valuable goal, and also opens up the application of human-computer interaction (HCI) to tasks where both system and user are contributing more equally to the task being addressed.

Problematic with the acceptance of mixed-initiative dialogue, however, is the radically increased interpretation load placed on the dialogue system. This flexibility impacts negatively on performance at all levels of system design, from speech recognition to intention interpretation. In particular, clarification questions initiated by the user are difficult to process because they may appear off-topic and can occur at any point. But preventing users from posing such questions leads to stilted interaction and a reduced sense of control over how things are proceeding.

In this paper we pursue a partial solution to the problem of user-initiated contributions that takes its lead from detailed empirical studies of how such situations are handled in human-human interaction. Most proposed computational treatments of this situation up until now rely on formalised notions of *relevance*: a system attempts to interpret a user contribution by relating it to shared goals of the system and user. When a connection can be found, then even an apparently off-topic clarification can be accomodated. In our approach, we show how the search space for relevant connections can be constrained considerably by incorporating the generic conversation analytic principle of *recipient design* (Sacks et al., 1974, p727). This treats user utterances as explicit instructions for how they are to be incorporated into the unfolding discourse and can itself be accomodated within a general orientation for much current discourse semantic work whereby potential discourse interpretation is facilitated by formulating tighter structural and semantic constraints (Scha and Polanyi, 1988; Webber et al., 1999; Asher and Lascarides, 2003). We extend this here to include constraints and conditions for the use of clarification subdialogues.

Our approach is empirically driven throughout. In Section 2, we establish to what extent the principles of recipient design uncovered for natural human interaction can be adopted for the still artificial situation of human-computer interaction. Although it is commonly assumed that results concerning human-human interaction can be applied to human-computer interaction (Horvitz, 1999), there are also revealing differences (Amalberti et al., 1993). We report on a targetted comparison of adopted dialogic strategies in natural human interaction (termed below HHC: human-human communication) and human-computer interaction (HCI). The study shows significant and reliable differences in how dialogue is being managed. In Section 3, we interpret these results with respect to their implications for recipient design. The results demonstrate not only that recipient design is relevant for HCI, but also that it leads to specific and predictable kinds of clarification dia-

logues being taken up by users confronted with an artificial dialogue system. Finally, in Section 4, we discuss the implications of the results for dialogic system design in general and briefly indicate how the required mechanisms are being incorporated in our own dialogue system.

2 A targetted comparison of HHC and HCI dialogues

In order to ascertain the extent to which techniques of recipient design established on the basis of human-human natural interaction can be transferred to HCI, we investigated comparable task-oriented dialogues that varied according to whether the users believed that that they were interacting with another human or with an artificial agent. The data for our investigation were taken from three German corpora collected in the mid-1990s within a toy plane building scenario used for a range of experiments in the German Collaborative Research Centre *Situated Artificial Communicators* (SFB 360) at the University of Bielefeld (Sagerer et al., 1994). In these experiments, one participant is the 'constructor' who actually builds the model plane, the other participant is the 'instructor', who provides instructions for the constructor.

The corpora differ in that the constructor in the HHC setting was another human interlocutor; in the other scenario, the participants were seated in front of a computer but were informed that they were actually talking to an automatic speech processing system (HCI).[1] In all cases, there was no visual contact between constructor and instructor.

Previous work on human-human task-oriented dialogues going back to, for example, Grosz (1982), has shown that dialogue structure commonly follows task structure. Moreover, it is well known that human-human interaction employs a variety of dialogue structuring mechanisms, ranging from meta-talk to discourse markers, and that some of these can usefully be employed for automatic analysis (Marcu, 2000) If dialogue with artificial agents were then to be structured as it is with human interlocutors, there would be many useful linguistic surface cues available for guiding interpretation. And, indeed, a common way of designing dialogue structure

---

[1] In fact, the interlocutors were always humans, as the artificial agent in the HCI conditions was simulated employing standard Wizard-of-Oz methods allowing tighter control of the linguistic responses received by the user.
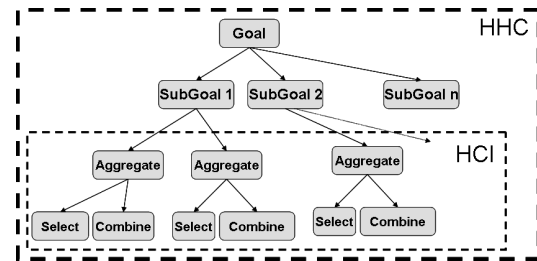

Figure 1: Contrasting dialogue structures for HHC and HCI conditions

in HCI is to base that structure on the structure of the task, since this defines the types of actions necessary and their sequencing.

Previous studies have not, however, addressed the issue of dialogue structure in HCI systematically, although a decrease in framing signals has been noted by Hitzenberger and Womser-Hacker (1995)—indicating either that the discourse structure is marked less often or that there is less structure to be marked. A more precise characterisation of how task-structure is used or expressed in HCI situations is then critical for further design. For our analysis here, we focused on properties of the overall dialogue structure and how this is signalled via linguistic cues. Our results show that there are in fact significant differences in HCI and HHC and that it is not possible simply to take the human-human interaction results and transpose results for one situation to the other.

The structuring devices of the human-to-human construction dialogues can be described as follows. The instructors first inform their communication partners about the general goal of the construction. Subsequently, and as would be expected for a task-oriented dialogue from previous studies, the discourse structure is hierarchical. At the top level, there is discussion of the assembly of the whole toy airplane, which is divided into individual functional parts, such as the wings or the wheels. The individual constructional steps then usually comprise a request to identify one or more parts and a request to combine them. Each step is generally acknowledged by the communication partner, and the successful combination of the parts as a larger structure is signalled as well. All the human-to-human dialogues were similar in these respects. This discourse structure is shwon graphically in the outer box of Figure 1.

Instructors mark changes between phases with

usage	goal		discourse marker		explicit marking	
	HHC	HCI	HHC	HCI	HHC	HCI
none	27.3	100	0	52.5	13.6	52.5
single	40.9	0	9.1	25.0	54.5	27.5
frequent	31.8	0	90.9	22.5	31.8	20.0

Percentage of speakers making no, single or frequent use of a particular structuring strategy.
HCI: 40 dialogues; CCH: 22 dialogues. All differences are highly significant (ANOVA p<0.005).

Table 1: Distribution of dialogue structuring devices across experimental conditions

signals of attention, often the constructor's first name, and discourse particles or speech routines that mark the beginning of a new phase such as *also* [**so**] or *jetzt's geht's los* [**now**]. This structuring function of discourse markers has been shown in several studies and so can be assumed to be quite usual for human-human interaction (Swerts, 1998). Furthermore, individual constructional steps are explicitly marked by means of *als erstes, dann* [**first of all, then**] or *der erste Schritt* [**the first step**]. In addition to the marking of the construction phases, we also find marking of the different activities, such as description of the main goal versus description of the main architecture, or different phases that arise through the addressing of different addressees, such as asides to the experimenters.

Speakers in dialogues directed at human interlocutors are therefore attending to the following three aspects of discourse structure:

- marking the beginning of the task-oriented phase of the dialogue;

- marking the individual constructional steps;

- providing orientations for the hearer as to the goals and subgoals of the communication.

When we turn to the HCI condition, however, we find a very different picture—indicating that a straightforward tuning of dialogue structure for an artificial agent on the basis of the HHC condition will not produce an effective system.

These dialogues generally start as the HHC dialogues do, i.e., with a signal for getting the communication partner's attention, but then diverge by giving very low-level instructions, such as to find a particular kind of component, often even before the system has itself given any feedback. Since this behaviour is divorced from any possible feedback or input produced by the artificial system, it can only be adopted because of the speaker's initial assumptions about the computer. When this strategy is successful, the speaker continues to use

it in following turns. Instructors in the HCI condition do not then attempt to give a general orientation to their hearer. This is true of all the human-computer dialogues in the corpus. Moreover, the dialogue phases of the HCI dialogues do not correspond to the assembly of an identifiable part of the airplane, such as a wing, the wheels, or the propeller, but to much smaller units that consist of successfully identifying and combining some parts. The divergent dialogue structure of the HCI condition is shown graphically in the inner dashed box of Figure 1.

These differences between the experimental conditions are quantified in Table 1, which shows for each condition the frequencies of occurrence for the use of general orienting *goal instructions*, describing what task the constructor/instructor is about to address, the use of *discourse markers*, and the use of *explicit signals* of changes in task phase. These differences prove (a) that users are engaging in recipient design with respect to their partner in these comparable situations and (b) that the linguistic cues available for structuring an interpretation of the dialogue in the HCI case are considerably impoverished. This can itself obviously lead to problems given the difficulty of the interpretation task.

3 Interpretation of the observed differences in terms of recipient design

Examining the results of the previous section more closely, we find signs that the concept of the communication partner to which participants were orienting was not the same for all participants. Some speakers believed structural marking also to be useful in the HCI situation, for example. In this section, we turn to a more exact consideration of the reasons for these differences and show that directly employing the mechanisms of recipient design developed by Schegloff (1972) is a beneficial strategy. The full range of variation observed, including intra-corpus variation that space precluded us describing in detail above, is seen to arise from

a common mechanism. Furthermore, we show that precisely the same mechanism leads to a predictive account of user-initiated clarificatory dialogues.

The starting point for the discussion is the conversation analytic notion of the *insertion sequence*. An insertion sequence is a subdialogue inserted between the first and second parts of an adjacency pair. They are problematic for artificial agents precisely because they are places where the user takes the initiative and demands information from the system. Clarificatory subdialogues are regularly of this kind. Schegloff (1972) analyses the kinds of discourse contents that may constitute insertion sequences in human-to-human conversations involving spatial reference. His results imply a strong connection between recipient design and discourse structure. This means that we can describe the kind of local sequential organisation problematic for mixed-initiative dialogue interpretation on the basis of more general principles.

The aspects that have been found addressed in insertion sequences are the following:

Location Analysis: Speakers check upon spatial information regarding the communication partners, such as where they are when on a mobile phone, which may lead to an insertion sequence and is also responsible for one of the most common types of utterances when beginning a conversation by mobile phone: i.e., "I'm just on the bus/train/tram".

Membership Analysis: Speakers check upon information about the recipient because the communication partner's knowledge may render some formulations more relevant than others. As a 'member' of a particular class of people, such as the class of locals, or of the class of those who have visited the place before, the addressee may be expected to know some landmarks that the speaker may use for spatial description. Membership groups may also include differentiation according to capabilities (e.g., perceptual) of the interlocutors.

Topic or Activity Analysis: Speakers attend to which aspects of the location addressed are relevant for the given topic and activity. They have a number of choices at their disposal among which they can select: geographical descriptions, e.g. *2903 Main Street*, descriptions with relation to members, e.g. *John's place*, descriptions by means of landmarks, or place names.

These three kinds of interactional activity each give rise to potential insertion sequences; that is, they serve as the functional motivation for particular clarificatory subdialogues being explored rather than others. In the HCI situation, however, one of them stands out. The task of membership analysis is extremely challenging for a user faced with an unknown artificial agent. There is little basis for assigning group membership; indeed, there are not even grounds for knowing which kind of groups would be applicable, due to lack of experience with artificial communication partners.

Since membership analysis constitutes a prerequisite for the formulation of instructions, recipient design can be expected to be an essential force both for the discourse structure and for the motivation of particular types of inserted questions in HCI. We tested this prediction by means of a further empirical study involving a scenario in which the users' task was to instruct a robot to measure the distance between two objects out of a set of seven. These objects differed only in their spatial position. The users had an overview of the robot and the objects to be referred to and typed their instructions into a notebook. The relevant objects were pointed at by the instructor of the experiments. The users were not given any information about the system and so were explicitly faced with a considerable problem of membership analysis, making the need for clarification dialogues particularly obvious. The results of the study confirmed the predicted effect and, moreover, provide a classification of insertion sequence types. Thus, the particular kinds of analysis found to initiate insertion sequences in HHC situations are clearly active in HCI situations as well.

21 subjects from all kinds of professions and with different experience with artificial systems participated in the study. The robot's output was generated by a simple script that displayed answers in a fixed order after a particular 'processing' time. The dialogues were all, therefore, absolutely comparable regarding the robot's linguistic material; moreover, the users' instructions had no impact on the robot's linguistic behaviour. The robot, a Pioneer 2, did not move, but the participants were told that it could measure distances and that they were connected to the robot's dialogue processing system by means of a wireless

```
usr11-1    hallo# [hello#]
sys           ERROR
usr11-2    siehst du was [do you see anything?]
sys           ERROR
usr11-3    was siehst du [what do you see?]
sys           ERROR 652-a: input is invalid.
usr11-4    miß den abstand zwischen der vordersten tasse und
              der linken tasse [measure the distance between
              the frontmost cup and the left cup]
```

Figure 2: Example dialogue extract showing membership analysis insertion sequences

LAN connection. The robot's output was either "error" (or later in the dialogues a natural language variant) or a distance in centimeters. This forced users to reformulate their dialogue contributions—an effective methodology for obtaining users' hypotheses about the functioning and capabilities of a system (Fischer, 2003). In our terms, this leads directly to an explicit exploration of a user's membership analysis.

As expected, very limited location analysis occured; however, as we can conclude from the results on global discourse structure reported on in the previous section, location analysis does depend on membership analysis, as the problems faced by the participants concern the establishment of what their shared situation actually consists of, which in turn is determined by what the robot can perceive. Topic analysis is also restricted; spatial formulations were chosen on the basis of what users believed to be 'most understandable' for the robot, which also leads back to the task of membership analysis.

In contrast, there were many cases of membership analysis. There was clearly great uncertainty about the robot's prerequisites for carrying out the spatial task and this was explicitly specified in the users' varied formulations. A simple example is given in Figure 2.

The complete list of types of questions related to membership analysis and which digress from the task instructions in our corpus is given in Table 2. Each of these instances of membership analysis constitutes a clarification question that would initiate an insertion sequence if the robot reacted to it.

4   Consequences for system design

So far our empirical studies have shown that there are particular kinds of interactional problems that will regularly trigger user-initiated insertion se-

quences. These might appear off-topic or out of place but when understood in terms of the membership and topic/activity analysis, it becomes clear that all such contributions are, in a very strong sense, 'predictable'. These results can be exploited in the following ways. One is to extend dialogue system design to be able to meet these contingently relevant contributions whenever they occur. That is, we adapt dialogue manager, lexical database etc. so that precisely these apparently out-of-domain topics or covered. A second strategy is to determine discourse conditions that can be used to alert the dialogue system to the likely occurrence or absence of these kinds of clarificatory subdialogues (see below). Third, we can design explicit strategies for interaction that will reduce the likelihood that a user will employ them: for example, by providing information about the agent's capabilities, etc. as listed in Table 2 in advance by means of system-initiated assertions. That is, we can guide, or *shape*, to use the terminology introduced by Zoltan-Ford (1991), the users' linguistic behaviour. A combination of these three capabilities promises to improve the overall quality of a dialogue system and forms the basis for a significant part of our current research.

We have already ascertained empirically discourse conditions that support the second strategy above, and these follows again directly from the basic notions of recipient design and membership analysis. If a user already has a strong membership analysis in place—for example, due to preconceptions concerning the abilities (or, more commonly, lack of abilities) of the artificial agent—then this influences the design of that user's utterances throughout the dialogue. As a consequence, we have been able to define distinctive linguistic profiles that lead to the identification of distinct user groups that differ reliably in their dialogue strategies, particularly in their initiation of subdialogues. In the human-robot dialogues just considered, for example, we found that eight out of 21 users did not employ any clarification questions at all and an additional four users asked only a single clarification question. Providing these users with additional information about the robot's capabilities is of limited utility because these users found ways to deal with the situation without asking clarification questions. The second group of participants consisted of nine users; this group used many questions that would have led

domain	example (translation)	
perception	VP7-3	[**do you see the cups?**]
readiness	VP4-25	[**Are you ready for another task?**]
functional capabilities	VP19-11	[**what can you do?**]
linguistic capabilities	VP18-7	[**Or do you only know mugs?**]
cognitive capabilities	VP20-15	[**do you know where is left and right of you?**]

Table 2: Membership analysis related insertion sequences

into potentially problematic insertion sequences if the system had been real. For these users, the presentation of additional information on the robot's capabilities would be very useful.

It proved possible to distinguish the members of these two groups reliably simply by attending to their initial dialogue contributions. This is where their pre-interaction membership analysis was most clearly expressed. In the human-robot dialogues investigated, there is no initial utterance from the robot, the user has to initiate the interaction. Two principally different types of first utterance were apparent: whereas one group of users begins the interaction with task-instructions, a second group begins the dialogue by means of a greeting, an appeal for help, or a question with regard to the capabilities of the system. These two different ways of approaching the system had systematic consequences for the dialogue structure. The dependent variable investigated is the number of utterances that initiate insertion sequences. The results of the analysis show that those who greet the robot or interact with it other than by issuing commands initiate clarificatory subdialogues significantly more often than those who start with an instruction (cf. Table 3). Thus, user modelling on the basis of the first utterance in these dialogues can be used to predict users' linguistic behaviour with respect to the initiation of insertion sequences. Note that for this type of user modelling no previous information about the user is necessary and group assignment can be carried out inobtrusively by means of simple key word spotting on the first utterance.

Whereas the avoidance of clarificatory user-initiated subdialogues is clearly a benefit, we can also use the results of our empirical investigations to motivate improvements in the other areas of interactive work undertaken by speakers. In particular topic and activity analysis can become problematic when the decompositions adopted by a user are either insufficient to structure dialogue ap-

usage	task-oriented beginnings	greetings
none	58.3	11.1
single	25.0	11.1
frequent	16.7	77.8

N = 21; average number of clarification questions for task-oriented group: 1.17 clarification questions per dialogue; average number for 'greeting'-group 3.2; significance by t-test p<0.01

Table 3: Percentage of speakers using no, a single, or frequent clarification questions depending on first utterance

propriately for interpretation or, worse, are incompatible with the domain models maintained by the artificial agent. In the latter case, communication will either fail or invoke rechecking of membership categories to find a basis for the understanding (e.g., 'do you know what cups are?'). Thus, what can be seen on the part of a user as reducing the complexity of a task can in fact be removing information vital for the artificial agent to effect successful interpretation.

The results of a user's topic and activity analysis make themselves felt in the divergent dialogue structure observed. As shown above in Figure 1, the structure of the dialogues is thus much flatter than the one found in the corresponding HHC dialogues, such that goal description and marking of subtasks is missing, and the only structure results from the division into selection and combination of parts. In our second study, precisely the same effects are observed. The task, to measure the distance between two objects, is often decomposed into 'simpler' subtasks; for example, the complexity of the task is reduced by achieving reference to each of the objects first before the robot is requested to measure the distance between the objects.

This potential mismatch between user and system can also be identified on the basis of the interaction. Proceeding directly to issuing low-level instructions rather than providing background general goal information is a clear linguistically recognisable cue that a nonaligned topic/activity analysis has been adopted. A successful dialogue system can therefore rely on this dialogue transition as providing an indication of problems to come, which can again be avoided in advance by explicit system-initiated assertions of information.

Our main focus in this paper has been on setting out and motivating some generic principles for dialogue system design. These principles could find diverse computational instantiations and it has not been our aim to argue for any one instantiation rather than another. However, to conclude, we summarise briefly the approach that we are adopting to incorporating these mechanisms within our own dialogue system [ref. anonymised].

Our system takes essentially an information-state based approach augmented by a distinguished vocabulary of discourse transitions between states. We attach 'conceptualisation-conditions' to these transitions which serve to post discourse goals whose particular function is to head off user-initiated clarification. The presence of a greeting is one such condition; the immediate transition to basic-level instructions is another. Recognition and production of instructions is aided by treating the semantic types that occur ('cups', 'measure', 'move', etc.) as elements of a domain ontology. The diverse topic/activity analyses then correspond to the specification of the *granularity* and *decomposition* of activated domain ontologies. Similarly, location analyses correspond to common sense geographies, which we can now model in terms of ontologies more recently developed for Geographic Information Systems (Fonseca et al., 2002).

The specification of conceptualisation-conditions triggered by discourse transitions and classifications of the topic/activity analysis given by the semantic types provided in user utterances represents a direct transfer of the implicit strategies found in conversation analyses to the design of our dialogue system. For example, in our case many simple clarifications like 'do you see the cups?,' 'how many cups do you see?' as well as 'what can you do?' are prevented by providing in advance users that use greetings

information on what the robot can perceive. Similarly, during a scene description where the system has the initiative, the opportunity is taken to introduce terms for the objects it perceives as well as the relevant way of describing the scene, e.g., by means of 'There are two groups of cups. What do you want me to do?' a range of otherwise necessary clarificatory questions are avoided. Even in the case of failure, users will not doubt those capabilities of the system that it has displayed itself, due to alignment processes observable also in human-to-human dialogical interaction (Pickering and Garrod, 2004). After a successful interaction, users expect the system to be able to process parallel instructions because they reliably expect the system to behave consistently (Fischer and Batliner, 2000).

In this paper, the discourse structure initiated by users in HCI situations has been investigated and the results have been three-fold. The structures initiated in HCI are much flatter than in HHC; no general orientation with respect to the aims of a sub-task are presented to the artificial communication partner, and marking is usually reduced. This needs to be accounted for in the mapping of the task-structure onto the discourse model, irrespective of the kind of representation chosen. Secondly, the contents of insertion sequences have also been identified as particularly dependent on recipient design. That is, they concern the preconditions for formulating utterances particularly for the respective hearer. Here, the less that is known about the communication partner, the more needs to be elicited in clarification dialogues: however, crucially, we now can state precisely which kinds of elicitations will be found (cf. Table 2). Thirdly, users have been shown to differ in the strategies that they take to solve the uncertainty about the speech situation and we can predict which strategies they in fact will follow in their employment of clarification dialogues on the basis of their initial interaction with the system (cf. Table 3).

Since the likelihood for users to initiate such clarificatory insertion sequences has been found to be predictable, we have a basis for a range of implicit strategies for addressing the users' subsequent linguistic behaviour. Recipient design has therefore been shown to be a powerful mechanism that, using the appropriate methods, can be addressed in user-adapted dialogue management design.

Information of the kind that we have uncovered empirically in the work reported in this paper can be used to react appropriately to the different types of users in two ways: either one can adapt the system or one can try to adapt the user (Ogden and Bernick, 1996). Although techniques for both strategies are supported by our results, in general we favour attempting to influence the user's behaviour without restricting it *a priori* by means of computer-initiated dialogue structure. Since the reasons for the users' behaviour have been shown to be located on the level of their conceptualisation of the communication partner, explicit instruction may in any case not be useful—explicit guidance of users is not only often impractical but also is not received well by users. The preferred choice is then to influence the users' concepts of their communication partner and thus their linguistic behaviour by *shaping* (Zoltan-Ford, 1991). In particular, Schegloff's analysis shows in detail the human interlocutors' preference for those location terms that express group membership. That is, in natural dialogues the speakers constantly signal to each other who they are, what the other person can expect them to know. Effective system design should therefore provide users with precisely those kinds of information that constitute their most frequent clarification questions initially and in the manner that we have discussed.

References

James Allen. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems*, Sept./Oct.:14–16.

R. Amalberti, 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.

Nicholas Asher and Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press, Cambridge.

Kerstin Fischer and Anton Batliner. 2000. What makes speakers angry in human-computer conversation. In *Proceedings of the Third Workshop on Human-Computer Conversation, Bellagio, Italy*.

Kerstin Fischer. 2003. Linguistic methods for investigating concepts in use. In Thomas Stolz and Katja Kolbe, editors, *Methodologie in der Linguistik*. Frankfurt a.M.: Peter Lang.

Frederico T. Fonseca, Max J. Egenhofer, Peggy Agouris, and Gilberto Câmara. 2002. Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3).

Barbara J. Grosz. 1982. Discourse analysis. In Richard Kittredge and John Lehrberger, editors, *Sublanguage. Studies of Language in Restricted Semantic Domains*, pages 138–174. Berlin, New York: De Gruyter.

L. Hitzenberger and C. Womser-Hacker. 1995. Experimentelle Untersuchungen zu multimodalen natürlichsprachigen Dialogen in der Mensch-Computer-Interaktion. *Sprache und Datenverarbeitung*, 19(1):51–61.

Eric Horvitz. 1999. Uncertainty, action, and interaction: In pursuit of mixed-initiative computing. *IEEE Intelligent Systems*, Sept./Oct.:17–20.

Daniel Marcu. 2000. The rhetorical parsing of unrestricted texts: a surface-based approach. *Computational Linguistics*, 26(3):395–448, Sep.

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

Martin J. Pickering and Simon Garrod. 2004. Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27(2):169–190.

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

Gerhard Sagerer, Hans-Jürgen Eikmeyer, and Gert Rickheit. 1994. "Wir bauen jetzt ein Flugzeug": Konstruieren im Dialog. Arbeitsmaterialien. Report sfb 360, University of Bielefeld.

Remko Scha and Livia Polanyi. 1988. An augmented context free grammar for discourse. In D. Vargha, editor, *Proceedings of the 12th. COLING*, volume 2, pages 573–577, Morristown. Association for Computational Linguistics.

E. A. Schegloff. 1972. Notes on a conversational practice: formulating place. In D. Sudnow, editor, *Studies in Social Interaction*, pages 75–119. The Free Press, New York.

Marc Swerts. 1998. Filled pauses as markers of discourse structure. *Journal of pragmatics*, 30:485–496.

Bonnie Webber, Alistair Knott, Matthew Stone, and Aravind Joshi. 1999. Discourse relations: a structural and presuppositional account using lexicalized TAG. In *Proceedings of the 37th. Annual Meeting of the American Association for Computational Linguistics (ACL'99)*, pages 41–48, University of Maryland. American Association for Computational Linguistics.

Elizabeth Zoltan-Ford. 1991. How to get people to say and type what computers can understand. *International journal of Man-Machine Studies*, 34:527–647.