Chapter 6

Discourse and Dialogue

6.1 Overview

Barbara Grosz

Harvard University, Cambridge, Massachusetts, USA

The problems addressed in discourse research aim to answer two general kinds of questions: (1) what information is contained in extended sequences of utterances that goes beyond the meaning of the individual utterances themselves? (2) how does the context in which an utterance is used affect the meaning of the individual utterances, or parts of them?

Computational work in discourse has focused on two different types of discourse: extended texts and dialogues, both spoken and written. Although there are clear overlaps between these—dialogues contain text-like sequences spoken by a single individual and texts may contain dialogues—the current state of the art leads research to focus on different questions for each. In addition, application opportunities and needs are different. Work on text is of direct relevance to document analysis and retrieval applications whereas work on dialogue is of import for human-computer interfaces regardless of the modality of interaction. A good sense of the current state of research in text interpretation can be gained from reading the papers on text interpretation in a recent special issue of *Artificial Intelligence* (hereafter, *AIJ-SI*), (Hobbs, Stickel, et al., 1994; Jacobs & Rau, 1994; Palmer, Passonneau, et al., 1994).

Text and dialogue have, however, two significant commonalities. First, a major result of early work in discourse was the determination that discourses divide into *discourse segments* much like sentences divide into phrases. Utterances group into segments, with the meaning of a segment encompassing more than the meaning of the individual parts. Different theories vary on the factors they consider central to explaining this segmentation; a review of the alternatives can be found in a previous survey (Grosz, Pollack, et al., 1989) (hereafter, *Discourse Survey*).[1] However, many of the implications for language processing are shared. For example, segment boundaries need to be detected; recent work suggests there are intonational indicators of these boundaries in spoken language (e.g., Grosz & Hirschberg, 1992 and the references cited in this paper) and can be used to improve speech synthesis (e.g., Davis & Hirschberg, 1988).

Second, discourse research on the interpretation of referring expressions, including pronouns and definite descriptions (e.g., *le petit chat*, *das grüne Buch*), and the event reference aspect of verb phrase interpretation (e.g., the relationship between the buying and arriving events in the sequence *John went to Mary's house; he had bought flowers at her favorite florist's* also is relevant to both text and dialogue. Work on these problems before 1990 is described in *Discourse Survey.*

6.1.1 Beyond Sentence Interpretation

The major lines of research on determining what information a discourse carries beyond what is literally expressed in the individual sentences the discourse comprises fall into two categories which, following Hobbs, we will refer to as *informational* and *intentional.* There are currently efforts to combine these two approaches (e.g., Kehler, 1994; Kehler, 1995; Moore & Pollack, 1992); this is an important area of research.

According to the informational approaches, the coherence of discourse follows from semantic relationships between the information conveyed by successive utterances. As a result, the major computational tools used here are inference and abduction on representations of the propositional content of utterances. *Discourse Survey* describes work in this area under *inference-based approaches*; more recent work in this area is presented in *AIJ-SI.*

According to the intentional approaches the coherence of discourse derives from the intentions of speakers and writers, and understanding depends on recognition of those intentions. Thus, these approaches follow Grice (1969); early work in this area drew on speech act theory (Searle, 1969). A major insight of work in this area was to recognize the usefulness of applying AI planning techniques; this work is described in *Discourse Survey.* Recently various limitations of this approach have been recognized. In particular, as originally argued by Searle (1990) and Grosz and Sidner (1990), models of individual plans are not adequate for understanding discourse; models of collaborative

[1] Many of the papers cited in this survey may be found in the collection *Readings in Natural Language Processing* (Grosz, Sparck Jones, et al., 1986).

plans or joint intentions are required. A variety of approaches to developing such models are currently underway (Grosz & Kraus, 1993; Sonenberg, Tidhar, et al., 1994; Cohen & Levesque, 1990) and used for dialogue (Lochbaum, 1993; Lochbaum, 1994; Lochbaum, 1995).

6.1.2 Interpretation and Generation in Context

Research in this area also splits into two approaches, those that examine the interaction of choice or interpretation of expression with focus of attention, and those that are coherence-based.

Focus of attention interacts with the interpretation and generation of pronouns and definite descriptions (Grosz & Sidner, 1986). The coherence-based approaches have been taken with the informational approaches described above. The main new issues in this area concern how to combine these approaches as it is clear that both kinds of consideration play roles both in determining which expressions to use and how to interpret expressions in context. The focus-based approaches have been applied cross-linguistically; because this is a cognitively-oriented approach it should have application to multi-media interfaces even when natural language is not being used, or when only a restricted subset can be handled.

6.2 Discourse Modeling

Donia Scott[a] **& Hans Kamp**[b]

[a] University of Brighton, UK
[b] University of Stuttgart, Germany

6.2.1 Overview: Discourse and Dialogue

A central problem which the development of dialogue systems encounters is one that it has inherited directly from contemporary linguistics, where one is still struggling to achieve a genuine integration of semantics and pragmatics. A satisfactory analysis of dialogue requires in general both semantic representation, i.e. representation of the content of what the different participants are saying, and pragmatic information - what kinds of speech acts they are performing (are they asking a question, answering a question that has just been asked, asking a question for clarification of what was just said, making a proposal, etc.?), what information is available to each of the participants and what information does she want; and, more generally, what is the purpose behind their various utterances or even behind their entering upon the dialogue in the first place. Determining the semantic representation of an utterance and its pragmatic features must in general proceed in tandem: to determine the pragmatic properties of the utterance it is often necessary to have a representation of its content; conversely, it is—especially for the highly elliptical utterances that are common in spoken dialogue—often hardly possible to identify content without an independent assessment of the pragmatic role the utterance is meant to play. A dialogue system identifying the relevant semantic and pragmatic information will thus have to be based on a theory in which semantics and pragmatics are (i) both developed with the formal precision that is a prerequisite for implementation and (ii) suitably attuned to each other and intertwined.

Current approaches to discourse and dialogue from the field of artificial intelligence and computational linguistics are based on four predominant theories of discourse which emerged in the mid- to late-eighties:

Hobbs (1985): A theory of discourse coherence based on a small, limited set of coherence relations, applied recursively to discourse segments. This is part of a larger, still-developing theory of the relations between text interpretation and belief systems.

Grosz and Sidner (1986): A tripartite organization of discourse structure according to the focus of attention of the speaker (the attentional state), the structure of the speaker's purposes (the intentional structure) and the structure of sequences of utterances (the linguistic structure); each of these three constituents deal with different aspects of the discourse.

Mann and Thompson (1987): A hierarchical organization of text spans, where each span is either the nucleus (central) or satellite (support) of one of a set of discourse relations. This approach is commonly known as Rhetorical Structure Theory (RST).

McKeown (1985): A hierarchical organization of discourse around fixed schemata which guarantee coherence and which drive content selection in generation.

No theory is complete, and some (or aspects of some) lend themselves more readily to implementation than others. In addition, no single theory is suitable for use on both sides of the natural language processing coin: the approaches advocated by Grosz and Sidner, and by Hobbs are geared towards whereas those of Mann and Thompson, and of McKeown are more appropriate for natural language generation. With the burgeoning of research on natural language generation since the late-eighties has come an expansion of the emphasis of computational approaches of discourse towards discourse production and, concomitantly, dialogue.

One important aspect of dialogues is that the successive utterances which make it up are often interconnected by cross references of various sorts. For instance, one utterance will use a pronoun (or a deictic temporal phrase such as *the day after*, etc.) to refer to something mentioned in the utterance preceding it. Therefore the semantic theory underlying sophisticated dialogue systems must be in a position to compute and represent such cross references. Traditional theories and frameworks of formal semantics are sentence based and therefore not suited for discourse semantics without considerable extensions.

6.2.2 Discourse Representation Theory

Discourse Representation Theory (DRT) (cf. Kamp, 1981; Kamp & Reyle, 1993), a semantic theory developed for the express purpose of representing and computing trans-sentential anaphora and other forms of text cohesion, thus offers itself as a natural semantic framework for the design of sophisticated dialogue systems. DRT has already been used in the design of a number of question-answering systems, some of them of considerable sophistication.

Currently, DRT is being used as the semantic representation formalism in VERBMOBIL (Wahlster, 1993), a project to develop a machine translation system for face-to-face spoken dialogue funded by the German Department of Science and Technology. Here the aim is to integrate DRT-like semantics with the various kinds of pragmatic information that are needed for translation purposes.

6.2.3 Future Directions

Among the key outstanding issues for computational theories of discourse are:

Nature of Discourse Relations: Relations are variously viewed as textual, rhetorical, intentional, or informational. Although each type of relation can be expected to have a different impact on a text, current discourse theories generally fail to distinguish between them.

Number of Discourse Relations: Depending on the chosen theoretical approach, these can range from anywhere between two and about twenty-five. Altogether, there are over 350 relations available for use (see Hovy, 1990).

Level of Abstraction at which Discourse is Described: In general, approaches advocating fewer discourse relations tend to address higher levels of abstraction.

Nature of Discourse Segments: A key question here is whether discourse segments have psychological reality or whether they are abstract linguistic units akin to phonemes. Recently, there have been attempts to identify the boundary features of discourse segments (Hirschberg & Grosz, 1992; Litman & Passoneau, 1993).

Rôle of Intentions in Discourse: It is well-recognized that intentions play an important rôle in discourse. However, of the four predominant computational theories, only that of Grosz and Sidner provides an explicit treatment of intentionality.

Mechanisms for Handling Key Linguistic Phenomena: Of the predominant theories, only RST fails to address the issues of discourse focus, reference resolution and cue phrases. Existing treatments of focus, however, suffer from terminological confusion between notions of focus, theme and topic, also rife in the text linguistics literature.

Mechanisms for Reasoning about Discourse: Cue phrases and certain syntactic forms are useful signals of prevailing discourse functions (e.g., discourse relations, discourse focus and topic) but do not occur with predictable regularity in texts.

Reasoning mechanisms for retriving and/or generating these discourse functions are thus required.

Recent advances have not involved the development of new theories but have been rather through the extension and integration of existing theories. Notable among them are:

- discourse as collaborative activity (e.g., Grosz & Sidner, 1990; Grosz & Kraus, 1993)
- the use of abduction as a mechanism for reasoning about discourse understanding and generation (e.g., Hobbs, Stickel, et al., 1993; Lascarides & Oberlander, 1992)
- integration of RST with AI approaches to planning (e.g., Hovy, 1991; Moore & Paris, 1993)
- introduction of intentions in computational approaches based on Hobbs' theory and on RST (e.g., Hobbs, 1993; Moore & Pollack, 1992)
- application of the theories to multimedia discourses (e.g., Wahlster, André, et al., 1993)
- application and extension of existing theories in the automatic generation of pragmatically-congruent multilingual texts (Delin, Scott, et al., 1993; Delin, Hartley, et al., 1994; Paris & Scott, 1994).
- extension of theories of monologic discourse to the treatment of dialogue (e.g., Cawsey, 1992; Moore & Paris, 1993; Green & Carberry, 1994; Traum & Allen, 1994
- identification of acoustic (suprasegmental) markers of discourse segments (Hirschberg & Grosz, 1992)

There are many implemented systems for discourse understanding and generation. Most involve hybrid approaches, selectively exploiting the power of existing theories. Available systems for handling dialogue tend either to have sophisticated discourse generation coupled to a crude discourse understanding systems or vice versa; attempts at full dialogue systems are only now beginning to appear.

6.3 Dialogue Modeling

Phil Cohen

Oregon Graduate Institute of Science & Technology, Portland, Oregon, USA

6.3.1 Research Goals

Two related, but at times conflicting, research goals are often adopted by researchers of dialogue. First is the goal of developing a theory of dialogue, including, at least, a theory of cooperative task-oriented dialogue, in which the participants are communicating in service of the accomplishment of some goal-directed task. The often unstated objectives of such theorizing have generally been to determine:

- what properties of collections of utterances and acts characterize a dialogue of the genre being studied
- what assumptions about the participants' mental states and the context need to be made in order to sanction the observed behavior as a rational cooperative dialogue
- what would be rational and cooperative dialogue *extensions* to the currently observed behavior

A second research goal is to develop algorithms and procedures to support a computer's participation in a cooperative dialogue. Often, the dialogue behavior being supported may only bear a passing resemblance to human dialogue. For example, database question-answering (ARPA, 1993) and *frame-filling* dialogues (Bilange, 1991; Bilange, Guyomard, et al., 1990; Bobrow & PARC Understander Group, 1977) are simplifications of human dialogue behavior in that the former consists primarily of the user asking questions, and the system providing answers, whereas the latter involve the system prompting the user for information (e.g., a flight departure time). Human-human dialogues exhibit much more varied behavior, including clarifications, confirmations, other communicative actions, etc. Some researchers have argued that because humans interact differently with computers than they do with people (Dahlbäck & Jönsson, 1992; Fraser & Gilbert, 1991), the goal of developing a system that emulates *real* human dialogue behavior is neither an appropriate, nor attainable target (Dahlbäck & Jönsson, 1992; Shneiderman, 1980). On the contrary, others have argued that the usability of current natural language systems, especially voice-interactive systems in a telecommunications setting, could benefit greatly from techniques that allow the human to engage in behavior found in their typical spoken conversations (Karis & Dobroth, 1991). In general, no consensus exists on the appropriate research goals, methodologies, and evaluation procedures for modeling dialogue.

Three approaches to modeling dialogue—dialogue grammars, plan-based models of dialogue, and joint action theories of dialogue—will be discussed, both from theoretical and practical perspectives.

6.3.2 Dialogue Grammars

One approach with a relatively long history has been that of developing a dialogue grammar (Polanyi & Scha, 1984; Reichman, 1981; Sinclair & Coulthard, 1975). This approach is based on the observation that there exist a number of sequencing regularities in dialogue, termed *adjacency pairs* (Sacks, Schegloff, et al., 1978), describing such facts as that questions are generally followed by answers, proposals by acceptances, etc. Theorists have proposed that dialogues are a collection of such act sequences, with embedded sequences for digressions and repairs (Jefferson, 1972). For some theorists, the importance of these sequences derives from the expectations that arise in the conversants for the occurrence of the remainder of the sequence, given the observation of an initial portion. For instance, on hearing a question, one expects to hear an answer. People can be seen to react to behavior that violates these expectations.

Based on these observations about conversations, theorists have proposed using phrase-structure grammar rules, following the Chomsky hierarchy, or equivalently, various kinds of state machines. The rules state sequential and hierarchical constraints on acceptable dialogues, just as syntactic grammar rules state constraints on grammatically acceptable strings. The terminal elements of these rules are typically illocutionary act names (Austin, 1962; Searle, 1969), such as request, reply, offer, question, answer, propose, accept, reject, etc. The non-terminals describe various stages of the specific type of dialogue being modeled (Sinclair & Coulthard, 1975), such as initiating, reacting, and evaluating. For example, the SUNDIAL system (Andry, Bilange, et al., 1990; Andry, 1992; Bilange, 1991; Bilange, Guyomard, et al., 1990; Guyomard & Siroux, 1988) uses a 4-level dialogue grammar to engage in spoken dialogues about travel reservations. Just as syntactic grammar rules can be used in parsing sentences, it is often thought that dialogue grammar rules can be used in *parsing* the structure of dialogues. With a bottom-up parser and top-down prediction, it is expected that such dialogue grammar rules can predict the set of possible next elements in the sequence, given a prior sequence (Gilbert, Wooffitt, et al., 1990). Moreover, if the grammar is context-free, parsing can be accomplished in polynomial time.

From the perspective of a state machine, the speech act become the state transition labels. When the state machine variant of a dialogue grammar is used as a control mechanism for a dialogue system, the system first recognizes the user's speech act from the utterance, makes the appropriate transition, and then chooses one of the outgoing arcs to determine the appropriate response to supply. When the system performs an action, it makes the relevant transition, and uses the outgoing arcs from the resulting state to predict the type of response to expect from the user (Dahlbäck & Jönsson, 1992).

Arguments against the use of dialogue grammars as a general theory of dialogue have been raised before, notably by Levinson (1981).

First, dialogue grammars require that the communicative action(s) being performed by the speaker in issuing an utterance be identified. In the past, this has been a difficult problem for people and machines, for which prior solutions have required plan recognition (Allen & Perrault, 1980; Carberry, 1990; Kautz, 1990; Perrault & Allen, 1980). Second, the model typically assumes that only one state results from a transition. However, utterances are multifunctional. An utterance can be, for example, both a rejection and an assertion, and a speaker may expect the response to address more than one interpretation. The dialogue grammar subsystem would thus need to be in multiple states simultaneously, a property typically not allowed. Dialogues also contain many instances of speakers' using multiple utterances to perform a single illocutionary act (e.g., a request). To analyze and respond to such dialogue contributions using a dialogue grammar, a calculus of speech acts needs to be developed that can determine when two speech acts combine to constitute another. Currently, no such calculus exists. Finally, and most importantly, the model does not say how systems should choose amongst the next moves, i.e., the states currently reachable, in order for it to play its role as a cooperative conversant. Some analogue of planning is thus required.

In summary, dialogue grammars are a potentially useful computational tool to express simple regularities of dialogue behavior. However, they need to function in concert with more powerful plan-based approaches (described below) in order to provide the input data, and to choose a cooperative system response. As a theory, dialogue grammars are unsatisfying as they provide no explanation of the behavior they describe, i.e., why the actions occur where they do, why they fit together into a unit, etc.

6.3.3 Plan-based Models of Dialogue

Plan-based models are founded on the observation that utterances are not simply strings of words, but rather are the observable performance of communicative actions, or speech acts (Searle, 1969), such as requesting, informing, warning, suggesting, and confirming. Moreover, humans do not just perform actions randomly, but rather they plan their actions to achieve various goals, and in the case of communicative actions, those goals include changes to the mental states of listeners. For example, speakers' requests are planned to alter the intentions of their addressees. Plan-based theories of communicative action and dialogue (Allen & Perrault, 1980; Appelt, 1985; Carberry, 1990; Cohen & Levesque, 1990; Cohen & Perrault, 1979; Perrault & Allen, 1980; Sadek, 1991; Sidner & Israel, 1981) assume that the speaker's speech acts are part of a plan, and the listener's job is to uncover and respond appropriately to the underlying plan, rather than just to the utterance. For example, in response to a customer's question of *Where are the steaks you advertised?*, a butcher's reply of *How many do you want?* is appropriate because the butcher has discovered that the customer's plan of getting steaks himself is going to fail. Being cooperative, he attempts to execute a plan to achieve the customer's higher-level goal of having steaks. Current research on this model is attempting to incorporate more complex dialogue phenomena, such as clarifications (Litman & Allen, 1990; Yamaoka & Iida, 1991; Litman & Allen, 1987), and to model dialogue more as a *joint* enterprise, something the participants are doing together (Clark & Wilkes-Gibbs, 1986; Cohen & Levesque, 1991b; Grosz & Sidner, 1990; Grosz & Kraus, 1993).

The major accomplishment of plan-based theories of dialogue is to offer a generalization in which dialogue can be treated as a special case of other rational noncommunicative behavior. The primary elements are accounts of planning and plan-recognition, which employ various inference rules, action definitions, models of the mental states of the participants, and expectations of likely goals and actions in the context. The set of actions may include speech acts, whose execution affects the beliefs, goals, commitments, and intentions, of the conversants. Importantly, this model of cooperative dialogue solves problems of indirect speech acts as a side-effect (Perrault & Allen, 1980). Namely, when inferring the purpose of an utterance, it may be determined that not only are the speaker's intentions those indicated by the form of the utterance, but there may be other intentions the speaker wants to convey. For example, in responding to the utterance *There is a little yellow piece of rubber,* the addressee's plan recognition process should determine that not only does the speaker want the addressee to believe such an object exists, the speaker wants the addressee to find the object and pick it up. Thus, the utterance could be analyzed by the same plan-recognition process as an informative utterance, as well as both a request to find it and to pick it up.

Drawbacks of the Plan-based Approach

A number of theoretical and practical limitations have been identified for this class of models.

Illocutionary Act Recognition is Redundant: Plan-based theories and algorithms have been tied tightly to illocutionary act recognition. In order to infer the speaker's plan, and determine a cooperative response, the listener (or system) had to recognize what *single* illocutionary act was being performed with each utterance (Perrault & Allen, 1980), even for indirect utterances. However, illocutionary act recognition in the Allen and Perrault model (Allen & Perrault, 1980; Perrault & Allen, 1980) was shown to be redundant (Cohen & Levesque, 1980); other inferences in the scheme provided the same results. Instead, it was argued that illocutionary acts could more properly be handled as complex action expressions, defined over patterns of utterance events and properties of the context, including the mental states of the participants (Cohen & Levesque, 1990). Importantly, using this analysis, a theorist can show how multiple acts were being performed by a given utterance, or how multiple utterances together constituted the performance of a given type of illocutionary act. Conversational participants, however, are not required to make these classifications. Rather, they need only infer what are the speaker's intentions.

Discourse versus Domain Plans: Although the model is capable of solving problems of utterance interpretation using nonlinguistic methods (e.g., plan-recognition), it does so at the expense of distinctions between task-related speech acts and those used to control the dialogue, such as clarifications (Grosz & Sidner, 1986; Litman & Allen, 1987; Litman & Allen, 1990). To handle these prevalent features of dialogue, *multilevel* plan structures have been proposed, in which a new class of discourse plans is posited, which take task-level (or other discourse-level) plans as arguments (Litman & Allen, 1987; Litman & Allen, 1990; Yamaoka & Iida, 1991). These are not higher level plans in an inclusion hierarchy, but rather are *meta*plans, which capture the set of ways in which a single plan structure can be manipulated. Rather than infer directly how utterances further various task plans, as single-level algorithms do, various multilevel algorithms first map utterances to a discourse plan, and determine how the discourse plan operates on an existing or new task plan. Just as with dialogue grammars, multi-level plan recognizers can be used to generate expectations for future actions and utterances, thereby assisting the interpretation of utterance fragments (Allen, 1979; Allen & Perrault, 1980; Carberry, 1985; Carberry, 1990; Sidner, 1985), and even providing constraints to speech recognizers (Andry, 1992; Yamaoka & Iida, 1991; Young, Hauptmann, et al., 1989).

Complexity of Inference: The processes of plan-recognition and planning are combinatorially intractable in the worst case, and in some cases, are undecidable (Bylander, 1991; Chapman, 1987; Kautz, 1990). The complexity arises in the evaluation of conditions, and in chaining from preconditions to actions they enable. Restricted planning problems in appropriate settings may still be reasonably well-behaved, but practical systems cannot be based entirely on the kind of first-principles reasoning typical of general-purpose planning and plan-recognition systems.

Lack of a Theoretical Base: Although the plan-based approach has much to recommend it as a computational model, and certainly has stimulated much informative research in dialogue understanding, it still lacks a crisp theoretical base. For example, it is difficult to express precisely what are the various constructs (plans, goals, intentions, etc.), what are the consequences of those ascribing those theoretical constructs to be the user's mental state, and what kinds of dialogue phenomena and properties the framework can handle. Because of the procedural nature of the model, it is difficult to determine what analysis will be given, and whether it is correct, as there is no independently stated notion of correctness. In other words, what is missing is a *specification* of what the system should do. Section 6.4 will discuss such an approach.

6.3.4 Future Directions

Plan-based approaches that model dialogue simply as a product of the interaction of plan generators and recognizers working in synchrony and harmony, do not explain why addressees ask clarification questions, why they confirm, or even, why they do not simply walk away during a conversation. A new theory of conversation is emerging in which dialogue is regarded as a joint activity, something that agents do *together* (Clark & Wilkes-Gibbs, 1986; Cohen & Levesque, 1991b; Grosz & Sidner, 1990; Grosz & Kraus, 1993; Lochbaum, 1994; Schegloff, 1981; Suchman, 1987). The joint action model claims that *both* parties to a dialogue are responsible for sustaining it. Participating in a dialogue requires the conversants to have at least a joint commitment to understand one another, and these commitments motivate the clarifications and confirmations so frequent in ordinary conversation.

Typical areas in which such models are distinguished from individual plan-based models are dealing with reference and confirmations. Clark and colleagues (Clark & Wilkes-Gibbs, 1986; Clark, 1989) have argued that actual referring behavior cannot be adequately modeled by the simple notion that speakers simply provide noun phrases and listeners identify the referents. Rather, both parties offer noun phrases, refine previous ones, correct misidentifications, etc. They claim that people appear to be following the strategy of minimizing the joint effort involved in successfully referring. Computer models of referring based on this analysis are beginning to be developed (Heeman & Hirst, 1992; Edmonds, 1993). Theoretical models of joint action (Cohen & Levesque, 1991b; Cohen & Levesque, 1991a) have been shown to minimize the overall team effort in dynamic, uncertain worlds (Jennings & Mamdani, 1992). Thus, if a more general theory of joint action can be applied to dialogue as a special case, an explanation for numerous dialogue phenomena, such as collaboration on reference, confirmations, etc.) will be derivable. Furthermore, such a theory offers the possibility for providing a specification of what dialogue participants should do, which could be used to guide and evaluate dialogue management components for spoken language systems. Finally, future work in this area can also form the basis for protocols for communication among intelligent software agents.

6.4 Spoken Language Dialogue

Egidio Giachin

CSELT, Torino, Italy

The development of machines that are able to sustain a conversation with a human being has long been a challenging goal. Only recently, however, substantial improvements in the technology of speech recognition and understanding have enabled the implementation of experimental spoken dialogue systems, acting within specific semantic domains. The refreshed interest in this area is represented by the numerous papers which appeared in conferences such as ESCA Eurospeech, ICSLP, and ICASSP, as well as by events such as the 1993 International Symposium on Spoken Dialogue and the 1995 ESCA Workshop on Spoken Dialogue Systems.

The need for a dialogue component in a system for human-machine interaction arises for several reasons. Often the user does not express his requirement with a single sentence, because that would be impractical; assistance is then expected from the system, so that the interaction may naturally flow in the course of several dialogue turns. Moreover, a dialogue manager should take care of identifying, and recovering from, speech recognition and understanding errors.

The studies on human-machine dialogue have historically followed two main theoretical guidelines traced by research on human-human dialogue. *Discourse analysis*, developed from studies on speech acts (Searle, 1976), views dialogue as a rational cooperation and assumes that the speakers' utterances be well-formed sentences. *Conversational analysis*, on the other hand, studies dialogue as a social interaction in which phenomena such as disfluencies, abrupt shift of focus, etc., have to be considered (Levinson, 1983). Both theories have contributed to the design of human-machine dialogue systems; in practice, freedom of design has to be constrained so as to find an adequate match with the other technologies the system rests on. For example, dialogue strategies for speech systems should recover from word recognition errors.

Experimental dialogue systems have been developed mainly as evolutions of speech understanding projects, which provided satisfactory recognition accuracy for speaker independent continuous speech tasks with lexicons of the order of 1000 words. The development of robust parsing methods for natural language also was an important step. After some recent experiences at individual sites (Siroux, 1989; Young & Proctor, 1989; Mast, Kompe, et al., 1992), one of the most representative projects in Europe that fostered the development of dialogue systems is the CEC SUNDIAL project (Peckham, 1993). The ARPA funded ATIS project in the United States also spurred a flow of research on spoken dialogue in some sites (Seneff, Hirschman, et al., 1991).

6.4.1 Functional Characteristics

The dialogue manager is the core of a spoken dialogue system. It relies on two main components, the *interaction history* and the *interaction model*. The interaction history is used to interpret sentences, such as those including anaphora and ellipsis, that cannot be understood by themselves, but only according to some existing context. The context (or, more technically, *active focus*) may change as the dialogue proceeds and the user shifts its focus. This requires the system to keep an updated history for which efficient representations (e.g., tree hierarchies) have been devised.

The interaction model defines the strategy that drives the dialogue. The dialogue strategy may lie between two extremes: the user is granted complete freedom of initiative, or the dialogue is driven by the dialogue manager. The former choice supports naturalness on the user's side but increases the risk of misunderstandings, while the latter provides easier recognition conditions, though the resulting dialogues can be long and unfriendly.

The *right* strategy depends on the application scenario and on the robustness of the speech recognition techniques involved. The design of a suitable strategy is a crucial issue, because the success of the interaction will depend mainly on that. A good strategy is flexible and lets the user take the initiative as long as no problem arises, but assumes control of the dialogue when things become messy; the dialogue manager then requires the user to reformulate his or her sentence or even use different interaction modalities, such as isolated words, spelling, or yes/no confirmations. The effectiveness of a dialogue strategy can be assessed only through extensive experimentation.

Several approaches have been employed to implement an interaction model. A simple one represents dialogue as a network of states with which actions are associated. The between-state transitions are regulated by suitable conditions. This implementation, used e.g., in Gerbino and Danieli (1993), enhances readability and ease of maintenance, while preserving efficiency at runtime through a suitable compilation. Architectures of higher complexity have been investigated. In the CEC SUNDIAL project, for example (see Peckham, 1993 and the references cited there), a dialogue manager based on the theory of speech acts was developed. A modular architecture was designed so as to insure portability to different tasks and favor the separation of different pieces of knowledge, with limited run time speed reduction.

6.4.2 Development of a Spoken Dialogue System

The development of an effective system requires extensive experimentation with real users. Human-human dialogue, though providing some useful insight, is of limited utility because a human behaves much differently when he or she is talking to a machine rather than to another human. The Wizard of Oz (WOZ) technique (Fraser & Gilbert, 1989) enables dialogue examples to be collected in the initial phase of system development: the machine is emulated by a human expert, and the user is led to believe that he or she is actually talking to a computer. This technique has been effective to help researchers test ideas, however, since it is difficult to realistically mimic the actual behavior of recognition and dialogue systems, it may be affected by an overly optimistic estimation of performance, which may lead to a dialogue strategy that is not robust enough. A different approach suggests that experimentation with real users be performed in several steps, starting with a complete, though rough, bootstrap system and cyclically upgrading it. This technique was used for the system in Seneff, Hirschman, et al. (1991). The advantage of this method is that it enables the system to be developed in a close match with the collected database.

The above methodologies are not mutually exclusive, and in practical implementations they have been jointly employed. In every case, extensive corpora of (real or simulated) human-machine interaction are playing an essential role for development and testing.

6.4.3 Evaluation Criteria

The difficulty of satisfactorily evaluating the performance of voice processing systems increases from speech recognition dialogue, where the very nature of what should be measured is complex and ill-defined. Recent projects nevertheless favored the establishing of some ideas. Evaluation parameters can be classified as *objective* and *subjective*. The former category includes the total time of the utterance, the number of user/machine dialogue turns, the rate of correction/repair turns, etc. The *transaction success* is also an objective measure, though the precise meaning of *success* still lacks a standard definition. As a general rule, an interaction is declared successful if the user was able to solve his or her problem without being overwhelmed by unnecessary information from the system, in the spirit of what has been done in the ARPA community for the ATIS speech understanding task.

Objective measures are not sufficient to evaluate the overall system quality as seen from the user's viewpoint. The subjective measures, aimed at assessing the users' opinions on the system, are obtained through direct interview by questionnaire filling. Questions include such issues as ease of usage, naturalness, clarity, friendliness, robustness regarding misunderstandings, subjective length of the transaction, etc. Subjective measures have to be properly processed (e.g., through factorial analysis) in order to suggest specific upgrading actions. These measures may depart from what could be expected by analyzing objective data. Since user satisfaction is the ultimate evaluation criterion, subjective measures are helpful to focus on weak points that might go overlooked and neglect issues that result of lesser practical importance.

Evaluation of state-of-the-art spoken dialogue technology indicates that a careful dialogue manager design permits high transaction success to be achieved in spite of the still numerous recognition or understanding errors (see e.g., Gerbino & Danieli, 1993. Robustness to spontaneous speech is obtained at the expense of speed and friendliness, and novices experience more trouble than expert users. Moreover, ease and naturalness of system usage are perceived differently according to user age and education. However, the challenge to bring this technology into real services is open.

6.4.4 Future Directions

The issues for future investigation can be specified only according to the purpose for which the spoken dialogue system is intended. If the goal is to make the system work *in the field*, then robust performance and real time operation become the key factors, and the dialogue manager should drive the user to speak in a constrained way. Under these circumstances, the interaction model will be simple and the techniques developed so far are likely to be adequate. If, on the other hand, immediate applicability is not the main concern, there are several topics into which a deeper insight must still be gained. These include the design of strategies to better cope with troublesome speakers, to achieve better trade-offs between flexibility and robustness, and to increase portability to different tasks/languages.

The performance of the recognition/understanding modules can be improved when they are properly integrated in a dialogue system. The knowledge of the dialogue status, in fact, generates expectations on what the user is about to say, and hence can be used to restrict the dictionary or the linguistic constraints of the speech understanding module, thereby increasing their accuracy. These *predictions* have been shown to yield practical improvements (see e.g., Andry, 1992), though they remain a subject for research. Since recognition errors will never be completely ruled out, it is important that the user can detect and recover from wrong system answers in the shortest possible time. The influence of the dialogue strategy on error recovery speed was studied in Hirschman and Pao (1993). It is hoped that the growing collaboration between the speech and natural language communities may provide progress in these areas.

6.5 Chapter References

ACL (1994). *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Allen, J. F. (1979). A plan-based approach to speech act recognition. Technical Report 131, Department of Computer Science, University of Toronto, Toronto, Canada.

Allen, J. F. and Perrault, C. R. (1980). Analyzing intention in dialogues. *Artificial Intelligence*, 15(3):143–178.

Andry, F. (1992). Static and dynamic predictions: a method to improve speech understanding in cooperative dialogues. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 1, pages 639–642, Banff, Alberta, Canada. University of Alberta.

Andry, F., Bilange, E., Charpentier, F., Choukri, K., Ponamalé, M., and Soudoplatoff, S. (1990). Computerised simulation tools for the design of an oral dialogue system. In *Selected Publications, 1988-1990, SUNDIAL Project (Esprit P2218)*. Commission of the European Communities.

Appelt, D. E. (1985). *Planning English Sentences*. Cambridge University Press.

ARPA (1993). *Proceedings of the 1993 ARPA Spoken Language Systems Technology Workshop*. Advanced Research Projects Agency, MIT.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press, London.

Bilange, E. (1991). A task independent oral dialogue model. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, Berlin. European Chapter of the Association for Computational Linguistics.

Bilange, E., Guyomard, M., and Siroux, J. (1990). Separating dialogue knowledge and task knowledge from oral dialogue management. In *COGNITIVA'90*, Madrid.

Bobrow, D. and PARC Understander Group, T. (1977). GUS-1, a frame driven dialog system. *Artificial Intelligence*, 8(2):155–173.

Bylander, E. (1991). Complexity results for planning. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 274–279, Sydney, Australia.

Carberry, S. (1985). A pragmatics-based approach to understanding intersentential ellipses. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pages 188–197, University of Chicago. Association for Computational Linguistics.

Carberry, S. (1990). *Plan recognition in natural language dialogue.* ACL-MIT Press Series in Natural Language Processing. Bradford Books, MIT Press, Cambridge, Massachusetts.

Cawsey, A. (1992). *Explanation and Interaction: The Computer Generation of Explanatory Dialogues.* ACL-MIT Press.

Chapman, D. (1987). Planning for conjunctive goals. *Artificial Intelligence*, 32(3):333–377.

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

Clark, H. H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22:1–39. Reprinted in Cohen, Morgan, et al. (1990).

Cohen, P. R. and Levesque, H. J. (1980). Speech acts and the recognition of shared plans. In *Proceedings of the Third Biennial Conference*, pages 263–271, Victoria, British Columbia. Canadian Society for Computational Studies of Intelligence.

Cohen, P. R. and Levesque, H. J. (1990). Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication.* MIT Press, Cambridge, Massachusetts.

Cohen, P. R. and Levesque, H. J. (1991a). Confirmations and joint action. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 951–957, Sydney, Australia.

Cohen, P. R. and Levesque, H. J. (1991b). Teamwork. *Noûs*, 25(4):487–512. Also Technical Note 504, Artificial Intelligence Center, SRI International, Menlo Park, California, 1991.

Cohen, P. R., Morgan, J., and Pollack, M. E., editors (1990). *Intentions in Communication.* MIT Press, Cambridge, Massachusetts.

Cohen, P. R. and Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.

Dahlbäck, N. and Jönsson, A. (1992). An empirically based computationally tractable dialogue model. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society (COGSCI-92)*, Bloomington, Indiana.

Dale, R., Hovy, E. H., Rösner, D., and Stock, O., editors (1992). *Aspects of Automated Natural Language Generation.* Number 587 in Lecture Notes in AI. Springer-Verlag, Heidelberg.

Davis, J. R. and Hirschberg, J. (1988). Assigning intonational features in synthesized spoken directions. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 187–193, SUNY, Buffalo, New York. Association for Computational Linguistics.

Delin, J., Hartley, A., Paris, C., Scott, D., and Vander Linden, K. (1994). Expressing procedural relationships in multilingual instructions. In *Proceedings of the Seventh International Workshop on Natural Language Generation*, pages 61–70, Kennebunkport, Maine. Springer-Verlag, Berlin.

Delin, J., Scott, D., and Hartley, A. (1993). Knowledge, intention, rhetoric: Levels of variation in multilingual instructions. In *Proceedings of the Workshop on Intentionality and Structure in Discourse Relations*, Columbus, Ohio. Association for Computational Linguistics.

Edmonds, P. G. (1993). A computational model of collaboration on reference in direction-giving dialogues. Master's thesis, Computer Systems Research Institute, Department of Computer Science, University of Toronto.

Eurospeech (1993). *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, Berlin. European Speech Communication Association.

Fraser, N. and Gilbert, N. (1989). Simulating speech systems. *Computer, Speech, and Language*, 3.

Fraser, N. M. and Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech and Language*, 5(1):81–99.

Gerbino, E. and Danieli, M. (1993). Managing dialogue in a continuous speech understanding system. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 3, pages 1661–1664, Berlin. European Speech Communication Association.

Gilbert, N., Wooffitt, R., and Fraser, N. (1990). Organising computer talk. In Luff, P., Gilbert, N., and Frohlich, D., editors, *Computers and Conversation*, chapter 11, pages 235–258. Academic Press, New York.

Green, N. and Carberry, S. (1994). A hybrid reasoning model for indirect answers. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Grice, H. P. (1969). Utterer's meaning and intentions. *Philosophical Review*, 68(2):147–177.

Grosz, B. and Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 1, pages 429–432, Banff, Alberta, Canada. University of Alberta.

Grosz, B. and Kraus, S. (1993). Collaborative plans for group activities. In *Proceedings of IJCAI-93*, volume 1, pages 367–373, Chambery, France.

Grosz, B., Pollack, M., and Sidner, C. (1989). Discourse. In Posner, M., editor, *Foundations of Cognitive Science*. MIT Press.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Grosz, B. J. and Sidner, C. L. (1990). Plans for discourse. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 417–444. MIT Press, Cambridge, Massachusetts.

Grosz, B. J., Sparck Jones, K., and Webber, B. L., editors (1986). *Readings in Natural Language Processing*. Morgan Kaufmann Publishers, Inc.

Guyomard, M. and Siroux, J. (1988). Experimentation in the specification of an oral dialogue. In Niemann, H., Lang, M., and Sagerer, G., editors, *Recent Advances in Speech Understanding and Dialog Systems*, volume 46. Springer Verlag, Berlin. NATO ASI Series.

Heeman, P. A. and Hirst, G. (1992). Collaborating on referring expressions. Technical Report 435, Department of Computer Science, University of Rochester, Rochester, New York.

Hirschberg, J. and Grosz, B. J. (1992). Intonational features of local and global discourse structure. In *Proceedings of the Fifth DARPA Speech and Natural Language Workshop*. Defense Advanced Research Projects Agency, Morgan Kaufmann.

Hirschman, L. and Pao, C. (1993). The cost of errors in a spoken language system. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 2, pages 1419–1422, Berlin. European Speech Communication Association.

Hobbs, J. R. (1985). On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Hobbs, J. R. (1993). Intention, information, and structure in discourse. In *Proceedings of the NATO Advanced Research Workshop on Burning Issues in Discourse*, pages 41–66, Maratea, Italy.

Hobbs, J. R., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.

Hobbs, J. R., Stickel, M. E., Appelt, D. E., and Martin, P. (1994). Interpretation as abduction. In Pereira, F. C. N. and Grosz, B. J., editors, *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Hovy, E. H. (1990). Parsimonious and profligate approaches to the question of discourse structure relations. In IWNLG, editor, *Proceedings of the Fifth International Workshop on Natural Language Generation*, Pittsburgh, Pennsylvania. Springer-Verlag.

Hovy, E. H. (1991). Approaches to the planning of coherent text. In Paris, C. L., Swartout, W. R., and Mann, W. C., editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic.

ICSLP (1992). *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff, Alberta, Canada. University of Alberta.

IJCAI (1991). *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Sydney, Australia.

IWNLG (1994). *Proceedings of the Seventh International Workshop on Natural Language Generation*, Kennebunkport, Maine. Springer-Verlag, Berlin.

Jacobs, P. S. and Rau, L. F. (1994). Innovations in text interpretation. In Pereira, F. C. N. and Grosz, B. J., editors, *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Jefferson, G. (1972). Side sequences. In Sudnow, D., editor, *Studies in Social Interaction*. Free Press, New York.

Jennings, N. R. and Mamdani, E. H. (1992). Using joint responsibility to coordinate collaborative problem solving in dynamic environments. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 269–275, Menlo Park, California. American Association for Artificial Intelligence, AAAI Press/MIT Press.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J., Janssen, T., and Stokhof, M., editors, *Formal Methods in the Study of Language*. Mathematisch Centrum, Amsterdam.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer, Dordrecht.

Karis, D. and Dobroth, K. M. (1991). Automating services with speech recognition over the public switched telephone network: Human factors considerations. *IEEE Journal of Selected Areas in Communications*, 9(4):574–585.

Kautz, H. (1990). A circumscriptive theory of plan recognition. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press, Cambridge, Massachusetts.

Kehler, A. (1994). Common topics and coherent situations: Interpreting ellipsis in the context of discourse inference. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Kehler, A. (1995). *Interpreting Cohesive Forms in the Context of Discourse Inference*. PhD thesis, Harvard University.

Lascarides, A. and Oberlander, J. (1992). Abducing temporal discourse. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 167–182, Trento, Italy. Springer-Verlag. Also in Dale, Hovy, et al. (1992).

Levinson, S. (1981). Some pre-observations on the modelling of dialogue. *Discourse Processes*, 4(1):93–116.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Litman, D. and Passoneau, R. (1993). Feasibility of automated discourse segmentation. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Ohio State University. Association for Computational Linguistics.

Litman, D. J. and Allen, J. F. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200.

Litman, D. J. and Allen, J. F. (1990). Discourse processing and commonsense plans. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 365–388. MIT Press, Cambridge, Massachusetts.

Lochbaum, K. E. (1993). A collaborative planning approach to discourse understanding. Technical Report TR-20-93, Harvard University.

Lochbaum, K. E. (1994). *Using Collaborative Plans to Model the Intentional Structure of Discourse*. PhD thesis, Harvard University.

Lochbaum, K. E. (1995). The use of knowledge preconditions in language processing. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence*, Montreal, Canada. In press.

Mann, W. C. and Thompson, S. A. (1987). Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, Information Sciences Institute, University of Southern California.

Mast, M., Kompe, R., Kummert, F., and Niemann, H. (1992). The dialog module of the speech recognition and dialog system EVAR. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 2, pages 1573–1576, Banff, Alberta, Canada. University of Alberta.

McKeown, K. R. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Studies in Natural Language Processing. Cambridge University Press.

Moore, J. D. and Paris, C. L. (1993). Planning texts for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–694.

Moore, J. D. and Pollack, M. E. (1992). A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544.

Palmer, M. S., Passonneau, R. J., Weir, C., and Finin, T. (1994). The kernel text understanding system. In Pereira, F. C. N. and Grosz, B. J., editors, *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Paris, C. and Scott, D. (1994). Stylistic variation in multilingual instructions. In *Proceedings of the Seventh International Workshop on Natural Language Generation*, pages 45–52, Kennebunkport, Maine. Springer-Verlag, Berlin.

Peckham, J. (1993). A new generation of spoken language systems: recent results and lessons from the SUNDIAL project. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 1, pages 33–42, Berlin. European Speech Communication Association. Keynote address.

Pereira, F. C. N. and Grosz, B. J., editors (1994). *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Perrault, C. R. and Allen, J. F. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3):167–182.

Polanyi, R. and Scha, R. (1984). A syntactic approach to discourse semantics. In *Proceedings of the 10th International Conference on Computational Linguistics*, pages 413–419, Stanford University, California. ACL.

Reichman, R. (1981). *Plain-speaking: A theory and grammar of spontaneous discourse.* PhD thesis, Department of Computer Science, Harvard University, Cambridge, Massachusetts.

Sacks, H., Schegloff, E., and Jefferson, G. (1978). A simplest systematics for the organization of turn-taking in conversation. In Schenkein, J., editor, *Studies in the Organization of Conversational Interaction.* Academic Press, New York.

Sadek, D. (1991). Dialogue acts are rational plans. In *Proceedings of the ESCA/ETRW Workshop on the Structure of Multimodal Dialogue*, Maratea, Italy.

Schegloff, E. A. (1981). Discourse as an interactional achievement: Some uses of unh-huh and other things that come between sentences. In Tannen, D., editor, *Analyzing discourse: Text and talk.* Georgetown University Roundtable on Languages and Linguistics, Georgetown University Press, Washington, DC.

Searle, J. R. (1969). *Speech Acts: An essay in the philosophy of language.* Cambridge University Press.

Searle, J. R. (1976). The classification of illocutionary acts. *Language in Society*, 5.

Searle, J. R. (1990). Collective intentionality. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication.* MIT Press, Cambridge, Massachusetts.

Seneff, S., Hirschman, L., and Zue, V. W. (1991). Interactive problem solving and dialogue in the ATIS domain. In *Proceedings of the Fourth DARPA Speech and Natural Language Workshop*, Pacific Grove, California. Defense Advanced Research Projects Agency, Morgan Kaufmann.

Shneiderman, B. (1980). Natural vs. precise concise languages for human operation of computers: Research issues and experimental approaches. In *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics*, pages 139–141, Philadelphia, Pennsylvania. Association for Computational Linguistics.

Sidner, C. and Israel, D. (1981). Recognizing intended meaning and speaker's plans. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 203–208, Vancouver, British Columbia.

Sidner, C. L. (1985). Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1(1):1–10.

Sinclair, J. M. and Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils.* Oxford University Press, London.

Siroux, J. (1989). Pragmatics in a realization of a dialogue module. In Taylor, M. M., Néel, F., and Bouwhuis, D. G., editors, *The structure of multimodal dialogue.* Elsevier Science, Amsterdam.

Sonenberg, E., Tidhar, G., Werner, E., Kinny, D., Ljungberg, M., and Rao, A. (1994). Planned team activity. Technical Report 26, Australian Artificial Intelligence Institute.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human/machine communication.* Cambridge University Press.

Traum, D. R. and Allen, J. F. (1994). Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Wahlster, W. (1993). Verbmobil, translation of face-to-face dialogs. In *Proceedings of the Fourth Machine Translation Summit*, pages 127–135, Kobe, Japan.

Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., and Rist, T. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, pages 387–427.

Yamaoka, T. and Iida, H. (1991). Dialogue interpretation model and its application to next utterance prediction for spoken language processing. In *Eurospeech '91, Proceedings of the Second European Conference on Speech Communication and Technology*, pages 849–852, Genova, Italy. European Speech Communication Association.

Young, S. J. and Proctor, C. E. (1989). The design and implementation of dialogue control in voice operated database inquiry systems. *Computer, Speech, and Language*, 3.

Young, S. R., Hauptmann, A. G., Ward, W. H., Smith, E. T., and Werner, P. (1989). High level knowledge sources in usable speech recognition systems. *Communications of the ACM*, 32(2).