ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

Technical Report No. 391

DISCOURSE UNDERSTANDING

R. J. H. Scha, B. C. Bruce, & L. Polanyi
Bolt Beranek and Newman Inc.
Cambridge, Massachusetts

October 1986

Center for the Study of Reading

TECHNICAL REPORTS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

CENTER FOR THE STUDY OF READING

Technical Report No. 391

DISCOURSE UNDERSTANDING

R. J. H. Scha, B. C. Bruce, & L. Polanyi
Bolt Beranek and Newman Inc.
Cambridge, Massachusetts

October 1986

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

The work upon which this publication is based was performed pursuant to Contract No. 400-81-0030 of the National Institute of Education. It does not, however, necessarily reflect the views of this agency. A version of this paper is to appear as an entry in S. Shapiro (Ed.), The Encyclopedia of Artificial Intelligence. New York: Wiley.

This page is intentionally blank.

Abstract

Research on natural language understanding has often focused on the problem of analyzing the structure and meaning of isolated sentences. To understand whole texts or dialogues, these sentences must be seen as elements whose significance resides in the contribution they make to the larger whole. A computer natural language understanding system must interpret each sentence with respect to both the linguistic context established by preceding sentences and the real-world setting. This paper reviews work on these issues, examining theories of the structure of discourse, the semantics of discourse, speech acts and pragmatics, and different communication modalities.

Discourse Understanding

1. Introduction

The term Discourse Understanding refers to all processes of Natural Language Understanding that attempt to understand a text or dialogue. For such processes, the sentences of natural language are elements whose significance resides in the contribution they make to the development of a larger whole, rather than being independent, isolated units of meaning. To understand discourse, one must track the structure of an unfolding text or dialogue and interpret every new utterance with respect to the proper context--taking into account the real-world setting of the utterance, as well as the linguistic context built up by the utterances preceding it. The problems of Discourse Understanding are thus closely related to those dealt with in the linguistic discipline of Pragmatics which studies the context dependence of utterance meanings.

Research on natural language understanding systems has often focused on the problem of analyzing the structure and meaning of isolated sentences. To deal with discourse instead, a system must have all the capabilities necessary for sentence understanding, but, in addition, it must be able to apply rules of discourse structure, which specify how sentences may be combined to form texts or dialogues.

Even with such discourse-level extensions, however, a purely linguistic approach can only construct the meaning of a text insofar as it follows from the meaning of its constituent utterances and the explicitly stated relations between them. In

Artificial Intelligence one tends to take a broader perspective, which emphasizes the role of world knowledge in discourse understanding. By taking into account common sense knowledge about the world, a system may derive semantic relations between constituents of the text that are not stated explicitly, but that may be plausibly assumed. By invoking scripts and frames, a system may analyze a text against the background of default assumptions about "normal" situations and "normal" courses of events, thereby filling in information left implicit in the text, and also noticing when something deviates from the usual pattern and is therefore worthy of special attention. In this way, a more complete understanding of the intended meaning of the text may be created.

A discourse understanding system worthy of that name should not only deal correctly with what is true or false in the world according to its input text, but should, at the same time, be able to distinguish between more and less important information--between what is crucial and what is mere background. With this capacity, a system would be able to generate adequate summaries of its input texts. A further level of understanding would involve the ability to infer what the "point" of a story or description is--to discover the more abstract, culturally relevant message which is instantiated by the text.

Much of the AI research on discourse understanding is oriented towards developing systems to exhibit reasonable and cooperative behavior in a goal-directed interaction with a human dialogue-partner. Such systems would do more than understand the

literal meanings of the utterances of their interlocutor; they would have to be able to assess, to some extent, the intentions and purposes behind these utterances. Methods to achieve this are usually based on the theory of Speech Acts: The system recognizes the goals which are conventionally associated with various types of utterances, such as assertions, questions, commands, and requests. Understanding an utterance at a deeper level is then viewed as establishing what goal the speaker wanted to achieve by performing the speech act, and what role the speech act plays in achieving that goal. Often, the goal can be seen as a subgoal which plays a role in achieving a higher level goal, and so on. By invoking plausible hypotheses about the goals the speaker may have, and about the methods she may employ to achieve them, a system may infer the intention behind a speech act.

Empirical studies of human discourse usually deal with real-time oral communication or with written texts. Discourse-understanding computer programs, however, will usually employ a video display terminal to communicate with their users in real time. They will thus use a new natural language interaction-mode which did not exist before. It is therefore of some interest to study how the properties of discourse depend on the interaction mode--e.g., on the amount of shared environment between the participants and on the sensory modality of the communication medium.

Each of the main topics mentioned above will be discussed below in some detail:

- the structure of discourse
- the semantics of discourse
- speech acts and pragmatics
- different I/O modalities

2. The Structure of Discourse

2.1 Introduction

To understand a text or dialogue, one must understand the relations between its parts. Clearly, these parts are not just the individual sentences; sentences are joined together to form larger units, which in their turn may be the building blocks of yet larger units. Discourse understanding must thus be based on some characterization of the way in which a discourse is built up out of constituent units.

Unlike the smooth, steady development of a central idea which characterizes the texts we read in books, journals and newspapers, everyday spoken discourse is characterized by interruptions, resumptions, backtracking and jumping ahead of oneself. Somehow, despite the apparent "disfluency" of everyday discourse, speakers and hearers manage to follow what is going on and to produce responses to one another which are situationally appropriate and which demonstrate an understanding of all of the "underspecified" items of meaning which are found in sentences. They understand who is being referred to by words such as "he" and "it," can recover the referent of phrases like "the one over there" or "the one we were just talking about," and manage to

orient themselves in "discourse time and space" by correctly assigning temporal and spatial referents to words like "now" and "here," whose meaning is totally context dependent.

2.1.1 Modeling discourse structure: The complexity of the problem. Faced with transcripts of natural interactions, analysts experience serious difficulties locating the "descriptions," "explanations," "stories," "plans" or other structural units which they may know have been "there" when the interaction was happening. With the move to the analysis phase, structural units become lost in all the "talk."

In order to illustrate the problem of locating a coherent discourse semantic unit in natural talk, let us take the following example modified from a corpus of Spatial Planning dialogues. There are five people involved: two primary speakers, A and B, who are jointly planning a journey in Europe in connection with a trip simulation in an experimental setting. C and D are researchers conducting the experiment, and E is a secretary who came by.

A. We are in Spain, o.k. So, let's go to France next. I love France anyway. We had a great time there last year. And then Italy did I tell you about the little restaurant we went to in Florence?

B. Yeah. I think you did. It was better than the place in Rome we ate at before we took the plane. But, anyway, no. Let's go to Belgium next. Then

C. Could you move closer to the camera, please.

D. You're out of range

A. O.K. yeah. But not if we have to go through Antwerp

B. Then Holland

A. When do we do Italy then? We can't miss it?

B. On the way back to

E. Sorry. I was looking for Dave

C. He's not here. We're running an experiment I'll talk to you later. You are still out of camera range, by the way

A. Good

B. Anyway. I saw the tulips last year. What about Italy?

A. On the way back to Spain. You taking a vacation this year? Or loafing at work as usual?

B. Haven't decided, you?

A. Might go to Spain again. Then Germany's next, right?

Competent language users would intuitively segment this discourse into sections in which A and B are planning--actually developing their plan--and other sections where they are commenting on places they have been, making small talk, or conversing with the researchers. In one exchange, neither A nor B are talking at all, but are listening in while C exchanges some quick words with the secretary who is looking for someone who is not there. In order to make it somewhat easier to find the "planning," we have arranged the text graphically as an outline, showing the "planning talk" in leftmost position and moving further to the right to represent the embedded or secondary status of the comments and other interruptions to the development of the plan. It should be noted that when other types of talk are completed, A and B return to developing their plan, which is the focus of their attention throughout this excerpt.

A. We are in Spain, o.k.
So, let's go to France next.
I love France anyway.
We had a great time there last year.

And then Italy
did I tell you about the little restaurant we went to in Florence?

B. Yeah.
I think you did.
It was better than the place in Rome we ate at before we took the plane . .
(But, anyway, no.)
Let's go to Belgium next.
Then

C. Could you move closer to the camera, please.

D. You're out of range

A. O.K. yeah.
But not if we have to go through Antwerp

B. Then Holland.

A. When do we do Italy then?
We can't miss it?

B. On the way back to

E. Sorry.
I was looking for Dave

C. He's not here.
We're running an experiment
I'll talk to you later
You are still out of camera range, by the way

B. (Anyway.)
I saw the tulips last year.
What about Italy?

A. On the way back to Spain.
You taking a vacation this year?
Or loafing at work as usual?

B. Haven't decided, you?

A. Might go to Spain for a few days.
Then Germany's next, right?

Although this outlining procedure may make it easier to see at a glance which clauses encode propositions which can be

interpreted as proposals relating to the sequence of actions to be taken in some future time "Plan Execution World," not all leftmost clauses represent proposals which were taken into the final plan decided upon. Some proposals were made and then accepted--like A's suggestion to visit "France" after "Spain" which was accepted by B without comment--while other suggestions, such as A's next proposal to visit "Italy" next, were not accepted and were not included in the final agreed-upon plan.

The plan at the point in the game we are considering, as finally agreed upon, consisted of a hypothetical itinerary which would take A and B to:

Spain, France, Belgium, Holland, Germany . . .
(Italy (Spain))

in this sequence.

It is important to notice how many different parameters must be monitored in order to recover this itinerary:

- Temporal reference points must be maintained and, if necessary, updated (to understand "when" in conceptual time an event would take place).
- Spatial reference points must be maintained and, if necessary, updated (to understand the speaker's orientation in conceptual space)
- The identity of the speaker and hearer must be available (to be able to recover the intended referents of "I" and You)
- The specific "world" in which events are to take place (or have taken place) must be known (in order to interpret a spatial location or temporal reference point in the "Game" world or in the "real" world i.e., A is planning to vacation in Spain "this year" in the "real" world; A had a great time in France "last year" in the "real" world. "A" and "B" tokens in the "Game" world are in Spain and "planning a trip" from Spain, to France, Belgium etc.)

In addition, it must be pointed out that correctly interpreting this discourse involves understanding the form and function of a number of linguistic and rhetorical structures, including:

- Narrative syntax-mechanisms, encoding update of reference points
- Sentential syntax and semantics
- Question/answer sequences
- Discourse "operators" such as "o.k.," "yes," "no," "well," "anyway" which do not add independent information but which either (1) affirm or deny information available elsewhere (2) indicate a digression or a "return" to another topic
- Joking conventions (such as insulting a hard worker by accusing him of "loafing on the job.")
- Discourse embedding and return conventions

2.2 Recent Directions In Modeling Discourse Structure

Recent advances in understanding the structure of natural language discourse make it possible to segment complex talk and recover the integrity of "discourse units," despite the complexity of the actual talk in which they occur. An important research focus within the past five years has been to capture the semantic or "coherence" relations among clauses and segments making up a text in which all of the constituent elements function together to communicate a set of mutually interconnected ideas (Halliday & Hasan, 1977; Hobbs, 1979; Hobbs, 1985; Mann & Thompson, 1983; and Polanyi, 1985). A second research focus has been to understand the structural relations obtaining even in discourses which are not coherent, but which are characterized by interruptions and resumptions, and even by hesitations and other

types of complex phenomena arising from the social and processing constraints on actual talk (Reichman, 1981; Polanyi & Scha, 1984; Grosz & Sidner, 1986; and Hinrichs & Polanyi, 1986).

Section 2.2.1 below reviews some discussions of coherence relations in discourse. Sections 2.2.2-2.2.6 discuss some frameworks which attempt to characterize the structure of discourse--accounting for coherence, while also allowing for digressions and interruptions.

2.2.1 Discourse coherence. It has been observed many times that not every sequence of sentences makes up a "text." In a well-formed text the sentences are perceived as working together to build up a unified whole, by expressing propositions which are related to each other in a limited number of specific ways.

A number of coherence relations which may obtain among the constituents of a well-formed text have been identified, for instance, by Hobbs (1979, 1985). He describes how a semantic structure for a whole discourse may be built up recursively by recognizing coherence relations between adjacent segments of a text. He addresses himself initially to why it is that we find discourses coherent at all--what are the sources of discourse coherence? Not surprisingly, one source of discourse coherence lies in the coherence of the world or object described. We can find a text coherent if it tells us about a set of objects or states or events which we know to be coherent. Thus, even a gasped out, highly-interrupted narrative of a disaster may appear "coherent" and be "understandable" when we bring to the text our belief that the disaster formed a coherent set of events, related

causally to one another and affecting in various ways the people, objects, and situations described. This relates closely to another source of discourse coherence: When we find that one assertion details the cause for the situation described by the next assertion, we view the sequence as coherent. We will also find a sequence of two sentences, two stories, or, generally speaking, two discourse constituents to be coherently related to one another if one tells us more detail about the other, offers an explanation, or otherwise gives more information about the proposition expressed by the other.

Hobbs provides a method for allowing the coherence relations in a discourse to emerge. He suggests segmenting the discourse "intuitively" and then labelling the various naturally occurring segments with the coherence relation(s) which tie them to immediately preceding constituents. There will be two types of relations: coordination and subordination relations. Coordinate coherence relations include parallel constructions and elaborations in which one discovers a common proposition as the assertion of the composite segment. Subordination relations obtain when one constituent provides background or explanatory information with respect to another. Hobbs' ideas of "coherence" allow us to see how even the subsequent moves in a conversation, which may appear incoherent to an outside observer, may be appropriate conversational moves for the participants--entirely coherent and describable with the relations which he has outlined (Hobbs & Evans, 1980; and Hobbs & Agar, 1985).

Mann and Thompson's work on rhetorical relations focuses exclusively on the relations which obtain within a coherent text (Mann & Thompson, 1983). They assign a phrase structure analysis to texts, in which two subsequent constituents can be related through each of a number of specific relations. Their inventory of coherence relations is more detailed than that provided by Hobbs. The relations they list are solutionhood, evidence, justification, motivation, reason, sequence, enablement, elaboration, restatement, condition, circumstance, cause, concession, background, and thesis-antithesis.

2.2.2 Discourse structure and pronoun resolution. In early work on the structure of Task Oriented Dialogues, Grosz (1974) provided an important demonstration of the hierarchical structure of natural texts. In the analysis of talk between an apprentice and an expert dismantling a water pump, she showed that the discourse could be represented as a tree or outline in which the relationships among the clauses could be chunked in a way which replicated the goal/subgoal structure of the original task. Perhaps not surprisingly, in taking apart one part of the pump, the talk would focus on that operation; when the apprentice had finished dealing with that aspect of the job, and moved on to the next subtask, the talk would move along, reflecting in its structure what was going on in the joint endeavor. What was surprising, and most significant, however, was that the choice of possible referents for pronouns in the text reflected the structure of the task as well. In discussing a part of the object involved in the task at hand, one could refer to it with a

pronoun; similarly, one could refer to the entire higher level unit with a pronoun, or even to the pump as a whole. It was not possible to use a pronoun to refer to the objects and subtasks involved in a part of the task which had already been completed. In the tree of the discourse task/subtask elements one was blocked from referring to a task element in a branch to the left of the branch currently being developed. Grosz's discovery, therefore, was that discourse has a structure in which the placement and semantic relations obtaining among the clauses making up the discourse plays a decisive role in the interpretation of given elements in that discourse.

Sidner (1983) has shown that a structurally analogous account of anaphora resolution also applies at a linguistic level of discourse structure which is independent of task structure. In her model, the candidates for anaphoric reference are stored in a stack. An incoming discourse constituent which is treated as embedded PUSHes new focused elements into this list, while the resumption of a suspended discourse constituent POPs the intervening focus elements off the stack.

In the next section we shall give brief overviews of three frameworks which build on this seminal work and provide more comprehensive accounts of the issues involved in understanding both "coherent" and "interrupted" discourse: Reichman's Contest Space Theory (Reichman, 1981), Discourse Structures Theory developed by Grosz and Sidner (1986), and Polanyi and Scha's Dynamic Discourse Model (Polanyi & Scha, 1984; Polanyi, 1985; and Hinrichs, & Polanyi, 1986).

2.2.3 Context space theory. Reichman's context space theory deals with the structure of conversation (Reichman, 1981). It associates with each topic of discussion a context space--a schematic structure with a number of slots. These slots hold the following information:

- a propositional representation of the set of functionally related utterances said to lie in this context space;
- the communicative function served by the utterances in this context space;
- a marker reflecting the foreground-background status of this context space at any given point in the conversation;
- focus level assignments to the discourse elements in this context space;
- links to preceding context spaces in relation to which this context space was developed; and
- specification of the relations involved.

The utterances that constitute the discourse are analyzed as "conversational moves" which affect the content of the various context spaces. Reichman has paid special attention to the conversational structures involved in arguments. Among the conversational moves she identifies, for instance, are assertion of a claim, explanation, illustration, support, challenge, interruption, and further development.

An important and influential part of Reichman's theory is her treatment of clue-words--devices which speakers use to indicate when their discourse shifts from one structural level to another. Clue-words are commonly divided into PUSH-markers and POP-markers. PUSH-markers are linguistic signals that indicate the initiation of a new embedded discourse constituent. Examples

are "like," "by the way," "for instance." POP-markers have the complementary function. They close off the currently active embedded unit and signal a return to a higher level of structure. Examples are: "Well," "so," "anyway," "OK."

An extensive study of clue words in spoken French is presented by Guelich (1970). Schiffrin (1982) did an extensive study for English. Merritt (1978) discusses the use of "OK" in service encounters. Cohen (1984) studied clue words from a computational perspective. She draws two important conclusions:

- clue words decrease the amount of processing needed to understand coherent discourse.
- clue words allow the understanding of discourse that would otherwise be incomprehensible.

While Reichman's work provided much important insight into the functioning of discourse, her Context Space formalism fails to distinguish between those cases in which one can return to a previous topic by use of a simple POP, for example, and those cases in which such a simple, purely structural, return is not possible and one must re-introduce the topic in order to continue talking about it. Reichman's Context Spaces are never "closed off" and inaccessible because one can always say anything one wishes, and continuing to talk about a matter dropped earlier is certainly possible. Discourse structural relations, in her account, are thus finally obscured by discourse semantic relations obtaining among the topics of talk in the various units.

The work of both Grosz and Sidner (1986) and Polanyi and Scha (1984; Polanyi, 1985; Hinrichs & Polanyi, 1986),

incorporates elements of Reichman's work--particularly her treatment of clue words--while separating structural and semantic relations between clauses. This separation allows for a treatment of "interruptions" and "resumptions" which is based on structural properties of the discourse, rather than being dependent on semantic relationships among topics of talk. These two frameworks generalize upon Grosz's early work by providing an account of discourse structure which is not task dependent.

2.3 The Discourse Structures Theory

In the view of Grosz and Sidner (1986), the structure of a discourse results from three interacting components: a linguistic structure, an intentional structure, and an attentional state. These three components deal with different aspects of the utterances in a discourse. Grosz and Sidner have particularly focused on the intentional and the attentional aspects of discourse.

The intentional structure is a hierarchical structure which describes relations between the purpose of the discourse and the purpose of discourse segments. These purposes (such as "Intend that a particular agent perform a particular talk" or "Intend that a particular agent believe a particular fact") are linked by relations of dominance (between a goal and a subgoal) or ordering (between two goals which must be achieved in a specific order).

The attentional state is an abstraction of the participants' focus of attention as their discourse unfolds. The attention state is a property of discourse, not of discourse participants. It is inherently dynamic, recording the object, properties, and

relations that are salient at each point in the discourse. The attentional state is represented by a stack of focus spaces. Changes in attentional state are modeled by a set of transition rules that specify the conditions for adding and deleting spaces.

A focus space is associated with each discourse segment; this space contains those entities that are salient--either because they have been mentioned explicitly in the segment, or because they became salient in the process of producing or comprehending the utterances in the segment (as in Grosz's, 1974, original work on focusing). The focus space also includes the discourse segment purpose; this reflects the fact that the discourse participants are focused not only on what they are talking about, but also on why they are talking about it.

Discourse Structures Theory provides a unified account of both the intentional and attentional dimensions of discourse understanding and makes explicit important links between the two. The Dynamic Discourse Model, on the other hand, is more limited in its scope. It provides an account of the discourse segmentation process on an utterance-by-utterance basis and is thus a more developed theory of the strictly linguistic aspects of the discourse understanding process.

2.3.1 The dynamic discourse model. The Dynamic Discourse Model (DDM) (Polanyi & Scha, 1984; Polanyi, 1985; and Hinrichs & Polanyi, 1986) is a formal theory of discourse syntactic and semantic structure which accounts for how a semantic and pragmatic interpretation of a discourse may be incrementally built up from its constituent clauses.

The DDM is presented as a discourse parser. The parser segments the discourse into linguistically and socially relevant units on a clause-by-clause basis by proceeding through the discourse, examining the syntactic encoding form of each clause, its propositional content, and its situation of utterance.

The Model consists of a set of discourse grammars which specify the constituents of possible discourse units, a set of recursive rules of discourse formation which specify how units may relate to one another, and a set of semantic interpretation rules which assign a semantic and pragmatic interpretation to each clause and to the discourse as a whole.

Each discourse is viewed as composed of discourse units which can be of many different types: jokes, stories, plans, question/answer sequences, lists, "narratives" (temporally ordered lists), and Speech Events, socially situated occasions of talk such as doctor/patient interactions, and everyday conversations (see Section 4.4 below). In the DDM every possible discourse unit type is associated with its own grammar which specifies its characteristic constituent structure and is interpreted according to specific rules of semantic interpretation.

The basic unit of discourse formation is the discourse constituent unit. For the purpose of joining with other clauses to create a complex discourse, each clause is considered an elementary discourse constituent unit (dcu). Dcu's are of three types: list structures (including narratives, which are sequentially ordered lists of events), expansion structures, in

which one unit gives more detail of some sort about some aspect of a preceding unit, and binary structures such as "if/the," "and," "or," "but"--relations in which there is a logical connective connecting the constituents.

Discourse Units (DU's) such as stories and descriptions, arguments and plans are composed of dcu's which encode the propositions which, taken together, and properly interpreted, communicate elaborate semantic structures.

Dcu's and DU's in their turn, are the means of realization of the information exchange which is so basic in Speech Events, which are constituents of Interactions.

The DDM provides an account of the coherence relations in texts by means of an explicit mechanism for computing the semantic congruence and structural appropriateness of strings of clauses (Polanyi, 1985; Hinrichs & Polanyi, 1986). Simultaneously, it provides an account of the complexities of interrupted or highly attenuated discourse by providing a uniform treatment of all phenomena which can interrupt the completion of an ongoing discourse unit: elaborations on a point just made, digressions to discuss something else, interruptions of one Speech Event by another or one ongoing Interaction by another. All of these phenomena are treated as subordinated or embedded relative to activities which continue the development of an ongoing unit--whether it be a list of some sort, a story, or a Speech Event or Interaction.

The structure which results from the recursive embedding and sequencing of discourse units with respect to one another has the

form of a tree. This Discourse History Parse Tree contains, for any moment in the discourse, a record of which units of what types have been completed, and which, having been interrupted before completion, remain to be completed.

In determining whether an incoming clause is to be coordinated, subordinated, or superordinated to the last clause added to the Tree, the first step is to assign a set of contexts of interpretation to the clause specifying to which Interaction, Speech Event and Discourse Unit (if any) it belongs. The propositional content of the clause is then parsed into a semantic frame with slots for recording the temporal, spatial, and participant parameters of the clause's interpretation, as well as other important information.

The process of discourse segmentation with the DDM is a process of clause parsing, assignment to contexts of interpretation and search of only the rightmost Tree nodes for a suitable partner. If a suitable partner is found, either a node exists which permits the two to be joined, or if no suitable node exists, a new node is created and labelled with the values of the label of the node computed. The resultant Tree is therefore an incremental description of the developing discourse which reflects the surface structure relations, if any, between the constituents. Interruptions are accommodated in the tree as discourse embeddings in a way not dissimilar to their treatment in the Discourse Structures Theory. However, in order to accommodate the fact that what may be an interruption to one participant--or from the point of view of one Interaction--may be

the ongoing discourse from another perspective; each participant in a discourse is associated with a unique Discourse History Parse Tree representing the individual's incremental analysis of the history of the discourse. The degree to which participants' trees are identical determines their ability to understand each other's references to underdetermined elements in the discourse such as pronomials, deictics, or definite noun phrases.

The structural aspects of the DDM just discussed are related to the enterprise of developing an adequate discourse semantics--one which would allow the meaning of a discourse to be built up on a left-to-right basis, along with the structural analysis of the discourse. Developing such a compositional semantics for discourse presupposes adequate ways of representing the semantics of both sentences and discourse, as well as effective ways of dealing with the context dependence of utterance meanings. These issues are discussed in Section 3.

3. The Meanings of the Text

3.1 Truth Conditions for Sentence and Text

Semantic studies in philosophical logic have focused on one important aspect of the meaning of indicative sentences: The truth conditions of the sentence, i.e., a characterization of what must be the case in the world for the sentence to be seen as true rather than false. The truth conditions of a sentence can be mathematically described as a function from states of affairs to truth values. Logical languages, such as First Order Predicate Calculus or Intensional Logic, provide formulas for expressing such functions. (In an extensional logic, states of

affairs are represented by "models" of the logical language; in an intentional logic, they are represented by elementary entities, called "possible worlds.")

This logical perspective on sentence meaning has had considerable influence in linguistics and AI. Many theories and systems account for the way in which the truth conditions of a sentence depend on its surface form, by providing a definition or procedure which translates a sentence into a formula of a logical language. The same paradigm can be applied to texts consisting of more than one sentence, since a report or description may also be said to be "understood" (though in a limited sense) by someone who knows what state of affairs in the world would make it "true."

Carrying over the logical perspective on meaning from the sentence level to the text level raises the question of how to build up a logical representation for the truth conditions of a text out of the logical representations of the truth conditions of its constituent utterances. To do this, a text understanding program must be able to recognize the structure of a text, and to apply the semantic operations which build meanings at the levels above the sentence. It must also deal correctly with the sentence-level text constituents. Instead of analyzing the meaning of isolated, independent sentences, it must determine the meaning of particular utterances of sentences, taking into account the context which has been set up by the previous discourse.

Processing an individual utterance in a discourse thus entails three distinct operations:

- determining the utterance meaning in the applicable context;
- integrating the utterance meaning with the meaning of the text as processed so far; and
- updating the context setting which will be used to interpret the next utterance.

The context-dependence of utterance interpretation is shown by several difficult phenomena. For instance, temporal, locative or conditional interpretive frameworks may be introduced in the first sentence of a discourse segment and have scope over all other constituents of that segment. The reference time in a narrative moves on as the narrative proceeds (Kamp, 1979; Hinrichs, 1986; and Polanyi & Scha, 1984). Anaphoric expressions may refer from a subordinate constituent to entities introduced by its superordinate constituent, or from a constituent of a coordinate paragraph to certain entities introduced by an earlier constituent of that same paragraph.

3.2 Consequences for Logical Formalisms

Context-dependence. The context-dependence of utterance-meanings in discourse can be dealt with by translating a sentence not directly into a proposition, but into a function from contexts to propositions, where by "context" one means a data structure that contains all the relevant information that may influence sentence interpretation: speaker, addressee, speech time, speech location, reference time, candidates for anaphoric reference, topic, etc. Formally, contexts are very similar to indices as employed in Montague's systems (Montague, 1968;

Bennett, 1978). The meaning of a particular utterance of a sentence is then constructed by evaluating the sentence meaning with respect to the proper context.

In processing an utterance, a discourse understanding system must therefore determine what its proper context is--and also how this utterance may create a new context, or modify existing ones, for the interpretation of subsequent utterances. Polanyi and Scha (1984) propose to use Woods' (1970) Augmented Transition Network formalism to formulate a recursive definition of discourse constituent structure which is coupled with semantic rules that build up meaning representations for discourse constituent units; the register mechanism of the ATN's is used to keep track of the correct contexts in this process.

3.3 Discourse Anaphora

Beyond adopting a Montague-style context mechanism, some other departures from standard logical practice may be necessary to build up meaning representations for texts from meaning representations for sentences.

Observations on anaphoric reference in discourse have motivated some proposals for significant innovations in representational formalisms, especially concerning the representation of the denotation of indefinite noun phrases. Several authors (including Karttunen, 1976) have argued that indefinite noun phrases should be translated into "indefinite entities" of some sort, as opposed to existential quantifiers.

For instance,

"John loves a woman."

would not be represented as

E x: Woman (x) and Love (J, x)

but rather as

Woman (u) and Love (J, u)

where u is a Skolem-constant--a constant whose denotation is undetermined, therefore behaving, for all practical purposes, like a variable which is implicitly existentially quantified. Leaving the existential quantifier implicit has an advantage when one deals with discourse anaphora.

"John loves a woman. Her name is Mary."

can be treated simply by conjoining the formula for "Her name is Mary," with the one for "John loves a woman," while resolving the pronoun "her" to corefer with the constant for "a woman:"

(Woman (u) and Love (J, u)) and name (u) = "Mary."

This procedure does not work if indefinite noun phrases are represented by existential quantifiers:

(E x: Woman (x) and Love (J, x)) and name (x) = "Mary"

is infelicitous because a variable is used outside the scope of its defining occurrence.

The perspective just sketched has been pushed furthest in a formalism devised by Hans Kamp (1979). The formulas used in this formalism are called Discourse Representation Structures (DRS's). They serve the role of logical formulas, representing the meaning of the text so far, as well as the role of contexts which set up

the right reference times and anaphoric reference candidates for the interpretation of next utterances.

DRS's differ from ordinary logical formulas in the way variables are used. A DRS is defined to be true if it is embeddable in a model which corresponds to the actual world. Embeddability of DRSs is recursively defined on the structure of the formulas.

An alternative approach to the problem of discourse anaphora is described by Webber, where the representation of sentence meanings is separated from the representation of "evoked entities" (Webber, 1982).

3.4 Background Knowledge and Plausible Inferences.

Understanding a text involves much more than understanding the literal meanings of its constituent utterances, and their explicitly stated relations. The message of a text is rarely completely explicit; the author relies on the fact that the hearer/reader will integrate the meanings of the utterances with an independently given set of background assumptions about the domain and about the author. All implications which follow in a simple and direct way from the combination of the explicit utterances and the presupposed background knowledge are considered to be implicit in the text.

For a system to be capable of discourse understanding in this more extended sense, its mechanisms must be augmented with a representation of the required background knowledge, and with a system that performs inferences on the basis of explicit text meanings and background knowledge, generating representations of

information that was implicit in the text. Different kinds of background information play a role. Ideally, a discourse understanding system should have a rather rich, encyclopedic knowledge base or, at least, a knowledge base comparable to the user's for the pertinent domain, and it should have particularly good coverage in knowledge which people consider "common sense." How to model common sense domains has therefore become a research area in itself (Charniak, 1977; and Hobbs & Moore, 1985).

An important set of background assumptions which has received a lot of attention concerns the characters in stories; unless told otherwise, story-recipients must assume the characters to be "normal," rational, purposeful people, and they must bring these assumptions to bear on the text in order to make sense of it. Various systems have been built which embody some knowledge of this sort and bring it to bear on the discourse-understanding process.

SAM (Cullingford, 1978, 1981; and Schank & Abelson, 1977), for instance, is a system for understanding narratives, which is based on the notion of a script. A script is a knowledge structure which represents a stereotypical sequence of events, such as taking a bus, going to a movie theatre, or going to a restaurant for dinner. SAM's representation of a script consists of a set of simple actions described as conceptual dependency structures, together with the causal connections between those actions. The actions in a script are further organized into a sequence of scenes, which, in the case of the restaurant script, includes entering the restaurant, ordering food, eating, paying,

and leaving. Each script also has a set of roles and props characterizing the people and objects that are expected to appear in the sequence of events.

In processing a narrative about eating in a restaurant, SAM first has to recognize that the restaurant script is the relevant context for interpreting the narrative. Once the script is chosen, SAM will try to interpret each new sentence as part of that script. It does this by matching the conceptual representation of the new sentence against the actions represented in the script. When it finds a match, it incorporates the sentence meaning into its representation of the narrative. It also fills in the script actions preceding the one matched. By this process, SAM infers actions that are implicit in the narrative it is reading. Thus, when it reads the narrative:

> John went to the Fisherman's Grotto for dinner.
> He ordered lobster. The bill was outrageous.

it includes in its representation that John actually ate his lobster, that he received a large bill, and that he paid it.

A later system, FRUMP (DeJong, 1979a, 1979b), pushes the idea of expectation-driven understanding a little further and dispenses with script-independent meaning representations altogether; it parses its input text directly into script-slots, and anything which does not fit is ignored. (FRUMP is presented as a model of human text skimming.) IPP (Levin & Moore, 1977; and Sidner & Israel, 1981), in its turn modifies the FRUMP approach by mixing script-based text skimming with a somewhat more careful semantic analysis of selected parts of the text.

Its meaning representations contain not only scripts with filled-in slots, but also representations of "unexpected events."

In a realistic application of the script approach, the scripts to be invoked must be selected from thousands of candidates; SAM chose from only three or four. Furthermore, one will have to drop SAM's assumption that each script contains one event that is always explicitly mentioned in the text in order to invoke the script. The task of finding which of the many candidate scripts matches the input sequence best thus presents computational problems which deserve further study.

The idea of a "script" is usually associated with the description of predefined sequences of events which constitute the "building blocks" of everyday life. Almost by definition, scripts are not sufficient to understand interesting stories. Real stories tend to involve somewhat more complex plots, arising from conflicts between the perceptions, ideas, and goals of the different characters. A program that interprets its input reports in terms of the goals and subgoals of the protagonist is PAM (Plan Applier Mechanism), designed by Wilensky (1981).

Later work derives plot structure from "interacting plans," that is, plans involving two or more participants in cooperative or competitive interaction. Such plans differ from single participant plans in several ways (see Bruce, 1986); the most significant being that they are produced, interpreted, and executed in a belief context, i.e., what participants believe about the interaction is significant, rather than any putative objective account of the events.

Thus, for example, in order for a system to make sense of a children's story such as "Hansel and Gretel," it must monitor the evolution of the children's, the parents', and the witch's beliefs about events as well as the events themselves (Bruce & Newman, 1978). When the parents tell the children that the family is going to "fetch wood," the system must note that the actions the parents subsequently take are designed to be interpretable by the children as simple wood fetching, but are simultaneously effecting the abandonment of Hansel and Gretel. Moreover, it must be able to compute embedded beliefs, e.g., the parents do not know that Hansel has overheard their plan and, hence, that he believes that they intend him to believe the actions contribute to wood fetching, but, in fact, are intended to lead to his and Gretel's death. Central to this belief monitoring is the computation of mutual belief (Allen, 1979; Bruce & Newman, 1978; and Cohen, 1984) i.e., those beliefs fully shared and known to be shared among the participants.

Mechanisms for interacting plans calculations have been outlined in some detail (Bruce & Newman, 1978), but not fully implemented in any current systems. Analyses in terms of interacting plans have proved useful in studies of conversations (Bruce, 1986), classroom interactions, skits (Newman & Bruce, 1986), and written stories (Bruce & Newman, 1978; Bruce, 1980a, 1980b).

3.5 Summarizing Stories

Understanding a story as a communicative object requires more than dealing with its explicit content and the associated

plausible inferences. When someone tells a story, not all the information reported is equally important. Truly understanding the story would mean, among other things, being able to see the distinctions between more important and less important information. Evidence of this kind of understanding would be a system's capability to generate adequate summaries of input texts.

Many approaches to the story understanding problem have been proposed. Four of them are discussed below; they are based, respectively, on surface text phenomena, on lot structure, on affective dynamics, and on the author-reader relationship.

The first approach implements the ideas formulated by Polanyi concerning the way in which human storytellers encode their information. She maintains that people explicitly mark the relative salience of different pieces of information in a text; they make sure that an important piece of information "stands out" against the surrounding information. They do this by means of meta-comments, of various evaluative devices: explicit markers, repetition, and the use of encoding forms which deviate from the "local norm" in the text (long versus short sentences, direct discourse versus narrated events, colloquial versus formal register, etc.) (Polanyi, 1985).

Based on these ideas, a system was developed that simply counts the number of evaluative devices used to highlight each proposition in a story and then puts the most highly evaluated states and the most highly evaluated events together in a summary of the input story. The system thus manages to construct a

reasonable summary on the basis of the surface appearance of the story, without understanding it in any sense; it shows that we must be careful in ascribing "understanding" capabilities to a system which performs a specific task.

The relevant work on plot structure originates with Propp (1968) and Rumelhart (1975). Lehnert (1981; Lehnert, Black, & Reiser, 1981; Lehnert, 1983; and Lehnert & Loiselle, 1985) developed a summarization algorithm based on the causal relations between the events and states reported in a story. By inspecting the network of causal connections, it concludes that certain events play a crucial role in the development of narratives, by moving the plot from one place to an essentially different place.

Closely related to Lehnert's work is Dyer's (1981, 1983) system, called BORIS, which attempts "in-depth understanding" of narratives. Such understanding should include being able to summarize the point or moral that the author intended the narrative to represent. This work moves beyond earlier work on plan-based understanding, such as Wilensky's (1981), by abstracting the communicative intent.

BORIS embodies thematic patterns, called Thematic Abstraction Units (TAUs). For example, TAU-DIRE-STRAITS encodes the pattern: x has a crisis goal; x can't resolve the crisis alone; x seeks a friend y to help out. TAUs arise from errors in planning or plan execution. They refer to a plan used, its intended effect, why it failed, and what can be done about the failure. As such, they allow BORIS to organize the narratives at

an intentional level, which leads naturally to an appropriate summarization or even drawing of a moral.

A contrasting approach is that of Brewer and Lichtenstein (1981, 1982). They argue that stories are a subclass of narratives whose purpose is to entertain. Thus, plan-based analyses ultimately miss the point of a story if they are not augmented by an effective component, one that shows how structural elements of the text influence the reader. For example, suspense is created when the author reveals that a negative outcome is in store for a central character and that the character is unaware of his or her fate. Thus, relations among the author's, the reader's, and the characters' belief states become essential to understanding, or being affected by, the story.

In the line of the Brewer and Lichtenstein approach, Bruce (1980b) outlines a central model of the author-reader relationship. The model makes explicit not only the author and the reader as participants in the communicative act, but also a constellation of other implied participants. For instance, in an ironic text, the author establishes an apparent speaker with beliefs and intentions which conflict in some respects with her own.

It is noteworthy that to date attempts such as those of Brewer, Lichtenstein, and Bruce have been purely theoretical; no working system addresses the interactions of author's and reader's goals at that level.

4. Plan Recognition

4.1 The Pragmatic Perspective on Discourse

Language, especially written language, is often viewed as a code for packaging and transmitting information from one individual to another. Under this view, a linguistic message is fully represented by the words and sentences it comprises; texts are thus objects that can be studied in isolation. By taking such a stance, one is led naturally, for instance, to regard words as referring back to other words. Concepts like coherence, relevance, and topic are then regarded as properties of texts, leading researchers to confine their search for these properties to words and sentences.

A contrasting view, proposed by Strawson (1950), Austin (1962), Searle (1969), and others is that speakers or writers use words to do things, for instance, to refer to things, or to get a hearer or reader to believe or do something. They are produced by a person, who is attempting to use them to produce certain effects on an audience (perhaps an imagined audience). According to this view, utterances are tools used in social interaction and should be studied in that light.

Morgan and Sellner (1980) suggest that properties like coherence, relevance, and text structure are likely to be obtained from a theory of plans and goals appropriately extended to linguistic actions. Properties like "relevance" would be epiphenomenal byproducts of the appropriate structuring of actions.

Pragmatics is the study of communication as it is situated relative to a particular set of communication demands, speakers, hearers, times, places, joint surroundings, linguistic conventions, and cultural practices. Including language in a theory of action, this suggests that "pragmatics" is just the application to verbal problems of general abilities for interpreting the everyday world (see Morgan, 1978, for fuller discussion). People tend to interpret the behavior of other humans in terms of the situation and the actor's intention and beliefs. Much of what has been discussed under the rubric pragmatics is most reasonably seen as the interpretation of linguistic behavior in similar terms.

The pragmatic perspective on language has three important implications for discourse understanding research. The first is that the meaning of a linguistic message is only partly represented by its content; its meaning for a hearer also depends on the hearer's construal of the purpose that the speaker had for producing it. The second is that the attribution of intentions to a speaker must be an integral component of the listener's comprehension process. The third is that a theory of language comprehension should determine the extent to which the same strategies people use to arrive at satisfactory explanations of the physical behavior of others can be employed in their comprehension of speech acts.

The way the meaning of a message is shaped by its producer's goals and beliefs is most obvious in a case such as propaganda, but it is no less critical for apparently straightforward

utterances. For example, a colleague at the office might say, "I brought two egg salad sandwiches today." Although the referential meaning of this statement might be simple to compute, its full meaning depends on whether the speaker's intention was, for example, to offer one of the sandwiches, to decline a luncheon invitation, or to explain why the office smelled bad. Whatever the speaker's goals, the meaning conveyed by the statement depends on the hearer's correctly inferring what they are (Adams & Bruce, 1982).

Thus, understanding discourse requires inferring the intentions and beliefs that led the speaker to produce the observed behavior. But as Grice (1957) points out, simply recognizing an actor's plan, as an unseen observer might do (cf. Schmidt, Sridharan, & Goodson, 1979; and, Wilensky, 1981), is insufficient as a basis for communication. Instead, hearers should attribute to speakers intentions that the speakers intend for them to infer. To ensure successful communication, speakers attempt to maximize the likelihood that hearers will make the inferences they were supposed to make by relying on what Lewis (1969) terms "conventions." Conventions are solutions to coordination problems--where any participant's actions depend on the actions of others--and themselves rely on "mutual knowledge" held amongst the parties involved. Mutual knowledge (see also Schiffer, 1972) occurs when two people know that a proposition P holds, that the other person knows as well that P holds, that the second knows that the first knows that P holds, and so on. In ordinary conversation, participants make assumptions about mutual

knowledge, signal their assumptions through the pragmatic presuppositions (Stalnaker, 1974) of their utterances, and negotiate misunderstanding of the developing mutual knowledge.

4.2 Speech Acts

From a pragmatic perspective, the goal of discourse understanding should not be to merely assess the truth conditions of one's interlocutor's utterances. Instead, one should be concerned with the goal which is being pursued through these utterances, and with the way in which every utterance contributes to that goal. From this perspective, every language utterance is viewed as a social act; it changes, be it perhaps on a small scale, the social relation between the speaker and his interlocutor. A simple assertion puts me under the obligation to defend it if challenged. A question creates for my interlocutor the obligation to answer it or to be prepared to justify his lack of an inclination to do so. And vows, promises and threats clearly extend beyond the micro-sociology of the interactional situation, creating commitments in the social world at large. The social acts performed by means of linguistic utterances are called Speech Acts (Searle, 1969).

The Speech Act types which play a role in current experimental dialogue systems are:

- Requests, typically formulated as questions of the form "Could you do X"
- Commands, directly expressed as imperative sentences. ("Do X") Notice that for most programs, which slavishly try to satisfy every whim of their human dialogue partner, there is no distinction between a request and a command. The program takes no responsibility for its actions.

- Assertions, directly expressed as indicative sentences. Assertions are usually interpreted as commands to store and/or evaluate the asserted information.

- Questions, directly expressed as interrogative sentence. A question is usually interpreted as a command to provide the answer.

4.3 Plan Recognition

If a system analyzes its input utterances as speech acts and has at its disposal a repertoire of plausible goals that its dialogue partner may pursue, it may be able to understand the purpose behind its input utterances by using a method which is reminiscent of the way in which a system like PAM (Wilensky, 1981) understands reports about goal-oriented behavior; it tries to guess the more encompassing goal that the speaker may be trying to accomplish by executing a plan which has the surface speech act as one of its subgoals.

A system that tries to derive the deeper intentions behind surface speech acts in exactly this way was developed by Allen (1979). His system exploits knowledge about what constitutes a rational plan, as well as beliefs about what goals the speaker is likely to have.

The plan inference process is specified as a set of inference rules and a control strategy. Rules are all of the form "If agent S believes agent A has a goal X, then agent S may infer that agent A has a goal Y." Examples of such rules are:

> If S believes A has a goal of executing action ACT, and ACT has an effect E, then S may believe that A has a goal of achieving E.
>
> If S believes A has a goal of knowing whether a proposition P is true, then S may believe that A has a goal of achieving P.

Of course, given the conditions in the second rule, S might alternatively infer that A has a goal of achieving not P; this is treated as a separate rule. Which of these rules applies in a given setting is determined by control heuristics, as follows.

The plan inference process can be viewed as a search through a set of partial plans. Each partial plan consists of two parts: one part is constructed using the plan inference rules from the observed action, and the other is constructed using the plan construction rules on an expected goal. When mutually exclusive rules can be applied to one of these partial plans, the plan is copied and one rule is applied in each copy. Each of these partial plans is then rated as to how probable it is to be the correct plan. The highest rated partial plan is always selected for further expansion using the inference rules. The rating is determined using a set of heuristics that fall into two classes: those that evaluate how well-formed the plan is in the given context and those that evaluate how well the plan fits the expectations. An example of a heuristic is:

> Decrease the rating of a partial plan if it contains a goal that is already true in the present context.

Allen argues that whenever the intended plan can be derived from mutual knowledge, i.e., from knowledge which is knowingly shared between speaker and hearer, the hearer is assumed to perceive the intended plan, and is expected to react to that plan, rather than to the surface speech act. The paradigm examples of such situations are known as indirect speech acts (Perrault & Allen, 1980): sentences like:

"Can you pass the salt?"

or

"Is the salt near you?"

uttered at the dinner table where the simple answer "Yes," without an accompanying action, would be experienced as a joke or an insult.

The idea also applies, however, to cases that are normally not classified as indirect speech acts. For instance, when at the information counter of a train station someone asks:

"Does the 4:20 train go to Toronto?"

the answer "No" is less helpful than the answer

"No, but the 5:10 train does."

which responds to the speaker's perceived goal of going to Toronto.

Allen's plan recognition paradigm has been developed in work by Sidner (Sidner & Israel, 1981; Sidner, 1983, 1985). Pollack (1986) has refined it to deal with situations where speaker and hearer have conflicting ideas about how certain goals may be achieved. Litman (1986; Litman & Allen, 1984), has introduced meta-plans which allow for clarification subdialogues and plan corrections; she also integrates an awareness of the surface structure of discourse, as discussed in section 2 above, into the plan-recognition process.

4.4 Speech Events

An "unframed" interaction between "uninterpreted" people is a rare event. People use a refined system of subcategorization to classify the social situations they engage in. These

subcategories, called Speech Event types (Hymes, 1967, 1972), often assign a specific purpose to the interaction, specify roles for the participants, constrain discourse topics and conversational registers, and, in many cases, specify a conventional sequence of component activities.

An awareness of what kind of Speech Event one is engaged in, thus helps the plan recognition process; the overall goals of the interaction, and often the steps to achieve them, are shared knowledge among the participants.

The most precisely circumscribed kinds of Speech Events are formal rituals. Speech Event types characterized by grammars which are less explicit and less detailed include service encounters (Merritt, 1978), doctor-patient interactions (Byrne & Long, 1976), and casual conversations. Schegloff (Schegloff & Sacks, 1973) has shown that the process of terminating a telephone conversation is a jointly constructed ending sequence unit with a predictable course of development.

The structure of talk which is exchanged in order to perform a task may follow the structure of some goal/subgoal analysis of this task (Allen, 1979). In Speech Event types which involve a more or less fixed goal, this often leads to a fixed grammar of subsequent steps taken to attain it. For instance, as described by Polanyi and Scha (1984), transcripts of the activities in Dutch butcher shops consistently display the following sequential structure in the interaction between the butcher and a customer:

1. It is established that it is this customer's turn.
2. The first desired item is ordered, and the order is dealt with, . . . , then the desired item is ordered and the order is dealt with.
3. It is established that the sequence of orders is finished.
4. The bill is processed.
5. The interaction is concluded.

Each of these steps is filled in in a large variety of ways--either of the parties may take the initiative at each step, question/answer sequences about the available meat, the right way to prepare it, or the exact wishes of the customer may all be embedded in the stage 2 steps, and clarification dialogs of various sorts may occur. In other words, we find the whole repertoire of possibilities admitted by the discourse grammar.

An important Speech Event type with characteristics slightly different from the types mentioned so far, is the casual conversation. In a casual conversation, all participants have the same role: to be "equals;" no purposes are pre-established; and the range of possible topics is open-ended, although conventionally constrained.

4.5 Dialogue Systems

Many dialogue systems have been designed to partake in specific types of speech events, in which the computer system and its human interlocutor each play a well-defined role. The assumption that every dialogue must fall within the patterns allowed by the speech event type makes it possible to resolve ambiguities in its input (anaphora, ellipsis) and to react to the intentions behind it, also when these are not explicitly stated.

Most systems of this sort play the role of the "professional" in a consultation interaction of some sort. Examples are:

- a system that teaches an assembly task (Hobbs, 1979)
- an information system at a train station (Allen, 1979)
- a travel budget manager (Bruce, 1980)
- a Rogerian psychotherapist

Such speech event types involve the participants cooperating towards a common goal. In doing this, they decompose the common task into subtasks, and, eventually, into elementary subtasks that can be executed by one or both of the participants without requiring further dialogue. For instance, as discussed earlier (Section 2.2.2), Grosz's original investigation of dialogues between a human instructor and an apprentice who was being told how to assemble a water pump, showed that the structure of such dialogues corresponds closely to the structure of the task (Grosz, 1974).

One should notice, however, that the description of the task structure does not predict one fixed tree structure (Grosz, 1974). A task may involve subtasks that must all be done, but can be done in any order. It is not difficult to imagine further complexities: alternatives, preconditions, etc. When a task does specify one fixed sequence of subtasks, the task structure degenerates into a script (cf. section 3.4).

5. Modes of Natural Language

We tend to think of language coming to us in one of two forms: oral or written. Thus, AI research on Discourse Understanding is conveniently divided between research on

understanding text and research on participating in interactive dialogues, which, although most often written rather than spoken, are thought of as analogous to oral conversations. That this division is inadequate and at times misleading, is shown by Rubin (1980), who postulates eight dimensions of variation among "language experiences."

The eight dimensions: (1) oral versus written modality, (2) interactiveness, (3) spatial commonality, (4) temporal commonality, (5) possibility of para-linguistic communication, (6) concreteness of referents, (7) audience specificity, and (8) separability of participants, define a range of communication modalities out of which AI research has focused on only a few, albeit significant ones.

Most AI research (notable exceptions being speech understanding work and some efforts at modeling real conversations (Reichman, 1981; Hobbs & Evans, 1980; Hobbs & Agar, 1985; Levin & Moore, 1977; and Hinrichs & Polanyi, 1986) has focused on written language and is thus clustered on one pole of Rubin's first dimension. What distinguishes the AI dialogue work from the AI text work then is that the former is interactive, and usually implies spatial and temporal commonality. On the other hand, neither of the two modes of language use includes para-linguistic communication, such as gestures, facial expressions, or body position cues. In some of the dialogue work, but not the text work, there are concrete referents, in the sense that objects are perceptually present to the user and the machine. The same holds for audience specificity; some of the dialogue work

assumes fairly detailed speaker models of the hearer. Neither of the modalities typically allows separability of participants. Indeed, most of the communication is one to one.

From the perspective of this dimensional analysis, the research directed at the implementation of interactive computer programs that display reasonable behavior in conducting a dialogue with a person amounts to the development of a new mode of natural language, rather than the analysis of an existing one: real-time alphanumeric interaction, usually without shared awareness of physical context. Other AI research has focused on text understanding, usually assuming a non-specific audience. (In contrast, note the many existing forms of text understanding, such as dealing with letters, memos, persuasive essys, etc., which do assume specific audience beliefs and plans).

Some studies (Cohen, Fertig, & Starr, 1982; Cohen, 1984; and Tierney, LaZansky, Raphael, & Cohen, 1983) have been devoted to the linguistic consequences of the use of different communication media. Cohen (1984), for example, used a plan-based model of communication to analyze dialogues in five modalities: face-to-face, telephone, linked CRT's, (non-interactive) audio tape, and (non-interactive) written text. He found that speakers in the face-to-face situation, for example, attempted to achieve more detailed goals in giving instructions than did users of keyboards. More specifically, requests that the hearer identify the referent of a noun phrase dominated spoken instruction giving discourse, but were rare in the keyboard dialogues.

These studies suggest that it is important to understand the constraints of the communication system as well as the texts per se when an AI system is being designed. Moreover, they imply a need for caution in interpreting results of AI research. Any form of language use is valid to examine and can be illuminating in a general way, but specifics of language processing must be interpreted in light of the communication modality in which they arise.

References

Adams, M. J., & Bruce, B. C. (1982). Background knowledge and reading comprehension. In J. Langer & M. T. Smith-Burke (Eds.), Reader meets author/bridging the gap: A psycholinguistic and sociolinguistic perspective (pp. 2-25). Newark, DE: International Reading Association.

Allen, J. F. (1979). A plan-based approach to speech act recognition (Tech. Rep. No. 131). Toronto, Canada: University of Toronto, Department of Computer Science.

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

Bennett, M. (1978). Demonstratives and indexicals in Montague grammar. Synthese, 39, 1-80.

Brewer, W. F., & Lichtenstein, E. H. (1981). Event schemas, story schemas, and story grammars. In J. D. Long & A. D. Baddeley (Eds.), Attention and performance IX (pp. 363-379). Hillsdale, NJ: Erlbaum.

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. Journal of Pragmatics 6, 473-486.

Bruce, B. C., & Newman, D. (1978). Interacting plans. Cognitive Science 2, 195-233.

Bruce, B. C. (1980). Analysis of interacting plans as a guide to the understanding of story structure. Poetics 9, 295-311.

Bruce, B. C. (1980). Plans and social actions. In R. Spiro, B. C. Bruce, & W. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 367-384). Hillsdale, NJ: Erlbaum.

Bruce, B. C. (1986). Robot plans and human plans: Implications for models of communication. In I. & M. Gopnick (Eds.), From models to modules: Studies in cognitive sciences from the McGill workshops (pp. 97-114). Norwood, NJ: Ablex.

Byrne, P. S., & Long, B. E. L. (1976). Doctors talking to patients. London: Her Majesty's Stationary Office.

Charniak, E. (1977). A framed PAINTING: The representation of a commonsense knowledge fragment. Cognitive Science, 1(4), 355-394.

Cohen, P. R., Fertig, S., & Starr, K. (1982). Dependencies of discourse structure on the modality of communication: Telephone vs. teletype. In Proceedings of The 20th Annual Meeting of the Association of Computational Linguistics (pp. 28-35). ACL.

Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. Computational Linguistics, 10, 97-146.

Cullingford, R. E. (1978). Script application: Computer understanding of newspaper stories. Unpublished doctoral dissertation, Yale University.

Cullingford, R. E. (1981). SAM. In R. C. Schank, & C. K. Riesbeck (Eds.), Inside computer understanding: Five programs plus miniatures (pp. 75-119). Hillsdale, NJ: Erlbaum.

DeJong, G. F. (1979). Prediction and substantiation: A new approach to natural language processing. Cognitive Science, 3, 251-273.

DeJong, G. F. (1979). Skimming stories in real time: An experiment in integrated understanding. Unpublished doctoral dissertation, Yale University.

Dyer, M. G. (1981). The role of TAUs in narratives. In Proceedings of the Third Annual Conference of the Cognitive Science Society (pp. 225-227). Berkeley, CA: Cognitive Science Society.

Dyer, M. G. (1983). In-depth understanding: A computer model of integrated processing and memory for narrative comprehension. Cambridge, MA: Massachusetts Institute Technology Press.

Grice, H. P. (1957). Meaning. Philosophical Review, 66, 377-388.

Grosz, B. [Deutsch] (1974). The structure of task oriented dialogs. In IEEE Symposium on Speech Recognition: Contributed Papers (pp. 250-253). Pittsburgh, PA: IEEE, Carnegie Mellon University, Department of Computer Science.

Grosz, B. J., & Sidner, C. L. (1986). The structures of discourse structure. In L. Polanyi (Ed.), Discourse Structure. Norwood, NJ: Ablex.

Guelich, E. (1970). Makrosyntax der Gliederungssignale im Gesprochenen Franzoesisch. Munich: Wilhelm Fink Verlag.

Halliday, M., & Hasan, R. (1977). Cohesion in English. London: Longmans.

Hinrichs, E. (1986). Temporal anaphora in discourse of English. Linguistics and Philosophy, 9(1), 63-82.

Hinrichs, E., & Polanyi, L. (1986). A unified account of a referential gesture. In Proceedings Parasession on Pragmatics. Chicago, IL: Chicago Linguistics Society.

Hobbs, J. (1979). Coherence and co-references. Cognitive Science, 3(1), 67-82.

Hobbs, J. R. (1985). On the coherence and structure of discourse (Tech. Rep. No. CSLI-85-37). Stanford, CA: Center for the Study of Language and Information.

Hobbs, J. R., & Agar, M. H. (1985). The coherence of incoherent discourse. Language and Social Psychology, 4, 213-232.

Hobbs, J., & Evans, D. (1980). Conversation as planned behavior. Cognitive Science, 4(4), 349-377.

Hobbs, J. R., & Moore, R. C. (1985). Formal theories of the commonsense world. Norwood, NJ: Ablex.

Hymes, D. (1967). Models of the interaction of language and social setting. Journal of Social Issues, 23(2), 8-28.

Hymes, D. (1972). Models of the interaction of language and social life. In J. Gumperz & D. Hymes (Eds.), Directions in sociolinguistics (pp. 35-71). New York: Holt, Rinehart and Winston.

Kamp, H. (1979). Events, instants and temporal reference. In U. Egli & A. van Stechow (Eds.), Semantics from a multiple point of view (pp. 376-471). de Gruyter, Berlin.

Karttunen, L. (1976). Discourse referents. In J. McCawley (Ed.), Syntax and Semantics, (Vol. 7). New York: Academic Press.

Lehnert, W. G. (1981). Plot units and narrative summarization. Cognitive Science, 5(4), 293-331.

Lehnert, W. G., Black, J. B., & Reiser, B. J. (1981). Summarizing narratives. In Proceedings of the Seventh International Joint Conference on Artificial Intelligence (pp. 184-189). Vancouver, B.C.: IJCAI.

Lehnert, W. G. (1983). An in-depth understander of narratives. Artificial Intelligence 20(1), 15-62.

Lehnert, W., & Loiselle, C. (1985). Plot unit recognition for narratives. In G. Tonfoni (Ed.), Artificial intelligence and text-understanding: Plot units and summarization procedures (pp. 9-47). Ed. Zara, Parma, Italy.

Levin, J. A., & Moore, J. A. (1977). Dialogue games: Metacommunication structures for natural language interaction. Cognitive Science, 1(4), 395-420.

Lewis, D. K. (1969). Convention: A philosophical study. Cambridge, MA: Harvard University Press.

Litman, D. J., & Allen, J. F. (1984). A plan recognition model for subdialogues in conversations (Tech. Rep. No. TR 141). New York: University of Rochester, Department of Computer Science.

Litman, D. J. (1986). Linguistic coherence: A plan-based alternative. In 24th Annual Meeting of the Association for Computational Linguistics (pp. 215-223). New York: ACL.

Mann, W. C., & Thompson, S. A. (1983). Relational propositions in discourse (Tech. Rep. No. RR-83-115). Marina del Rey, CA: Information Sciences Institute.

Merritt, M. (1978). On the use of o.k. in service encounters. Austin, TX: Southwest Educational Development Lab, Texas Working Papers in Sociolinguistics 42.

Montague, R. (1968). Pragmatics. In R. Klibansky (Ed.), Contemporary philosophy: A survey (pp. 102-22). Florence, Italy: La Nuova Italia Editrice.

Morgan, J. L. (1978). Two types of convention in indirect speech acts. In P. Cole (Ed.), Syntax and semantics, (Vol. 9: Pragmatics (pp. 261-280). New York: Academic Press.

Morgan, J. L., & Sellner, M. (1980). Discourse and linguistic theory. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 165-200). New York: Erlbaum.

Newman, D., & Bruce, B. C. (1986). Interpretation and manipulation in human plans. Discourse Processes, 9, 167-195.

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

Polanyi, L. (1985). A theory of discourse structure and discourse coherence. In 21st Regional Meeting of the Chicago Linguistic Society (pp. 306-322). Chicago, IL: University of Chicago, Chicago Linguistic Society.

Polanyi, L. (1985). Telling the American story. Norwood, NJ: Ablex.

Polanyi, L., & Scha, R. (1984). A syntactic approach to discourse semantics. In Proceedings of International Conference on Computational Linguistics (pp. 413-419). Stanford, CA: Stanford University.

Pollack, M. E. (1986). A model of plan inference that distinguishes between the beliefs of actors and observers. In 24th Annual Meeting of the Association for Computational Linguistics (pp. 207-214). New York: ACL.

Propp, B. (1968). Morphology of the folktale. Austin, TX: University of Texas Press.

Reichman, R. (1981). Plain-speaking: A theory and grammar of spontaneous discourse. Cambridge, MA: Doctoral Thesis, Harvard University, Department of Computer Science.

Rubin, A. D. (1980). A theoretical taxonomy of the differences between oral and written language. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 411-438). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding (pp. 211-236). New York: Academic Press.

Schank, R. C., & Abelson, R. (1977). Scripts, plans, goals, and understanding. Hillsdale, NJ: Erlbaum.

Schegloff, E., & Sacks, H. (1973). Opening up closings. Semiotica VIII, 4, 289-327.

Schiffer, S. (1972). Meaning. London: Oxford University Press.

Schiffrin, D. (1982). Discourse markers: Semantic resource for the construction of conversation. Unpublished Doctoral Dissertation, University of Pennsylvania.

Schmidt, D. F., Sridharan, N. S., & Goodson, J. L. (1979). The plan recognition problem: An intersection of artificial intelligence and psychology. Artificial Intelligence, 10, 45-83.

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

Sidner, C. L., & Israel, D. J. (1981). Recognizing intended meaning and speaker's plans. In Proceedings of the International Joint Conference in Artificial Intelligence (pp. 203-208). Vancouver, BC: IJCAI.

Sidner, C. L. (1983). What the speaker means: The recognition of speakers' plans in discourse. International Journal of Computers and Mathematics, Special Issue in Computational Linguistics, 9(1), 71-82.

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

Stalnaker, R. C. (1974). Pragmatic presuppositions. In M. K. Munitz, & P. K. Unger (Eds.), Semantics and philosophy. New York: New York University Press.

Strawson, P. F. (1950). On referring. Mind, 59, 320-344.

Tierney, R. J., LaZansky, J., Raphael, T., & Cohen, P. R. (1983). Author's intentions and readers' interpretations. In R. J. Tierney, P. Anders, & J. N. Mitchell (Eds.), Understanding readers' understandings. Hillsdale, NJ: Erlbaum.

Webber, B. L. (1982). So what can we talk about now. In M. Brady (Ed.), Computational approaches to discourse. Cambridge, MA: MIT Press.

Wilensky, R. (1981). PAM. In R. C. Schank & C. K. Riesbeck (Eds.), Inside computer understanding: Five programs plus miniatures (pp. 136-179). Hillsdale, NJ: Erlbaum.

Woods, W. A. (1970). Transition network grammars for natural language analysis. CACM, 13(10), 591-606.