American Journal of Computational Linguistics *Microfiche 35 : 19*

Discourse Models and Language Comprehension

Bertram C. Bruce

Bolt Beranek and Newman Inc.
50 Moulton Street, Cambridge, Massachusetts 02138

ABSTRACT

Higher order structures such as "discourse" and "intention" must be included in any complete theory of language understanding. This paper compares two approaches to modeling discourse. The first centers on the concept of a "discourse grammar" which defines the set of likely (i.e. easily understood) discourse structures.

A second approach is a "demand processing" model in which utterances create demands on both the speaker and the hearer. Responses to these demands are based on their relative "importance". The length of time they have been around, and conditions attached to each demand. The flow of responses provides another level of explanation for the discourse structure.

These two approaches are discussed in terms of flexibility, efficiency, and of their role in a more complete theory of discourse understanding.

*

1. Introduction

As has been said many times, understanding anything a problem, an action, a word - demands some knowledge of the context in which it appears. Certainly this is true of language, where an utterance s meaning may depend upon who the speaker is, when he is talking, what has just been said, who the listeners are, what the purpose of the conversation is, and so on. It is reasonable to define language understanding as the process of applying contextual knowledge to a sound (or string of symbols) to produce a change in that context. Successful language understanding occurs whenever the changes in the hearer s context (model of the world) coincides with changes the speaker intended.

Of course, stating a problem in a different way does not solve it. Instead it suggests a series of subsidiary questions such as:

(1) What is a context? What does it look like? What are its components, its structural characteristics?

(2) How does a new utterance change an existing context? What is the assimilation process? What must be kept; what can de discarded?

(3) How does a model of changing context account for observed phenomena such as the ability to switch contexts, and to return later (but not too much later)?

(4) How does the domain of conversation influence the

structure of a "context"? Do different mechanisms operate when the subject matter is tightly constrained?

It may be quite a while before questions of this type can be answered fully. This paper is a discussion of some of the issues and of the characteristics required of a solution. To do this, we will examine two classes of discourse models which have been proposed. The first is a "discourse grammar" approach which attempts to define the set of likely sequences of utterances. the second is a "demand processor" approach which attempts to account for utterances as responses to internal demands. Before discussing the specifics of these models, a sample dialogue, from the much restricted world of man-machine communication, is presented and discussed in Section 2.

2. An Example from Travel Budget Management

The problem of discourse understanding is complex enough that a complete analysis is certainly premature. We will examine here a micro-discourse, restricted to two conversants, one of them a machine, and, a conversation restricted in scope to that of managing a travel budget. Imagine that a person responsible for a travel budget is trying to decide how many people to send to an upcoming conference. He has available a computer system which can answer questions about the budget, conferences, trips, and so on. In talking to such a system, the travel budget

manager is engaging in an admittedly simple dialogue, but one which nevertheless exhibits many interesting properties of discourse understanding. (The following dialogue is hypothetical; however a system of this type is being developed at BBN (Woods, et al (1974, 1975))). The (human) manager begins:

(1) WHEN IS THE NEXT ASSP CONFERENCE?

May 15-18, 1975

(2) IS THAT IN OUR CURRENT BUDGET YEAR?

Yes.

(3) HOW MUCH MONEY DO WE HAVE LEFT?

$316.41 remains in the budget.

(4) WHAT IS THE COST OF THE ASSP CONFERENCE?

The registration fee is $25. Would you like me to estimate the cost of a trip to the conference?

(5) YES. FOR THREE PEOPLE.

Assuming they travel by air ...

Assuming each person spends three days in Philadelphia ...

Assuming a per day expense of $25 ...

The cost would be $371.50.

(6) WILL THE BUDGET COVER THAT?

No, the deficit would be $55.09.

(7) ASSUME THEY GO BY BUS.

The cost would be $327.40. The budget deficit would be $10.99.

(8) OK. SCHEDULE A TRIP FOR THREE PEOPLE BY BUS TO THE ASCP MEETING.

Done. That leaves $ -10.99 in the budget.

Even in a simple dialogue like the one above we see important discourse phenomena. For example, the manager's input (4) is ambiguous to the system. "Cost" could mean "registration fee" or "cost of a trip to the conference". The system has to ask about this ambiguity, producing an embedded question-answer pair within the original question-answer pair (cf. Schegloff s (1972) "insertion sequences"). However, there is more than a Q-1, Q-2, A-2, A-1 pattern in (4) and (5). The system does not just say, "Did you mean fee cost or 'trip cost'?". Assuming that "fee cost" is a reasonable interpretation, and useful to know besides, it simply states that fact, and asks about the other interpretation, which would demand more computation. A discourse model should account for this apparent awareness of computational difficulty, which is exhibited in human conversation and between a human and our idealized machine above.

Another phenomenon worth noting in this dialogue is the variation in detail and precision among the utterances. Sentence (8) is fairly precise and complete. Since alternatives have been considered to the trip he has decided upon it is important.

stress those aspects of the trip - "three people", "by bus" - which have been in question. On the other hand, sentence (3) is clearly elliptical. This is all right since the question is merely exploratory. Furthermore, the previous question insures that "money ... left" refers to money in the current budget. An adequate discourse model should account as well for our apparent ability to accommodate for the speech channel capacity, to minimize transmission errors through the use of redundancy and stress, and in general to attempt to optimize the communication.

One way to account for these and related phenomena is to postulate a discourse grammar. The grammar might say that part of a dialogue is a "question-answer" pair, and that it may be recursive in the sense that question-answer pairs may be embedded within it. This approach is discussed in the next section. A contrasting approach is to say that each utterance produces "demands" in the heads of the listeners. Responses to these demands may take the form of subsequent utterances. This latter model is discussed in Section 4.

3. Grammar Models of Discourse

Upon reading a dialogue like the example in Section 2, most of us readily form an opinion about its structure. In any dialogue we see this kind of structure: one person is asking another to do something; two people are arguing about politics, or discussing a novel. There is almost always a structure higher

than the individual sentences. In the example of Section 2, the travel budget manager seems to be entering into a "schedule a trip" dialogue. His question about a future conference is one of the cues to a bundle of information known by both him and the system about scheduling trips. Such a bundle has been variously referred to as a "frame" (Minsky (1975), Winograd (1975)), a "script" (Abelson (1975), Schank and Abelson (1975)), a "theme" (Phillips (1975)), a "story schema" (Rumelhart (1975)), and a "social action paradigm" (Bruce (1975a, 1975b)).

The information associated with scheduling a trip includes facts about dates and times, about the budget, about travel, about conferences, and so on. It also includes "plans", that is, time ordered structures of beliefs about achieving "goals". In this case, the goal is scheduling a trip to a conference. (See also Bruce and Schmidt (1974), Schmidt (1975)). One such partially instantiated plan might be -

1. Find out to which budget the trip should belong.
2. Determine how much is in the budget (budget).
3. Figure the cost of the trip (tripcost).
4. Decide whether (budget - tripcost) is acceptable.
5. If acceptable, schedule the trip and stop.
6. If not acceptable, determine if trip can be modified to be cheaper.
    a. If modifiable, go to 3.
    b. If not modifiable, stop.

The steps (1 - 6) above are ordered, though nothing is said about their relative lengths. Also, there are variants on the plan where the order might be changed, e.g. step 3 might come before step 2 in some other plan. The structure of such a plan, coupled with the by now commonplace observation that a discourse is structured, leads to the natural idea of representing a discourse by a grammar. Such a grammar may be large; it may be probabilistic; it may apply in only limited domains. Nevertheless it does give some idea of what to expect in a dialogue and may play a central role in language comprehension.

A portion of the grammar for our example dialogue is shown in Figure 1. This is an Augmented Transition Network Network (ATN) in which the arcs may refer to other networks (PUSH arcs), may signify direct transitions to other states (JUMP arcs), or may signify conclusion of the path (POP arcs). For example, in addition to this "SCHEDULE" network there is an "ENTER" network wherein the manager describes a new trip to be entered and the system asks him questions to complete the description.

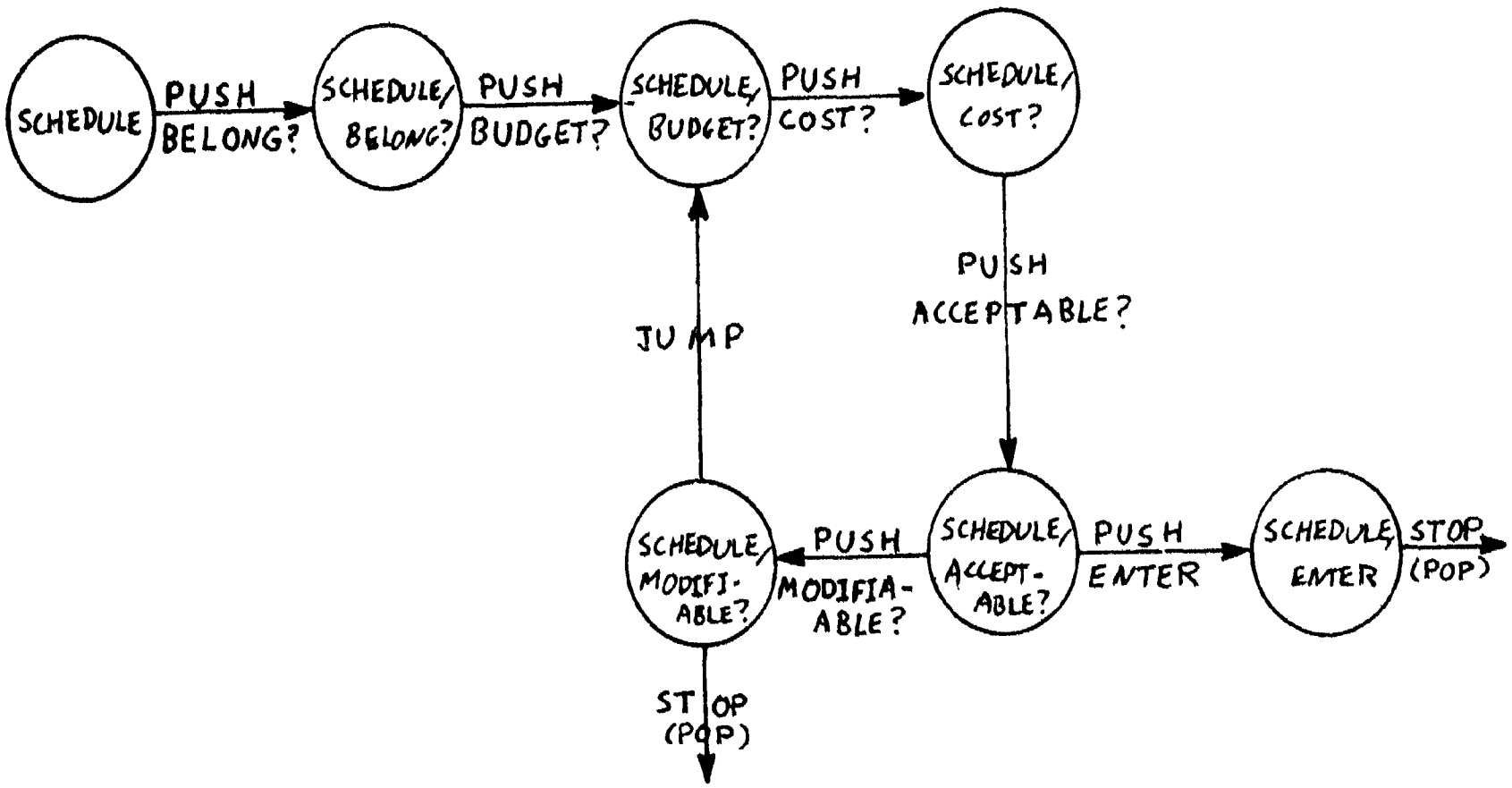

Fig. 1. ATN for scheduling a trip.

A discourse or dialogue grammar can be used with a modified ATN parser to "parse" a dialogue, generating both analyses of the current utterance and predictions about the one to come. In fact, one such modified parser and grammar has been implemented for the BBN speech system (Bruce(1975c), Woods, et al (1975)). For many dialogues, the grammar applies quite well, testing for the head verb in the utterance, the mood, and checking presuppositions of the action implied. When successful, it makes

corresponding predictions for application to the next utterance. Unfortunately, when the grammar fails it is not very good at recovering from its error.

Discourse grammars seem to be most effective in tightly constrained domains, more for instance in a discussion about how to cook a turkey, where there are specific subproblems to analyze, than in the travel budget management domain, and less still in a general question answering context. (Cf. Deutsch (1974, 1975)).

Lest it be thought that discourse parsing is just sentence parsing for "big sentences", I should emphasize some of the differences, differences which some would say preclude the use of terms like "grammar", "ATN", and "parsing". First, discourse parsing proceeds in a mode of partial parse, then output, then partial parse, etc. In other words, the goal is to derive information from the partial discourse which has occurred to suggest what may follow and to explicate the role of the current utterance. The parse is never completed, no structure is built. Since the entire discourse is not available to the parser (as the entire sentence is to a sentence parser), it is necessarily probabilistic. One can never know how the next utterance may alter the current interpretation of the trend of the dialogue. Another important difference is that PUSH's and POP's in the discourse grammar are "sloppy". That is, the participants in a dialogue may descend several levels ("Before you finish, let me

tell you about ...", "Before that ...") and never "pop" back up to the original level of the discourse. A discourse parser is faced with the peculiar phenomenon that a PUSH usually implies a POP but not always.

Some, but not all of these oddities of a discourse grammar are resolved by an approach which emphasizes internal models of the speaker and the listeners. This approach is discussed in the next section.

4. Demand Models of Discourse

One obvious characteristic of a discourse is that many processes may be occurring at once. A person cannot, nor does he wish to respond at one time to all unanswered questions, extend each unfinished line of thought, or deal with every inconsistency. While a grammar may predict the most likely action for a given point in a dialogue, it is not very good at suggesting alternatives out of the main line. There appears to be an additional mechanism of roughly the following form:

An event in a discourse (or prior to it) sets up a number of internal demands. Examples of such demands are to confirm what was said, explore its consequences, dispute it, answer it, etc. For any given event (such as an utterance) there may be none, one, or many demands created. A person's own action may place demands upon himself. If X asks a question of Y, then Y normally establishes an internal demand to answer the question. But X may

also establish a demand of the form, "check to see if the question has been answered". This latter demand may generate a later utterance such as, "Why haven t you answered me?".

Simple demand models already exist in a few systems. In general, they suggest that utterances are produced in response to conditions in the (internal model of the) environment rather than as units in a larger linguistic form. (See also Stansfield (1975)). It would be premature to argue that either a demand model or a grammar model is sufficient by itself. Instead, what follows is simply a description of a demand model for the travel budget management domain mentioned above.

Internal demands on the travel budget system help to explain how one computation of a response can be pushed down, while a whole dialogue takes place to obtain missing information, and how a computation can spawn subsequent expectations or digressions. Associated with each demand is a priority, a pointer (purpose) to the demand which spawned this one (if any), and a time marker indicating how long the demand has been around. An active unanswered question is a typical demand with high priority. Demands of lower priority include such things as a notice by the system that the manager is over his budget. Such a notice might not be communicated until after direct questions had been answered. The fact that some questions cannot be answered without more information leads to the

User-makes-query

System-asks-question

User-clarifies

System-answers-query

kind of embedding which is typically represented in a discourse grammar by a PUSH to a "clarification" state.

Counter-demands are questions the system has explicitly or implicitly asked the user. While it should not hold on to these as long as it does to demands, nor expect too strongly that they will be met, the system can reasonably expect that most counter-demands will be resolved in some way. This is an additional influence on the discourse structure.

A demand model also includes a representation of the current topic, the active focus of attention in the dialogue. For the travel budget system, it could be the actual budget, a hypothetical budget, a particular trip, or a conference. The current topic is used as an anchor point for resolving references and deciding how much detail to give in responses. Again, this structure leads to certain modes of interaction. For example, if the manager says "Enter a trip," the system notes that the current topic has changed to an incompletely described trip. This results in demands that cause standard fill-in questions to be asked. If the manager wants to complete the trip description later, then the completion of the trip description becomes a low

priority demand.

5. Synthesis?

Discourse has been an object of study for many both in and out of the field of computational linguistics. Especially worth noting is the work of sociolinguists such as Labov (1972), Sacks, Schegloff, and Jefferson (1975), and Schegloff (1972). Linguists (e.g. Grimes), sociologists (e.g. Goffman (1971)), and philosophers (e.g. Austin (1962), Searle (1969)) have important direct or related contributions. I certainly can't presume in this short paper to give the definitive solution to all the problems revolving around the discourse question. What I have tried to do is to emphasize a distinction in approach between looking at a discourse as a linguistic whole with subparts being individual utterances, and as a side effect of responses to task demands.

Both approaches are useful in exemplifying ways in which the otherwise hazy area of discourse might be modeled. The grammar approach makes the strongest statement about actual discourse structure and can best be used where the structure is well known or can be tightly constrained, e.g. in generating a discourse or in a man-machine system where the computer imposes control on the dialogue. A grammar and a discourse parser can be very efficient in such situations. When the dialogue is less predictable the (more bottom-up) demand processing approach may be more resistant

to "surprises" in the dialogue.

The ultimate discourse model probably contains aspects of both goal-directed grammars and of localized responses to demands. What should be particularly interesting to see is how characteristics of the model are affected by the type of discourse, human-machine v. human-human, problem-oriented v. information-exchanging, or new domain v. old.

REFERENCES

Abelson, Robert. "Concepts for Representing Mundane Reality in Plans". In Representation and Understanding: Studies in Cognitive Science (Ed: D. Bobrow and A. Collins), Academic Press, New York, 1975.

Austin, J. L. How to Do Things with Words. Clarendon Press, Oxford, 1962.

Bruce, Bertram. "Belief Systems and Language Understanding". BBN Report No. 2973, 1975a.

-----. "Generation as a Social Action". In Theoretical Issues in Natural Language Processing (Ed: B. L. Nash-Webber and R. C. Schank), ACL, 1975b.

-----. "Pragmatics in Speech Understanding". Proc. 4th IJCAI, Tbilisi, 1975c.

----- and C. F. Schmidt. "Episode Understanding and Belief Guided Parsing". Presented at 12th ACL Meeting, Amherst, 1974. (Also Rutgers Computer Science Dept. Report

CBM-TR-32).

Deutsch, Barbara G. "The Structure of Task Oriented Dialogues". Contributed Papers, IEEE Symposium on Speech Recognition, CMU, Pittsburgh, 1974.

-----. "Discourse Analysis and Pragmatics". In Speech Understanding Research (D. Walker, W. Paxton, J. Robinson, G. Hendrix, B. Deutsch, and A. Robinson), Annual Technical Report, SRI, 1975.

Goffman, Erving. Relations in Public. Basic Books, New York, 1971.

Grimes, Joseph. The Thread of Discourse. Mouton, Paris, in press.

Labov, William. "Rules for Ritual Insults". In Studies in Social Interaction (Ed: David Sudnow), The Free Press (Macmillan), 1972.

Minsky, Marvin. "A Framework for the Representation of Knowledge". In The Psychology of Computer Vision (Ed: P. Winston), 1975.

Phillips, Brian. Topic Analysis. Ph. D. Thesis, SUNY Buffalo, 1975.

Rumelhart, David. "Notes on a Schema for Stories". In Representation and Understanding: Studies in Cognitive Science (Ed: D. Bobrow and A. Collins), Academic Press, New York, 1975.

Sacks, Harvey, Emanuel Schegloff and Gail Jefferson. "A Simplest Systematics for the Organization of Turn-Taking for

Conversations". Semiotica, 1974.

Schank, Roger and Robert Abelson. "Scripts Plans and Knowledge". Proc. 4th IJCAI, Tbilisi, 1975.

Schegloff, Emanuel A. "Notes on a Conversational Practice: Formulating Place". In Studies in Social Interaction (Ed: David Sudnow), The Free Press (Macmillan), 1972.

Schmidt, Charles F. "Understanding Human Action: Recognizing the Motives and Plans of Other Persons". Carnegie Symposium on Cognition: Cognition and Social Behavior, CMU, Pittsburgh, 1975.

Searle, J. R. Speech Acts. Cambridge University Press, London, 1969.

Stansfield, James L. Programming a Dialogue Teaching Situation. Ph. D. Thesis, U. of Edinburgh, 1974.

Winograd, Terry. "Frame Representations and the Declarative-Procedural Controversy". In Representation and Understanding: Studies in Cognitive Science (Ed: D. Bobrow and A. Collins), Academic Press, New York, 1975.

Woods, William, M. Bates, B. Bruce, J. Colarusso, C. Cook, L. Gould, D. Grabel, J. Makhoul, B. Nash-Webber, R. Schwartz, J. Wolf. "Natural Communication with Computers, Final Report - Vol. I, Speech Understanding Research at BBN". BBN Report No. 2976, 1974.

Woods, William A., R. Schwartz, C. Cook, J. Klovstad, L. Bates, B. Nash-Webber, B. Bruce, J. Makhoul. "Speech Understanding Systems: QTPR 3". BBN Report No. 3115, 1975.