PHACMATICS TH SPEECH UNDERSTANDINO

Bertram Bruce
Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Mass. 02138

Abstract

When a person speaks he is using words to achieve a goal, whether that be to gain information, to threaten, to promise, or to reassure. Recognition of that goal is an essential part of speech understanding, both in determining what was said and in deciding what was meant. The term "pragmatics" is used here to mean the procedure which applies knowledge about the speaker, the previous dialogue, and the domain of discourse to interpret utterances and respond appropriately. This procedure invokes definitions of intents (speech acts) and modes of interaction to recognize the goal of a speaker and consequently to understand his utterance.

1• Introduction

In order to understand speech it is necessary to use knowledge of the speaker and his purpose. In the BBN speech understanding system (known as SPEECHLIS) this knowledqe is incorporated in the Pragmatics component. Pragmatics interacts with other components such as Syntax, Semantics, and Control to build an interpretation of an utterance (see Woods, et al [8]). This paper discusses the structure and function of Pragmatics in speech understanding.

The pragmatics component of SPEECHLIS is a process which applies various facts about the speaker, the previous dialogue, and the domain of discourse to interpret utterances and respond appropriately. For example, an early version of the system operates as a question answerer for the domain of lunar geology. In that domain one is usually not concerned with the particular scientists who investigated the samples, but rather with the samples themselves. Thus verbs which allow agent deletion in the passive voice are usually expressed that way, rather than in the active voice. One says "Which new minerals were discovered in the lunar breccias?" and not "Which new minerals did the investigators discover in the lunar breccias?"

In the current travel budget management system we have recognized similar effects of pragmatics in simulated dialogues. For instance, following a supposition, a speaker tyoically asks a question. This question usually concerns future events and is related to the content of the supposition. Another example is that use of the verb "cancel" implies that the speaker believes that the object of "cancel" has previously been entered into the system data base.

There is no doubt that pragmatics information can be helpful in certain cases. We are currently exploring the use of a user/task model to generalize and structure the pragmatics rules we have discovered. This model provides a focus on a central issue in pragmatics, the recognition of the speaker's purpose.

A person uses a speech system to accomplish some purpose, whether that be to obtain information, to gain assistance in planning and decision making, or to control some process. His purpose is reflected in both the vocabulary and syntax of the language and in the interpretations which are assigned to utterances. An at least implicit recognition of the purpose behind an utterance is necessary for complete speech understanding. Related work by Deutsch [41 has also shown the importance of a model of the user's purpose in understanding spoken dialogue.

We have formulated a set of structures which can be used to represent the concept of intention in language use. These structures are based on analyses of simulated dialogues with the travel system, and on general considerations of what it means to communicate with a purpose. Discussion of the general considerations can be found in Bruce [2], Bruce and Schmidt [3], and Schmidt [5]. This paper is primarily concerned with the more specific application of user and task knowledge to the travel budget SDeech understanding system.

In section 2 we consider a set of intents derived from examination of simulated uses of the travel budget system. Section 3 covers the organization of these intents into modes of interaction. Section 4 is a 3Tscussion of organization and implementation issues.

2. Intention in Speech

We can describe actions at many different levels. For example, the action -

> Susan said to Mary, "I hope you come tonight".

could be described as -

> Susan was facing Mary and uttered the sounds typically associated with the sentence, "I hope you come tonight".

On the other hand, a purpose oriented description might be -

> Susan urged Mary to come.

or in another context -

> Susan threatened Mary about coming.

The ability to generate purpose oriented descriptions for utterances is crucial for speech understanding because the speech act is always part of some plan directed towards a goal. General speech communication relies on the ability of the communicators to maintain an awareness of the other's purposes. In general an utterance can express any of several intents and an intent can be realized by"many different utterances.

Before describing some intents we should sketch the context in which they are used. Imagine an observer of, or a participant in a dialogue. When he hears a sentence he immediately makes some interpretation. This interpretation may simply be that the speaker has chosen to inform his listeners that X, where X is some proposition.

Whatever interpretation he makes, a rational observer commits himself to various beliefs. For example, the interpretation, "the speaker informed the hearer of X," commits him to the belief that the speaker believes X and that the hearer was not aware of X at the time of the utterance. Different beliefs correspond to different interpretations, e.g. "the speaker lied to the hearer about X" entails the belief that the speaker does not believe X. Beliefs of this kind are called preconditions since they refer to conditions prioF to the utterance. There are also outcome conditions which refer to conditions after the utterance. For example, at least one sense of "inform" has the outcome condition that the hearer is aware of x. Both preconditions and outcome conditions are subject to later verification. if the observer later concludes that one or more of the conditions does not hold then he may change his interpretation of the utterance .

Each condition can be expressed as a formula consisting of a predicate with its arguments. Typically the oredicates are such things as "believe" and "want", and the arguments (or cases [1]) are such things as the speaker, the hearer, the time, and embedded propositions. (An embedded proposition might be the "X" in "the speaker believes X".)

Realization of an intent is needed to determine what was meant from what was said, and, in the case of speech, even to determine what was said. For example, consider the following "word lattice" (Woods, et al [8]) for a sentence spoken to our travelbudget management system:

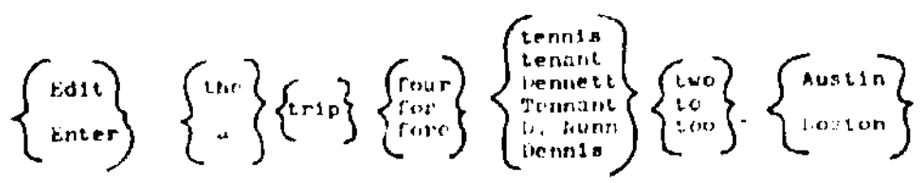

Figure 1. A Word Lattice

The word lattice represents several possible strings of words which might be present in the utterance. Often the uncertainty can be resolved by a combination of acoustic-phonetic, syntactic, and semantic information about the words and the usual ways they fit together. Sometimes, though, a correct interpretation requires the use of knowledge about intents.

In this example our knowledge of what it means to "edit" tells us that the object of the sentence should be a previouslv stored item, or, more precisely, that interpreting the first portion of the sentence as "edit" commits us to the belief that the speaker believes that the latter portion of the sentence describes a data item which exists in the data base. Thus the "edit" interpretation forces (via its preconditions) constraints on the sets of possible names and cities. Similarly,

the "enter" interpretation requires us to believe such things as: The shaker believes that a "trip" is something which can be added to the data base, and at least one of tennis, tenant, Dennett, Tennant, 0. Nunn, and Dennis is a oerson whose trips are recorded by the system. The definitions of EDIT and ADD NEW STRUCTURED ITEM *TO* DATA BASE (see below) are compared with resDect to the dialogue history and our model of the user to select the best interpretation.

A full definition of an intent consists of its case structure, preconditions, outcome conditions, and *a* set of pointers to typical expressions of the intent in language. In the examples given here the case structures are all the same. There *is* an agent (the speaker), a recipient (the hearer), a time of utterance, and a proposition. We will symbolize these as A, R, T, and X respectively. For the sake of readability the preconditions and outcome conditions for each intent are expressed in English. It is, however, possible to formalize these expressions (see Bruc [2], Bruce and Schmidt [3]). The following are two of the intents found in the travel budget management domain. (Square brackets indicate conditions believed by the agent).

ADD NEW STRUCTURED ITEM TO DATA BASE

(A "structured item" is a concept such as "trip" which is known to have specific components such as cost, travelers, destination, etc.)

Preconditions:
PI. A is useF : R is system : X is a structured item
P2. [X is true]
P3. (X was not added before]
P4• (There is a standard set of questions based on the structure of X]
P5. [X is the kind of data item appropriate to the data base]

Outcome conditions:

01. X is added to data base
02. R knows that A added X

Instances:

Add a trip for Bill to Berkeley-
Insert a new budget item.

EDIT

(The user wants to change the data base of the system in some way.)

Preconditions:

PI. A is user : R is system : X is command to change an item in the data base
P2. IX refers to previously stored item]
?3. [The effect of X is consistent with data base]

Outcome conditions:

01. R applies X to data base if its effect is not inconsistent

Instances:

Change the registration fee to $75.
Add Bonnie to the list of people going to Chicago.

Other intents have also been defined and are used in characterizing the modes of interaction. These include ASK STANDARD QUESTION, REPLY TO STANDARD QUESTION, ASK AGAIN, CONFIRM DATA ITEM, CONFIRM STRUCTURED ITEM, POINT OUT A CONTRADICTION, *REASSURE*, STRONG EDIT, QUESTION, CLARIFY, QUERY, INFORM, PRESENT A SUPPOSITION, NAME SUPPOSITION, SUSPEND, TEST, and RESPOND. (See [8] for more definitions.)

There are two preconditions applying to all intents which are not listed explicitly in the examples above. First, the agent of the intent must intend to express that intent, i.e. he must be sincere. Regardless of the utterance, a given intent is realized only when the utterance is deliberately chosen (and not said as a joke, under duress, in a play, etc.). Second, the agent must believe that the recipient of the intent believes that the agent is sincere. If he does not then he has an obligation to supply additional information. Together these conditions imply what Searle [5,6] calls, "normal input/output conditions" for the speech act. Since one of the participants in the dialogues we are describing is SPEECHLIS itself, such notions as "sincerity" and "belief in sincerity" must be built into the user model and the system's programmed interactions.

There is also a general outcome condition which says that if an observer (speaker, hearer, or third party)

believes that an intent is expressed, then he may compute any consequence of the preconditions or outcome conditions. For example, a sincere "promise" has a precondition that the agent believes he can do the action promised. An observer of the promise might infer that the agent also believes that he has all the appropriate equipment and skills to do the action.

In addition to the preconditions associated with each intent, there are assumptions which can be made about all communication within the travel budget world. These latter assumptions are essentially global presuppositions about utterances as opposed to the local presuppositions expressed as preconditions. One such global presupposition is that the travel budget system is helpful. While it may fail to assist the user in a particular case, its overall design is to help the user, not hinder, or ignore him. Another presupposition is that the user is bona fide, i.e. that he has the right to use the system and will not deliberately enter false information, nor attempt to foil the system. Certainly a system might not make these presuppositions and its actions would differ accordingly. However, the system's performance will benefit to the extent that global rules can be established and used.

3. Modes of Interaction

A direct consequence of the recognition of an utterance's intent is an expectation concerning the possible utterances which may reasonably follow. For example, if the travel budget system points out a contradiction in the data base then it can expect the user to respond with an utterance which realizes one of a few intents. He may rectify the data base, may assert that the contradiction is of no consequence, or may begin making tests of the data base to ascertain the reason for the discrepancy. Completely ignoring the system's comment is also a possibility, but it is not likely, especially in light of the global presuppositions that the system is trying to help and the user wants the system to be effective. An organization of intents into a larger structure expressing expectations is callcd a mode of interaction. Modes consist of (expectation} links between intents and (possibly) other modes.

Each mode is defined by a header and a body. The header determines whether or not the mode body is applicable in a given situation. In addition, it binds variables within the mode body to entities in the situation. The mode body is a graph in which the nodes are either intents or other modes, and the arcs are directed links between nodes, labelled by likelihood. In general, there is a small number (often one) of starting nodes in the mode body. The header requires that the preconditions for the starting mode intents be met. It may also impose other more general constraints, e.g. that the mode occurs only at the beginning of a session.

Examples of some modes of interaction which occur in the travel budget management domain are the following:

ADD - the user is attempting to add new information to the data base.

CONFLICT - the system has pointed out a contradiction between some statement or assumption made by the user and its own information. The user should then respond to it.

EDIT - the user is attempting to change some information already in the data base.

QUERY - the user is attempting to get information from the system.

QUESTION/CLARIFY - the system does not understand the user 's utterance and asks for clarification.

SUPPOSITION - the user is making hypothetical changes to the data base to see where they will lead.

TEST - the user is attempting to ascertain that the system's knowledge about some past or future event conforms with his own.

The body for the mode ADD is shown in Figure 2. Its header is omitted since it simply checks the preconditions on the starting node. In the figure, an oval means the user is talking to the system; a rectangle means the system is talking to the user; a pointed box represents a call to another mode. Likelihood ratings are not given, but it can be assumed that the arcs given represent the most likely, if not the only possible transitions.

The user enters ADD mode when he wishes to add some item with an internal structure to the data base. For example, a trip has an agent, a destination, a time, a purpose, and other components. When the user states his intention to add such an item the system can be expected to present him with a set of usually standardized questions in order to fill out the structure of the item, e.i.

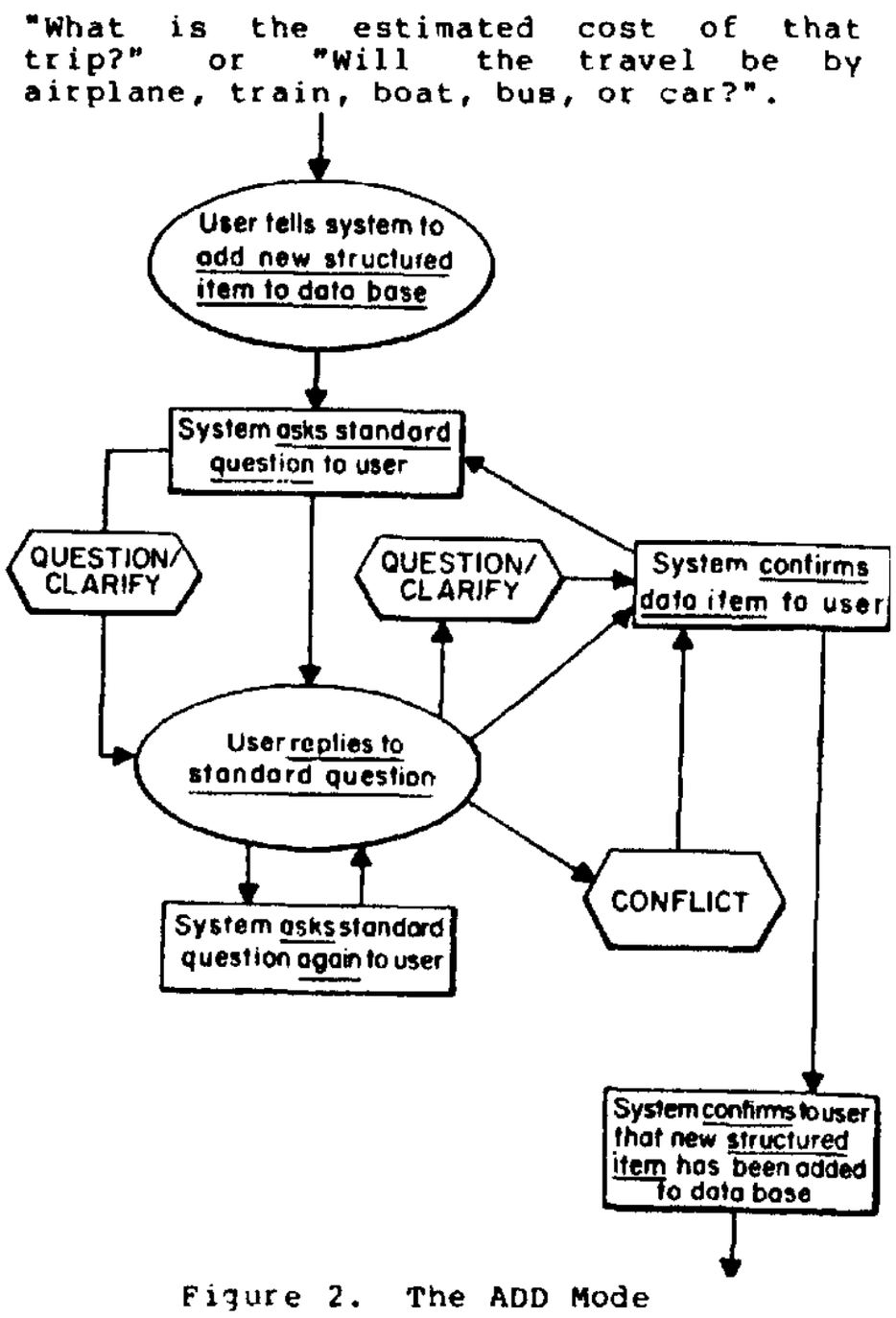

Figure 2. The ADD Mode

Following each question by the system, the user may give a reply or enter the QUESTION/CLARIFY mode. The latter occurs if the user does not understand the system's question. Once the question is clarified the user should reply. The system then responds to the user's reply in one of four ways:

(1) It may confirm the addition of the given information to the data base. Then, if there are further questions the next one is asked. Otherwise, the system confirms the addition of the entire item and a new mode may be initiated.

(2) It may enter QUESTION/CLARIFY if it does not understand the user's reply. Following the clarification things proceed as in (1).

(3) The system may ask the question again if the answer is insufficient or inappropriate. This is not just clarification. The system believes that it has understood the user's utterance correctly, but that the user JL not understand the standard question. Thus the system rewords the question and asks it again.

4) If the user's reply contradicts previously stored information then the CONFLICT mode is entered. When the conflict is resolved, the system confirms as in (1).

It may be worthwhile to contrast the ADD mode with the EDIT mode. A user enters EDIT mode (Figure 3) with the intention of changing some information in the data base. As a result of his utterance,

1) The system may ask for clarification. That is, the mode may switch to QUESTION/CLARIFY. Upon successful clarification, things proceed as in (3) below.

2) The system may point out a contradiction. For example, the user may have a mistaken assumption about what is actually in the data base. Here the mode switches to CONFLICT.

3) The system may make the requested change and confirm to the user that it has made it. At this point, the user may want to make another change, remaining in EDIT mode, or 1eave that mode for another one.

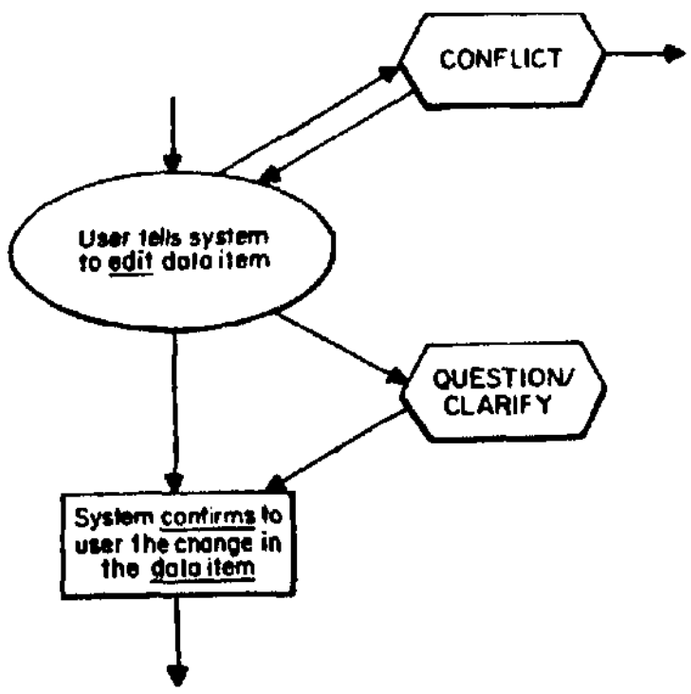

Figure 3. The EDIT Mode

Mones ot interaction represent additional pragmatics information which can be used in understanding spoken dialogue. Together with intents they facilitate interpretation of otherwise

ambiguous utterances. Returning to out example in Section 2 we can see that an awareness of the current mode is a useful complement to the previously mentioned intentional knowledge. The beginning of the utterance could be either "Edit the" or "Enter a". It is not possible to resolve this ambiguity using the intent definitions alone. Knowledge based on modes might provide enough information, together with word match scores, to select the correct interpretation. For example:

(1) If the mode is CONFLICT then the "edit" interpretation is much more likely.

(2) If a mode has just been completed then "enter a" has its greatest likelihood , particularly if there has been a succession of ADD modes.

(3) If the mode is QUERY then the content of the rest of the sentence is likely to be related to recent auestions. This may provide the basis for rejection or confirmation of one of the intents.

4. Organization and Implementation Issues

A prototype pragmatics component has been added to the BBN SPEECHLIS system. This component (called "Pragmatics") interacts with other components such as Syntax and Semantics, and also with the Control component.

In order to recognize intents and modes it is necessary to have a model of the speaker. The model includes such things as the speaker's presumed knowledge, his previous purposes, idiosyncratic pronunciation, vocabulary or syntax, and his role or position. Such a user model must be subject to change on the basis~of interactions with the system.

The combination of a task model, expressed through modes and intents, and i user model can be a powerful aid to speech understanding. It can help first by providing expectation- which structure the space of possibilities for utterances. For example, if the user says, "Suppose we cancel the upcoming Pittsburgh trip", the system can expect a question to follow, either immediately or after further suppositions. The question should be related to the supposition and should refer to future possibilities. The fact that expectations are never certain does not invalidate their importance in suggesting possibilities. Thus Pragmatics can use the user/task model to indicate likely classes of morphemes (e.g., future tense indicators following a supposition), or structures for the next utterance.

Secondly, Pragmatics can use its user/task model to express preferences for certain readings over other ones. People certainly take into account what they suppose is the speaker's Durpose when they hear an utterance. For example, when a gas station attendant says, "Fill'er up?", it is one's understanding of his purpose which selects "Fill'er up?" over "Phil Rupp?".

Thirdly, Pragmatics can ensure that the actions of the system are appropriate to the goals of the user. If a user of the travel budget system were to say, "The cost of a flight to L.A. is two hundred dollars", he could be asking a question, attempting to insert new information into the system, or deliberately trying to change information in the data base. The system's response might be either:

(1) No, it's $250.
(2) My data base has $250 as the cost of a trip to L.A. Is that in error?
or (3) OK.

depending on what it discerns to be the user's purpose.

The functions of Pragmatics suggest that it needs to communicate with SPEECHLIS Control and, perhaps, directly with Syntax, Semantics, and the factual data base. We are currently exploring the establishment of these communications channels.

Pragmatics itself requires a control structure which allows access to varied sets of data. A preliminary design is shown in Figure 4. It is essentially a single coordinating process called the Pragmatics Control plus a set of knowledge sources and a context representation. The knowledge sources include the definitions of intents and modes. The context representation consists of the mode status (the current mode and state within the mode), a representation of the facts of the dialogue (i.e. the system s knowledge), the system's representation of the user's facts (the user's knowledge), and a dialogue history, which contains such things as information about likely ways of referring to objects. This latter element is esoecially important for problems of anaphora and ellipsis.

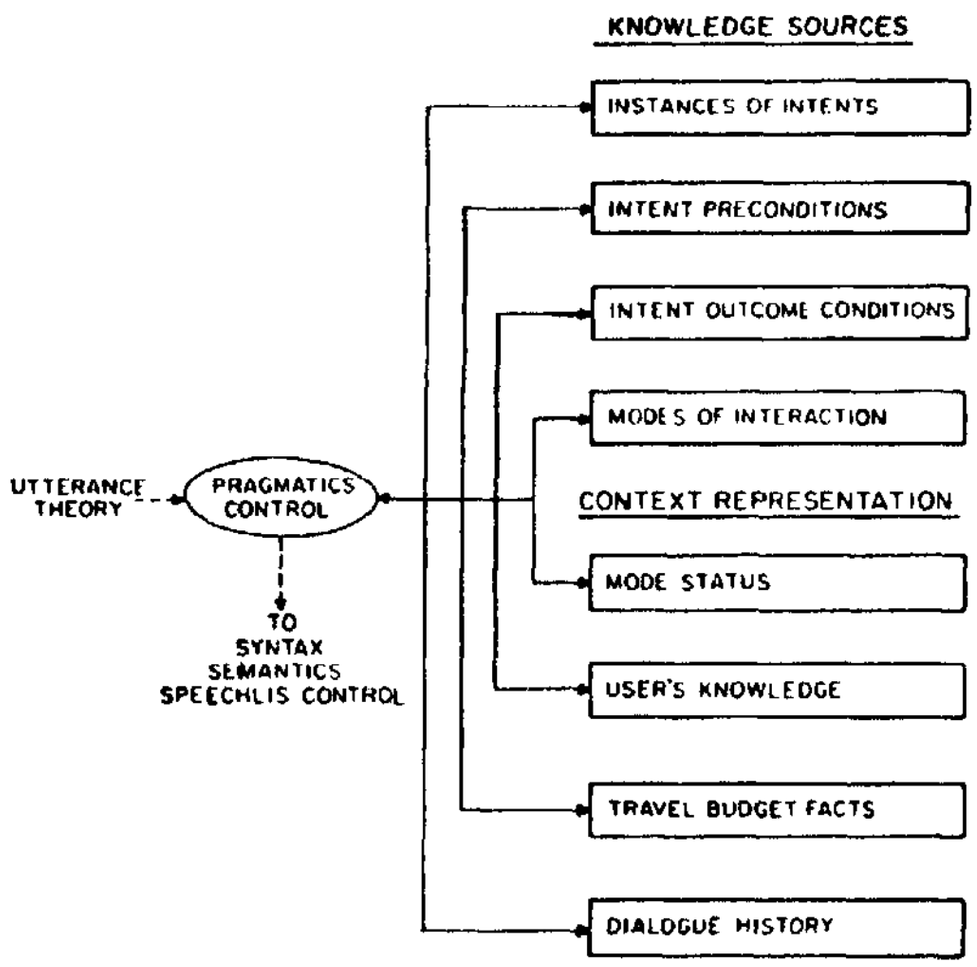

Figure 4. The Pragmatics Component

Pragmatics can be called whenever an interpretation of an utterance (or portion of an utterance) is to be evaluated or responded to by the system. Pragmatics Control first looks at instances of intents. This is a knowledge source which defines a maDping of words and phrases into intents. Using simple pattern matching rules various intents may be suggested. These suggestions can be supported or rejected by consideration of the mode status. Pragmatics Control looks there to determine what, if any, intents can be expected in the current context. Given this filterinq of the possible intents, Pragmatics Control can then begin to select the probable intent using the intent preconditions. These may require significant computations on both the travel budget facts and the user's knowledge. Once an intent is selected, Pragmatics Control processes its outcome conditions, changing the context representation as needed. The output of Pragmatics control can be a message to SPEECHLIS Control, such as a verification of the utterance interpretation, a request for actions on the data base, or notes to Semantics or Syntax concerning the subsequent utterance, i.e. words or classes of words to look for.

The current work on Dragmatics within the speech system represents a compromise between the ideal of a general pragmatics component which truly understands human motivations and the reality of a working system. Further development of the user/task model outlined above will provide a framework in which otherwise ad hoc pragmatics principles can be implemented.

References

1. Bruce, B., "Case Systems for Natural Language," BBN Report No. 3010, Bolt Beranek and Newman Inc., Cambridge, Mass., (1975).

2. Bruce, B. "Belief Systems and Language Understanding," BBN Report No. 2973, Bolt Beranek and Newman Inc., Cambridge, Ma. (1975).

3. Bruce, B. and C.F. Schmidt, "Episode Understanding and Belief Guided Parsing," Computer Science Department, Rutgers University, NIH ReDort, CBM-TR-32 (1974) .

4. Deutsch, Barbara G., "The Structure of Task Oriented Dialogs," IEEE Symposium on Speech Recognition, (1974) .

5. Schmidt, C.F., "Recognizing Plans and Purposes," Computer Science Department, Rutgers University, NIH Report, CBM-TR-34 (1974).

6. Searle, J.R., Speech Acts; An Essay in the Philosophy of"Language, Cambridge University Press, Cambridge, England (1969) .

7. Searle, J.R., (Ed.) The Philosophy of Language, Oxford University Press

8 Woods, W.A., M. Bates, B. Bruce, J. Colarusso, C. Cook, L. Gould, D. Grabel, J. Makhoul, B. Nash-Webber, R. Schwartz and .1. Wolf, "Natural Communications with Computers Pinal Report - Volume I Speech Understanding Research at BBN," BBN Report No. 2976, Bolt Beranek and Newman Inc., Cambridge, Ma. (1974) .