Pragmatic sensitivity in NL interfaces and the structure of conversation

Tom Wachtel

Scicon Ltd, London
and
Research Unit for Information Science & AI, Hamburg University

1. Introduction

The work reported here is being conducted as part of the LOKI project (ESPRIT Project 107, "A logic oriented approach to knowledge and data bases supporting natural user interaction"). The goal of the NL part of the project is to build a pragmatically sensitive natural language interface to a knowledge base. By "pragmatically sensitive", we mean that the system should not only produce well-formed coherent and cohesive language (a minimum requirement of any NL system designed to handle discourse), but should also be sensitive to those aspects of user behaviour that humans are sensitive to over and above simply providing a good response, including producing output that is appropriately decorated with those minor and semantically inconsequential elements of language that make the difference between natural language and natural natural language.

This paper concentrates on the representation of the structure of conversation in our system. We will first outline the representation we use for dialogue moves, and then outline the nature of the definition of well-formed dialogue that we are operating with. Finally, we will note a few extensions to the representation mechanism.

2. The MOVE frame

We are assuming a seven-slot frame for the representation of moves within a dialogue, with the following slots: MOVE, USER, AGENT, TURN, ACT, BASE, CONT. Every move in a conversation is represented by such a frame. The MOVE slot uniquely identifies a particular move in the conversation by an arbitrary integer. The USER slot identifies the current user of the system. The AGENT slot specifies whether it is a user move or a system move. The TURN slot has the value OPEN n or CLOSE n, where n is a number that refers to a particular exchange. Typically, a move with a value OPEN n for the TURN slot might be a request, and one with a CLOSE n value the corresponding response.

The ACT slot specifies what act is performed by the utterance. This will be either a speech act, or the value ACTION, since not all moves need be speech acts. The range of speech acts that the system will have to recognise or produce is clearly smaller than that which occurs in conversations between humans. Furthermore, certain speech acts will be of primary importance given the domain of application of the system, namely, as a front end to an expert system. We have therefore produced an initial hierarchy of potentially relevant speech acts (Wachtel 1985a), where the major classification is into requests, assertions and commentaries. Some of these are referred to below. Many of the speech acts we use go one level below what is traditionally viewed as a speech act (i.e. in the sense of Austin (1962), Searle (1969), etc.) and may be compared with distinctions that McKeown (1985: 9ff.), for example, discusses under the category of "rhetorical predicates", though they are by no means the same. The only speech acts discussed below are referred to by the following abbreviations:

REQACT	request-for-action
REQAFF	request-for-affirmation
SUPPAFF	supply-of-affirmation
REQCONST	request-for-constant
SUPPCONST	supply-of-constant
REQCONF	request-for-confirmation
SUPPCONF	supply-of-confirmation

The BASE slot specifies the current topic, in the very restricted sense of a pointer to the node in the semantic network that corresponds to the object what the current exchange is about. This simplistic view of topic is adopted here as a first step only, and serves to illustrate the points discussed below.

The CONT slot specifies the semantic representation of the utterance, and we envisage using the same representation for semantics and for actions, so that all possible ACT types can be represented uniformly in the CONT slot. In particular, we will define an exchange as a pair of utterances with the same value for the CONT slot, for the time being. This is of course too strict. Other functions specifying "local coherence" in the sense of Hobbs (1982: 227) are also relevant here. The particular illocutionary force of an utterance will be a function of the value of the ACT slot and the CONT slot. Subdialogues that are not in the mainstream of the conversation will be identified by particular relationship of values for the TURN slot between adjacent moves, enhanced by the values of the ACT slots for the moves.

Some examples of the use of this frame to represent sequences of utterances in conversations can be found in Wachtel (1985b, 1985c), including its use to identify shifts of topic, subdialogues and relevance, as well as the contextual disambiguation of speech acts, which is the main topic of these working papers.

3. The structure of conversation

We assume that it is possible to define the structure of a possible conversation by rule. Actual human-human conversations may defy such analysis, as illustrated by, for example, the work of Sachs, Schegloff & Jefferson (1974). However, the possible ways in which the conversations we are dealing with may go are severely limited by three factors: (a) this is an interface to an expert system (or some similarly specific software), which delimits the possible range of topics; (b) one of the participants in the dialogue is a machine, which means that it will not suddenly want to indicate that, for example, Albert's niece is a friend from school, but this fact has no bearing on the supportive things being said about her; and (c) the other participant knows that his interlocutor is a machine, and will behave accordingly. Therefore, what we need to model is not a typically natural open human conversation, but a restricted type of conversation that also occurs between humans in certain well-circumscribed contexts. For example, a conversation between a would-be passenger and a ticket clerk at a railway station is closer to what we need to model, and in such cases it is possible to define what is or is not a well-formed conversation by rules of an abstract nature that may well be inadequate for other naturally occurring converations.

We therefore propose three rules that define the notion of well-formed conversation in the present context, making the following assumptions. The structure of a conversation can be represented as a tree structure. The wellformedness of such trees can be defined by rewrite rules. The maximal number of levels of embedding in such trees is six (see below). In particular, subdialogues can be embedded within dialogues, but there can be no embedding within subdialogues. The last restriction conflicts with what people do. It is one of the restrictions we consider necessary, and which can be handled in such a way that the user will not notice that any such restriction exists.

We assume that the following four categories are sufficient for the representation of the structure of conversation. The symbols used serve as mnemonics for their approximate counterparts in English, but they should not be strictly equated with them: CONV (conversation), DIAL (dialogue), EXCH (exchange) and MOVE (as discussed above).

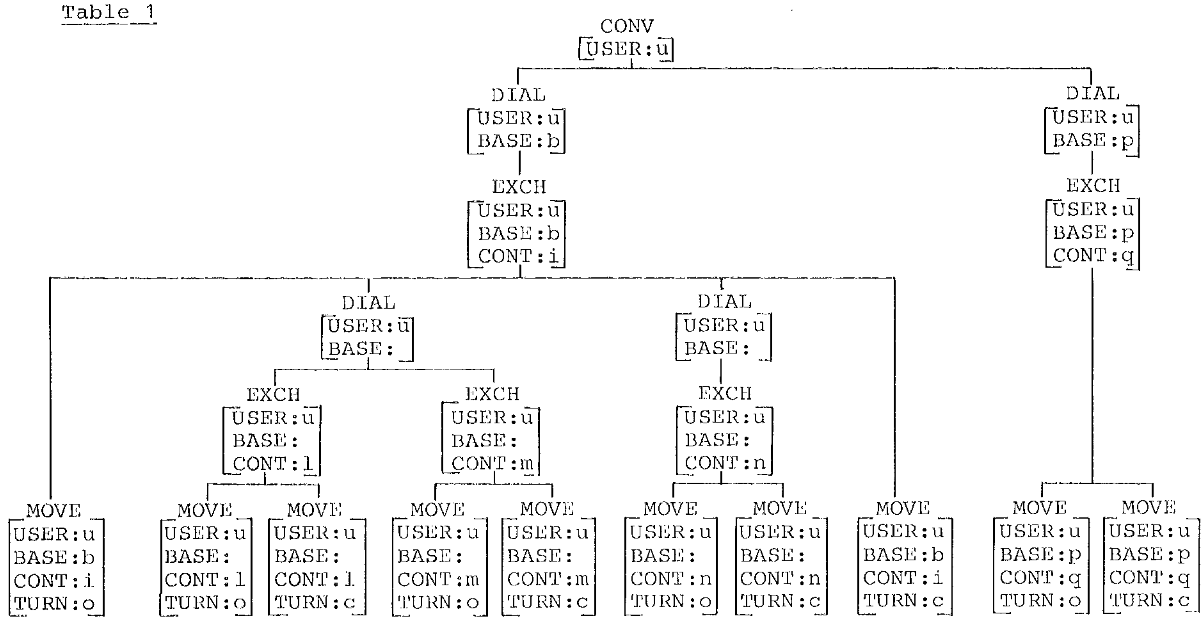
To formulate in informal terms the general style and atmosphere of the rules that we will propose more formally below, let us say that a CONV may consist of one or more DIALs, a DIAL may consist of one or more EXCHs, and an EXCH consists of two MOVEs, with each of these MOVEs followed by an optional DIAL.

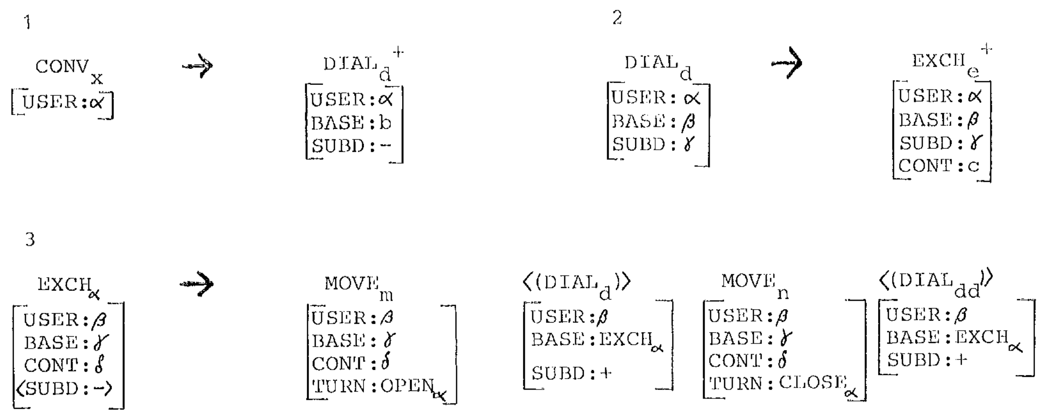
A major point about conversations that must be handled in a grammar of this type is the fact that although MOVEs are the only terminal nodes, and are therefore the nodes that correspond to the utterances that are actually produced, with all other nodes representing more abstract elements, certain features of conversation need to be associated with these abstract nodes. For example, although each MOVE is specified for who the current user of the system is and each MOVE also has a particular topic, as discussed above, these notions properly belong to more abstract levels of conversational structure. Who the user is can be defined at the CONV level (i.e. we define a CONV as a conversation with one user). The topic of an utterance can be defined at the DIAL level (i.e. a CONV can consist of one or more dialogues, each on a single topic). Furthermore, a DIAL can consist of one or more EXCHs, and it is at this point that the content of the utterances that form part of that EXCH is defined.

Let us now be more precise. We assume that some of the slots mentioned above in the MOVE frame are represented as features on the nodes in the trees representing the structure of the conversation of which the moves described by the MOVE frames are part. This association of features with nodes, plus the assumption that all features trickle down, with a few exceptions discussed below, provides for trees of the general form shown in Table 1. The lower case letters are constants. Note that the values of the BASE feature on the subdialogue nodes have not been specified. We return to this point below. Table 1 represents a goal: the sort of structures we want the rules to produce. The following three rules generate trees of this type. Kleene Plus notation is used.

The notation should be interpreted as follows. Roman letters as feature values are constants. Greek letters are variables ranging over possible feature values, and are to be interpreted consistently within a rule, but not necessarily between rules. They are used to ensure that the correct

Table 1

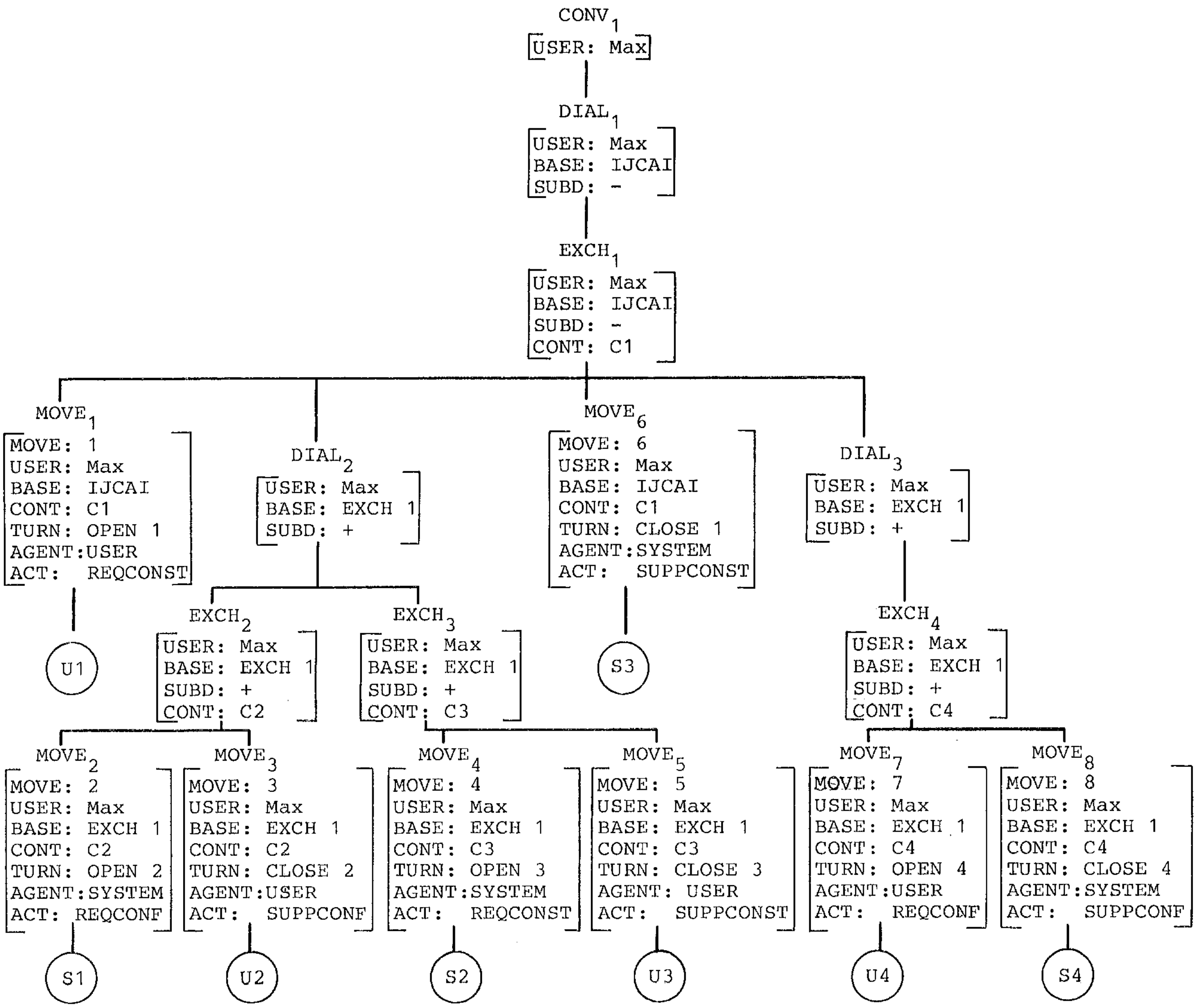




feature values trickle down in the right cases. Node subscripts distinguish between different instances of the same node. Note that Greek letters are also used as variables for node subscripts (rule (3)). Round brackets indicate optionality in the usual way. Angle brackets (rule (3)) are used in the same way as used by e.g. Labov (1972: ch. 8) in the study of sociolinguistic variation. They indicate a relationship between elements on the left and on the right of the rule. The expansion on the right is only possible if the feature on the left of the rule has the value specified, if both are enclosed in angle brackets. In the present case, they are used to prevent the expansion of a subdialogue as a further subdialogue. The feature [SUBD:-] is introduced by the rule that expands CONV. The rule expanding DIAL copies this feature with this value. The rule expanding EXCH allows an expansion of EXCH to include (optionally) one or two DIALs, but if the DIALs appear, then they carry the feature [SUBD:+]. The expansion of such a DIAL by rule (2) copies this feature with this value, as before, when the DIAL is expanded to one or more EXCHs. However, since the EXCHs so generated carry the feature [SUBD:+], the rule that expands EXCH will not allow the possibility of further DIALs, because any such expansion is conditional upon the EXCH having the feature [SUBD:-], as specified in rule (3).

The value of the feature TURN is either OPEN or CLOSE plus a constant that refers to the relevant EXCH. Note the use of the Greek

Table 2



variable. The same constant is used as the value of the feature BASE in subdialogues. What this amounts to is a stipulation that the topic of a subdialogue is the EXCH that it is part of, which seems to be about right intuitively. This is what makes them metalinguistic in character. Furthermore, note that this is a case where a feature/value pair does not trickle down. This is tantamount to stipulating that DIAL is a BASE-bounding node: it creates "islands" with BASEs that do not extend upwards to the main dialogue, but without overwriting the BASE of the current main dialogue. Again, this seems intuitively correct.

Let us now provide a concrete example of the structure that these rules assign to a dialogue such as (4).

4
U1 When is the next JICAI meeting?
S1 I presume you mean "IJCAI"
U2 Yes
S2 Do you mean the next conference or the next conveners' meeting?
U3 Conference
S3 12 August
U4 1985?
S4 Yes

The structure is given as Table 2. The values for the feature CONT are given as constants rather than as full semantic representations, and the constants rather than as full semantic representations, and the constant IJCAI is used for the BASE, which is the actual conference due to take place on 12 August 1985. This value has been given in all relevant cases, thus glossing over the fact that the BASE could not be identified immediately, which is what triggered the subdialogues. We add, however, certain features that were discussed earlier, such as MOVE, AGENT and ACT, to clarify how the final form of the frame representing each MOVE is derived.

This is a conversation in which there is only one main dialogue and only one main exchange within that dialogue. I hope that it is clear how these additional elements would be incorporated into the structure, and how the appropriate values for BASE, CONT and SUBD would be maintained or changed.

It is interesting to note that the nodes in conversation trees of this sort have a conceptual validity, in that different node types correspond to different aspects of a conversation. Thus a CONV node corresponds to "this conversation with this user", a DIAL [SUBD:-] node corresponds to "this topic", an EXCH [SUBD:-] node corresponds to "this point", a DIAL [SUBD:+] node corresponds to "a point that needed clarification", and an EXCH [SUBD:+] node corresponds to "what was unclear". Each MOVE node represents an utterance, of course. The set of MOVE nodes dominated by EXCH [SUBD:-] corresponds to "what was said; the general line of the conversation", and the set of MOVE nodes dominated by EXCH [SUBD:+] corresponds to "the subdialogues". Likewise, sets of other nodes correspond to other broader elements of a conversation. The set of CONV nodes corrends to "all the different conversations I had in this session with different users", the set of DIAL [SUBD:-] nodes corresponds to "the topics covered", the set of EXCH [SUBD:-] nodes corresponds to "the points discussed", and so on. By taking into account configurations of features at nodes, one can isolate, therefore, such elements as "the last but one topic discussed by the previous user" or "the first point in this conversation that needed clarification".

Let us now turn to two extensions of the above system, hypothetical moves and anticipatory moves, required by certain dialogue phenomena.

4. Hypothetical moves

There are apparently innocuous conversations such as (5) which may cause trouble on formal grounds.

5 U: Can you print the minutes of the last meeting?
S: Yes. Do you want to see them?
U: Yes
S: (prints)

This is a five-move conversation (the system's "Yes" counts as a separate move). The grammar would assign the structure shown informally as (6) to the first four moves.

6 MOVE 1: USER, OPEN 1, REQAFF
MOVE 2: SYSTEM, CLOSE 1, SUPPAFF
MOVE 3: SYSTEM, OPEN 2, REQAFF
MOVE 4: USER, CLOSE 2, SUPPAFF

On formal grounds, the conversation is closed after the fourth move (U: Yes), with all OPENed moves having been CLOSEd. What, then, triggers the system's printing of the required text (MOVE 5), and what happens after that, since an odd number of moves cannot constitute a well-formed conversation according to the grammar presented above? It is clear that it is the meaning of MOVE 3 that is the key. To handle this formally, we propose the use of the notion of "hypothetical move" and the representation speech acts not as atomic elements but as structures. Thus we will represent a supply-of-affirmation in response to a request-for-affirmation as SUPPAFF(REQAFF).

A hypothetical move is a move that does not actually occur in the conversation, but which the system constructs on the basis of very specific clues, and which allow it to continue the conversation appropriately. They correspond in some way to a representation of Grice's (1975) notion of implicature. For example, a more detailed analysis of (5) reveals that (6) omits several important details. The first move is actually ambiguous between a request-for-affirmation and a request-for-action. What we would like the system to do is to supply the affirmation to the request-for-affirmation part, and to request affirmation concerning the request-for-action part. The important point is that a proper analysis of "Do you want to see them?" should represent the fact that this is response to the potential request-for-action interpretation of "Can you print the minutes of the last meeting?". The upshot of this is that a more precise representation of the first four moves of (5) is (7), rather than (6).

7

MOVE 1: U, OPEN 1, REQAFF/REQACT
MOVE 2: S, CLOSE 1, SUPPAFF(REQAFF)
MOVE 3: S, OPEN 2, REQAFF(REQACT)
MOVE 4: U, CLOSE 2, SUPPAFF(REQAFF(REQACT)

We now have a way of accounting for the system's next more (printing), and for explaining why anything at all happens. MOVE 4 is a SUPPAFF(REQAFF(REQACT)), i.e. a supply of affirmation in response to a request for affirmation in response to a request for action. It seems quite clear intuitively that this complex structure is equivalent to a REQACT, and we propose that this type of reduction should take place by rule.

However, this rule must not over-write the original interpretation of the illocutionary force of the move, which must be retained for the dialogue to be well-formed with respect to the grammar. We propose that the effect of this type of rule (an implicature redundancy rule) is to create a hypothetical move immediately following it of the appropriate type. Its effect is to alter the structure of the conversation in exactly the same way as if the user (in this case) had actually uttered something like "I request you to print the minutes now", except for the fact that it is noted that this is a hypothetical move. We now have a formal entity that can trigger the printing of the required text, since this is a CLOSure of the hypothetical move. If no printing took place, then the dialogue would be ill-formed, since it would contain one OPEN that had not been CLOSEd. This, the system is behaving as if the user had made a particular move that did not actually occur. (The notion 'as if' is central to Vaihinger's (1935) theory of fictions. It is also crucial to Gricean implicature.)

The result is that (4) is now analysed as a six-move dialogue, with the structure shown as (8).

8

MOVE 1: U, OPEN 1, REQAFF/REQACT
MOVE 2: S, CLOSE 1, SUPPAFF(REQAFF)
MOVE 3: S, OPEN 2, REQAFF(REQACT)
MOVE 4: U, CLOSE 2, SUPPAFF(REQAFF(REQACT))
MOVE 5: U, OPEN 3, REQACT, hypothetical
MOVE 6: S, CLOSE 3, ACTION

5. Anticipatory moves

Another type of irrealis move is an anticipatory move, where on the basis of specific clues the system anticipates what the user's next move will be. The difference between these and hypothetical moves is that no action is taken by the system until there has been a reaction from the user that either confirms or disconfirms the correctness of the move that has been anticipated. The use of such moves will be of assistance in the interpretation of cryptic follow-ups, as in (9).

9

U: Can you provide progress reports on LOKI subprojects?
S: Yes. Do you want to see them?
U: LOKA
S: (prints)

The user's second utterance must be interpreted as a request-for-action, which is difficult on formal grounds. Without going into too much detail, we propose that in such cases ths system should have formulated an anticipation of a following request-for-action before the user's next utterance. This could either be an explicit request-for-action ("Please print the LOKA progress report"), or simply "Yes" (i.e. a SUPPAFF(REQAFF(REQACT))), which would trigger a system request for clarification, perhaps, or anything else at all that can serve to identify the BASE of the anticipated request-for-action. This is the important point about the anticipation. Anything at all that can fill in the unspecified slots in the BASE of the anticipated request-for-action will confirm that this utterance is intended as a REQACT. For this reason, the bare name LOKA is enough to get the report printed. Any other sufficiently identifying description of the relevant subproject would have achieved the same, such as any of the following (as appropriate): the one based in Hamburg, Hamburg, NL, Max's project, most recent, etc.

6. Conclusions

The processes and formalisms outlined above are all tentative in nature, and represent part of an approach to the problem of pragmatic sensitivity, rather than purported solutions to the problem. We envisage then as being part of a system that uses a multi-level parsing technique, with mutual assistance between different subcomponents of the parser, so that pragmatic information can immediately be used to assist parsing for syntax, and so on. We also see that parsing will involve not only sentence parsing, but also conversation parsing, in that the appropriate structure of a conversation must be built up at each step. This is simply one further part of the general parsing process, but one that we envisage as being of assistance to other parser subcomponents, as well as fulfilling its primary function of making sure that the system is something of a conversationalist, rather than just being a communicative plodder.

REFERENCES

Cole, P. & J. L. Morgan (eds.) 1975. Syntax & Semantics, volume 3: Speech acts. New York: Academic Press

Findler, N. V. (ed.) 1979. Associative networks: representation and use of knowledge by computers". New York: Academic Press

Givon, T. (ed.) 1979. Syntax & Semantics, volume 12: Discourse and syntax. New York: Academic Press

Grice, H. P. 1975 "Logic and conversation". In: Cole & Morgan (eds.), 1979, 41-58

Hobbs, J. R. 1982. "Towards an understanding of coherence in discourse". In: Lehnert & Ringle (eds.), 1982, 223-243

Labov, W. 1972. Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press

Lehnert, W. & M. H. Ringle (eds.) 1982. Strategies for natural language processing. London: Erlbaum

McKeown, K. R. 1985. "Discourse strategies for generating natural-language text". In: Artificial Intelligence 27, 1-41

Vaihinger, H. 1935. The philosophy of 'As if' -- a system of the theoretical, practical and religious fictions of mankind. London: Routledge & Kegan Paul

Wachtel, T. 1985a. "Discourse structure". LOKI Report NLI-1.1, University of Hamburg

Wachtel, T. 1985b. "Guided speech act assignment in 'can you' utterances". Unpublished LOKI working paper TW-11, University of Hamburg

Wachtel, T. 1985c. "Contextual factors in speech act assignment". Unpublished LOKI working paper TW-13, University of Hamburg