THE REPRESENTATION AND USE OF FOCUS IN DIALOGUE UNDERSTANDING

Technical Note 151

July 1977

By: Barbara J. Grosz
Artificial Intelligence Center

SRI Project 5844

The research reported in this paper was supported by the National Science Foundation under Grant No. MCS76-22004 and by the Advanced Research Projects Agency of the Department of Defense under Contract DAAG29-76-C-0011 and Contract DAHC04-75-C-0005 administered through the Army Research Office.

ABSTRACT

This report develops a representation of focus of attention that circumscribes discourse contexts within a general representation of knowledge. Focus of attention is essential to any comprehension process because what and how a person understands is strongly influenced by where his attention is directed at a given moment. To formalize the notion of focus, the need for and the use of focus mechanisms are considered from the standpoint of building a computer system that can participate in a natural language dialogue with a user. Two ranges of focus, global and immediate, are investigated, and representations for incorporating them in a computer system are developed.

The global focus in which an utterance is interpreted is determined by the total discourse and situational setting of the utterance. It influences what is talked about, how different concepts are introduced, and how concepts are referenced. To encode global focus computationally, a representation is developed that highlights those items that are relevant at a given place in a dialogue. The underlying knowledge representation is segmented into subunits, called focus spaces, that contain those items that are in the focus of attention of a dialogue participant during a particular part of the dialogue.

Mechanisms are required for updating the focus representation, because, as a dialogue progresses, the objects and actions that are relevant to the conversation, and therefore in the participants' focus of attention, change. Procedures are described for deciding when and how to shift focus in task-oriented dialogues, i.e., in dialogues in which the participants are cooperating in a shared task. These procedures are guided by a representation of the task being performed.

The ability to represent focus of attention in a language understanding system results in a new approach to an important problem in discourse comprehension -- the identification of the referents of definite noun phrases. Procedures for identifying referents are developed that take discourse structure into account and use the distinction between highlighted items and those that are not highlighted to constrain the search for the referent of a definite noun phrase.

Interpretation of an utterance also depends on the immediate focus established by the linguistic form of the preceding utterance. The interpretation of elliptical sentence fragments illustrates the effect of immediate focus. Procedures that interpret elliptical sentence fragments are developed. They use a representation that superimposes syntactic information about an utterance on the interpretation of the

underlying meaning of that utterance to minimize the processing required to expand a fragment into a complete sentence.

CONTENTS

ILLUSTRATIONS

ACKNOWLEDGMENTS

This report is a slightly revised version of a dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Computer Science in the Graduate Division of the University of California, Berkeley. I would like to thank my adviser, Martin Graham, for giving me the opportunity to work on this topic. The other members of my committee, Wallace Chafe and Alan Kay, provided helpful suggestions. Alan Kay deserves special thanks for suggesting language understanding as an area for study.

Discussions with Gary Hendrix were extremely helpful in developing the focus representation. Jane Robinson provided many insights into the nature of discourse and the importance of syntax. Bill Paxton provided a system framework in which to test my ideas about language. Ann Robinson put all the pieces together. Don Walker brought all of us together. Most of these people read multiple drafts of the dissertation. Their comments have improved it greatly. Richard Waldinger, Bonnie Tenenbaum, Earl Sacerdoti, David Levy, and Peter Deutsch also provided helpful comments.

Thanks to my friends; their support and encouragement made all the difference.

The research reported here was supported by the National Science Foundation under Grant No. MCS76-22004 and by the Advanced Research Projects Agency of the Department of Defense under Contract DAAG29-76-C-0011 and Contract DAHC04-75-C-0005 administered through the Army Research Office.

I INTRODUCTION

The great thing about human language is that it prevents us from sticking to the matter at hand.

Lewis Thomas, The Lives of a Cell

CONTENTS:

A. THE PROBLEM

To understand the sentences in a discourse, a computer system, like a person, must have knowledge about the domain of discourse. However, the knowledge required to solve problems in even simple real-life domains is so extensive that it will overwhelm any knowledge-based system that does not apply it selectively. This means that the ability to focus on the subset of knowledge relevant to a particular situation is crucial. The need for focus is present in problems ranging from understanding an utterance or interpreting a visual scene to problems like designing a building or solving a differential equation.* This report addresses the problem of focus from the perspective of building a computer system for understanding dialogue. Its major concern is the incorporation of a representation of focus in a system that participates in a dialogue. A focus representation is developed that highlights

----

* It might seem that creative thinking and innovative problem solving derive from an ability to turn off the normal focusing mechanisms and look at a problem in a different way, but viewing a problem from a new perspective does not eliminate focusing; the focusing capability is not turned off; the default connections about what to focus on are overridden, and a new and different focus is chosen.

those items in the knowledge base (i.e., the encoding of that portion of the world the system knows about) that are relevant at a given point in a dialogue and includes mechanisms for changing focus as the dialogue progresses. A simplified version of the focus representation was implemented in the SRI speech understanding system (Walker, 1976) and used by the discourse component to resolve definite noun phrases.

The following hypothetical conversation between two people illustrates several facets of how focus operates in a discourse.

(1)P1: I'm going camping next week-end. Do you have a two-person tent I could borrow?

(2)P2: Sure. I have a two-person backpacking tent.

(3)P1: The last trip I was on there was a huge storm.

(4)P1: It poured for two hours.

(5)P1: I had a tent, but I got soaked anyway.

(6)P2: What kind of tent was it?

(7)P1: A tube tent.

(8)P2: Tube tents don't stand up well in a real storm.

(9)P1: True.

(10)P2: Where are you going on this trip?

(11)P1: Up in the Minarets.

(12)P2: Do you need any other equipment?

(13)P1: No.

(14)P2: OK. I'll bring the tent in tomorrow.

Since most objects do not have proper names, definite noun phrases are a primary means of identifying objects. However, the same noun phrase may be used to describe (and hence identify) different objects at different times. For example, in the last utterance (14) of the hypothetical conversation, the noun phrase "the tent" refers to the tent introduced in (2). Even though the tent discussed in (5) to (7) has been mentioned more recently than the tent in (2), it is no longer in focus and hence is not considered as the referent of the noun phrase "the tent" in (14). This example illustrates the fact that the most recently mentioned object that matches a noun phrase may not be the object identified by that noun phrase. Shifts in focus in the dialogue must be taken into account.

In this dialogue, the statements in (1) introduce into focus a camping trip and the need for some equipment (a tent). The response in (2) brings a particular tent into focus. Statement (3) shifts the focus to a previous camping trip. The tent used on that trip is brought into focus in (5) and leads to a discussion of tube tents in (6) through (9). The focus shifts back to the trip being planned in (10). Utterance (12) shifts the focus back to the need for equipment on this trip. As a result, when "the tent" is used in (14), the only tent that is in focus is the tent first mentioned in (2).

Focus also affects the interpretation of word senses. The "soaking" in (5) does not involve someone paying too much money. The influence of focus on the choice of word sense is usually quite subtle; alternative senses do not occur to most people. For example, when discussing the steps in a folkdance, the sense of "step" that corresponds to steps in a house never arises.

Statements (7) and (11) illustrate a more local effect of focus. The focus of the preceding utterance supplies the information necessary to interpret an elliptical expression. The phrase "a tube tent" is not a syntactically complete sentence, but is sufficient to convey "It was a tube tent (that I had on the last trip)" following the question "What kind of tent was it?" Similarly, "up in the Minarets" makes no sense out of context, but is a completely understandable statement following the question in (10).

The importance of focus in language understanding became clear in the course of analyzing several dialogues that involved communication between two parties cooperating to complete a task. These dialogues were collected in situations simulating direct interaction between a person and a computer. The key result of the analysis was that task-oriented dialogues subdivide into units just as a task subdivides into subtasks. The segmentation of dialogues reflects the shifts in focus with time that occur as a dialogue progresses. As a result, the structure of the task provides a guide to shifts in focus in these dialogues. The collection and analysis of these dialogues is described in the next chapter to provide a background for the discussion of the representation and use of focus presented in the remainder of the report.

B. FOCUS IN DISCOURSE

The choice of the term focus as the theme of this report reflects a concern with the importance of the role of attention in any comprehension or reasoning process. What and how a person understands is strongly influenced by what he is thinking about at a given moment, by what his attention is directed towards. The focus of attention that influences the interpretation of an utterance in a discourse results from a combination of contextual factors. In fact, what is usually meant by "the context of an utterance" is precisely that set of constraints which together direct attention to the concepts of interest in the discourse in which the utterance occurs. Both the preceding linguistic context -- the utterances that have already occurred -- and the situational context -- the environment in which an utterance occurs -- affect the interpretation of the utterance. For a dialogue, the situational context includes the physical environment, the social setting, and the relationship between the participants in the dialogue. Hence, focus refers to the effect of a composite of contextual influences.

It is useful to separate the influence of focus into two ranges: immediate and global. Immediate focus refers to the influence of a listener's memory for the linguistic form of an utterance (the actual words and the syntactic structure) on his interpretation of a subsequent utterance. It influences both the ordering of constituents in sentences and the interpretation of sentence fragments. For instance, in the hypothetical conversation presented above, immediate focus causes the elliptical response "up in the Minarets" to be understood as meaning "We are going up in the Minarets on this trip." In contrast, global focus refers to the influence of memory for the more general meaning conveyed by all of the preceding utterances in a discourse on the interpretation of subsequent utterances. Global focus is determined by the total discourse and situational setting of an utterance. It influences the choice among different senses of a word, the interpretation of noun phrases and actions, and the overall interpretation of an utterance. The influence of global focus on language is illustrated in the example conversation by the reference in (14) to a tent that not only is mentioned much earlier in the dialogue, but also is not the most recently mentioned tent.

The most crucial requisite of a focus representation is that it differentiate among the items in the knowledge base on the basis of relevance. By highlighting those items that are relevant to the current discourse, the focus representation enables the system to access more important information first during its retrieval and deduction operations. The representation of focus presented in this report is based on segmenting the knowledge base into subunits. Each subunit, called a focus space, contains those items that are in the focus of attention of the dialogue participants during a particular part of the dialogue. This segmentation is structured by ordering the spaces in a hierarchy that corresponds to the structure of the dialogue.

Corresponding to this static requirement on the focus representation there is a dynamic requirement. The focus representation must include mechanisms for shifting focus. As successive utterances in a discourse are processed, the items in focus change. What indicates a shift in focus depends both on the kind of discourse being processed and on the topic of discourse. Shifts in focus in task-oriented dialogues are closely tied to the task. Mechanisms are developed specifically for detecting shifts in such dialogues. They use a representation of the task to decide when and how to shift focus.

The process of identifying the object referred to by a definite noun phrase illustrates the use of the focus representation in discourse processing. Definite noun phrases both affect and are affected by the focus of attention of a discourse. The identification of the referent of a definite noun phrase requires some model of both the situational and linguistic contexts in which the noun phrase occurs. In turn, definite noun phrases can indicate a change in focus. When the resolution of definite references is considered from the perspective of focus, questions like how far back in a discourse to look for a referent are no longer relevant. Instead, the problem is how long an item stays in focus and what can cause a shift in focus.

The major portion of the report is concerned with the representation and use of global focus. The effect of immediate focus and the processes needed to use it are considered only as they arise in the interpretation of elliptical utterances. The syntactic structure of an utterance (along with some additional syntactic and semantic characteristics of its phrases) provides the immediate focus for the utterance that follows. Interpretation of an elliptical sentence fragment requires splicing the fragment into the (possibly transformed) structure of the preceding utterance at the appropriate place.

C. GUIDE TO THE REMAINDER OF THE REPORT

Chapter II describes the collection of several kinds of dialogues and presents analyses of some of their discourse characteristics. The structure of the dialogues and its importance for understanding definite noun phrases is described. The results of these analyses were used in designing the focus representations presented in the remainder of the report. Chapter III presents the representation of focus and describes its use in the retrieval of information from a knowledge base. It contains the core ideas of the report. Chapter IV describes one use of the focus representation in the interpretation of utterances, namely to guide procedures that identify the referents of definite noun phrases. Chapter V describes mechanisms for deciding when to shift focus so that the focus representation is updated as a dialogue progresses. Chapter VI describes the role of immediate focus in the interpretation of elliptical utterances. Representations and procedures for handling a limited set of elliptical expressions are presented. Chapter VII discusses how the representations developed in this report can be extended. Both extensions to generalize the representation and other uses of the representation are presented. Appendix A contains a brief summary of research in linguistics, psychology, philosophy and computer science that is related to and has had an influence on the research described in this report.

II DISCOURSE ANALYSIS

CONTENTS:

A. INTRODUCTION

In this chapter we examine several dialogues collected in situations simulating those in which a person, using a computer as a problem solving aid, interacts with the system in natural language. From the point of view of building a natural language understanding system, the main purpose of the collection and analysis of such dialogues is to characterize the language used when people communicate for the purpose of solving a problem. Since the goal of the dialogue analysis is to determine a person's language needs when using a computer system, the ideal context for collection would be one in which a person is in fact interacting with a computer. But this is a 'Catch-22' situation since the data are needed to guide the design of the system. The best that can be done initially is to simulate this situation by using the computer as a communication medium.

The next section of this chapter describes the method of collection of two kinds of dialogues. The major portion of the analysis is concerned with a set of task-oriented dialogues: the conversation that ensues when two people work cooperatively on a task that requires knowledge each of the participants alone has. In addition to these dialogues, a set of dialogues resulting from one person's querying a data base in natural language is examined. This set differs from the task-oriented dialogues in several ways; examination of both the similarities and the differences is of interest.

The remaining sections of the chapter contain analyses of the dialogues. The results of these analyses were the starting point for the research described in the remainder of this report. Familiarity with the results is important for understanding the relevance of this work to the problem of building a computer language understanding system. The analysis is presented at three different levels. At the global discourse level, the structure of the dialogues is examined. This structure reflects the shifts in focus as a dialogue progresses and influences descriptions and referential expressions. At the more local discourse level, the influence of focus on closely contiguous utterances is examined from the point of view of elliptical expressions. Finally, at the level of constituents of individual utterances, we examine the kinds of words appearing in the dialogues and the different types of utterances used.

B. <u>COLLECTION OF THE DIALOGUES</u>

1. OVERVIEW

The first set of dialogues we collected were task-oriented dialogues. These dialogues occur when two people work cooperatively on a task, where a 'task' is some real-life activity that is directed toward achieving a particular goal and that can be broken down into small steps, each having its own goal. Examples of tasks include repairing faulty equipment, building a house, carrying out a chemistry experiment, and solving algebra word problems. Task-oriented dialogues occur normally when a master craftsman instructs an apprentice, when two mechanics work together to repair a car, and when a teacher guides a student in a chemistry lab. The major characteristics of these dialogues are that both participants are aware of the task to be performed and that communication between the participants is necessary to accomplish it.

The tasks considered in this research have one further characteristic: they are tasks for which it is feasible to consider a computer taking the role of one of the participants sometime in the not-too-distant future. In particular, we have investigated situations in which the computer guides a person performing a task. Interest in such dialogues arose in part from considering the language requirements of a computer-based consultant system. A description of initial steps toward building such a system may be found in Hart (1975). The goals of this system were to build a computer system that could guide a person in the performance of a complex task with which (s)he had little experience. Natural language communication was a key element of the system.

In addition to the task-oriented dialogues, we collected a set of question-answering dialogues. Question-answering dialogues occur when one person asks another (or a computer system) a series of questions in order to help solve some problem. They are distinguished from task dialogues mostly in that the answerer cannot be viewed as sharing a goal in common with the questioner. Although short question-answering dialogues occur frequently in everyday conversation, extended sequences (more than five or so questions) are more frequent in communications with computers, for example, in a sequence of queries to a computer data base. In the dialogues that we collected, a person queried a data base in order to solve an assigned problem that required interaction with the data base. To avoid confusion with other kinds of question-answering dialogues, these dialogues will be referred to as data base dialogues in the remainder of the discussion.

Task-oriented dialogues are a good source of unbiased data on discourse. Concentration on the performance of a task keeps the participants from becoming self-conscious about their language. The resulting dialogues are spontaneous and unrehearsed. The data base dialogues are somewhat less spontaneous. The less realistic nature of the assigned problems made the subjects in these dialogues more self-conscious than those in the task dialogues.

The dialogues described in this report were both written and spoken. To simplify the following discussion, the term speaker will refer to the transmitter of a message and hearer to the receiver even though some of the transmissions were typed.

2. TASK DIALOGUES

The main task used for collection of data on task-oriented language was the assembly of part of an air compressor. In addition, two dialogues were collected in which an expert plumber provided guidance in the repair of a leaky faucet. A sketch of an air compressor is shown in Figure II-1. For the purposes of understanding the dialogue fragments in this report, it is important to note the pump, the pump pulley, the platform, the aftercooler, the belt-housing frame and cover, and the connections between these parts. Tasks involving both high-level assembly -- installing the pump and belt -- and lower-level assembly -- putting the pump together -- were used.

The participants in each of the dialogues were an expert (E) and an apprentice (A). The experts, in addition to being skilled at

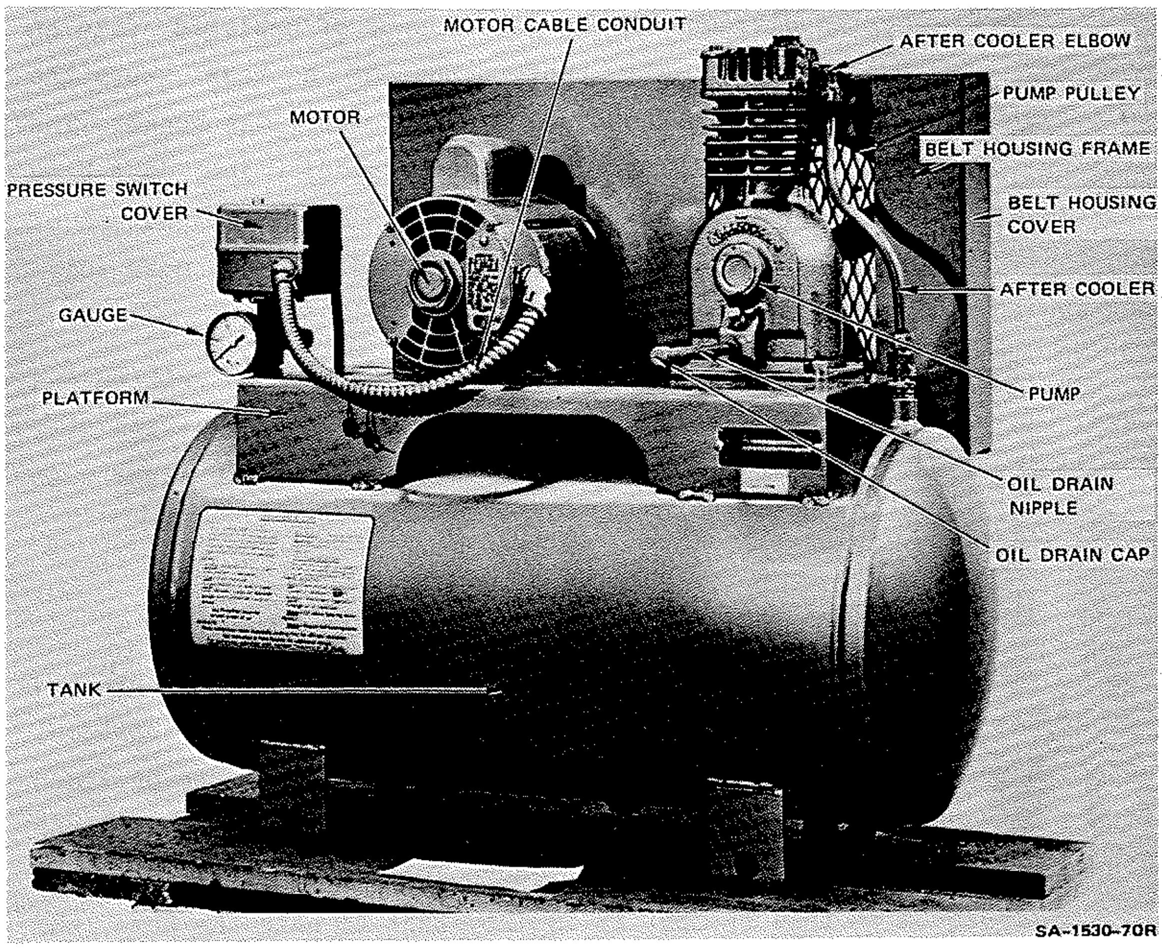


Figure II-1. A SMALL AIR COMPRESSOR

mechanical tasks, were familiar with the compressor and the tools used in assembling and disassembling it. Before participating in the dialogues, the experts performed the task themselves and then had a practice session instructing someone else. None of the apprentices was familiar with the air compressor; in their general mechanical knowledge, they ranged from complete novices to amateur auto mechanics.

Dialogues were collected under a variety of conditions. The amount of visual contact between participants was varied to determine the effects of limited vision and to collect data on descriptions. In the first experiments, E and A were allowed to communicate freely, and they interrupted each other frequently. For the next set of experiments, the ability to interrupt was removed to see what effect this would have on communication and task accomplishment. Finally, the information given to the apprentice about the expert was varied.

The dialogues fall into four classes:

(a) Free, with vision: E and A were in the same room; they were able to see each other; verbal communication was spoken; no restrictions were placed on language use. The only instructions were to complete the task. The only restriction was that E could not help DO the task; he could only instruct A. In this setup, then, E could see A, monitor what A was doing, and notice where A put tools and parts. E and A were free to interrupt one another.

(b) Free, with no vision: the conditions were the same as (a) except that E was not able to see what A was doing.

(c) Restricted and aware: both visual and verbal communication were restricted in these dialogues. The experimental set-up is shown in Figure II-2. Verbal communication passed through a monitor who was responsible for assuring that E and A did not interrupt each other. In these dialogues A spoke, and the monitor typed the message; E typed

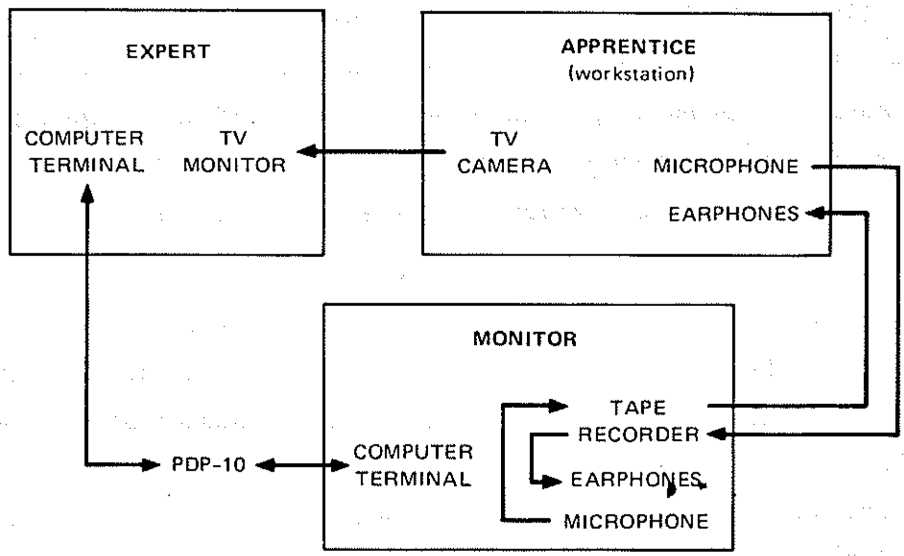


Figure II-2. EXPERIMENTAL SETUP FOR RESTRICTED DIALOGUES

a response and the monitor read it to A. Computer terminals were used solely so that transcripts could be easily obtained. E was able to get 'still' pictures from the television camera, but they had to be requested; normally, the camera was focused on a blank wall. In these experiments, A was informed that the experiment was a simulation of a computer system. Hence, A was aware that E was a person.

(d) Restricted and unaware: the experimental setup was the same as in Condition (c), but A was told that E was a computer system. In each case we determined after the dialogue was collected and before explaining the true nature of the experiment that A believed that a computer system was serving as expert.

3. DATA BASE DIALOGUES

The data base experiments were designed to collect samples of the language people would use if they had verbal access to a data base. (Detailed descriptions of the procedures for collecting the samples together with examples are in Deutsch, 1974 and Silva, 1975.) In order to collect realistic samples, it was necessary to provide people with a specific problem, requiring information from the data base. Again the purpose was to make their language as unself-conscious as possible.

The data base used for these dialogue experiments contained information about the ships of the United States, British, and Russian fleets. In the first set of dialogues, the subjects were given tables to fill out (similar to the ones found in naval reports), and two short problems to solve. They were instructed to ask for information from an analyst, who answered using material from the data base. The subjects and analysts were in the same room but were not allowed to interrupt one another or to view each other's materials. For these problems, no additional information could have been obtained by either subject or analyst, if they had been allowed visual contact.

The second set of dialogues used a revised data base containing information on U.S. and Russian ships in the Mediterranean. Subjects were given one long problem to solve for which they needed information in the data base. Again, the subjects were not restricted in their use of language. Their questions were translated into data base queries and typed to a computer data base system by an <u>operator</u>. The answers were read back to the subject.

C. <u>BACKGROUND FOR THE ANALYSIS</u>

The emphasis of the analyses presented here will be on discourse-level phenomena: those features of utterances in the dialogues that come from the utterances being part of a cohesive unit of discourse. The relation between dialogue and task, the structure of the dialogues, and the influence of an utterance on the utterance that follows will be examined.

Chapanis (1975) has been interested in characterizing differences in language use across different modes of communication. For example, he investigated differences in measures such as number of sentences, number of words, and number of 'noun-like' words across modes such as handwriting, typing, and speaking. In addition, he examined the differences in time required for problem solution across the different modes of communication. His analyses are statistical; they provide information about how the language used in each mode differs. Although such statistical measures provide some indication of the advantages of one mode over another as a means of communication and of the effect of the mode on the language used, they do not provide certain information required for building a computer language-understanding system. For that purpose, information is needed on the particular words used and on how they are put together in utterances to provide meaningful communication.

The analysis reported here has a different emphasis: it is concerned with taking a single mode (actually a small number of very similar modes) of communication and characterizing the range of language devices used to achieve successful communication of an idea. The analysis will be concerned with when and how different language devices are used; with what particular types of occurrences there are rather than with comparisons of numbers of occurrences. Many different questions can be asked along these lines. They include (1) sentence-level questions like "What different sentence structures occur?", "Do some occur more frequently than others?", and "In what context?"; (2) intersentential questions like "What links are there from one utterance to another?"; and (3) more global questions like "Does a dialogue have some overall structure?"

1. <u>INFLUENCE OF THE RESTRICTIONS ON VISION AND SPEECH</u>

Ten task dialogues were collected: one under Condition (a), and three each under Conditions (b), (c), and (d). The major distinction between the free dialogues and the restricted dialogues was the frequent occurrence of interruptions in the free dialogues. Expert and apprentice cooperated on completing utterances as well as on completing the task. The dialogue segments in Figure II-3 illustrate this cooperative aspect of the interruption. Lines (5)-(6), (9)-(13), and (17)-(18) are the most direct examples. In the first two cases, E is pausing in search of the 'right' phrase when A fills it in. In (17)-(18), E gives a similar kind of aid to A. Lines (2) and (4) are typical of the kind of ongoing mutual support of the two participants. A indicates an understanding of what has been said so far, so E may continue. This support is also evident in the echoing of (14)-(16). The kind of fragment resulting from these interruptions was more than we wanted to attempt to handle in an initial speech understanding system. We surmised that not allowing the participants to interrupt would not seriously hamper problem solution. Chapanis (1973) has empirical evidence that supports this assumption. The restricted dialogues were designed to eliminate interruptions. The design of the experiment for restricted dialogues closely resembles Chapanis' setup but was arrived at independently.

The different visibility conditions had several different effects on the dialogue. Robinson (1975) discusses some of these. The most pronounced difference was in the kind of descriptions that resulted. Figure II-4 shows the most blatant contrast found in the dialogues. If visual information is shared, it can be used in descriptions. In the protocols with restricted dialogue and limited vision, E often asked for a still picture in order to use this kind of information. The dialogue fragment in Figure II-5 is an example. The difficulty of giving descriptions without the aid of shared visual information is best illustrated by the fragment in Figure II-6. A more extensive discussion of the descriptions found in the dialogues and some of their characteristics is presented later in Section D.5.

2. CORE DIALOGUES

Four of the ten task dialogues form the core data of the

(1) E: . . . and those are to be inserted in the side of the motor . . . in the side of the rear of the motor . . .

(2) A: Uh hm.

(3) E: . . . and . . .

(4) A: . . . I see it . . .

(5) E: O.K. and each wire is to be attached to a . . .

(6) A: One of those bolt things here?

(7) E: bolt? . . . yes.

* * *

(8) A: . . . now should I unscrew the nuts from the bolts?

(9) E: No. The wire goes on top of that . . . on top of the nuts that are on there . . .

(10) A: I see . . .

(11) E: . . . and there're . . .

(12) A: Other nuts.

(13) E: . . . there are other nuts . . .

* * *

(14) E: The washer will be the last thing that . . .

(15) A: The washer will be last . . .

(16) E: The last item that will be on it.

(17) A: O.K. Then this little plastic thing

(18) E: With the holes in it.

Figure II-3. FRAGMENTS OF COOPERATIVE DIALOGUES

analysis: two each of the two kinds of dialogues in which a monitor prohibited interruptions [i.e., dialogues under conditions (c) and (d)]. These conditions were selected because they were closest to the

WITH VISION:

E: You have a top piece with a KNURLED section that you can take ahold of.

A: What's a knurled section?

E: You've got your fingers on it.

WITHOUT VISION:

E: Now underneath is what they call a cap assembly. It has a KNURLED face around it.

A: What does knurled mean?

E: Little lines running up and down on it so you can take ahold of it.

Figure II-4. DESCRIPTION OF "KNURLED" WITH AND WITHOUT VISION

E: Use the ratchet wrench on the top and hold the nut stationary on the bottom with a box wrench.

A: What is a ratchet wrench?

E: Show me the table.

E: The ratchet wrench is the object lying between the wheel puller and the box wrenches on the table.

Figure II-5. USING VISION TO HELP WITH A DESCRIPTION

situations that would occur in any person-computer interaction in the near future. Since each of the dialogues took between forty minutes and two hours and consisted of between 120 and 250 lines, this constitutes a considerable body of data.

In addition to the ten task-oriented dialogues, five data base dialogues were analyzed. Two dialogues were chosen as representative of the dialogues collected during the first experiment. All three dialogues from the second set were analyzed. Again, although the number

E: O.K., uh . . now, we need to attach the um . . conduit to the motor. .. the conduit is the uh . . the covering around the wire that you . . uh . . were working with earlier. Um, there is a small part um . . oh brother . . .

A: Now, wait a s . . . the conduit is the cover to the wires?

E: Yes. and . . .

A: Oh, I see, there's a part that . . a part that's supposed to go over it . . .

E: Yes . .

A: I see . . it looks just the right shape, too. Ah hah! yes . . .

E: Wonderful, since I did not know how to describe the part!

Figure II-6. DIFFICULTIES IN EXPLAINING AN UNFAMILIAR COMPLEX OBJECT

of dialogues is small, the amount of data in each dialogue is quite large. The dialogues in the first set are over 100 lines long and represent approximately thirty minutes of speaking time. The dialogues from the second set each represent over an hour of dialogue. It was necessary to look at long segments of dialogue to get the data needed, since many interesting phenomena occur infrequently in any given dialogue. That such phenomena occur at all is important; the infrequency with which they occur is irrelevant.

D. DIALOGUE STRUCTURE AND ITS INFLUENCE

The structure of a discourse reflects the shifts of focus occuring in it. A key use of the structure of a discourse in an understanding system is to provide keys to the current context, and thus to help establish expectations and interpret object and action references. Dynamically determining where an utterance fits in the structure is a crucial part of its interpretation. Correspondingly, determining where the utterance fits helps determine the structure of the discourse and how to shift focus. It is this aspect of structure that we will examine.*

1. THE STRUCTURE OF THE DIALOGUES

In general, the task dialogues exhibit more structure than the data base dialogues. These differences in structure arise mainly from differences in inherent structure of the problems being solved. The task dialogues involved tasks that decompose into subtasks. The relationship between subtasks is well defined. As a result, successive utterances in the task dialogues had strong links. In contrast, the information needed for solution of the data base problems could be asked for in a variety of ways (i.e., a variety of question sequences). There was no necessary dependence of a query on what preceded or followed it. The following sections examine indications of structure* in the two kinds of dialogue.

a. TASK-ORIENTED DIALOGUES

Task-oriented dialogues have a structure that closely parallels the structure of the task being performed. The whole dialogue is segmented into subdialogues, which themselves may break down into subdialogues, just as the task breaks down into subtasks, which themselves may be decomposable. For example, the task of making a cake has subtasks of preparing the batter, actually baking the cake, and icing the cake. A recipe (or television cooking program description) contains distinct parts for each of these subtasks. Likewise, the compressor task of installing the pump decomposes into attaching the pump, attaching the pump pulley, attaching the belt, and several other tasks. Attaching the pump decomposes into positioning the pump and actually securing it. An analysis of the dialogues for the pump installation task reveals that they fall into subdialogues paralleling these subtasks.

----

* The concept of structure used here is similar to that in Halliday and Hasan (1976) (see especially p. 327), but different from that occurring elsewhere. We are not producing a dialogue or text grammar (cf., vanDijk, 1972; Rumelhart, 1975). In particular, we are not interested in either generating or recognizing a valid dialogue (and hence in using such a grammar as sentence grammars are used).

The task hierarchy imposes a hierarchy on the subdialogue segments. As different parts of the task are performed, different objects and actions come into focus. When a subtask is completed, it fades from focus. However, the higher level (parent) task remains in focus. Hence, when a sibling subtask is performed, the concepts in the parent -- but not those in the completed subtask -- are in focus and affect the use of referring expressions like pronouns. This correspondence between task structure and dialogue structure plays a crucial role in determining the focus in which an utterance is interpreted. It is particularly important for the interpretation of references (see Section D.4 below).

Several linguistic devices indicate the segmentation of a dialogue. As an example, consider the use of "when". The subdialogue corresponding to a task ends, or is closed, when the task it parallels is completed. If the context that existed during a subdialogue that has been closed needs to be re-established (for example, so that actions and objects that appeared in the subdialogue can be discussed as they occurred in that context), the subdialogue must be reopened. "When" provides one means of accomplishing this. The utterance, "A little metal semicircle fell off when I took the wheel off" is meant to reinvoke the entire context of taking the wheel off in order to determine the meaning of the metal semicircle falling out.

Another indication of the segmentation phenomenon is the use of pronouns whose referents lie far back in the previous discourse. In every case, the pieces of dialogue skipped over are whole segments relating to some distinct subtask or subtasks. This is the case in the dialogue example of Figure II-7. The completion of the belt housing cover attachment closes the subtask of installing the cover. The "it" in the last utterance refers to the air compressor last mentioned over a half-hour before. This use of "it" is not unique. In fact, similar expressions containing "it" references to the air compressor occurred in three of the four core dialogues. There were also several instances of pronoun references skipping over smaller pieces of dialogue.

E: Good morning. I would like for you to reassemble the compressor.

. . .

E: I suggest you begin by attaching the pump to the platform.

. . . (other subtasks)

E: Good. All that remains then is to attach the belt housing cover to the belt housing frame.

A: All right. I assume the hole in the housing cover opens to the pump pulley rather than to the motor pulley.

E: Yes that is correct. The pump pulley also acts as a fan to cool the pump.

A: Fine. Thank you.

A: All right the belt housing cover is on and tightened down.

(30 minutes + 60 utterances after beginning)

E: Fine. Now let's see if it works.

Figure II-7. PRONOUN USE REFLECTING DIALOGUE STRUCTURE

The segmentation of dialogues into subdialogues may also be seen by considering a dialogue with groups of lines removed. If a whole subdialogue is removed, the dialogue remains coherent. Although it is sometimes possible to delete some utterances that are not whole subdialogues without damaging coherency, such removals often result in dialogue fragments that do not make sense. Removing a question and its answer may not affect coherency. Removing an utterance that opens or closes a subdialogue (these kinds of utterances will be discussed shortly) does.

In summary, a subdialogue forms a cohesive subunit of a higher level dialogue. Closure of the subdialogue entails closure of the focus corresponding to that subdialogue and a return to the focus present before the subdialogue was entered. As a result, references, including pronoun references, may be used to refer to objects in this higher level focus. The relationship between the segmentation of dialogues and the interpretation of referential expressions makes representations of the task structure and the shifts of focus in the dialogues crucial to a language understanding system.

b. THE DATA BASE DIALOGUES

The data base dialogues did not exhibit the same kind of segmentation, but there was definite evidence of groups of closely related utterances. The amount of segmentation evident in these dialogues differed according to the problem being solved.

The dialogues for the table-filling-out problems had no global structure although there were sequences of related utterances. The sentence-to-sentence links were most evident from the use of elliptical sentence fragments. The sequence in Figure II-8 illustrates how one utterance can provide sufficient context so that only a phrase suffices as a complete subsequent utterance; i.e, the phrase conveys a whole question in the context of the preceding utterance. As Chapter VI describes, the use of ellipses is a local discourse feature; it operates only between adjacent utterances.

S: What's the surface displacement of the Lafayette class?

A: 7300 tons.

S: What's the submerged displacement?

A: 8200 tons.

S: The length?

A: 425 feet.

S: Number of torpedo tubes?

Figure II-8. A SEQUENCE OF ELLIPTICAL SENTENCE FRAGMENTS

The dialogues for the other problems exhibit slightly larger groupings of utterances. Some evidence of shifting of focus over subproblems appears. The dialogue fragment in Figure II-9 is a self-contained unit. The immediately preceding utterance was about British diesel patrol submarines. The utterances following this subdialogue were about submarines other than the Yankee and the Hotel II. The subdialogue itself narrows from considering all Soviet submarines to asking about attributes of two particular submarines. There is a short subdialogue inside the subdialogue itself. The two starred utterances form a clarification-question/answer pair. Only a

S: What classes of USSR submarines are there?

A: <answer>

S: How many of those are nuclear ballistic missile submarines?

A: Two.

S: What are they?

A: Yankee, Hotel II.

S: How many tubes does the Yankee have?

A: Eight.

S: *That's torpedo tubes, right?

A: *Eight.

S: And, how many torpedo tubes and missile launchers for the Hotel II?

A: Ten torpedo tubes, three missile launchers.

S: What is the submerged speed for the Yankee and Hotel II?

A: <answer>

Figure II-9. A DATA BASE QUERY SUBDIALOGUE

few such segments appear in the dialogues for the short problems. This is the longest sequence that appears; the others are only six to eight utterances long. Most of the dialogues consist of sequences of utterances related locally but without structure. Long segments are more common in the dialogues for the long problems, but openings and closings of the subdialogues are often hard to detect; they are much less clear than those in the task dialogues. As a result, the segmentation is harder to detect.

What distinguishes the data base dialogues most from the task dialogues is the lack of any discernible intermediate structure. There are local discourse phenomena which tie adjacent utterances together, and there is some structure provided by the overall problem, but there is little relating the local segments together into bigger segments. As the problems posed to the subjects get larger, intermediate level organization appears. What seems to happen with these problems is that a solution breaks down into some recognizable substeps and the dialogues fall into segments according to these substeps. There seems to be a continuum, of which we have only a few samples, from the totally unstructured table-filling dialogues to the highly structured task dialogues.

2. KINDS OF SUBDIALOGUES

The subdialogues we have discussed so far are task or problem related; they can be linked directly to some substep of the task being attempted. Several other kinds of subdialogues occur related to general questions, requests for clarification, and communication channel checks. Some of these are quite short, only a pair of utterances, but they are all distinguishable as separate from the surrounding dialogue and cohesive as a unit. Distinguishing among these kinds of subdialogues is important for comprehension because each kind establishes different expectations about the subsequent utterance and because the closure of each kind of subdialogue is different.

General question-and-answer subdialogues include subdialogues related to identifying objects in the domain (e.g., "What's a motor bolt?"), describing tool use ["How is this (wheelpuller) used?"], identifying the right tool to be used or seeing if a better tool is available (e.g., the expert asking "What tools are you using?"), making sure no blatant error occurs in performing the task (e.g., the apprentice asking, "Will this require some effort?"), and testing whether a task was performed correctly (e.g., "How tight should the bolts be?").* The data base dialogues contain only a few general question-answering dialogues; they are all concerned with terminology, e.g., "What do you mean by deployment?"

Two kinds of subdialogues fall inbetween subtask and general

---

* Since understanding about objects involved in a task is important to the performance of the task, these subdialogues may also be viewed as task-related. They differ from the task-related subdialogues in that they are not as directly tied to the particular task.

question answering. They are clearly related to the task being performed but are also general questions. First, there are questions about why a certain part or step is needed (e.g., "What is the key for?"). Second, there are requests by the apprentice for alternative ways of doing some task (e.g., "Do you have another way to get the nuts in underneath the platform?").

Both the task and the data base dialogues contain pairs of exchanges whose purpose is to determine that the previous message was heard correctly or to have a missed message retransmitted. The middle two lines of the dialogue in Figure II-10 are an example of this kind of subdialogue. Requests for retransmission include statements like "What was that again?" and "Please repeat the last instruction."

A: One of them is at 14 degrees E, 34 degrees N.

S: 34 degrees you said?

A: Yes.

S: O.K.

Figure II-10. A SUBDIALOGUE CHECKING PREVIOUS MESSAGE

There are also subdialogues where one participant wants to make sure that the other participant means the same thing as he does. This kind occurs in the starred sequence of the dialogue fragment of Figure II-9.

3. SUBDIALOGUE TRANSITIONS

a. OPENING AND CLOSING OF SUBDIALOGUES

Detection of subdialogue units and hence knowing when to shift focus, are crucially dependent on detecting statements that open and close subdialogues. Task subdialogues may be opened by either expert or apprentice. In the dialogues that were examined, opening statements made by the expert were always statements of the subtask goal. Sometimes the statement was augmented by a sequencing expression such as "next" or "now". Subdialogues opened by apprentices also included subtask goal statements. These were embedded either in statements indicating the task was being, or was about to be, performed, or in statements requesting information on how to perform the task. Frequently, a pair of utterances serves to open a subtask. This happens when A asks for the next task, as in the following:

A: What should I do now?

E: Remove the pump.

Alternatively, a pair may result from A asking how to do some task, leading to E giving a subtask specification, as in the pair:

A: How do I remove the pump?

E: First remove the flywheel.

Such pairs occurred both when A knew what task was next but not how to do it and when E gave the task and A needed more specification. As an example, consider the preceding four utterances as part of a single dialogue.

Task subdialogues that occurred when the apprentice ran into trouble were opened by a statement of the problem. Similarly, subdialogues for checking task performance were opened by the expert asking if some goal had been achieved or was in the process of being achieved.

The most typical closings of subdialogues were through statements like "O.K." or ones indicating that a task goal was completed. Often a combination of these was used. These closings are explicit; implicit closings also occurred quite frequently. Typically, A would indicate that a subtask was finished by asking for the next subtask. In these cases, the same statement might serve both to close an old subdialogue and to open a new one.

Question-answering subdialogues are always opened by a question about some part, tool, task, or problem. In the dialogues collected, some of these subdialogues were closed with a direct answer. In other cases, a long series of exchanges occurred before the answer was arrived at. Only some short sequences contained a closing "O.K."

or other explicit indication from A. Almost all of the longer sequences ended with such a communication.

b. MULTIPLE USES OF O.K.

Robinson (1975) pointed out the use of "O.K." as an acknowledgment that the preceding message has been received. This is only one of four meanings this interjection took on in the dialogues. In particular, "O.K." was used at different times to mean:

* I heard you.

* I heard you and I understand.

* I heard you, I understand, and I am now doing (or will do) what you said.

* I'm finished (O.K. what next?).

Figure II-11 contains an example of each of these meanings.

O.K. -- I HEARD YOU:

E: Loosen the motor bolts and slide the motor toward the pump.

A: O.K. What's a motor bolt.

O.K. -- I HEARD YOU AND I UNDERSTAND:

E: I need to know what kind of wrench you're using.

A: O.K. (no further spontaneous communication).

O.K. -- I HEARD YOU, I UNDERSTAND, AND I AM DOING WHAT YOU SAID:

E: First loosen the two allen head setscrews holding it to the shaft, then pull it off.

A: O.K.

A: I can only find one setscrew. Where's the other one.

O.K. -- I'M FINISHED:

A: O.K. All the bolts are off.

Figure II-11. DIFFERENT USES OF "O.K."

Each of these uses of "O.K." requires a different response from the hearer. Often the indication of which one is meant comes from the next statement in the dialogue. Although the time between the preceding statement and the "O.K." is often a clue to which meaning is intended, it is not always a reliable indication. For example, if the expert directs the apprentice to a task requiring a lot of time to complete, then an immediate "O.K." cannot mean the task is done. However, if the apprentice misunderstands and does a shorter-length task, then his "O.K." may mean he is done.

The main problem in interpreting an "O.K." is to distinguish the first three uses of "O.K." from the fourth. In the task domain, use 2 never occurred where use 3 was applicable (though one can imagine it in some situations, like a child being told to make his bed). The distinction between use 1 and uses 2 and 3 is immediately evident from the utterance that follows the "O.K." Furthermore, no ambiguity problems can arise from this distinction since it does not have any impact on change of focus. Use 4, on the other hand, does indicate a change of focus: once a task is completed, focus shifts to a new task. At present, the best strategy for interpreting "O.K." seems to be to wait for the next utterance to determine if a shift of focus is intended.

Figure II-12 contains a dialogue fragment illustrating one of the problems that arise from the use of "O.K." for closing a subdialogue. In line (4), A indicates completion of part of the 'open-valve' task. In line (5), E gives the next task; he has closed the whole 'open-valve' task. However, from line (6) it is clear that A thinks another subtask may be involved in the 'open-valve task'. To answer (6), E must re-open the closed (for him) 'open-valve' task and its corresponding subdialogue.

(1) E: Open the top of the valve and let the water out. Just open the faucet up on top. Just like you were going to turn the water on.

(2) A: Oh, like I'm going to turn the water on. O.K.

(3) E: Now, that'll relieve the pressure.

(4) A: O.K. some water came out.

(5) E: Now the next thing you do, you take an allen wrench . . .

(6) A: Do I leave it on or turn it back off?

(7) E: It doesn't make any difference.

(8) A: O.K.

Figure II-12. A MISUNDERSTOOD "O.K."

c. MULTIPLE OPEN SUBTASKS

The preceding discussion has centered around the idea of only one task being under discussion at any time and hence providing focus for the dialogue. However, some examples of more than one focus being active at a time were encountered in the dialogues analyzed. These fell into two categories: hypothetical and competition. In the hypothetical case, one task was being performed, but a future one was being considered. Although the task being performed was a lengthy one, there were no problems, so the apprentice asked about how to perform some future task, or what would happen if some task were performed differently. In all such instances, both A and E seemed comfortable with the multiple foci. In the competition case, however, E and A appeared to be competing for who would determine what would get discussed. Although both could handle the dual foci, at least one of the two always seemed annoyed. The annoyance was manifest both through repetition of statements and from the tone of message communicated orally. In all cases, the maintenance of multiple foci did not last more than two or three exchanges.

4. REFERENCE

The importance of the link between task structure and dialogue structure and the need for representing focus of attention are most clearly seen when examining the use of definite noun phrases. Determining the information and processes needed to identify the object referred to by a definite expression (i.e., resolving a reference) was a primary goal of the dialogue analysis. For some of the analysis it will be useful to distinguish two kinds of definite noun phrases: pronouns and nonpronominal definite noun phrases. In the following discussion, the term DEFNP will be used to refer to nonpronominal definite noun phrases only. The basis of this distinction arises from the different amounts of information carried by DEFNPs and pronoun references and from the different processes needed for resolving these two kinds of reference (see Chapter IV, Section B for more details).

This distinction may be compared to the distinction that Chafe (1976) makes between givenness and definiteness. Givenness (as in the given/new distinction of Halliday, 1967) relates to an item being in the consciousness of the hearer (Chafe, 1974; Chafe, 1972, uses the term "foregrounded"). Givenness is usually expressed by pronominalization or low pitch or weak stress. Definiteness concerns whether or not the speaker believes the hearer can select the referent from among all the other items he knows about. In English, definiteness is expressed through the definite determiner. Although focus, as described in this report, affects both givenness and definiteness, its influence is different for each. Focus is always a factor in the resolution of DEFNPs. It provides the set of objects from which the item being referred to must be distinguished. In contrast, the importance of (global) focus to pronoun resolution is only evident when a shift of focus establishes a different set of items as given. It is this use of global focus that enables the resolution of pronominal references that refer back over long portions of dialogue (e.g., see Figure II-7.

There are several ways in which the object referred to by a

DEFNP may be evident in the discourse context. The simplest case is when the object was explicitly mentioned in a preceding utterance. DEFNPs are also used to refer to objects that are not explicitly mentioned in the discourse but are so closely coupled to some object which has been that they can be easily identified by the hearer (see Chafe, 1972, 1974; Karttunen, 1968). Such objects may be considered "implicitly focused." For example, in the sequence,

E: Are you using the socket wrench?

A: Yes. The socket fell off ...

"the socket" has not been previously mentioned but is clearly identifiable once "the socket wrench" is identified.

A problem of particular interest in resolving references is determining where to search for referents: how far back in the dialogue is it necessary to go? Searching the whole preceding discourse may be quite time consuming. The necessity of considering implicitly focused concepts as well as those explicitly mentioned makes searching the whole dialogue unfeasible.

Although the time between utterances (or its analog, distance, in a text) affects whether or not a definite reference can be used, it is not clear how much discourse can occur before an object ceases to be in focus.* Discourse structure provides a clue to the solution to this problem. In a structured discourse, both time and structure need to be taken into account in resolving references. Most language understanding systems use some time measure as the sole basis for considering objects as referents of definite noun phrases. The system of Norman et al. (1975) has a concept of <u>working memory</u>, which could be used to accommodate structure, but is not. Objects must be explicitly rementioned in order to stay in this memory. These systems have dealt either with edited text or with unstructured tasks. For example, in Winograd's (1971) block manipulation task any instruction can be followed by any other. Although there is utterance-to-utterance

----

* Chafe (1972) discusses this problem in relation to foregrounding.

cohesion in such discourses, there is little global cohesion. This is exactly what happens in the data base domain, too. Time provides the only basis of reference in such cases, but this use of language is atypical and differs markedly from the language people use in direct communication. The examples presented above in Section D.2 of this chapter illustrate that time alone is not a sufficient determiner. Whole segments of dialogue may be skipped over, and objects not mentioned for a long time may be referred to by definite noun phrases, even by pronouns.

Examination of the references occurring in the task dialogues showed that references operate within subdialogues. That is, as long as a subdialogue is open, objects introduced into it are referred to by definite noun phrases. We consider these objects in focus, because the successful use of definite reference depends on the object referred to being in the focus of attention of the hearer. When a subdialogue is closed, the objects inside it leave focus and require different kinds of references (unless the whole subdialogue is reopened or they are first reintroduced in some other subdialogue). When a subtask is completed, the definite noun phrases may refer to objects in higher level tasks. For illustrative purposes, consider the simple tree task structure of Figure II-13. When task T6 is completed, there is a return to the focus of T2 and possibly directly to T1. Objects that participate only in T4 or T5 are not in focus. Similarly, objects in T2 or T4-T6 cannot be directly referenced from T7 or T8. When T8 is completed, there may be a 'pop' up to T3 or T1.

Although most references can be resolved in terms of the preceding utterances within the subdialogue, this is not of itself sufficient for establishing the existence of the segmentation of dialogues. Since the preceding utterances in the same subdialogue are also the most recent utterances, an alternative explanation of the reference resolution process, which is simpler, is that the referent is the most recently mentioned object matching the DEFNP. The references that occur after a subdialogue has been closed illustrate the

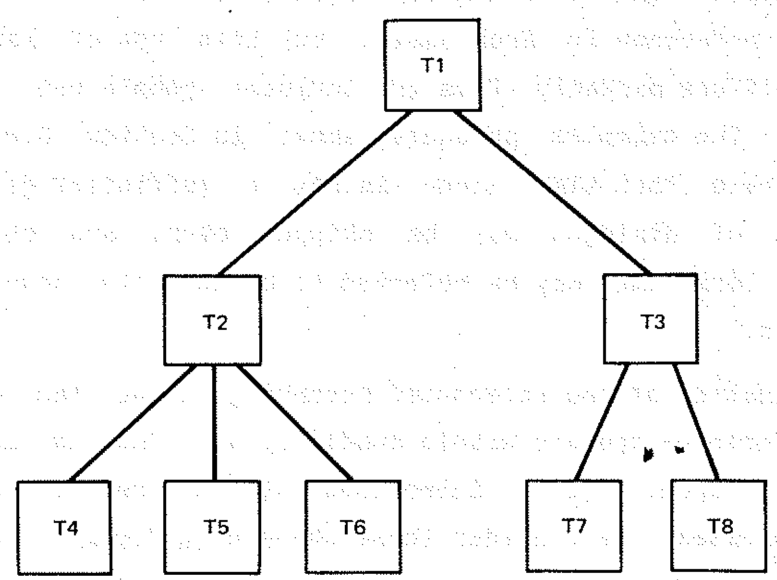


Figure II-13. A SIMPLE TASK MODEL FOR ILLUSTRATING DIALOGUE POPS

insufficiency of this recency explanation. When a subdialogue is closed and focus shifts back to a higher level task, the objects in that higher task get referred to definitely even though they have not been mentioned recently. The use of DEFNPs in this way might be expected, but the use of pronouns for objects not recently mentioned is certainly striking. The example in Section D.1.a is hard to account for if task and dialogue structure are ignored.

A second indication of the necessity of considering structure in any reference resolution process comes from the use of plural DEFNPs. Consider again the task structure of Figure II-13 and suppose that some bolts, B2, are involved in task T2 and another set, B3, in task T3. Then, even if some utterance in the end of the subdialogue for T2 contains the phrase "the bolts", any reference to "the bolts" once T2 is closed and T3 opened will be taken to mean the set B3. This is true with a combination of singular and plurals also. So if T2 involves a single bolt, B, the phrase "the bolts" inside of T3 will not be taken to include B. As an example, consider the dialogue fragment in Figure II-14. Even though the two screws have been mentioned within one exchange of the wheelpuller screw, the phrase "the screw" is totally unambiguous.* Completion of the tightening task has closed one subdialogue and removed those two screws from focus.

A: How do I remove the flywheel?

E: First loosen the two allen head setscrews holding it to the shaft, then pull it off.

A: The two screws are loose but I'm having trouble getting the wheel off.

E: Use the wheel puller. Do you know how to use it?

A: No.

E: Loosen the screw in the center and place the jaws around the hub of the wheel, then tighten the screw . . .

Figure II-14. EFFECT OF SHIFT IN SUBDIALOGUE ON DEFNPS

In this connection, the dialogues reveal that people are sensitive to the distinction between singulars and plurals. In the subdialogue of Figure II-15, E indicates the ambiguity of the phrase "the allen screw" by pointing out the fact that there are two. (In addition, he indicates that they both need to be tightened).

5. DESCRIPTIONS

The previous section described the role of structure, as a reflection of focus, in the resolution of definite noun phrases. In this section, we examine a companion problem. In a language understanding system, the problem of generating a good description of an object is just as important as the problem of identifying an object from its description. The linguistic description of an object must distinguish it from all others in the context of speaker and hearer in

---

* The modifying phrase, "in the center" adds information about where to find the screw, but is not necessary to avoid ambiguity. This may be seen by considering the same dialogue fragment with "three" replacing "two" in the phrase "two allen head setscrews" in the second utterance.

E: Check the alignment of the two pulleys before you tighten the setscrews.

A: Yes. I'm doing that now.

E: O.K.

A: Tightening the allen screw now.

E: O.K. Thank you.

A: That's finished.

E: By the way, there are two setscrews.

Figure II-15. SINGULAR/PLURAL DISTINCTIONS

order for any communication to be possible. For this reason, the descriptions that appeared in the dialogues were examined in an initial attempt at characterizing the information and processes involved in generating descriptions.

a. <u>SPECIFICATION</u>

Olson (1970) has shown that the description of an object changes depending on the surrounding objects from which it must be distinguished. So, for example, the same flat round white object was described as "the round one" when a flat square object of similar size and material was present, but as "the white one" when a similarly shaped but black object was present. The importance of contrast for distinguishing objects is well established in vision research (e.g., Gregory, 1966; Tenenbaum, 1973; Garvey, 1976). Comparison of differences has also played a crucial role in computer programs that reason analogically (Evans, 1963; similar strategies are used in Winston, 1970).

It is clear from the task dialogues and from other data (Freedle, 1972) that the description of an object seldom contains only the minimal amount of information necessary to distinguish it. Descriptions, like the rest of language, are redundant. (Olson, 1970, p.266, comments on this phenomenon and on the need for further investigation of it.) What appears to be the case is that the speaker describes an object not in the minimum number of 'bits' of information, but rather in a manner that will enable the hearer to locate the object meant as quickly as possible. Clear distinguishing features (e.g., color, size, and shape) are part of a description precisely because they eliminate large numbers of wrong objects and hence help the hearer to isolate the correct object more quickly.

The use of redundant information (and not just distinguishing information) to speed up the search for a referent can be easily seen from an example. If A asks "What tool should I use?", the response, "The red-handled one.", is not satisfactory even if there is only one red-handled tool in the workstation,*. Processing such a description requires considering too many alternatives. Although A might eventually find the tool, he would certainly question E's choice of description. "The red-handled screwdriver" is more helpful, because it limits the search to screwdrivers. Olson's descriptions were probably as minimal as they were because of the bare environment in which the distinguishing had to be done. In giving a description that minimizes search time (i.e., the time it takes the hearer to determine the referent of a referring expression), a balance must be reached. Too much information is as harmful as too little, since all parts of the description must be processed to make sure the object is the correct one. Furthermore, the hearer may wonder whether he is mistaken if he thinks he has determined the referent but there is more description to process. Rather than minimize either just the communication time (including processing of the description) or just the search time, the combination of communication time and search time must be minimized. A description can be redundant only to the degree that redundancy speeds up the search; anything further is confusing.

Because the goal of most descriptions in the task dialogues was to enable the hearer to locate an object, the descriptions in the task dialogues were, to some extent, procedural. Either

----

* Even if there are only a few tools in the workstation, this response is awkward, though perhaps satisfactory.

implicitly or explicitly, they described how to locate an object, rather than what the object was in general. For example, the response to "What's a nutdriver?" was "It looks like a screwdriver and is in the yellow case by the wall", rather than the (nonprocedural) definition description, "A tool with a handle on one end and the end shaped to fit over a nut, used for tightening and loosening nuts." This combination of description of the object itself coupled with locational information was quite common in response to questions (e.g., "What's an x?"). In a sense, the speaker was saying, "Keep these properties in mind and look at place Y." It is interesting that the descriptions of the object itself preceded the locational information more often than following it. The location provides a narrowing of focus. What is not clear is why this narrowing occurs after and not before the object properties are given. Possibly, even though narrowing of focus is useful for identification, the question "What is an x?" demands mention of inherent properties of the object x first.

b. <u>CATEGORIES OF FEATURES</u>

The features used in the descriptions of objects in the dialogues fell into four categories: physical characteristics, location, analogies, and function. A class name of the object always appeared in initial introductions, but it is not included in this list. Otherwise, the list contains items used in initial introductions as well as in response to questions concerning object identification.

The physical characteristics of the object itself included color, shape (often including the word "shape" as in "the little half-moon shaped part"), size (either absolute or relative), and material of which the object is composed (e.g., "metal").

Location, both physical and in time, of the object were often used. Physical location was specified in response to a "What's a" question. Time references occurred when an object description was embedded in some higher-level statement. For example, "Use the two screws you mentioned earlier", "... the cover to the wires you were working with earlier".

Analogy provides a lot of information in a small package. It occurred most often when any other description would have been long and involved. In addition to the above screwdriver example, there was "it looks like a pocketknife", "it looks like ears sticking out", or "it looks like a y".

Closely related to analogy is the use of "function" to describe an object. Functional descriptions also enable bypassing other more complex descriptions (e.g., of shape). The statement "it is used for doing x" or "it has the right shape for doing x" may be used to communicate complex shapes and structures. The success of such descriptions depends on the hearer's ability to determine what such an object is like, or to pick out the object from a set. The combination of analogy and functional description often occurs with the phrase "it looks like it does x" (and, in fact it does do x!). Functional descriptions implicitly convey this concept of "looks like" even when it is not explicitly stated.

Finally, there is a set of miscellaneous distinguishing features that are best characterized as the absence of something usual or the presence of something atypical. For example, "[you can tell where it goes] by where there is no paint", or "the side with writing on it".

c. PERSPECTIVE

In order for a description to work, it is crucial that it take into account the hearer's point of view. The role of the hearer's physical location is well established. The well-known 'Empire State Building' question (you give a different answer to the question "Where is it" to a person in Moscow and a person in New York City) illustrates this point. In the task domain, words like "left" and "front" must take into account both canonical orientations (the front of a car is the same no matter where you stand relative to it) and hearer orientation.

There is also a nonlocational aspect of the hearer's orientation. Descriptions must be given to a level of detail pertinent for the hearer's skill level. Concepts unfamiliar to the hearer may be introduced, but they must be explained in terms familiar to him. Indications of such sensitivity to user skill in the dialogues came both from the level of detail of task described and from the description of parts and tools. These are evident in the differences between naive apprentice and experienced apprentice dialogues. The same object might be described differently to a naive apprentice and an experienced one. Alternatively, one way of determining skill level is from the descriptions that must be explained or elaborated upon.

E. IMMEDIATE FOCUS: ELLIPSIS

The preceding analyses have concerned how the global focus in which an utterance occurs affects the interpretation of the utterance. This section examines a more local aspect of focus: how the immediate focus of one utterance affects the interpretation of the following utterance. In particular, the use of immediate focus in the interpretation of elliptical sentence fragments will be examined.

Elliptical sentence fragments are phrases that function in context as full sentences, although they are only parts of what would constitute a complete sentence. The use of fragments in the task dialogues was quite different from that in the data base dialogues. In the data base dialogues, the fragments all formed part of a series of questions. In each case, the meaning of the fragment could be obtained by finding a similar phrase in the preceding question and substituting the new phrase for the old. An algorithm for handling this kind of fragment is presented in Chapter V. In the task dialogues, fragments occurred as responses to previous requests for information and as qualifying phrases on immediately preceding utterances. As a result, the fragments in the task dialogue were patterned on and needed to be interpreted in terms of the immediately preceding utterance.

The most common form of fragment used in response to a request was the one that fit into the WH-phrase of the preceding question. This occurs, for example, in

E: What tools are you using?

A: My fingers.

A's response "my fingers" matches the phrase "what tools". Arriving at a complete utterance requires a set of standard syntactic transformations like changing the "you" to an "I" and changing word order.* Secondly, a fragment may occur in response to a choice question; this is the case in the pair:

E: Does the side of the pump pulley with the holes face away from the pump or towards it?

A: Away from the pump.

(In a sense, this is a restricted form of a WH-question. The WH-phrase is replaced by a choice phrase. This could be phrased as a "Which way ..." question).

The use of a fragment to qualify a preceding utterance is illustrated by the sequence

E: Place the key in the slot.

A: Flat side upward?

The apprentice is really asking, "Should I place the key in the slot with the flat side upward?"

In each of these cases, the full sentence needed to get an interpretation of the fragment can be derived from transformations on the preceding utterance. When fragments appear as answers to questions (the first two examples), the questions themselves provide an indication of where the fragment fits in. In the last example, this is not the case. There is no place marked by a WH-phrase to indicate a slot for the fragment. Instead, the fragment fills an optional slot in the sentence structure (for verb complements), which was not used in the first utterance of the pair.

----

* Robinson (1975) contains a description of the transformations required to interpret this kind of fragment.

F. SENTENCE LEVEL ANALYSES

1. KINDS OF UTTERANCES: PURPOSE AND TYPE

There are marked differences in the kinds of utterances occurring in the task dialogues and in the data base dialogues. Syntactic differences include such things as differences in the number and kinds of WH-questions and differences in the ratios of questions, imperatives, and declaratives. Several of these are enumerated in Section IV, The Language Definition, in Walker et al. (1975). Differences occurred along two other dimensions that we will call utterance purpose and utterance type. Utterance purpose refers to the overall reason for the utterance (e.g., to convey task information). Utterance type refers to the form in which the utterance conveys information (e.g., a request or a response). It is important to distinguish utterances along these two dimensions both for detecting where an utterance fits in the discourse structure and for setting up expectations or determining a response. The purpose of the utterance establishes the kind of subdialogue the utterance belongs in. The utterances in the dialogues that were examined were used for three purposes: to convey task information, to convey sensory information (as a substitute for some missing sensory channel), and to check the communication channel. The type tells the role of the utterance in the subdialogue. Five types of utterances occurred in the dialogues: requests, responses, reports, imperatives, and acknowledgments.*

Almost all of the utterances in the data base dialogues are questions whose purpose is to get information out of the data base (that being the nature of a data base query). In the task domain, there was a wider variety of utterance purposes and also of utterance types. Utterances served three purposes. The majority were task related; they involved such things as describing task steps, identifying parts and

----

* An examination of other kinds of dialogues would clearly yield both other dimensions in which differences occurred and other categories in both of these dimensions. The concept of speech act (Searle, 1969) is particularly relevant to the dimension of utterance purpose.

tools, and describing progress on a task. Task specific utterances have places in the dialogue structure hierarchy that correspond to the related task's position in the task hierarchy. General task-related questions go into the dialogue hierarchy immediately below the subdialogues they occur within. Secondly, utterances served as sensory substitutes; these included requests from E, such as "Show me ...", and statements by A, such as "I'm pointing at ...". Finally, some utterances served to establish that the communication channel was still open, for example, the question "Can you hear me?" In addition, several of the "O.K.s" served as channel checkers as well as providing task information.

Of the five types of utterances, most were requests for information or responses to such requests. These included questions about task steps, which tool to use, how a task step was progressing, and the answers to such questions. Often, however, information was offered without being requested. Some apprentice utterances were reports of progress, quite similar to answers to requests like "What are you doing now?" but different in that they also indicate A's need to communicate his progress. Similarly, E imperatives are quite similar to answers to the question "What should I do next?" but convey E's feeling of task progress rather than A's. Both reports and imperatives are often followed by utterances that serve merely to acknowledge that a message has been received. "O.K." and "Yes" often function in this way.

Each type of utterance may be followed only by a subset of the other types, as shown in Figure II-16. (The two 'special' entries are described below). Responses are an exception: they may be followed by an utterance of any type. This is a reflection of the fact that a response is a local closure. (Correspondingly, the table shows that a request can be preceded by any of the utterance types, reflecting the local opening aspect of requests.) Imperatives and reports may be followed by either acknowledgments or combinations of an acknowledgment and a request. In the latter case, if the request immediately follows

UTTERANCE TYPE \ REPLY TYPE	REQUEST	RESPONSE	REPORT	IMPERATIVE	ACKNOWLEDGEMENT
REQUEST	S	•			
RESPONSE	•	S	•	•	•
REPORT	•			•	•
IMPERATIVE	•				•
ACKNOWLEDGEMENT	•		•	•	

• Indicates Reply Type May Follow Utterance Type
S Indicates Special Kind of Follow-On.

Figure II-16. CORRESPONDENCE BETWEEN UTTERANCE AND REPLY TYPES

the imperative or report, the acknowledgment is implicit and may be omitted. Typical requests following imperatives involve questions about parts of the task; typical requests following reports involve checking that some subtask has been done correctly. Reports may also be followed by imperatives. Again, the acknowledgment is implicit.

With one exception, requests and responses come in pairs. In the usual case, requests are followed by a response. The response may be followed by anything other than another response. The exception (see the 'special' entries of Table 2) occurs with embeddings of questions and answers as in the dialogue of Figure II-17. In this case a request is followed by another request. Correspondingly, the response is followed by another response. Finally, acknowledgments may be followed by imperatives, requests, or reports. In a sense, an acknowledgment signals that the acknowledging person is ready to receive another message.

A: Should I put the belt on next?

E: Are the setscrews tight?

A: Yes.

E: (OK)(Then) you can put on the belt.

Figure II-17. EMBEDDINGS OF REQUESTS AND RESPONSES

Figure II-18 contains a segment of dialogue containing the five types of task utterances. In this example, each of the imperatives and reports is followed by an acknowledgment. In one case, the acknowledgment following the imperative is immediately followed by a request. In this case, the acknowledgment itself is optional. There are similar examples in other dialogues of imperatives being followed by requests for information. In this respect, reports resemble imperatives; although in this fragment all reports are followed only by acknowledgments, it is also possible to follow them with requests or with a combination of acknowledgment and request.

The utterances in the dialogues vary along another dimension that might be called <u>response constraint</u>: the amount of influence an utterance has on the form and content of the utterance that follows. It is difficult to identify all of the factors influencing this dimension and although many utterances are clearly marked, others are neutral with respect to it. Consider the two sets of utterances in Figure II-19. Utterance A1 is neutral with respect to response constraint. Either party could take over the dialogue at this point; neither the form nor the content of the next utterance is indicated. Utterance B1, on the other hand, puts responsibility for the form of the following utterance on E. Both utterances A2 and B2 are neutral; they are quite similar in what they convey. The responses to them are quite different, though. Utterance A3 exhibits strong influence over the response to it. One of the two alternatives must be picked or some explanation of why neither was given. The preferred response is a simple phrase choosing one of

E: The pump pulley should be next.
IMPERATIVE (this direction follows a report indicating completion of the preceding task)

A: Yes uh does the side of the pump pulley with the holes face away from the pump or towards it?
ACKNOWLEDGMENT FOLLOWED BY A REQUEST FOR INFORMATION

E: Away from the pump.
RESPONSE

A: All right.
ACKNOWLEDGMENT

E: Did you insert the key, i.e., the half-moon shaped piece?
REQUEST

A: Yes I did.
RESPONSE

E: Be sure and check the alignment of the two pulleys before you tighten the setscrews.
IMPERATIVE

A: Yes I'm just now fiddling with that.
ACKNOWLEDGMENT FOLLOWED BY A REPORT

E: O.K.
ACKNOWLEDGMENT

A: Tightening the allen screw now.
REPORT

E: O.K. Thank you.
ACKNOWLEDGMENT

A: That's finished.
REPORT

Figure II-18. UTTERANCE TYPES IN A SAMPLE DIALOGUE FRAGMENT

the two options. Utterance B3 is harder to classify. It does not seem entirely neutral since it indicates no choice or narrowing of alternatives by A, but it is not as clearly an abdication as is B1. Imperatives and yes/no questions exhibit strong influence over the form of responses to them.

Set A: 1. A: I've finished installing the strap.

2. E: The pump pulley should be next.

3. A: Yes uh does the side of the pump pulley with the holes face away from the pump or towards it?

Set B: 1. A: Now what should I do?

2. E: Install the pulley on the shaft.

3. A: What is the first thing to do in installing the pulley?

Figure II-19. TWO SIMILAR DIALOGUE FRAGMENTS FOR COMPARING RESPONSE CONSTRAINT

Subjective evaluation of the dialogues indicates the lack of response-constraining utterances from apprentices who were unsure of the task, and a higher presence (and more constraints) in the dialogues with experienced apprentices. Before this kind of information can be utilized in a language understanding system, more analysis is needed both on how the information is conveyed and how it is used. That the information is important is clear since it provides one indication to the hearer of the extent of the speaker's knowledge about the problem.

2. LEXICON

Analysis of the words occurring in the dialogues is necessary to determine both the size of lexicon and the breadth of concepts present. A description of the kinds of words found in the data base dialogues may be found in section IV, The Language Definition, in Walker et al. (1975) In this section of this report, only the task-oriented dialogues will be considered. In the following analysis, different forms of the same root were not distinguished. For example, "bolt", "bolted", and "bolts" were treated as identical.

One of the most interesting results was that only 520 different words occurred in the four core dialogues. (There were approximately 8000 words in the dialogues -- not including occurrences of the articles "a" and "the"). Malhotra's (1975) results confirm our finding that only a small number of words seem to be required for communication in a limited domain. This finding is different, but not inconsistent with, the underlying tenet of Basic English. Basic English maintains that a small number of words are sufficient to convey any idea. Our results suggest that, in a given discourse context, even if people are allowed unrestricted use of language, they will use only a small number of words.

Of the 520 words occurring in the four core dialogues, only 100 are used more than ten times. Although this suggests that most of the communication is achieved by a small core lexicon, it is important to realize that many words occurring only once or twice are crucial to conveying events that occur and objects that are used only a few times. Examples are "clamp" (as in "clamp the cylinder head casting ...") and "lockwasher". Half of the words are unique to a particular dialogue. Many of these words are simple differences in expressing similar concepts. For example, "slip" was used only in one dialogue (in "The aftercooler is too long to slip easily into place"); other dialogues used "slide" to convey similar situations. In contrast, 90 words occur in all four of the dialogues. Of these 90, 74 are among the 100 words used more than ten times. A list of these 90 words appears in Figure II-20. The starred words were used fewer than ten times. Since the number of different words in each dialogue ranged from 236 to 303, approximately one-third of the words in each dialogue occurred in each of the other three dialogues as well. If the dialogues are separated into pairs according to task, then the pairs in each grouping share over half of their words (142 and 154). These results suggest both a large overlap in concepts, and a large variety in how concepts are expressed.

The two 'naive apprentice' dialogues share 60% of their words. Correspondingly, only 20% of the words in each of the naive apprentice dialogues are unique to that dialogue. The other two dialogues each had approximately 30% unique words.

If we add a fifth dialogue to the analysis that covered a

a	*again	all	allen	*also	and	at	back
be	belt	bolt	box	by	can	do	*easy
*enough	*fit	from	get	go	good	*hand	*hard
have	hold	how	I	if	in	it	just
key	know	*like	*long	loose	*more	motor	no
not	now	of	off	ok	on	one	or
out	*over	place	plate	please	pulley	pump	put
screw	see	*seem	should	show	slide	so	*some
tank	that	the	then	there	they	tight	to
top	*toward	turn	two	up	use	way	we
what	*when	where	which	will	with	*work	wrench
you	yes						

Figure II-20. WORDS OCCURRING IN ALL FOUR DIALOGUES

different task but also used an inexperienced apprentice, similar results occur. The number of different words increases to from 520 to 580. Again, over half of the words are unique to some particular dialogue. Only 61 words are shared by all of the dialogues. These words, grouped by category, appear in Figure II-21. If we consider the three naive apprentice dialogues, the number of shared words is 88. Twenty-six of these words, listed in Figure II-22, are missing from at least one of the experienced apprentice dialogues. The words shared by the naive apprentice dialogues suggest two characteristics of these dialogues. First, words applicable to low level task descriptions (e.g., specific simple tools, like screwdrivers) get used more often in these dialogues, because more low level tasks get talked about. Second, the presence of "thing" and "tool" on the list suggest that extremely general terms are also more likely to occur, probably because more specific ones are not known to the naive apprentice.

It is dangerous to generalize from such a limited sample; speaker idiosyncrasies cannot be filtered out. However, there are some clear trends, giving indications for system building and suggestions for future studies. Approximately 140 of the words in the dialogues were task-dependent words; as the task shifts, the need for these words changes. Although the overlap of words is interesting, it is important not to ignore the large number of words that are unique to some one of

AUXILIARY AND PRO-VERBS

be	can	do	have	should	will

DOMAIN-RELATED WORDS

bolt	box	fit	go	hold	place
plate	pump	put	show	tight	top
turn	use				

FUNCTION WORDS

a	also	and	by	from	how
if	in	no	not	now	of
on	out	over	so	that	the
then	there	to	up	what	when
which	with				

MISCELLANEOUS

good	just	like	ok	one	please
see	two	way	yes		

PRONOUNS

it	they

SPEAKER/HEARER IDENTIFIERS

I	we	you

Figure II-21. WORDS OCCURRING IN ALL FIVE DIALOGUES, GROUPED BY CATEGORY

align	around	both	bottom	but
down	end	face	first	groove
hammer	metal	onto	other	remove
right	round	screwdriver	shaft	side
slot	socket	sure	take	thing
took	wheelpuller			

Figure II-22. WORDS IN ALL NAIVE APPRENTICE DIALOGUES BUT MISSING IN AT LEAST ONE OF THE OTHERS

the dialogues. The overlap means that, for a given task, a relatively small number of words (significantly fewer than the 1000 often taken as a benchmark for a computer language understanding system; e.g., see Newell et al., 1973) will suffice to cover almost all of what almost every speaker says. The 'unique words' indicate that although many of the concepts being expressed by the performers of the task are the same, there is a wide variability in just how to express those concepts.

Analysis at the lexical level is important, but it must be used in conjunction with higher-level syntactic, semantic, and discourse analyses.

G. MISCELLANEOUS OBSERVATIONS

There were several areas that are important for understanding the choices made in generating an utterance and the information conveyed by that utterance but for which only limited data are available from the dialogues. There were clear indications of the influence of one speaker on another, differences in formality, and influence of apprentice skill level.

One question of importance in constructing natural language-understanding systems is the influence of the system's output on the language with which it has to deal. For example, how the form in which the system asks a question influences the form of the response. Since only two different experts were used in the task dialogues, only one of whom worked with more than two apprentices, it is hard to conclude much from the dialogues. Still there are indications that apprentices adopt the experts' language. Adoption of common names is the most frequent example. "The half-moon shaped piece" gets referred to as "the (woodruff) key" once the name is introduced by the expert. Similarly "the screws holding the pulley on" become "the (allen head) setscrews". The transference may be from the apprentice to the expert as well. In one dialogue with an experienced apprentice, the expert adopted terms (such as "pressure register") used by the apprentice.

One of the confounding factors in determining language influences is that in the case of two of the dialogues, the apprentices thought that the expert was a computer. In both, the apprentice's language is more 'formal' than in the other dialogues. In the dialogue in which the apprentice is most formal, the expert's responses are more formal. It is not clear in this case how much of the difference is due to the expert's speech and how much to the apprentice's preconceptions about what a computer could understand. Although there are clear differences between the <u>computer-expert</u> dialogues and the others, it is hard to point at exactly what aspect of an utterance makes it seem more formal. For example, the utterance,

> "Is it correct that the strap is attached to the pump by one of the cylinder head bolts?"

seems more formal than a question that starts simply, "Is the strap ...". Similarly, "I've finished attaching the tubing to the elbow." is less formal than "The elbow and tubing installation is completed." Unfortunately, there are too few data here to decide what is speaker-idiosyncratic and what comes from preconceived notions of computer capabilities. Still, there are enough indications of differences when a computer is thought to be a participant in the dialogue to mark this as an important area for study. Furthermore, although the apprentices thought they were being helpful by being more formal, in fact the resulting sentences often were more complex and would have been harder for a computer language understanding system to process. However, it is possible that such differences would disappear after repeated exposure to a system that understood natural language.

Experts' instructions to apprentices varied according to the perceived skill level (previous knowledge about similar tasks) of the apprentice. In almost all cases, the expert did not know initially how skilled the apprentice was. Although the first few instructions to all apprentices were quite similar, subsequent instructions varied substantially. Not only was the amount of detail presented different but also the way in which instructions were given. Dialogues with inexperienced apprentices contained more requests and fewer spontaneous reports. In the dialogues with more experienced apprentices, there were more imperatives that checked that steps had been done and fewer giving directions. The clearest example of an expert moderating his interactions as he determines the skill level of an apprentice is in a dialogue with an experienced apprentice. Up to a particular point in the dialogue, most of the expert's utterances are directions or answers to requests. Then the expert starts to give a direction and changes his 'tone'. He types

"OK. Tig XXX OK. Make sure ... are tight."

The XXX indicates an erasure to the monitor. The expert changes from directing the apprentice to perform a step (i.e., tighten the bolts) to asking him to check that the step has been done. The change indicates that the expert needs to know the step has been done rather than that he thinks the apprentice needs to be told what to do. The important question for builders of computer systems is what information the human expert is using to base his impressions of skill level on. There are clearly several factors involved. A comparison of the few dialogues we have collected indicates that the apprentice's terminology, the level of detail of instruction the apprentice asks for, and the apprentice's own indication of skill level contribute. More data need to be collected and examined to determine how skill impressions are transmitted and generalized.

Finally, there were a few examples in the dialogues of the kinds of ambiguity that people are and are not willing to tolerate. For example, the phrase "allen bolts" in the context of attaching the pump pulley was accepted as meaning "allen head screws". Quite often the use of "nut" and "bolt" interchangeably was accepted, but in the dialogue of Figure II-23 the misuse of "bolt" is not acceptable since it causes confusion about which task is being done.

A: Should I unscrew at the top of the airhose or at the bottom and which of the bolts at the bottom?

(by bolts, A means nuts)

E: Loosen the pipe at the tank (bottom) end and unscrew it completely at the top end.

A: End of what, the pipe or the bolts?

("bolts", really nuts)

E: We're working on the pipe now. Don't worry about the bolts yet.

Figure II-23. BOLT/NUT CONFUSION

H. CONCLUSIONS

The purpose of this part of the research was to determine the scope of discourse phenomena in dialogues with computers, and to provide a basis for initial attempts to incorporate discourse capabilities in a language understanding system. Dialogue analysis is an important tool. Close examination of a single dialogue reveals a multitude of language phenomena that present problems for current language understanding systems. However, it would be a mistake to look only at a single dialogue, because it is not possible to separate out idiosyncratic behavior. Performing statistical analyses on a large variety of dialogues suffers from the opposite problem; it ignores the variety of language phenomena by concentrating on similarities. It seems particularly appropriate to focus the analysis on a small number of features by collecting several dialogues on similar tasks. In examining dialogues, the emphasis can be either on different ways of expressing the same idea (e.g., how different people describe the same complex operation) or on different uses of the same linguistic device (e.g., what kinds of conjunction appear).

The dialogue analyses reported here provide some initial data on the characteristics of language that occurs in task-related communication between a person and a computer. There are many dimensions along which much further analysis must be done. For example, further research is needed on the operation of focus in other tasks with different degrees of structure, the influence of one speaker on another (from word choice to the form of response), and how people handle ambiguity and error. The development of strategies for generating descriptions and resolving ambiguities in a language understanding system can be aided greatly by examining successful and unsuccessful occurrences of these phenomena in natural communication.

III FOCUS SPACES: A REPRESENTATION OF THE FOCUS OF ATTENTION OF A DIALOGUE

CONTENTS:

A. INTRODUCTION

This chapter describes a representation for focus in a language understanding system. The representation highlights that part of the knowledge base relevant at a given point in a dialogue by grouping together those concepts that are in the focus of attention of the dialogue participants. The representation has several distinguishing features. It is designed so that the structure of the dialogue (see Chapter II, Section D.2) can be represented and used in discourse processing. It is linked with representations of associated task situations. Finally, the representation has the potential for two kinds of extensions that are important to natural language understanding: focusing on different attributes of the same object under different circumstances and forgetting information no longer relevant to a discourse.

To meet the focus requirements of discourse described in Chapter II, a focus representation must satisfy the following four criteria:

1. separates out relevant part of knowledge representation
2. dynamically changes with the discourse.
3. accounts for implicitly focused items.
4. provides for reinvoking old foci of attention.

Criterion 1 indicates the most important function of the focus representation: to separate the basic knowledge base (i.e., the encoding of that portion of the world the system knows about) into subparts so that those items relevant to the current discourse are distinguished from all other items. The focus representation must highlight those items in the knowledge base that are relevant to the current discourse. This highlighting enables the system to access more important items first in its retrieval and deduction operations.

Criterion 2 reflects the dynamic nature of discourse. As successive utterances in a discourse are interpreted, the items in focus change. Shifts of focus occur both gradually with time and more drastically with change of topic. In addition, not only the objects in focus, but also the particular way of viewing them can change. For example, a doctor can be viewed as a member of the medical profession or as having a role in a family.

Criterion 3 reflects the fact that focusing on a concept entails focusing on other closely related concepts (see Chapter II, section D.4; also Karttunen, 1968). Specific mention of an object brings not only the object, but also certain associated items, into focus. For example, mention of "the house" brings into focus such associated objects as "the roof", "the living room", and "the owner". Parts of actions as well as objects may enter focus in this way. For example, "sewing a dress" brings into focus "cutting out the skirt." The focus representation and the processes that use it must account for these implicitly focused items as well as explicitly focused ones.

Criterion 4 is necessary because reference may be made to a focus situation that was in effect in the past (see Chapter II, section D). Although it is possible to consider reconstructing the situation (or, at least, constructing the most probable situation) from general information (and people may do this rather than actually recalling the situation directly), this would entail a substantial amount of computation and would introduce the possiblity of error.

This chapter describes a focus representation that satisfies these requirements (discussion of the shift mechanisms for Criterion 2 is postponed to Chapter V). The representation is divided into two parts; one part corresponds to explicit focus, the other to implicit focus. The explicit focus data structure contains those items that are relevant to the interpretation of an utterance because they have participated explicitly in the preceding discourse. Implicit focus consists of those items that are relevant because they are closely connected to items in explicit focus. Concepts that are implicitly focused are separated from those that are explicitly focused (i.e., they are not simply added to the explicit focus data structure) for two reasons. First, there are a large number of implicitly focused items, many of which are never referenced in a dialogue. Including these items in the explicit focus data structure would clutter it, weakening its highlighting function. Second, references to implicitly focused items are considered as indications of shifts of focus.

The focus representation presented here uses the partitioned network formalism developed by Hendrix (1975a,b). Section B gives a brief introduction to partitioned semantic networks. Sections C and D describe how focus may be represented using partitioning. In a computer system with a network based knowledge representation, the analog of the human process of identifying and retrieving an item from memory is identifying a piece of network structure. The central process is matching of network structures. Section E describes the general process of structure matching. Section F describes how focus is used to constrain the matching process. The representation of explicit focus was implemented and used in the SRI speech understanding system. The representation of implicit focus is designed, but has not yet been incorporated in a system. Section G describes several extensions to the implemented procedures.

B. PARTITIONED SEMANTIC NETWORKS

A semantic network is a directed graph: a set of nodes and a set of (labelled, directed) arcs connecting pairs of those nodes. Networks have been used in several previous language understanding systems (e.g., Quillian, 1968; Simmons, 1973). Conventions about the use and meaning of nodes and arcs vary. The networks described here use the conventions of Hendrix (1975a,b): nodes are used to represent objects, where object includes such things as physical objects, events, relationships, and sets. Arcs are used only to encode those binary relationships that do not change over time. Most arcs encode element, subset, or case relationships. Figure III-1 shows a sample semantic network. The node 'UNIVERSAL' represents the set UNIVERSAL, the universal set of all objects. (Single quotation marks denote node names.) The s arc from 'PHYSOBJS', the node representing the set of all physical objects, to 'UNIVERSAL' indicates that the set PHYSOBJS is a subset of the set UNIVERSAL. Similarly, the s arc from 'BOLTINGS' to 'SITUATIONS' indicates that BOLTINGS, the set of all bolting operations, is a subset of the set of all situations. The e arc from 'B1' to 'BOLTINGS' indicates that B1 is an element, or particular instance, of the set of all boltings. The other arcs emanating from 'B1' indicate that this particular bolting took place between times T1 and T2 and involved bolting the minor-part OBJ1 to the major-part OBJ2 with the bolts and nuts in B/N1. The de arcs from 'OBJ1' and 'OBJ2' to 'PHYSOBJS' indicate that OBJ1 and OBJ2 are distinct elements of the set of physical objects. The e arc from 'OBJ3' to 'PHYSOBJS' indicates that OBJ3 is also a member of that set although not (necessarily) distinct from OBJ1 and OBJ2. (The mutually distinct aspect of de arcs, and their analog for subsets, ds arcs, is used in the matching process.)

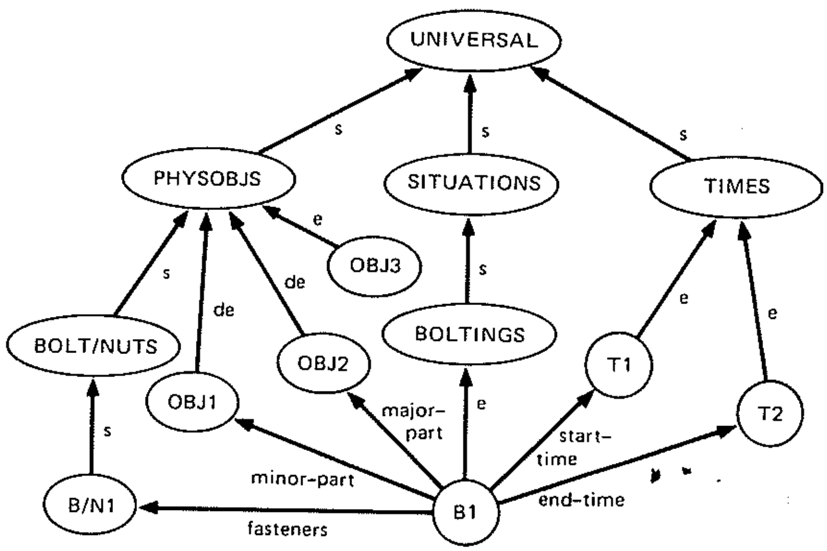


Figure III-1. A SAMPLE SEMANTIC NETWORK

Partitioning adds to the structure of a semantic network by segmenting the nodes and arcs of the network into subnets called spaces. Hendrix (1975a,b) introduces the notion of network partitioning and describes its use for encoding quantification, abstraction, and hypothetical worlds. In addition to separating the nodes of a network into spaces, partitioning provides for grouping the spaces into ordered sets called vistas. Vistas are typically used to restrict the network entities that are seen by procedures that reference the network (i.e., to impose visibility constraints). A procedure, when given a vista, can operate as though the only nodes and arcs in the network are those contained in some space in the vista. Although any set of spaces may be collected into a vista, vistas are typically used to group spaces hierarchically. The conventions adopted for figures are that spaces are represented by boxes, a node lies on the space inside of which it is drawn and an arc lies on the space inside of which its label appears.

If the boxes representing two spaces overlap, but neither contains the other, then the nodes and arcs in the overlap lie on both spaces.

One of the uses of partitioning is to encode quantified statements (see Hendrix, 1976). An example of the use of partitioning to encode an implication appears in Figure III-2. The node 'WRENCHES' represents the set of all wrenches. The node 'I' is an element of the set of all implications. The ante (antecedent) and conse (consequent) arcs from 'I' point at <u>supernodes</u>, spaces that have been given node-like properties. The nodes and arcs lying in the ante space are universally quantified; those in the conse space are existentially quantified. Thus, the node 'I' encodes the quantified statement, "for every WDE, there exists ET and SZ such that if WDE is in WRENCHES, then ET is in SHAPES and SZ is in LINEAR.MEASURES and ET is the endtype of WDE and SZ is the size of WDE."

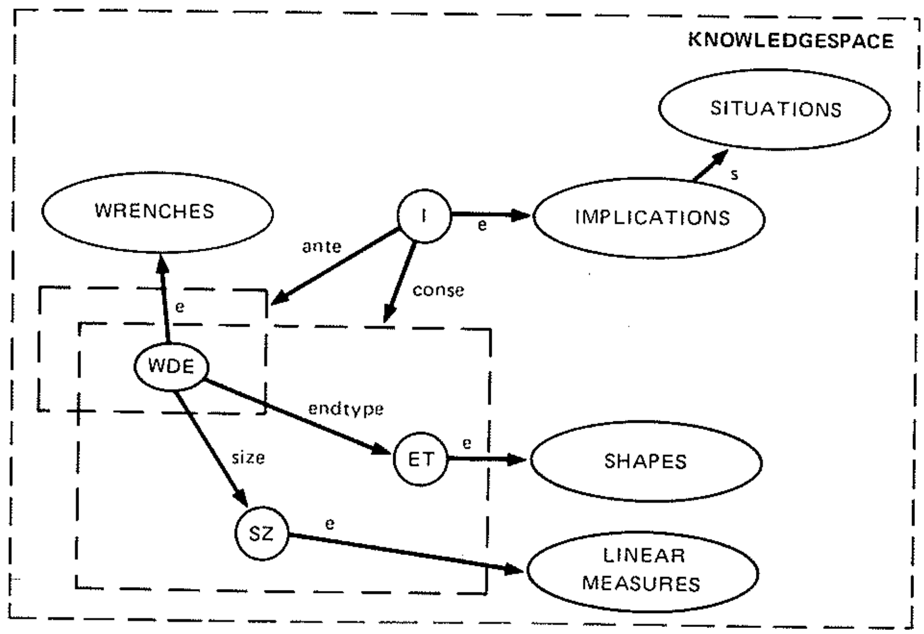


Figure III-2. THE DELINEATION OF WRENCHES ENCODED AS AN IMPLICATION

A particular use of implications that will occur in the ensuing discussion is to represent delineating information. For any given set, the delineating element is a hypothetical element of the set that is used to encode properties common to all real members of the set; i.e., properties possessed by the delineating element are common to all other elements. For example, in Figure III-2, WDE is the delineating element of WRENCHES.

Figure III-3 illustrates the same information in a shorthand that will be used in figures in the remainder of this report. The delin arc from the node 'WDE' to the node 'WRENCHES' represents the fact that WDE is the delineating element of WRENCHES. The properties of WDE are the properties of the prototypical element of the set WRENCHES: all wrenches have a size (represented by the node 'SZ'), which is some linear measure, and an endtype (represented by the node 'ET'). In this particular example, there are two real wrenches, W1 and W2. W1 is a 1 cm open-end wrench. W2 is a box-end wrench; its size is not given in the network fragment shown.

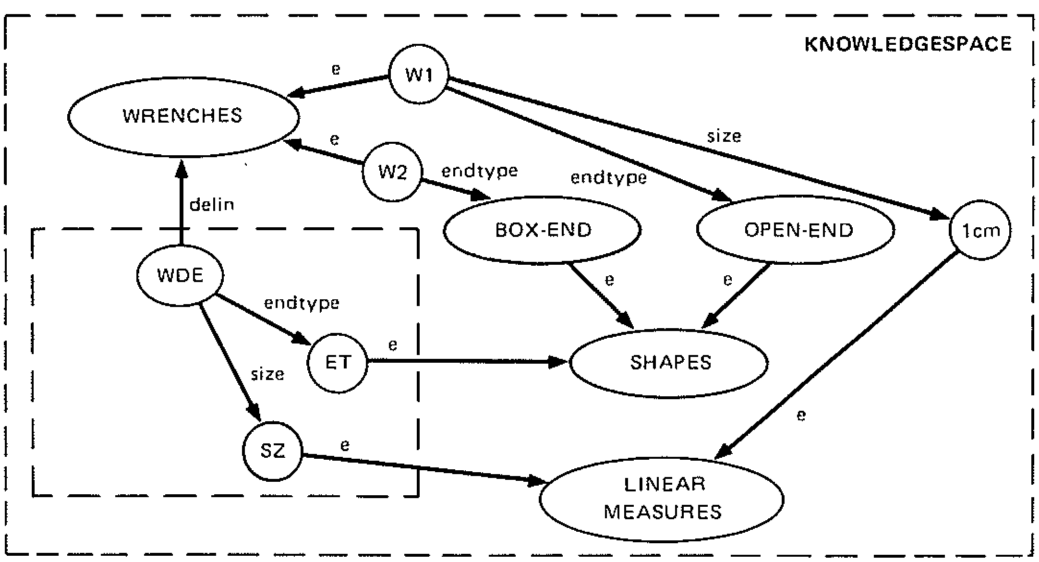


Figure III-3. THE DELIN SHORTHAND

C. FOCUS SPACES -- A REPRESENTATION OF EXPLICIT FOCUS

To encode focus, Hendrix's notion of partitioning has been extended to allow a network to be partitioned in more than one way. The nodes and arcs are separated into different sets of segments for different purposes. In particular, in addition to partitioning the network to encode quantification, it is also partitioned to encode focus. The former partitioning is referred to as the logical partitioning and is represented by dashed lines in the figures in this report. The latter is referred to as the focus partitioning and is represented by solid lines. The spaces in the focus partitioning are used to highlight items that become focused in a discourse. The focus spaces are related in a hierarchy that reflects the structure of the discourse.

As an example, consider the network portrayed in Figure III-4. The network is divided into four spaces, S0, S1, S2, and S3. Space S0 groups together the nodes representing EXCHANGES (the set of all exchange situations), ATTACHINGS, BOLTS, PUMPS, and PLATFORMS (the sets of all attach operations, bolts, pumps, and platforms, respectively). Space S1 contains a specific exchange, represented by the node 'EX1' (node names are enclosed in single quotes), in which the set of bolts represented by the node 'B1' is exchanged for the amount of money represented by the node '$1'. Space S2 contains a specific attaching operation, A1, of the minor part PU1 and the major part PL1. Space S3 also contains the specific attaching operation A1, but it shows this operation involves the specific set of bolts, B1.

The hierarchy of spaces in Figure III-4 is shown by the heavy arrows between spaces. Each space is associated with a particular vista that is the orthodox vista for that space. In the example, the orthodox vista associated with each space S is composed of the space S itself and all spaces that can be reached from S by following the heavy arrows. For instance, the orthodox vista of S0 is (S0) and the orthodox vista of S3 is (S3 S2 S0).

The visibility constraints that result from this partitioning may

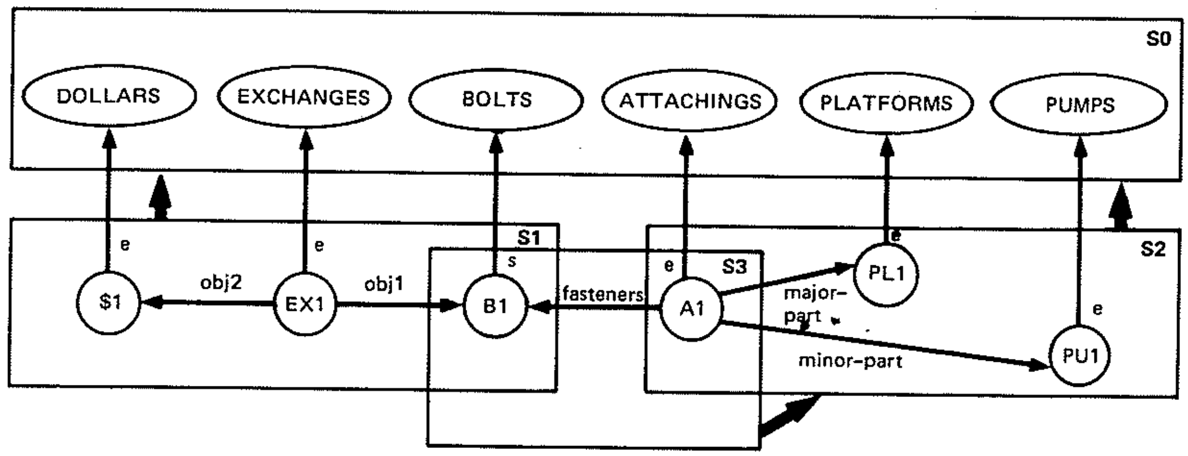


Figure III-4. A SAMPLE PARTITIONED SEMANTIC NETWORK

be seen by considering different views of the bolts B1 and the attaching operation A1. B1 is shown as taking part in two different events. A1 is a single operation shown at two different levels of detail. From the vista (S1 S0), the set of bolts B1 are seen only to be involved in the exchange EX1. However, from the vista (S3 S2 S0) B1 are seen as the fasteners in the operation of attaching PU1 to PL1. The two vistas give two alternative views of B1. A similar situation occurs with A1. From the vista (S2 S0) A1 is seen only as an attaching between two parts, with the fasteners left unspecified. When S3 is added as the bottom space in the vista, A1 is seen to involve the specific fasteners B1.

The focus partitioning makes it possible to highlight the particular way of looking at a concept that is germane to a given point in a dialogue. When the same object enters the dialogue twice, in two different subdialogues (e.g., a tool used in two distinct subtasks), the node corresponding to that object will appear in two distinct focus spaces. If different aspects of the object are focused on in the two subdialogues, different relationships in which the object participates will be in the two focus spaces. For example, in Figure III-4, B1 is focused on in S1 as a part of an exchange. In contrast, in S3 it is focused on as part of an attaching operation.

The main reason for providing the ability to focus on different attributes of an object is to allow differential access to the properties of the object, and hence to order the retrieval of facts that are derivable about that object. Differential access is important for events and relationships as well as for physical objects. For example, when quilting is considered as a kind of sewing, the subactions of cutting and pinning are accessed first, but when quilting is considered as a social gathering, then the subactions of talking and eating are more important and selected first.

There are two rules governing what is contained in a focus space. First, if a concept is in focus, type information about that concept must also be in focus. This information indicates the aspect of the concept being focused on. It provides the key index to additional knowledge about the concept. In the network representation, this rule corresponds to requiring that every node in focus have one outgoing element or subset arc also in focus. Second, if a concept's participation in some situation (e.g., a book's being the object of an owning relationship) is in focus, then the situation itself (i.e., the particular owning relationship) also must be in focus. In the network representation, this rule corresponds to requiring that the from node of any focused case arc be in focus.

New focus spaces are created as the focus of a discourse shifts. At any point in a dialogue, only one focus space is <u>active</u>, but several may be considered <u>open</u>. The active focus space reflects the focus of attention at the current point in the dialogue. The open focus spaces reflect previous active spaces that contain some unfinished topics and hence may become active again; they are possible areas to which the dialogue may return. The relationship between focus spaces is determined by (and hence reflects) the structure of the particular discourse being processed. For task dialogues, the task hierarchy provides a framework for this structure (see Chapter II). When a focus space is first created, it is considered open. The focus space is closed when an utterance indicates a shift to a new topic (in the task dialogues, this corresponds to a shift of task). Chapter V discusses some strategies for shifting focus and deciding when to close a focus space. A closed focus space records where the focus of attention was at some previous point in the discourse. The combination of all focus spaces for a dialogue together with a time line records the shifts of focus in attention over time of the dialogue. At any point in a dialogue, several focus spaces may be open. Although exceptions occur in naturally occurring dialogues, in the remainder of this report, multiple openings are restricted to spaces that are related in a strict linear hierarchy. That is, there is a top-most focus space and each of the other open focus spaces is the child of precisely one focus space. The hierarchy of currently open focus spaces is called the open focus space hierarchy.

D. IMPLICIT FOCUSING THROUGH A TASK REPRESENTATION

The representation of implicit focus requires a decision about what information associated with a concept should be put in focus when that concept is introduced. The bounds on this information depend on the knowledge and expectations about the concept that are shared by speaker and hearer (see Karttunen, 1968; Maratsos, 1976). The tradeoff between how much information to associate with a given concept and how many levels of associations to consider for implicit focusing must be resolved. In general, these problems entail basic issues about representation. They will be addressed here only as they occur for events.

For physical objects, the subparts of the object are among the concepts that must be implicitly focused when the object is in focus.

For events, the situation is somewhat more complicated. The direct analogy of subparts of an object is subevents of an event. However, the participants in the subevents of an event are also implicitly focused. The following dialogue fragment illustrates this point.

S: Attach the lid to the container.
R: Where are the bolts?

The statement S implicitly focuses on the bolts involved in the attaching as well as the subevent of fastening the lid down.

To enable implicit focusing on both the subevents and the objects involved in them, the representation of an event indicates both its subevents and the participants in its subevents. Figure III-5 shows a network representation that accomplishes this for the task step of attaching a pump to a platform.* The logical space KNOWLEDGESPACE, only part of which is shown here, contains representations for all items in the knowledge base. The set of ATTACHINGS.PUMP.PLATFORM is shown to be a subset of all ATTACHINGS. The delin arc from 'APP' to 'ATTACHINGS.PUMP.PLATFORM' indicates that APP is the prototypical element of the set of such attachings (see Hendrix, 1975a,b for a discussion of delineations). The two nodes 'APP' and 'APPD' together with the other structures inside the delineation space, DS, describe the nature of events in which a pump is attached to a platform. APP relates the participants in the event. The outgoing arcs from 'APP' indicate that these attachings involve a minor part, which is an element of the set PUMPS, and a major part, which is an element of the set PLATFORMS.

APPD is the event descriptor for APP. It relates the preconditions, effects, and substeps of the event. The two constituents of APPD that are most relevant here are the plot space and the binding space. The plot space, PS, contains the breakdown of APP into two substeps, S1 and S2, specifying a POSITION operation OP1 and a SECURE operation OP2. The suc arcs indicate successor links between substeps. (Although not shown here, the representation allows for partial ordering

----

* This representation has been developed jointly with Gary G. Hendrix and Ann E. Robinson.

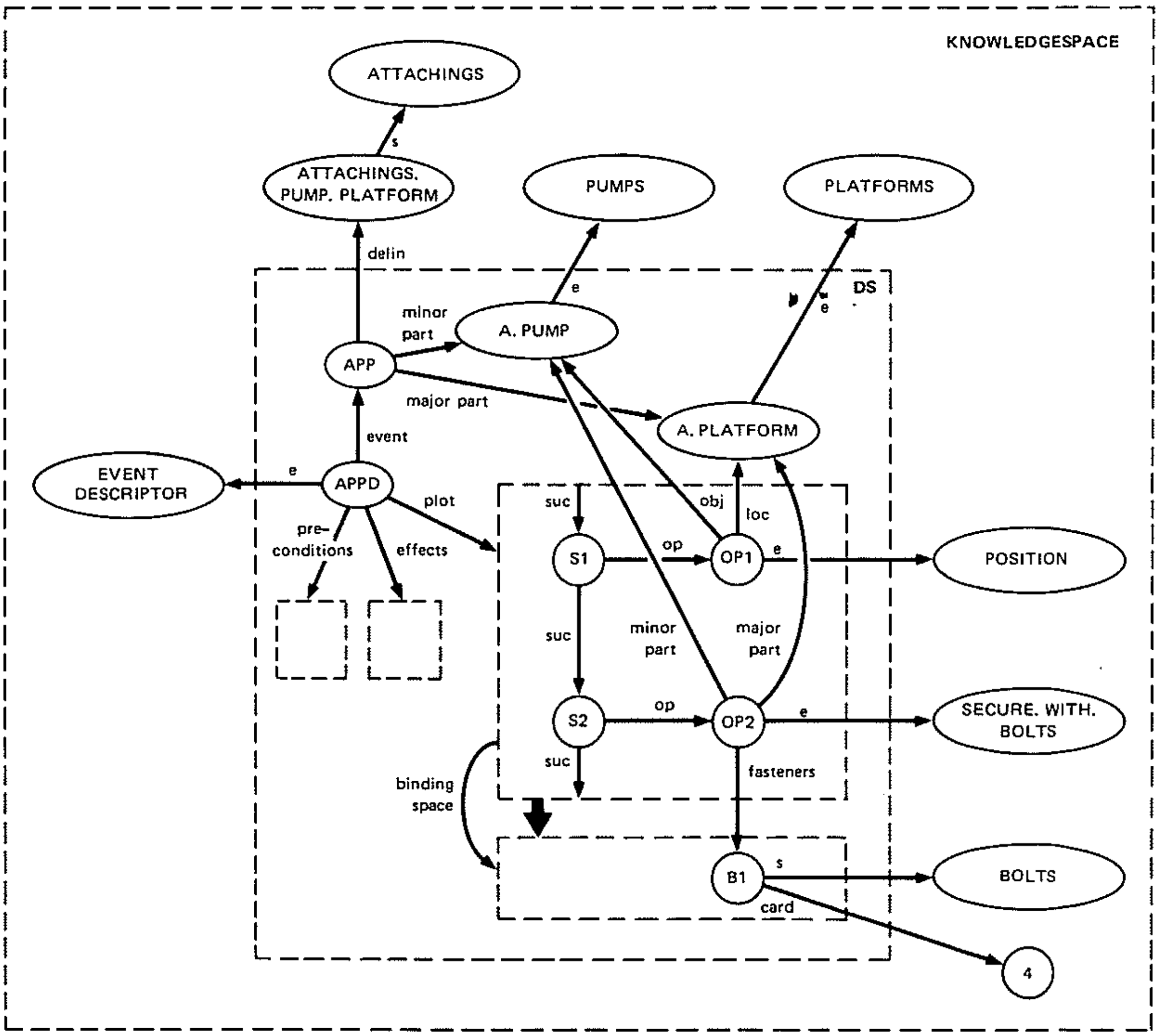


Figure III-5. EVENT ENCODING SHOWING IMPLICIT FOCUS of substeps, as in Sacerdoti, 1975). The binding space, BS, contains a set of four bolts that take part in the securing substep. When the task step of attaching a particular pump to a particular platform is in (explicit) focus, then the corresponding substeps for S1 and S2 and the set of bolts in the binding space are considered implicitly in focus.

In general, the binding space contains all of the participants in any subevent that are at too low a level of detail to be mentioned explicitly as participants in the main event. The implicit focus for an event consists of the vista of the plot space and binding space and thus contains both the subevents and the participants in those subevents. Because more inferencing is required if more levels of associations (e.g., deeper levels of the task hierarchy) are referenced, when retrieval requires a search of implicit focus (e.g., the concept sought is not in explicit focus), a breadth-first search is done. Subconcepts of all relevant concepts are examined before any sub-subconcepts are examined.

Implicit focus is used for the interpretation of both object and action references (cf. Rieger, 1975; the implicit focus of the task representation provides the same task context as conceptual overlays). For example, if the current task is attaching the pump to the platform, then "the bolts" refers to the bolts that participate in the securing operation and "put" refers to the positioning subevent.

E. NETWORK STRUCTURE MATCHING

The retrieval of items from memory is one of the most frequent operations any knowledge-based system must do. In a system with a semantic network knowledge base, the central process involved in retrieval is matching a network fragment containing variables with the knowledge base. This matching process typically entails considerable search that is guided only by local constraints. A major use of the focus representation is to constrain the search on the basis of discourse information. In this paper, the system component that performs this matching process will be called the matcher. Fikes (1976) describes in detail how this component works.* Only enough detail will be given here to elucidate the need for and the role of the focus representation in this process.

Structure matching in the memory representation is the basic process involved both in resolving DEFNPs and in finding the answer to a question. These two problems are related: DEFNP resolution may be viewed as finding the answer to a simple "what is" question. For example, finding the wrench referred to by the phrase "the long-handled wrench" is the same as answering the question "Which wrench is long-handled?" or (at least for one interpretation) "What is the long-handled wrench?" Note that DEFNP resolution may involve identifying nodes representing actions (e.g., the node corresponding to "the testing" in "the testing took three days") as well as nodes representing physical objects. Although most of the examples in this chapter are DEFNPs referring to physical objects, the procedures described also pertain to these other uses of matching.

The matcher works with two (logical) vistas: a QVISTA (question vista) and a KVISTA (knowledge vista). The QVISTA is a set of spaces collectively containing a piece of network for which a match is sought. The KVISTA represents the set of all knowledge in which the match is sought. For example, when the matcher is called as part of the procedure for resolving a definite noun phrase (e.g., the red bolts), the QVISTA is a piece of network structure that describes the object referred to by the noun phrase, as it is described by the noun phrase (i.e., for the example, a net structure for a subset of bolts that are colored red). The KVISTA is the whole knowledge base. The match of the QVISTA fragment to the KVISTA corresponds to finding a real object (i.e., an object that 'exists' in the knowledge base) that can be described by the definite noun phrase.

In the process of arriving at a match, the matcher binds each item

----

* In the SRI speech understanding system, this component was implemented by Richard E. Fikes and was called the deduction component.

(i.e., each node and arc) in the QVISTA to an element of the KVISTA. Two kinds of decisions affect the amount of computation done in arriving at a match. First, at each step of the match, an item must be selected for matching from the QVISTA. The order of selection influences the efficiency of the matching computation. Second, once a QVISTA element is selected, the matcher must select an element of the KVISTA for trial binding to the QVISTA element. In general, there are many candidates and only local information is available to guide the selection.

For example, consider Figure III-6 which portrays the QVISTA corresponding to the DEFNP, "the 1 cm wrench" and a sample KVISTA. The matcher must determine that node 'WP' matches node 'W3'. In the process of arriving at this match, it matches the arc WP--size-->1cm with the arc W3--size-->1cm and deduces that W3 is an element of WRENCHES by following the e arc from 'W3' to 'WS' and the s arc from there to 'WRENCHES'. As a result of making this deduction, a new arc W3--e-->WRENCHES is added to the KVISTA and the QVISTA arc WP--e-->WRENCHES is bound to this newly deduced arc. This chaining of e and s arcs is the simplest kind of deduction. More complex deductions arise from delineation elements and theorems in the knowledge base. An example of the first kind of decision in this match is the choice of looking first for a match for the arc WP--e-->WRENCHES or for a match for the arc WP--size-->1cm. An example of the second kind of decision is the choice between the candidate arcs W3--size-->1cm and B1--size-->1cm as a match for the arc WP--size-->1cm.

Each binding of a QVISTA and a KVISTA element is only tentative. First, side effects of the binding must be checked. For example, if a node is bound, the matcher must establish that unbound element or subset arcs in QVISTA from that node are consistent with the arcs in KVISTA. The match will be carried further only if such consistencies hold. Even so, the binding may be rejected later if a match of the remainder of the QVISTA is not found. Hence, the number of bindings attempted is a significant element of the cost of arriving at a match. Optimally, for both kinds of decision, the matcher will choose the most constraining

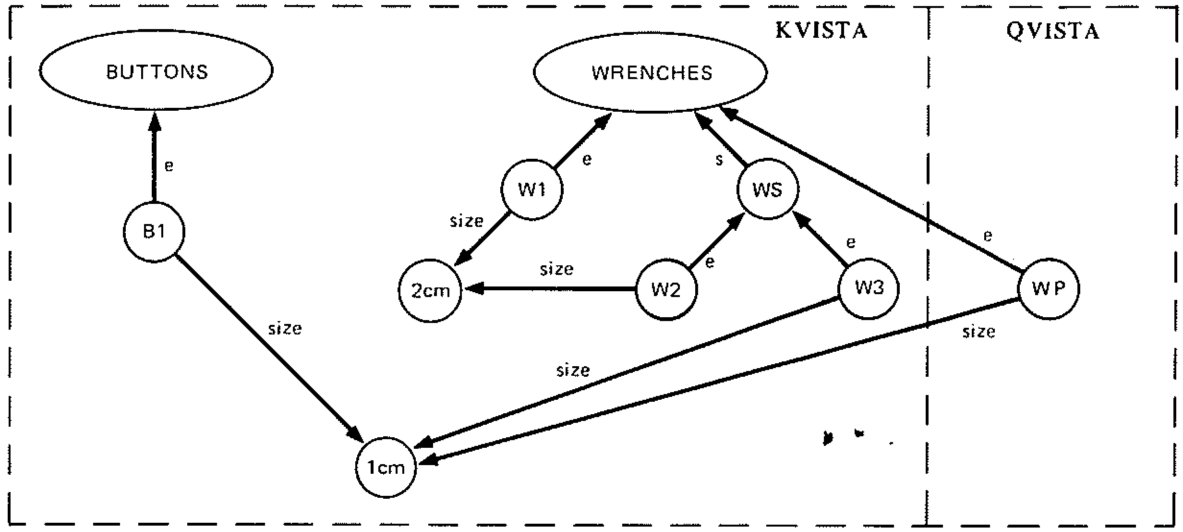


Figure III-6. A SAMPLE KVISTA AND QVISTA

element. In an unfocused match, the choice can be made only on the basis of local structural information. For example, in the match of Figure III-6, the choice of whether to bind the e arc or the size arc from 'WP' first is made on the basis of whether the number of e arcs into 'WRENCHES' is smaller than the number of size arcs into '1cm'. In essence, the matcher chooses between trying to find the wrench that WP matches or the 1 cm object that WP matches. It makes this choice on the basis of whether there are more wrenches or more 1 cm objects in the knowledge base. One of the goals of the focus representation is to guide these match decisions on the basis of discourse information.

F. <u>MATCHING IN FOCUS</u>

The focus representation is used to order the candidates considered for binding by the matcher. The term <u>focused match</u> is used to denote matches that are constrained by focus. Focusing on certain concepts (both nodes and arcs) constrains the matcher to consider only objects germane to the dialogue. Since arcs provide indices from focused items into general network (KVISTA) information, focusing on an arc also guides the matcher in establishing properties about nodes being matched. That is, focused arcs provide a means of differential access to unfocused information. Using the arcs in focus for differential access does not rule out considering a concept differently than it has already been portrayed. Instead, it orders the way in which aspects of the concept are to be examined in looking for new (to the dialogue) information about the concept.

When a focused match is requested, the matcher is passed two arguments in addition to the usual QVISTA and KVISTA: a focus vista and a forced-in-focus list. The focus vista represents the set of nodes and arcs considered to be in focus. Different calls on the matcher are made for explicit and implicit focus matches. For explicit focus, the focus vista may be either the active focus space alone, or the entire vista of open focus spaces. For implicit focus, the focus vista is the composite of the implicit focus vistas for all items in explicit focus (e.g., for each event, the vista of plot space and binding space). The forced-in-focus list contains those items in the QVISTA that must be bound to items in the focus vista. As an example of the use of the forced-in-focus list, consider the requirement that the referent of a definite noun phrase be in focus. This requirement corresponds to a focused match in which the forced-in-focus list contains the QVISTA node corresponding to the head noun of the noun phrase.

Forcing a QVISTA item to be in focus provides a strong constraint on the search for a matching KVISTA item. Hence, forced-in-focus items are selected as the first candidates from the QVISTA to be matched. If a successful match is obtained for such an item, it constrains other items in the QVISTA. If no match can be found for a forced-in-focus item, then no focused match of the QVISTA is possible.

The focus vista is used to order the selection of KVISTA items for trial binding to a QVISTA item. Each step of the matching algorithm first selects relevant items in the focus vista both for explicit matches (the item in the QVISTA is bound to an item that explicitly exists in the KVISTA) and for derived matches (application of a general rule produces a new KVISTA element). Hence, focus influences the order in which deductions are made in the process of arriving at a match.

1. EXAMPLES OF FOCUSED MATCHES

This section contains several examples that illustrate the use of focus in constraining a match. For the purposes of these examples, a focus space is assumed and the problems of obtaining a successful match are examined. Chapter V discusses the problems of deciding what items get moved into a focus space and when focus shifts. To simplify the discussion, definite noun phrases (DEFNPs) will be used to describe most of the network structures in the examples.

The following example illustrates the use of focus to reduce the number of candidates considered for binding by the matcher. Consider the KVISTA of Figure III-7 and the QVISTA (q.w1) of Figure III-8. The KVISTA contains several wrenches: W1 is a box-end wrench that is in focus FS1; W2 is a box-end wrench in focus FS2; W3 is an open-end wrench also in focus FS2; W4 is another open-end wrench not in focus at all. There is another object, O1, with a box end. The QVISTA contains the structure that corresponds to the description "box-end wrench". In an unconstrained match, the matcher would consider all of the nodes with e arcs to 'WRENCHES' or all of the nodes with endtype arcs to 'BOX-END' (depending on which set is smaller) as candidates for binding to QW1. Eventually, it would try 'W1' or 'W2' and obtain a successful match. In the worst case, this would entail one node and two arc bindings for each of the candidate nodes that fails as a complete match. In general, there may be many such unsuccessful candidates (e.g., tens of wrenches that are not box-end wrenches, but are considered by the matcher before it selects W1 or W2).

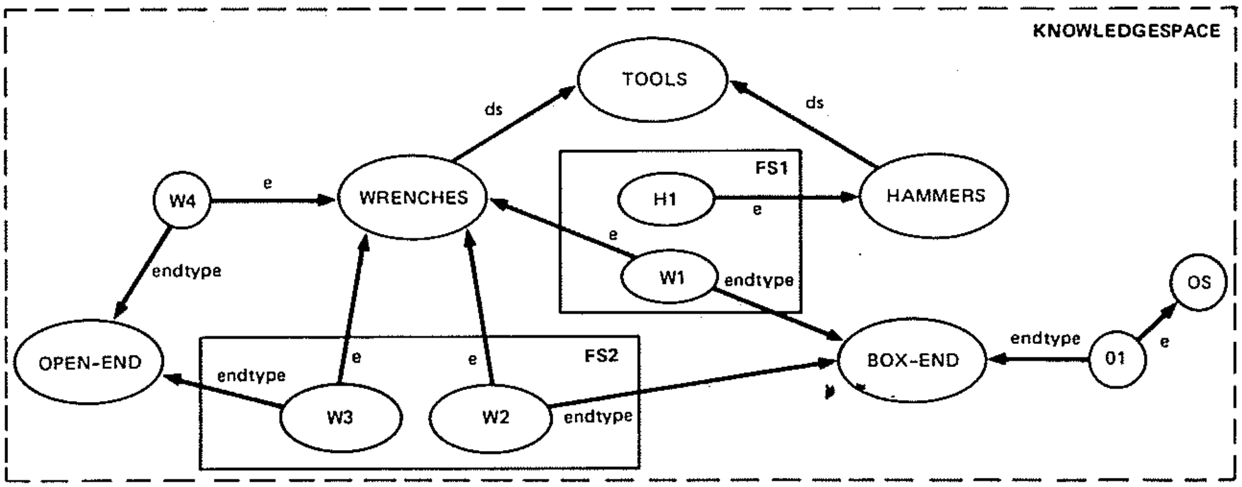


Figure III-7. A SIMPLE KVISTA WITH TWO FOCUS SPACES

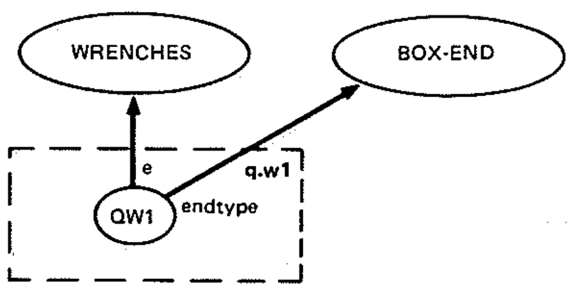


Figure III-8. QVISTA FOR "THE BOX-END WRENCH"

The focused match is able to avoid all this searching. If focus space FS1 is used, only nodes 'H1' and 'W1' are considered. 'H1' will be rejected immediately because the e arc to 'HAMMERS' is incompatible with the e arc from 'QW1' to 'WRENCHES'. (The matcher knows that the sets HAMMERS and WRENCHES have no intersection from the ds arcs from 'WRENCHES' and 'HAMMERS' to 'TOOLS'.) With focus space FS2 as the constraint, both 'W3' and 'W2' are considered, but 'W3' is eliminated because of its incompatible endtype. In the worst case, one set (one e arc and one node) of unnecessary bindings is made.

Even greater savings are obtained when deduction is necessary to achieve a match; i.e., when general rules -- chunks of information stored in the net as applicable to whole sets of concepts -- must be applied. In such cases, focus constrains the application of such rules, avoiding a combinatorial explosion of trial bindings. To illustrate such a match, consider the KVISTA of Figure III-9. Here the set of wrenches has two subsets, B-E, the set of all box-end wrenches, and O-E, the set of all open-end wrenches. The (logical) space oew.desc represents the fact that all elements of the set O-E have endtype OPEN-END; bew.desc represents a similar rule. For purposes of this discussion, assume that 'WRENCHES' has fewer elements than 'BOX-END' has incoming endtype arcs. (This assumption simplifies the discussion, and is reasonable, considering that objects other than wrenches may be classified as "box-end".) The unconstrained match for 'QW2' proceeds by considering all nodes with e arcs to 'WRENCHES'. The e arcs are all implicit in this case; they must be derived by following e-and-s chains. For each element of wrenches proposed as a match for 'QW2', the matcher attempts to establish an endtype arc to 'BOX-END'. In particular the delineating element descriptions for 'B-E' and 'O-E', contained in the logical spaces bew.desc and oew.desc, respectively, represent applicable general rules. The rule in bew.desc states that every element of the set B-E has an endtype BOX-END. Suppose W2 is selected as the element of WRENCHES to try as a match for QW2. The relevant rule in this case is represented by the box labeled oew.desc. Since W2 is in O-E, and since every element of O-E has endtype OPEN-END, an endtype arc from 'W2' to 'OPEN-END' will be constructed; only then will the matcher realize 'OPEN-END' is not 'BOX-END', and hence 'W2' will not match. In general, there may be many nodes like 'W2' that appear to be candidates

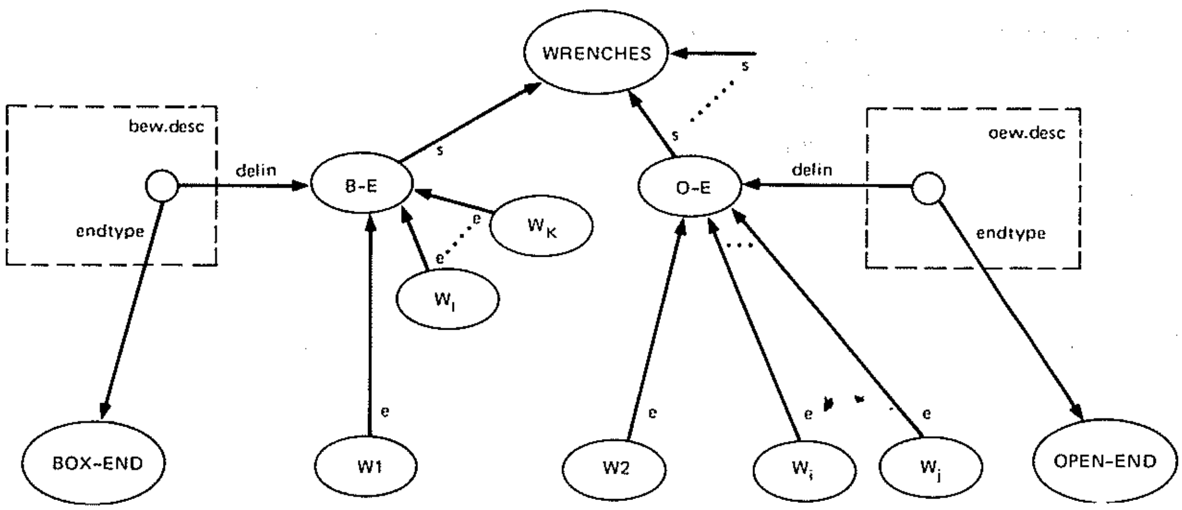


Figure III-9. A KVISTA WITH THE SET OF WRENCHES DIVIDED INTO SEVERAL SUBSETS

but do not match, and many rules that will apply. The work done before considering W1 may be extensive: tens of wrenches may exist in the KVISTA and be tried as candidates before selecting one with a box end.

By constraining the search to nodes in focus, a considerable reduction can be achieved. The matcher only looks in any detail at the wrenches that are in focus (other nodes will be dismissed immediately, because a node binding entails immediately checking its e or s arc). In general only one or two nodes in focus will be elements of WRENCHES and endtype theorems will only be invoked for those nodes.

2. SPECIAL USE OF FOCUS FOR DEFNP RESOLUTION

Resolution of DEFNPs often entails a particularly simple kind of match that corresponds to finding an element of a set. The primary role of focus in these matches is to enable the matcher to find the element of the set that is relevant to the current discourse. The influence of focus on the time it takes to arrive at such a match comes from reducing the number of candidates the matcher considers initially. For example, reconsider the situation portrayed in Figure III-7 and the QVISTA (q.w2) of Figure III-10. The QVISTA corresponds to the DEFNP "the wrench". If the matcher were asked to find a match without focus for this QVISTA, any of the W nodes would do. This corresponds to the fact that, without focus, the phrase "the wrench" is four ways ambiguous. However, if the matcher is provided with QVISTA q.w2 and focus vista FS1 (and the node QW2), it will find that W1 is the only match. In arriving at this solution, it may consider H1 but will discard this possibility when realizing that hammers and wrenches are mutually disjoint subsets of tools. The attempt to match "the wrench" in focus space FS2 will result in both W2 and W3 matching, reflecting the fact that, for the discussion at that point, two wrenches were relevant, and the DEFNP, "the wrench" is ambiguous.

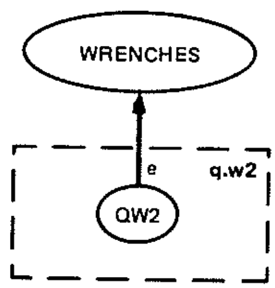


Figure III-10. QVISTA FOR "THE WRENCH"

A similar situation exists for the KVISTA of Figure III-9 and QVISTA (q.w1) of Figure III-10. Finding a match for QW1 in this KVISTA entails following the e and s chain from, say 'W1', to 'WRENCHES'. This process is the simplest form of deduction. Again, the unfocused match will consider all elements of WRENCHES equally (and may spend some amount of computation time realizing this). Some extra mechanism must be added (e.g., an indication of the last time the node was referenced) to enable the correct resolution of DEFNPs. This mechanism will have to take into account discourse structure as well as time to be sufficient (see Chapter II, Section D.4). Focus spaces provide this mechanism in addition to minimizing the search for candidates for binding in the KVISTA.

G. EXTENSIONS

The use of the focus representation to direct structure matching for such things as answering questions and resolving DEFNPs is only one of its roles in language understanding. Focus is relevant for several other problems that arise in building a language understanding system. This section explores the use of the focus space representation in the solution of two such problems. First, there is a space/time tradeoff between storing derived information and recomputing the information. Ideally, the information would be stored only as long as it was needed and then erased from the knowledge base. This issue is closely related to the general issue of forgetting in a language understanding system. Second, any given object may be viewed from several different perspectives. Highlighting a particular view relates to the companion problems of deriving information about the object and capturing the information conveyed by the particular way an object is described in a given utterance.

1. DERIVED INFORMATION AND FORGETTING

In the discussion of matching there were examples that illustrated the need for deducing information about particular objects from general rules in the knowledge base. In the process of matching network structures, the matcher may create new network structure. If the network structure is permanently stored in the knowledge base, the deduction will never have to be repeated. However, making the structure permanent uses up valuable storage. Focus spaces provide a mechanism for determining how long to store such information. When the new structure is derived, it can be added to the current focus space. When the focus space is closed, the new information can be erased.

As an example of this use of focus spaces, consider the modified version of Figure III-9 portrayed in Figure III-11. Suppose that initially the nodes 'W1' and 'W2' were in focus as elements of the sets B-E and O-E respectively (e.g., the wrenches were selected from two boxes each containing one type of wrench). If the matcher is given the structure for "box-end wrench" (see Figure III-8) to match, it will create two new arcs, an endtype arc from W1 to 'BOX-END' and an explicit e arc from 'W1' to 'WRENCHES'. These new arcs are added to the focus space, FS, as shown in the figure. Any further matches sought for "the box-end wrench" while the focus is FS will be able to take advantage of this explicitly stored information. When FS ceases to be open, the arcs will be erased (from the logical space as well as from the focus space they are on). If the deduction had resulted in new nodes being created, they too could be erased. Using focus spaces in this way results in both having the information available when it is relevant and allowing it to be 'garbage collected' or 'forgotten' after it ceases to be relevant.

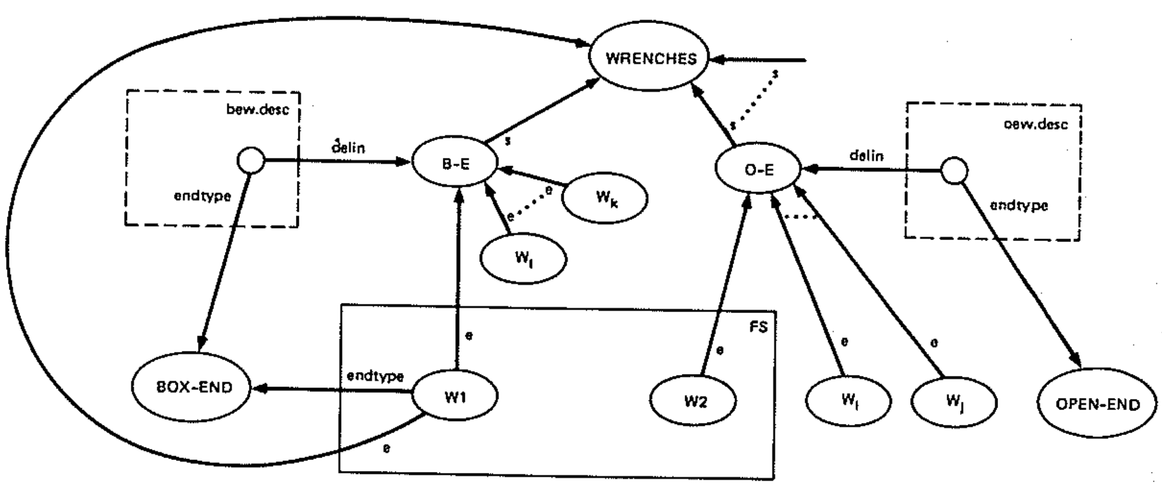


Figure III-11. THE WRENCHES KVISTA WITH FOCUS ADDED

2. DIFFERENTIAL ACCESS AND DESCRIPTION

The representation of some concept C may include descriptions of C as an instance of several different categories.* Focusing allows the particular way of looking at C germane to a given point in a dialogue to be highlighted. The arcs from focused items to unfocused items provide the matcher with preferential access to information that is most likely to become relevant to a discourse. Using the arcs in focus for differential access does not rule out considering a concept differently than it has already been portrayed. Instead, it orders the way in which aspects of the concept are to be examined in looking for new (to the discourse) information about the concept.

As an example, consider the portrayal of C in Figure III-12 (the dots inside the delin spaces indicate that some delineating information has been omitted). C is shown to be a doctor friend of K's who backpacks. If C is discussed in her role as doctor, then the phrase "her bag" will be seen to refer to the bag containing her medical supplies. However, if she enters a discussion as a backpacker, then the same phrase will be taken to mean her sleeping bag. The matcher can be led to these different deductions, by differentially following the two e arcs from 'C' in focus spaces FS1 and FS2 respectively.

This use of focusing addresses one part of the 'mayor of San Diego' problem posed in Norman et al. (1975). Consider the situation portrayed in Figure III-13. The person represented by the node 'MNMSD' is shown both to be D's neighbor and the mayor of San Diego. If MNMSD is referred to by D either as "the mayor of San Diego" or "D's neighbor", then node 'MNMSD' represents the individual referred to. The problem is that only looking at that node provides no reflection of the differences in the two references to MNMSD, even though the surface DEFNPs do express this difference. Focus spaces provide a means of representing this difference. Even though node 'MNMSD' will be in focus

---

* Description from multiple perspectives is the basis of the representation of entities in the representation language presented in Bobrow and Winograd, 1977.

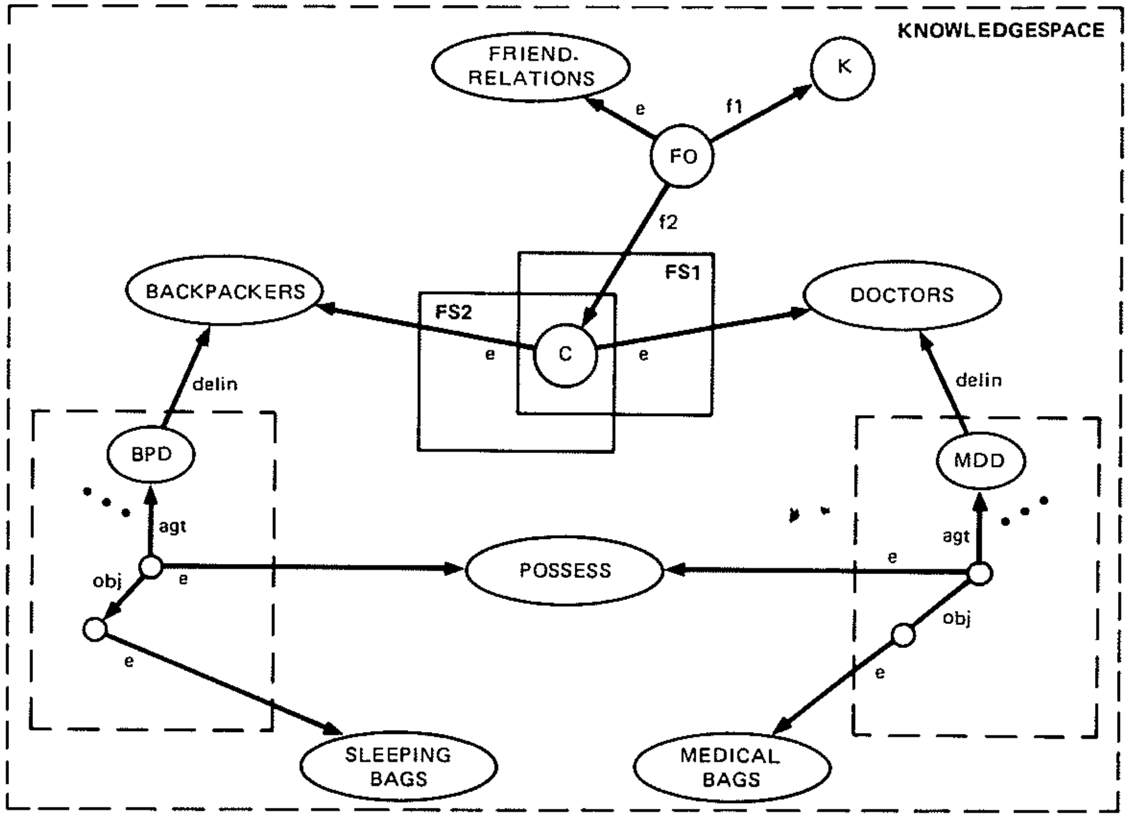


Figure III-12. C -- FRIEND, DOCTOR, AND BACKPACKER

no matter which reference is used, arcs from 'MNMSD' that are in focus in the two cases will differ. Focus spaces FS1 and FS2 illustrate this difference.

H. SUMMARY

The focus representation groups together items relevant to a particular point in a discourse, providing a small subset of the knowledge base for the understanding system to concentrate on. In particular, the focus representation may be used to guide the retrieval of information from the knowledge base. It reduces the size of the

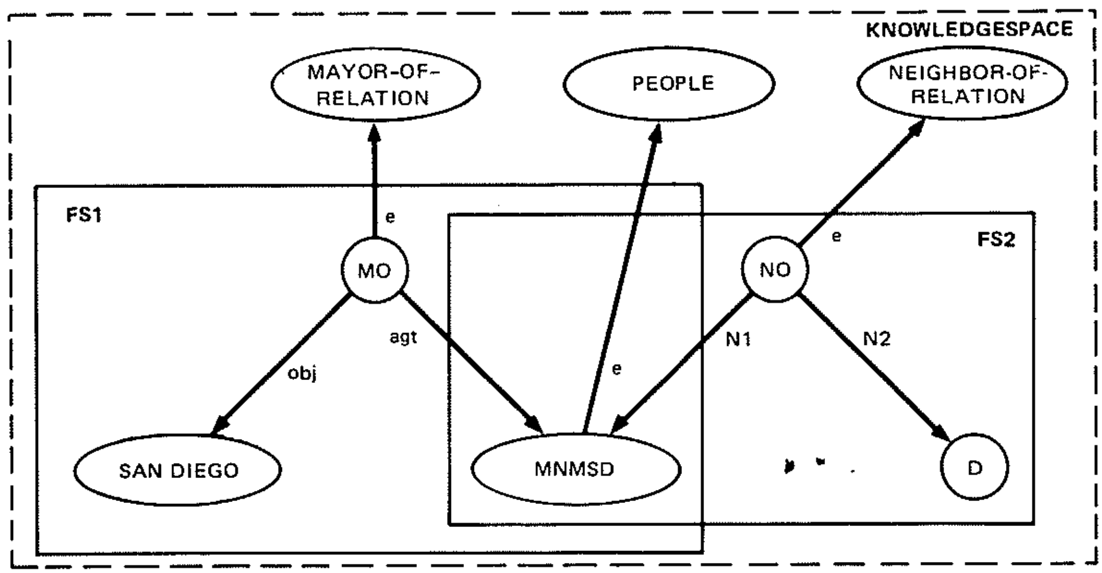


Figure III-13. MY NEIGHBOR THE MAYOR OF SAN DIEGO

search space that the retrieval mechanism must traverse. The representation of explicit focus in focus spaces also appears to be useful for related understanding system problems such as describing objects and forgetting information. Although the representation presented is in terms of a semantic network, partitioning a memory representation for the purpose of reflecting focus of attention is a general mechanism which may be used in other representation schemes as well.

IV RESOLVING DEFINITE NOUN PHRASES

CONTENTS:

A. INTRODUCTION

Definite noun phrase resolution and the maintenance of a focus representation are synergetic processes. The resolution of a definite noun phrase requires a model of the focus of the discourse in which the noun phrase occurs. In turn, the definite noun phrases that occur in a discourse often indicate shifts of focus in the discourse. Hence, this chapter provides a link between the preceding chapter's discussion of focus and the next chapter's discussion of shifting focus and noun phrase resolution in a task situation.

Section B describes the differences between pronominal and non-pronominal definite noun phrases, emphasizing the different roles of sentential and global focus in the resolution of these two forms of reference. Section C addresses the inference problems that arise in noun phrase resolution and shows how these relate to matching problems and the focus space representation. Section D discusses several categories of definite noun phrase references and procedures for interpreting them. The section covers the processing that must be done to build a representation of a particular definite noun phrase, given that noun phrase and a representation of the focus in which it appears.

B. SENTENTIAL AND DIALOGUE CONTEXT: A COMPARISON OF PRONOUNS AND DEFNPS

As in Chapter II, it will be useful here to divide definite noun phrase references into two categories: pronouns and nonpronominal definite noun phrases (DEFNPs). Although referring expressions in both categories depend on the context in which they occur for their interpretation, the nature of this dependence is quite different in each case. Similarly, although some of the processing required for building interpretations of pronouns and DEFNPs may be shared, there is other processing that is unique to each of these forms of reference. Both the global dialogue context and the immediate context of the preceding utterance play roles in interpreting each of these forms of reference, but the former is more important for DEFNPs, the latter for pronouns.*

Reference resolution entails selecting the item referred to from a set of candidate items. For a DEFNP, the candidate set is delineated by the focus in which the DEFNP appears. The head noun of the DEFNP specifies the class of the object being referred to and additional descriptive and distinguishing information is provided by modifiers. The focus in which the DEFNP appears delineates the set of objects from which the referent must be distinguished. Both the surrounding non-linguistic environment and the global linguistic context of the preceding discourse are part of this focus and, hence, crucial to the process of resolving the DEFNPs in the utterance. The immediate linguistic context and, especially, the sentential context of the referent itself (outside of the phrase in which the referent occurs) are

----

* See Chapter II, Section D.4 for a discussion of how this distinction is related to the distinction Chafe (1976) makes between givenness and definiteness.

usually not important (as will be shown shortly with an example). It is misleading to use a DEFNP that requires immediate context for the identification of its referent, because sufficient information for distinguishing the item from other candidates is supposed to be contained in the DEFNP itself.

Unlike DEFNPs, pronouns carry almost no information themselves. Hence, for most pronouns, the roles of global and immediate focus are reversed from that for DEFNPs. Pronouns are slot fillers; they depend on the sentential context in which they occur to provide most of the clues needed for identifying the referent. The immediate linguistic context of the preceding utterance and preceding clauses in the same utterance supply candidates for the referents; sentential context provides restrictions for choosing among them. Global focus is less important than for DEFNP resolution because of this dependence on immediate context.

An exception to this description of the roles of global and immediate focus occurs with certain pronominal references. In a structured discourse, a pronoun may refer back over long portions of the discourse. The dialogue fragment of Figure II-7 (Chapter II) provides one example. In such instances, the global focus supplies candidates for the referent and the process of establishing a candidate set resembles that for DEFNPs. However, the lack of semantic information in the pronoun makes sentential context necessary for choosing among the candidates. This use of pronouns is similar to the 'pragmatic anaphora' discussed in Hankamer and Sag (1976). In both instances, the surrounding (nonlinguistic) global focus provides sufficient constraints on the candidate set to allow for successful use of a pronominal (rather than a DEFNP) reference.

The relative role of sentential context in resolving DEFNPs and pronoun references can be seen by considering an example from Charniak (1972) and some variations of it. The original dialogue is presented in Figure IV-1. The "it" in (7) can be resolved only when

(1) Today was Jack's birthday.

(2) Penny and Janet went to the store.

(3) They were going to get presents.

(4) Janet decided to get a top.

(5) "Don't do that" said Penny.

(6) "Jack has a top.

(7) He will make you take it back."

Figure IV-1. THE TOP STORY

the context of "take ... back" is considered and even then a large amount of inferencing must be performed; e.g., see Charniak (1972) and Hobbs (1976).

Note that it is misleading to use the DEFNP "the top" in place of this "it".* The problem stems from the fact that the focus in which the utterance appears includes two tops, but use of the phrase "the top" implies there is only one. Although the sentential context of "take ... back" can also be used here, the use of the DEFNP strongly implies no need of recourse to such information. Finally, if instead of (7) the sentence were

"If you get Jack a top, he will make you take (it / the top) back" ,

either "it" or "the top" may be used and the reference to the hypothetical top of the if-clause is clear. The difference between the use of "the top" here and in (7) is that here the if-clause sets up a new focus in which there is only one top: the hypothetical one.

In many respects pronoun reference is closer to ellipsis, which will be discussed in the Chapter VI, than to DEFNP reference, and in a sense, the use of pronouns and ellipsis are duals. To see this,

----

* The reader might argue that, for pragmatic reasons, no person would use "the top" in this instance. The point of this discussion is to bring out some of the reasons.

consider a sentence, S, composed of constituents A, B, and C; i.e., assume that a context-free part of a language definition rule for S is S->A B C. Let a, b, c be respective instances of the particular phrase types A, B, C. Pronoun reference entails substituting a pronoun for one of these constituents; the remaining constituents provide selectional restrictions on what the referent of the pronoun is. For example, in the 'sentence', "it b c", properties of b and c constrain the referent of "it". Ellipsis, on the other hand, entails providing only one of the constituents and, depending on context, to supply the others. So, if a' is also an instance of A, the 'sentence' "a'" in the context of the previous utterance, "a b c", may be expanded to "a' b c". Elliptical expressions can always be resolved in terms of the immediately preceding utterance.

Elliptical DEFNPs (e.g., "the four by the door") and DEFNPs with the word "one" substituted for the head noun (e.g., "the red ones") are a hybrid of reference and ellipsis. An examination of these DEFNPs illuminates the different roles of global and local focus in the interpretation of the individual kinds of expressions. These references are like pronouns in that a slot (or a slot holder) is given, and the immediate sentential context and the preceding utterance are used to 'fill out' the phrase (e.g., "the four by the door" to "the four boxes by the door"). Once the phrase is filled out, these references are like other DEFNPs. In particular, the role of global focus in their resolution is identical.

C. THE INFERENCE PROBLEM

The simplest form of DEFNP resolution occurs when a DEFNP refers to an object that has been introduced into the discourse by an indefinite noun phrase. This kind of reference occurs in the second sentence of the sequence:

I bought a new wrench today.

The wrench is on the table.

However, restricting the use of DEFNPs to such cases results in rather boring discourse since it requires explicit statement of obvious facts. For example, the second sentence of the following sequence

Susan bought a car today.

The car has seats.

The seats . . .

is totally unnecessary and makes for awkward reading. Such redundant information usually is left out of a discourse. Comprehension then requires that the hearer be able to fill in the missing information from what he knows about the objects and actions being discussed. As a result, the resolution of DEFNPs often requires inferencing on the part of the listener.

Two kinds of inferences are needed for resolving DEFNPs. First, resolution may entail establishing additional properties of an object already in focus. This kind of inference is required when a later reference to a concept differs from the way the concept was originally introduced into focus. Second, resolution may depend on general information about objects, events, and relationships in the domain of discourse. This kind of inference is required when a definite reference occurs to an object that has been brought into focus only implicitly.

The first kind of inference is illustrated in the sequence:

I took your coats to the cleaners.

The blue coat will be ready tomorrow.

To understand the DEFNP, the hearer must infer that one of the coats is blue. A second instance in which this kind of inference is required occurs when an object already in focus is referred to in more general terms than those in the description first used to bring it into focus. Resolving the reference entails establishing that the new description is true of the old object. In the sequence:

I got another novel and some records at the library today.

The book is on the coffee table.

the fact that novels are books must be inferred to understand the DEFNP, "the book."

The problem posed for resolution here is not the difficulty of the inferences themselves, but rather restricting the number of objects considered. Even though the the chain of inferencing is not complex, the number of times it is applied must be minimized. If resolution of "the book" in the preceding example requires consideration of the possibly hundreds of books known to the hearer, understanding the second utterance will take a long time. The analogous case holds for a computer system. The representation of focus presented in the preceding chapter distinguishes, from among all those items known to the system, those that are relevant to the discourse. The system must only determine which of those objects (namely, the novel) is a book.

The second kind of inferencing required for DEFNP resolution arises because an object implicitly brings certain associated items into focus when it is brought into focus (see Chafe, 1972,1974; Karttunen, 1968). For example, mention of "the living room" brings into focus items such as "the ceiling" and "the furniture". In the ensuing discourse, these associated items may be referred to by DEFNPs. In the sequence:

E: Use the crescent wrench.

A: The handle is too long.

the phrase "the handle" can be resolved because the handle of a wrench is brought into focus when the wrench is. Parts of actions as well as objects may become focused in this way. For example, in the sequence:

E: Attach the pump to the platform.

A: Where are the bolts?

"the bolts" become focused because they are a part (namely the fasteners) of this attaching operation.

The problem in handling this kind of inference is deciding how much information related to a concept should get brought into focus when that concept is introduced. The bounds on this information depend on shared knowledge about the concept. In particular, the successful use of a reference requiring this second kind of inference depends on shared expectations about items associated with the concept (see Karttunen, 1968; Maratsos, 1976). This issue is clearly related to the question of what goes into the 'frame' (Minsky, 1974; Winograd, 1975) for a concept. Chapter III, Section D and Chapter V examine this problem in the limited context of the task dialogues. In these dialogues, the hierarchical structure of the task and the correspondence between task structure and dialogue structure combine to guide implicit focusing.*

D. DEFNP RESOLUTION IN CONTEXT

The focus representation described in the preceding chapter provides a framework for determining when a DEFNP can be resolved and when it is ambiguous. Different types of DEFNPs use the bounds provided by the focus representation in slightly different ways. This section examines several types of DEFNPs and shows the role of focus in their resolution. The aim of this section is not to provide a comprehensive study of DEFNPs, but rather to illustrate the different ways in which the use of focus affects resolution. As a result, several problems that arise in resolving DEFNPs (e.g., differentiating between restrictive and nonrestrictive relative clauses and between specific and nonspecific noun phrases) will not be addressed.

The typical noun phrase has several constituents. For the purposes of this discussion we will consider the structure of an NP to be:

(1) (DET/QUANT)[NUM] NOM

(This rule does not correspond to an actual rule in the SRI speech understanding system language definition. However, the grouping is convenient for purposes of discussing discourse processing.) In this notation, the slanted line indicates a choice of one or the other constituent, parentheses are used for grouping, and brackets indicate an optional constituent. DET is the category containing determiners; it

----

* This kind of inferencing was not handled in the speech understanding system because of the lack of structure in the data base dialogues.

contains words such as "the", "this", and "which". QUANT is the category of all quantifiers; e.g., "all", "any", "some". NUM is the set of number expressions; e.g., "one" and "three hundred fifty." NOM, the set of nominal expressions, contains unmodified nouns, premodified nouns, postmodified nouns, and nouns that are both pre- and postmodified. Respective examples of such NOMs are "wrench", "box-end wrench", "wrench with the red handle", and "box-end wrench with the red handle."

The emphasis of this section is on the processing done on definite noun phrases to go from the semantic interpretation to the identification of the referent. The effect of number, determiners, and quantifiers on the final interpretation of NPs is discussed. To avoid complications, several forms of NP have been omitted from the discussion; for example:

(2) (DET/QUANT)[NUM] (e.g., "those two")
(3) NUM (e.g., "two")
(4) QUANT of NP (e.g., "two of the bolts")

The elliptical aspects of forms (2) and (3) complicate their interpretation. Form (4) can be handled by semantics alone. The discourse aspects of the phrase are all handled when resolving the embedded NP (e.g., "the bolts").

There are many syntactic and semantic problems associated with parsing and building representations for the group of phrases in the category NOM. For example, it takes semantic knowledge to determine the difference between "the big ship" and "the British ship". For the purposes of this section, these problems can be ignored. We will assume that any NOM has been checked syntactically and that a semantic representation has been built for it. The first section below describes the interface between semantics and discourse. It is only when looking for the concept described by the NOM that discourse processing is really needed. There are several dimensions that influence the interpretation of DEFNPs; these are discussed in subsections of Section D.2 below. The simplest NPs, from the discourse point of view, are unquantified, unmodified NPs. These are discussed in the first subsection. The following subsection looks at some of the problems introduced by adding modifiers. Since genitives present special problems, they are discussed separately. Finally, some problems that arise from introducing quantifiers are addressed.

1. FROM SEMANTICS TO DISCOURSE

The semantic interpretation for the NOM constituent of a noun phrase encodes the relationships among the concepts that are conveyed by the constituents of the NOM in the underlying knowledge representation. In essence, it provides a representation of the typical item described by the NP. For example, the representation for "American sub" in the partitioned semantic network notation is shown in SPACE P1 of Figure IV-2. Note that the 'ownership' relation conveyed by "American" in this particular construction is represented in this network structure. The discourse component contributes to building an interpretation of an NP only if the determiner or quantifier for the NP indicates definiteness. The basic problem for the discourse routines is to locate the object or set currently in focus that corresponds to the description in the NOM part of the NP. When an instance of NUM is included in the NP, discourse processing is influenced only insofar as a check on the set found is required to be sure the set has the correct cardinality. "One" is an exception and is treated like "a" rather than other, plural, NUMs. For the NP, "the American sub", an individual submarine owned by the U.S. must be found in focus. For the NP, "all six American subs", a set of (exactly) six subs, all owned by the U.S., must be found.

2. INTERPRETING COMPLETE NPS

The matcher, when augmented for focus matches as described in the preceding chapter, performs the central function in the process of interpreting complete NPs. Given the semantic interpretation of a DEFNP and a focus vista, it determines which, if any, object in that focus matches the DEFNP. Note that the first kind of inferencing discussed in

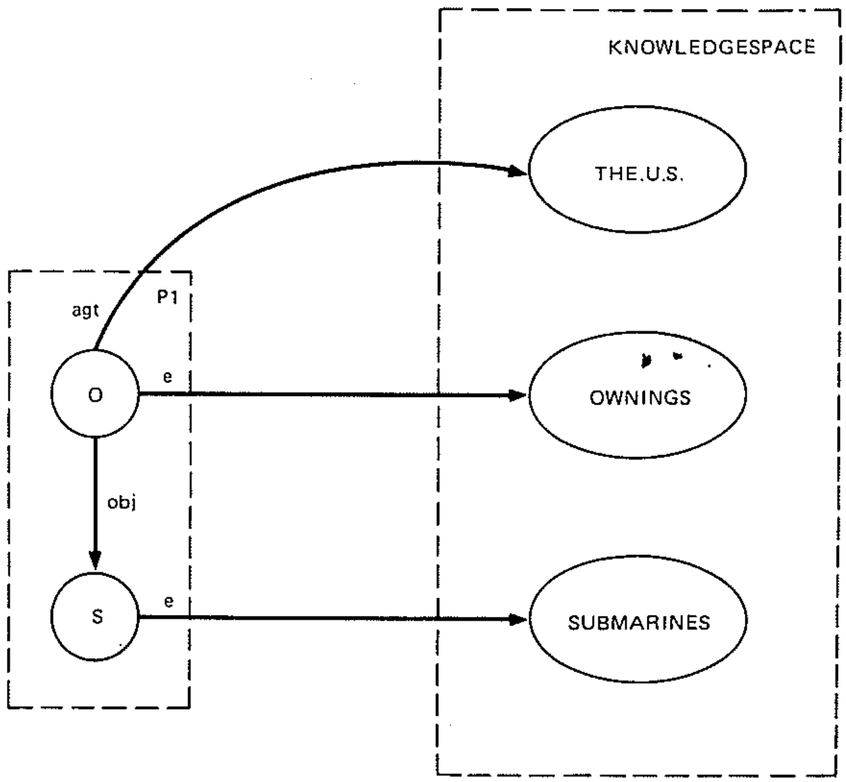


FIGURE IV-2 PARSE LEVEL SEMANTIC NET REPRESENTATION FOR "AMERICAN SUB"

Section C occurs at this stage of the processing. The matcher in determining whether a given object in focus is the referent of the DEFNP, follows the subset hierarchy and deduces information from theorems in the network. The restriction of the search to a focus vista is crucial; generally, the number of objects in focus is quite small and contradictions (e.g., if the candidate focus space node and the node corresponding to the head of the DEFNP are elements of mutually exclusive sets) can be reached quickly for many of the objects. At present, this matching procedure is carried on depth-first. In the limited data base domain for which resolution has been done, this strategy is sufficient. A parallel search has the advantage of finding the match more quickly, on the average. However, it is still necessary to establish that no other object matches in order to rule out ambiguities.

a. UNMODIFIED UNQUANTIFIED DEFNPS

The search for the referent of unmodified unquantified DEFNPs starts by examining explicit focus.* If a match is found, the node matching the node that corresponds to the head noun of the DEFNP indicates the referent. If a match is not found, one of three possibilities still exists (assuming the NP can be resolved!): the DEFNP may refer to a concept implicitly, but not explicitly, in focus; the concept may be unique (e.g., "the sun"); or the DEFNP may contain a genitive or a modifier containing new information (e.g., the DEFNP, "the red coat" when several coats are in focus, but none is known to be red).

The uniqueness check requires determining whether more than one object fitting the DEFNP description exists in the knowledge base. This check is done after the search of focus, because context may in fact overrule the usual uniqueness conditions. The phrase "the sun" in the sequence,

Rose has a beautiful sunset picture.

The sun is teetering above the mountain.

refers to the image of the sun in the picture, not the real sun; i.e., the sunset picture creates a context with a special sun.

A plural DEFNP may refer to a set in the same way that a singular DEFNP refers to an individual. However, the resolution of a plural DEFNP may encounter an additional problem. The DEFNP may create a new set by grouping together objects already in focus. In the sequence,

You will need the wrench, the screwdriver, and the hammer.

----

* For relational DEFNPs, e.g., "the height of the building", a unique result always is obtained (through the relation) and the focus mechanism is not used.

Should I put those tools in the tool box?

the DEFNP "those tools" (note that the pronoun "them" could also have been used) refers to the set of three individual tools in focus. The set itself, however, does not exist as a node in the network. The resolution routines handle this problem by looking for individual objects in focus that satisfy the DEFNP. If it finds more than one such object, a new set is created and added to focus.

b. MODIFIED NPS

Modifiers may be used in three ways. The simplest case is the use of modifiers to select among individual objects in focus. This case entails a straightforward match (although some inferencing may be required). Modifiers are also used to select an element of a set in focus (when the individual elements of the set are not explicitly in focus) and to supply new information about an object in focus. The last two cases each present problems, and the existing DEFNP routines will fail to find a match for instances of either.

An example of selection from a set occurs in the sequence,

A high school class came to visit the hospital.

The brightest student . . .

The DEFNP "the brightest student" singles out an element of the high school class. An example of new information being added by the DEFNP occurs in the third sentence of the sequence,

Jane got some books today.

They're on the coffee table.

The new book by Haley is on top.

The DEFNP "the new book by Haley" singles out one of the set of books. The information that Haley wrote it and that it is new is introduced by the DEFNP.

Resolution of such references requires both implicitly focusing on the individual elements of sets in focus (e.g., the individual students in a class are implicitly focused when the class itself is explicitly focused) and using the modifier(s) to select one element. The ability to remove modifiers from the DEFNP until a match can be found is required. That is, if a match of the complete DEFNP cannot be made, successively less restrictive matches must be tried. For these more complicated searches, the use of a focus representation to constrain the search is crucial. When a match is found, the removed modifiers may be asserted as new information about the matching concept. If the match is to a whole set, information may be asserted about one of the members of the set. For instance, in the second sequence above, the information about Haley is asserted of one of the books that Jane got. The use of network partitioning to reflect the parse structure of a DEFNP in the semantic interpretation of the phrase (used for ellipsis, see Chapter VI; also used by semantics, see Hendrix, 1976) provides a means of removing modifiers from DEFNPs. The problem that remains (a major reason such DEFNPs were not handled in the speech understanding system) is deciding how (i.e.,in what order) to strip modifiers.

c. <u>GENITIVES</u>

Two kinds of problems can arise when a genitive is used as a (preposed) modifier in a DEFNP. First, several similar items may be in focus and the genitive used to choose among them. When used this way, a genitive may cause the same problems as other modifiers. As an example, consider the use of the DEFNP "Peter's car" when a set of cars, one of which is owned by Peter, is in focus. This use of the genitive may be handled exactly like other modifiers. If no car is known to be owned by Peter (i.e., the genitive supplies new information), ownership by Peter can be asserted of one of the cars (as long as Peter is identifiable in focus).

The second and more interesting problem arises when only the genitive constituent of the DEFNP is in focus. In this case, the genitive constituent supplies the old information in the phrase. That is, the genitive constituent of a DEFNP may refer to an object in focus, while the object referred to by the complete DEFNP may not be in focus. For example, assume a focus in which there are two people, a boy and a girl. Then the phrase "the boy's mother" is unambiguous and resolvable because the boy is in focus and mother-of is a unique relation. That is, even though there is no mother in focus, there is a boy in focus, and the relation conveyed by the genitive can be used to determine, via the link to the boy, which person is being referred to.

In a sense, a DEFNP with a genitive has two heads: the head of the genitive, as well as what is usually considered to be the head noun. For this reason, if a DEFNP with a genitive cannot be resolved, the genitive constituent of the noun phrase alone must be considered. The genitive must be resolvable. If the remainder of the NP is not resolvable in focus, then the genitive relationship must be used to determine uniqueness. This processing is identical to that done if the genitive is expressed by an embedded noun phrase. That is, if "the y of the x" were used instead of "the x's y", then "the x" would be resolved to some particular concept, say X1, and then "the y of X1" found using properties of X1 and the y-of relation.

d. QUANTIFIED DEFNPS

At the discourse level, the processing of quantified DEFNPs is the same as that for unquantified plural DEFNPs except for the consideration of a generic interpretation. The interpretation depends on whether the optional NUM (number) constituent is present in the DEFNP and on the particular quantifier used.

For constructions not including NUM, the question of interpreting a phrase generically depends on whether or not a referent can be found in focus. If a referent is found, then the quantified DEFNP inherits the generic property of the referent. In sequence G1, the DEFNP "both dogs" is generic; in sequence G2, it is not.*

G1: The collie and the Labrador are good pets. Both dogs are gentle.

---

* These examples were suggested by B. Nash-Webber.

G2: Rose has a collie and a Labrador. Both dogs are gentle.

If no referent can be found, then a generic interpretation of the quantified DEFNP may be used in certain cases. Quantifiers implicitly conveying a set of size two ("both", "either", "neither") are never generic in this way. There must be a referent in focus for these quantifiers to be meaningful. In contrast, "all" can always be interpreted generically; in fact, construction (4) (i.e., QUANT of NP) is usually used to limit the restriction of "all" to some local set (e.g., "all of the bolts").* "Some" and "every" also tend to convey the generic, but less strongly than "all".

The heuristic used in the speech understanding system was to assume the generic for "all" (and force use of form (4) if a local set was meant) and when "some" and "every" were used with unmodified NOMs. In all other cases, a referent was looked for first. If no referent could be found, then the generic interpretation was assumed for quantifiers other than "both", "either", and "neither". There are clear counter-examples to this rule; e.g., in the utterance, "Some tall trees are killed by lightning", the generic is intended even if there is some particular set of trees in focus. Such cases are not currently handled by the discourse routines.

Inclusion of a NUM in the NP limits the quantifier to one of "all", "some", or "any".** For this construction, it is always the case that a local referent must be found, with the correct cardinality, over which the quantification holds. To see this, contrast "All subs have beams over 30 feet." with "All five subs have beams over 30 feet." In the first utterance, the generic interpretation (all of the subs in

----

* At first there seems to be some ambiguity between expressions involving "all" meaning "all in the computer knowledge base" and "all in the world". However, this ambiguity can be seen only from a frame of reference outside the computer model. Inside, the two are, by definition, equivalent.

** For semantic reasons, the use of "any" and "some" in this construction were not handled in the speech system.

the world) is clearly preferred. In the second, a referent must be identified in focus and the DEFNP is interpreted generically only if the referent is generic.

E. SUMMARY

This chapter described the role of the (global) focus representation in the resolution of certain kinds of definite noun phrases. The resolution of definite noun phrases entails a number of problems ranging from deciding what items in the knowledge base to consider as possible referents to determining when a referent has been found or when a phrase is ambiguous. Given a representation of focus, a number of different questions arise that depend on the particular kind of definite noun phrase being resolved. A subset of these problems was examined to illustrate the importance of the focus representation to the resolution of DEFNPs.

V SHIFTING FOCUS

CONTENTS:

A. INTRODUCTION

This chapter links the preceding three chapters and completes the description of the focus representation with a discussion of a mechanism for shifting focus. Chapter III presented a representation of focus and described its use for constraining the retrieval operations of a knowledge based system. The discussion in Chapter IV assumed the presence of such a focus representation and described its role in the process of identifying referents of definite noun phrases. The problem then becomes deciding what objects should be in focus at any given point in a discourse. For task-oriented dialogues, the relationship between focus and both discourse and task structure that is described in Chapter II provides a key to the solution.

A shift in focus may be directly stated by some utterance in a discourse (e.g., "I've finished that step. What's next?" or "Let's change the topic"), but usually the cues are more subtle. For example, when the discussion of some activity turns to a discussion of one of the participants in the activity, the focus shifts from the overall activity to that participant. What constitutes a shift in focus depends on both the kind of discourse and the topic of discourse. The shift strategy described in this chapter is specific to task-oriented dialogues. It reflects the task as the major topic of such dialogues and, hence the major indicator of shifts of focus. Although the rest of the focus representation is general, this aspect would need modification for application to other kinds of discourse.

It is important to distinguish here between different kinds of discourse and different domains. The important point to be taken from this chapter is that some top-down model of the structure of discourse is needed to guide the decision about whether a particular utterance shifts the focus of a discourse and how. The shift of focus in task-oriented dialogues is closely tied to the particular subject domain of the dialogues (i.e., the task) because the structure of the dialogues parallels the structure of the task. As a result, domain information is used by the shift strategy described here. However, the use of domain information should not be taken to mean that shifts of focus are domain dependent nor that switching tasks requires switching shift strategies. Shifts of focus in other kinds of discourse (e.g., novels, newspaper stories) are often not as closely related to the particular domain.

The major portion of this chapter describes a mechanism for shifting focus that has not been implemented yet because it requires a task representation that is currently being designed. A much simpler shifting strategy that was implemented as part of the discourse component of the SRI speech understanding system is described briefly first.

B. THE LINEAR CASE

A simple shift strategy was implemented in the SRI speech understanding system to test the use of the focus representation for resolving definite noun phrases. This strategy is linear; it does not take discourse structure into account. Basically, the concepts in an utterance are considered in focus until a small number of subsequent utterances have been interpreted.

After an utterance is parsed, the concepts in the (accepted) interpretation of the utterance are entered in a focus space. The focus spaces are arranged in a first-in first-out queue of a fixed size. The distinction between the active focus space and other open focus spaces is captured by considering the focus space corresponding to the utterance processed last as the active focus space and other focus spaces in the queue as open.

This strategy results in a reference resolution mechanism that is similar to those of previous systems (e.g., Winograd, 1971; Norman et al., 1975) in which those items that occur in a fixed number of (or all) preceding utterances are considered as possible referents. This strategy is not adequate for resolving the references that occur in many interesting kinds of discourse. An adequate reference mechanism must take into account the overall structure of a discourse and the way individual utterances fit in that structure. Recall from Chapter II that the task-oriented dialogues were more structured than the data base dialogues. In the remainder of this chapter a more sophisticated shift strategy which uses the additional structure available for task-oriented dialogues is presented.

C. THE INFLUENCE OF TASK STRUCTURE

The structure of a task provides a framework for the structure of a dialogue concerning that task because (performance of) the task is the topic of the dialogue. Chapter II presented several examples that illustrate the relationship between the structure of a task-oriented dialogue and the structure of its corresponding task. It is important to recognize that the use of the structure of the task as a framework for the structure of the dialogue does not result in a static model of dialogue structure. The task model does not prescribe the exact form of a dialogue. Rather it provides a description of the pieces (i.e., subtopics) that can enter the dialogue and relates these pieces in a hierarchy. Only some of these pieces will enter any particular dialogue. Similarly the order in which pieces are invoked, although partially constrained, varies from dialogue to dialogue. Hence, the use of the task structure as a framework does not restrict the system to understanding dialogues that it has heard before (or ones whose precise structure have been built in). Part of the interpretation of an utterance is the determination of how the utterance influences the focus of the dialogue. For task-oriented dialogues, the structure of the task provides a top-down guide for these decisions.

In task dialogues, a shift in focus takes place whenever a new task is entered or an old one completed. A narrowing of focus takes place whenever a subtask of the active task is opened for discussion. The focus shifts back up to the higher level task when that subtask is completed. Hence, when a subtask of the current task is referenced, a new active focus space is created below the current active focus space. When the subtask is completed, the new focus space is closed and the old space (i.e., the higher space) becomes the active focus space again. The top of the focus space hierarchy is the focus of the overall task. In addition, new focus spaces may be created by other kinds of subdialogues (e.g., a general question about some tool or procedure).

D. DETECTING SHIFTS IN FOCUS

Intuitively, a shift in focus takes place in a task-oriented dialogue whenever the particular subtask that is being performed changes. The shift may be either to a subtask of the current subtask, to another subtask at the same level as the current subtask, to a general subproblem like identifying a part or using a particular tool, or back up to a higher level task (i.e., to a supertask of the current task). The major problem is to decide whether a particular utterance entails a shift in focus and, if so, what the new focus is. In general, a new subtask is entered by an utterance (from either apprentice or expert) that references the goal action or objects involved in the subtask. When this happens, a new focus space is established. Initially the only concepts explicitly focused on are the concepts mentioned in this utterance and any objects associated with this level of the subtask but not explicitly mentioned in the utterance (they are assumed in focus even though elided from utterance). As more utterances concerning the subtask are processed, any new (i.e., not focused) concepts associated with them are added to focus.

A shift of focus may be cued by any part of an utterance: a noun phrase, a verb phrase, or modifying phrases. Although an individual constituent (e.g., noun phrase) may indicate a shift of focus, the constituent alone cannot be used to determine the shift, as will be shown shortly, because the remainder of the utterance influences the shift. For example, the utterance may include some higher-level embedding predicates (e.g., need, belief) that affect whether or not a shift in focus is needed (and, how that shift is to be handled). Furthermore, time information (e.g., tense of a verb) influences the decision. For instance, modification in a noun phrase can indicate a previous context (e.g, the screws that you bought yesterday). The following discussion looks first at the relationship between identifying the referent of a definite noun phrase and shifting focus and then at the interaction of the various noun phrase and verb phrase constituents of an utterance in determining a shift. The discussion will be restricted to task-related utterances.

1. INDIVIDUAL DEFNPS AND SHIFTS OF FOCUS

A shift in focus may be foreshadowed by an individual DEFNP.* For example, a DEFNP that refers to an item implicitly focused by some

----

* The same shift may be indicated by other information in the utterance, e.g., the main verb. The point here is that the resolution of the DEFNP may provide information that must be considered in the shift.

explicitly focused task is an indication of a possible shift in focus to the subtask involving that item. An item may be implicitly in focus either because it participates in some subtask, either of the current task or of some other uncompleted task, or because it is associated with some object in explicit focus. Its connection to explicit focus must be examined to see if a shift in focus is indicated. In particular, in the task dialogues, if the connection to implicit focus comes from participation in some subtask, the reference is considered to indicate a shift in focus to that subtask (unless the indication is overridden by other information in the utterance).

To illustrate how a DEFNP may indicate a shift in focus, consider the task hierarchy of Figure V-1 and the focus environment portrayed in Figure V-2. The task hierarchy is only for the reader's benefit; it does not reflect any structure in the computer representation. This task structure is part of the information encoded in the process model described in Chapter III. The dotted lines show the task hierarchy and the solid lines show time sequencing. Suppose that task T2 (in Figure V-1), installing the aftercooler, is the current task. The focus spaces FS0, FS1, and FS2 correspond to subtasks T0, T1, and T2. FS2 is the active focus space; the vista (FS1 FS0) is the hierarchy of open focus spaces.

A reference to an item in either the active focus space or one of the open focus spaces does not cause a shift in focus. Those items in the active focus space are considered first when resolving a reference because the currently active task is more in focus than its embedding tasks. The phrases "the aftercooler", "the wrench", and "the crescent wrench" all refer to objects in FS2, the active focus space. Hence, the use of any of these phrases does not affect focus of attention. The referent can be retrieved immediately. The use of "the air compressor", "the pump", or "the ratchet wrench" also does not cause a shift in focus. Since these objects are in open focus spaces, they are also in focus, but are accessed only after considering the objects in FS2. Note that the noun phrase "the wrench" is not ambiguous because

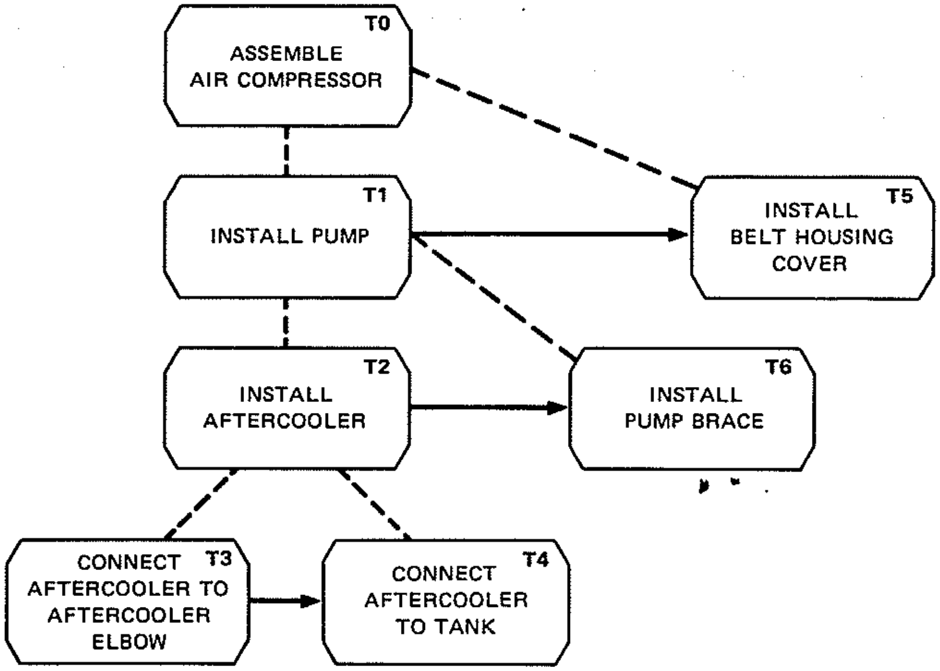


FIGURE V-1 PARTIAL TASK HIERARCHY FOR ASSEMBLING AIR COMPRESSOR

of the distinction between the active focus space and the open focus spaces.*

References to either a new subtask or a new parallel or higher task, or to subtasks of any of these, do change focus. As an example, consider the expansion of Figure V-2 in Figure V-3. Space IADS contains the delineation of the process for installing the aftercooler. The plot space of this delineation is the implicit focus for node 'IAC1'. It shows that this installation has two substeps (corresponding to T3 and T4 in Figure V-1). The first substep, I1, involves a connection operation, IO1, between the aftercooler and one of its subparts, an aftercooler elbow (represented by the node 'ACEX' in the figure). The phrase "the aftercooler elbow" indicates a possible shift in focus to task T3 since there is no aftercooler elbow in explicit focus, but there is one in the implicit focus for IAC1. The

----
* See the example in Chapter II, Figure 15.

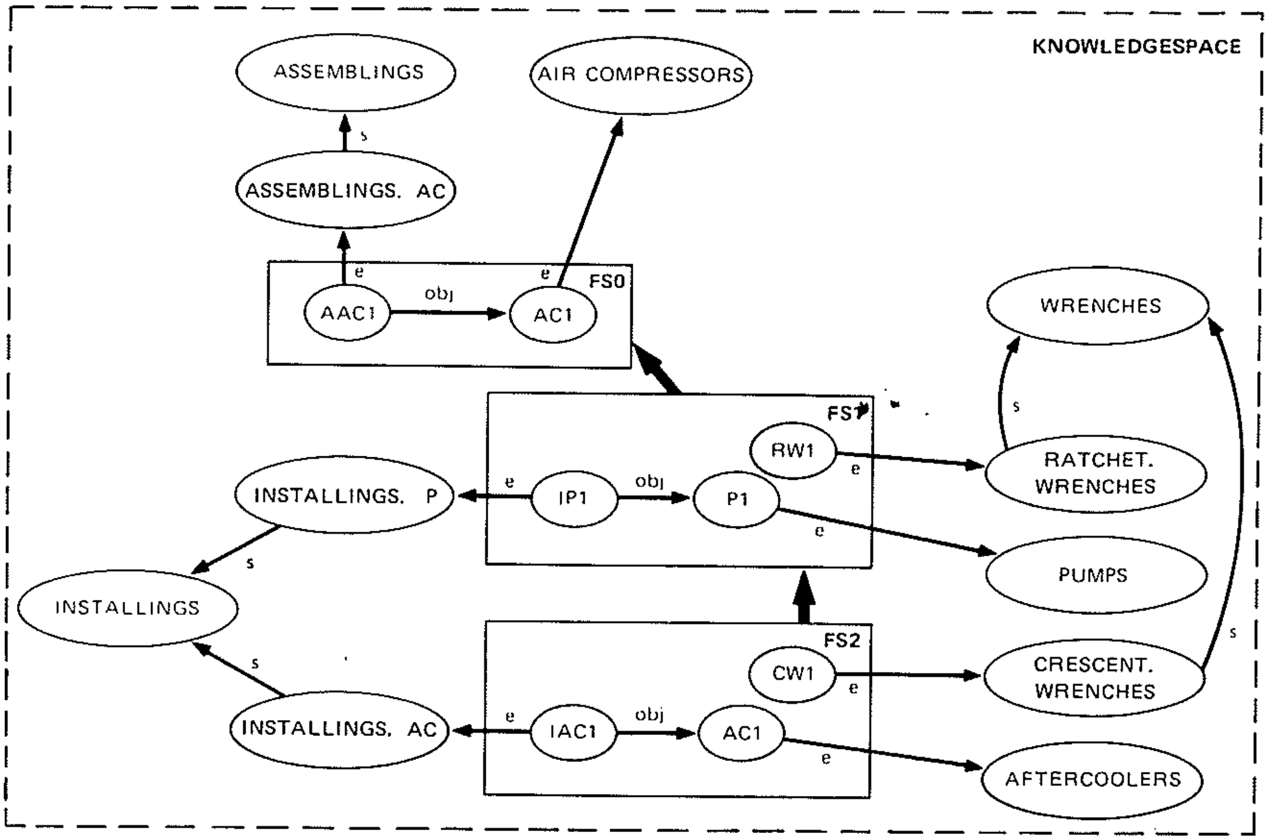


FIGURE V-2 FOCUS SPACES FOR ASSEMBLY TASK

node 'INSTALLINGS.P' has a similar delineation that includes substeps for tasks T2 and T5. The substep for T5 involves a pump brace. Hence, the occurrence of the phrase "the pump brace" would suggest a shift to task T5. A decision about whether to shift focus in either of these cases depends on the remainder of the utterance. If T3 is opened, a new focus space is created below FS2 in the hierarchy. If T5 is opened, T2 is closed and and a new focus space is created below FS1 in the hierarchy.

A shift of focus may entail instantiating new entities or identifying real entities corresponding to hypothetical entities in

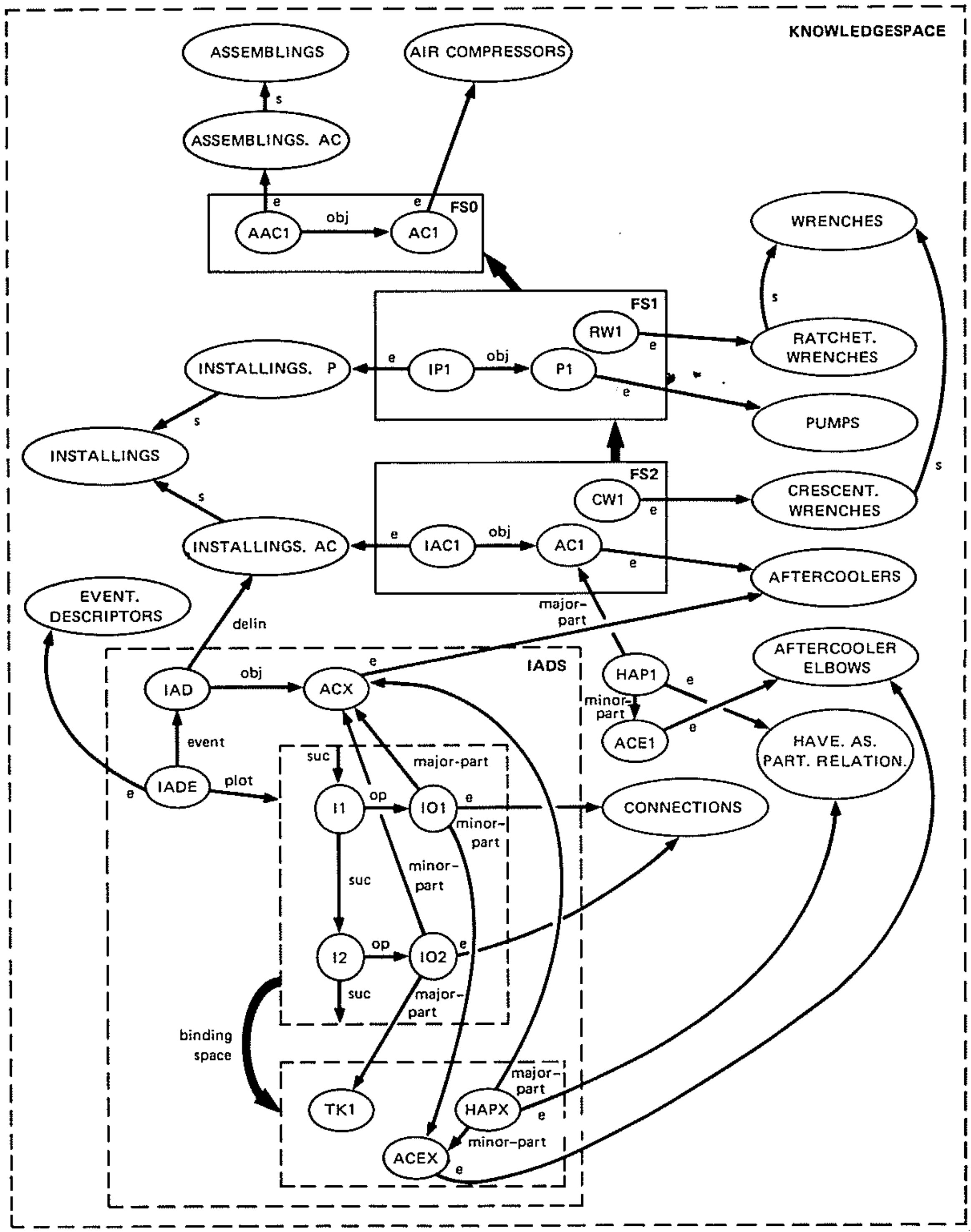


FIGURE V-3 FOCUS SPACES AND IMPLICIT FOCUS FRAGMENT FOR SHIFTING FOCUS implicit focus. For example, if focus is shifted to task T3 (i.e., an instantiation of node 'I01'), the aftercooler elbow ACE1 is brought into focus and the noun phrase "the aftercooler elbow" is identified with it. This identification comes from determining that ACE1 is in the same relationship to the currently focused aftercooler, AC1, as the hypothetical ACEX is to the hypothetical ACX.

The search for the referent of a DEFNP takes into account the difference between those items that do and those that do not shift focus. Items that do not cause a shift in focus are checked first to determine if they are referents. Hence, items in explicit focus are always checked before items in implicit focus. In the task dialogues, only those items that are implicitly in focus because they are participants in a substep of some explicitly focused process can result in a shift in focus. Hence, those items are the last items in focus to be checked.

2. INTERACTION BETWEEN THE DEFNPs IN AN UTTERANCE

The final decision of what focus an utterance has must wait until the entire utterance is processed. If a shift in focus is indicated by more than one constituent of the utterance, then the final shift must be consistent with all of these indicators. That is, if the constituents that require a shift do not all indicate a shift to the same focus (i.e., subtask or subtopic), then a search of implicit focus must be made for a shift that will satisfy all the different shift indicators.

To illustrate, consider the set of possible next tasks in Figure V-4. In a real situation, some of these subtasks might be subtasks of the current task and others might be new higher level subtasks. For the purposes of this example, the notation object[i] means some object that might be called "the object[i]"; that is, object[i] in task T1 and object[i] in task T2 are not necessarily the same object though they are the same kind of object. For example,

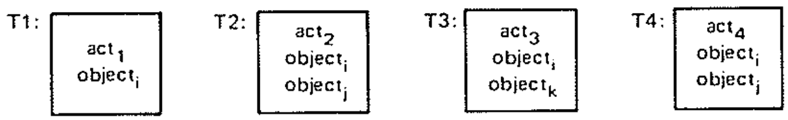


FIGURE V-4 A SET OF POSSIBLE NEXT SUBTASKS

object[i] could be a crescent wrench in T1 and an open-end wrench in T2; the phrase "the wrench" could refer to either. Assume that the search for referents of DEFNPs described in the preceding section considers tasks in the order T1, T2, T3, T4. Also, for the purposes of this example, assume that no object meeting the description object[i] or object[j] is in focus (implicit or explicit) except as a participant in these tasks.

At the phrase level, the DEFNP "the object[i]" will be resolved to the object in task T1. Even if no other DEFNPs occur in the utterance, this match can only be considered tentative. For the match to be final, and a shift to a new focus corresponding to T1 to occur, the action of the utterance must correspond to 'act[1]'. Hence, when the complete utterance has been parsed, a match of the action with the task is checked before a shift in focus to task T1 is made. If the utterance action is not 'act[1]' then a search is initiated to find an implicitly focused subtask involving both an object[i] and the correct action.

If the DEFNP "the object[j]" occurs after "the object[i]" has been resolved, then task T1 is rejected as a possible next focus and a search for a task involving both objects is carried out. Task T2 is selected. Note that if the DEFNP "the object[j]" had occurred first in the utterance, T1 would have been rejected immediately and task T2 would have been the proposed focus. Then when "the object[i]" occurred, the search for a referent could start from task T2. Again the whole utterance must be checked since either 'act[2]' or 'act[4]' can occur with these two objects.

3. THE DEFNP TABLE

A table of the noun phrases that occur in an utterance and their referents is built as the utterance is parsed, both to facilitate coordinating the shifts indicated by the different DEFNPs and to enable instantiation of implicitly focused items. As each noun phrase is resolved, it is entered in the table. If the referent is found in explicit focus, nothing further is done. If the referent is in implicit focus and the connection to implicit focus indicates a shift, the shift is compared to other entries in the table that indicate shifts. This table acts as a cache when parsing left to right; in particular, the table can be used for resolving intrasentential references (see Limitations and Extensions section).

The general form of the table is shown in Figure V-5. The parse vista (i.e., semantic interpretation) of the DEFNP is kept to enable a new search for a referent to be found in case there is a conflict between items indicating a shift in focus. The RESTYPE (resolution type) entry is used to distinguish referents in implicit focus. This distinction is used for instantiating implicitly focused items (e.g., "the aftercooler elbow" in Figure V-3) as well as for coordinating references that indicate a shift of focus. The list of matcher bindings is kept for updating focus after the entire utterance is processed. If there is a shift in focus, then all of the information in the utterance is entered in the new focus space. If no shift is required, the new information in the utterance (i.e., any information not already in focus) is added to the active focus space.

E. EXAMPLES

This section examines several different sentences from the

PVISTA	RESTYPE	WHERE	MATCHLIST
parse vista of DEFNP	one of {CFS HFS LT NT NTIF}	focus space or implicit focus pointer	list of bindings returned by the matcher

CFS = current focus space

HFS = higher space in open focus space hierarchy

LT = subtask of current task

NT = new (sub)task

NTIF = implicit focus, not task-related

FIGURE V-5 THE DEFNP TABLE

perspective of how they interact with a given focus. Consider the task of constructing a carrying case that consists of a box with a lid that has a handle. The hierarchy of subtasks for this task is given in Figure V-6. (This structure does not correspond to any proposed internal representation; it is provided only to clarify the following discussion). Figure V-7 shows a sample focus environment at the substep of attaching the handle to the lid (i.e., T1). Focus space FS1 contains items in focus for task T1: a lid, a handle, and an attaching operation. Focus space FS3 contains a particular fastening, F1, of handle H1 to lid L1 using the fasteners S1. The fastening is a substep of attaching the handle to the lid. This substep relationship is represented by the ep (for event part) arc from node 'F1' to node 'AHL1'. The plot structure associated with AHLS shows the canonical relationship between such fastenings and attachings.

Consider a focus environment in which T3 is the current task, FS3

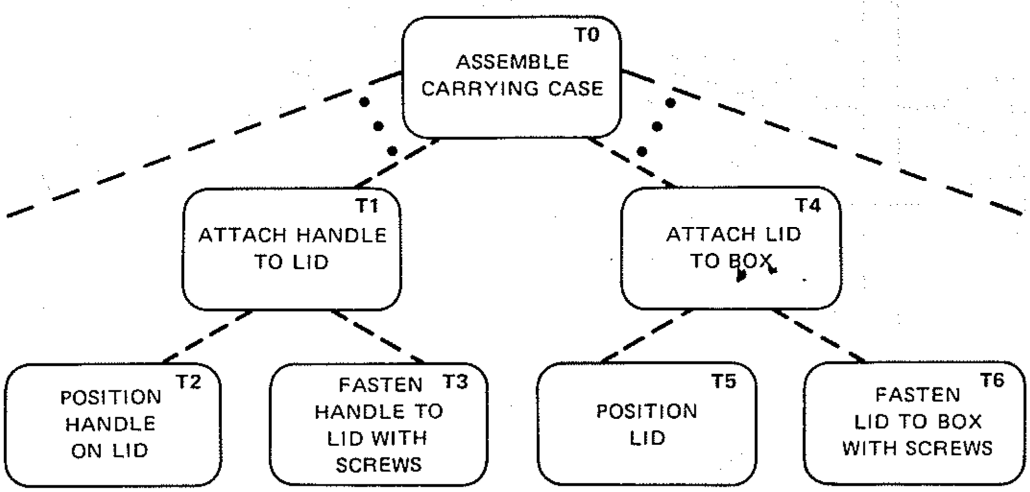


FIGURE V-6 PARTIAL TREE OF SUBTASKS FOR ASSEMBLING A CARRYING CASE

is the active focus space, and FS1 is an open focus space. In this environment, the noun phrase, "the screws" is resolved to the set S1. If the phrase occurs in the utterance, "The screws are one inch long," then this new information about S1 is added to FS3. If the phrase occurs in the utterance, "The screws don't fit", then a new focus space is created below FS3 to contain the items in any dialogue that ensues concerning this problem. If the phrase occurs in the utterance, "The handle is fastened down with the screws", then focus space FS3 is considered closed; the utterance indicates the subtask is completed.

Suppose now that FS3 is closed (e.g., by the utterance, "The handle is fastened down"). Focus shifts back up to the complete task (constructing the carrying case) and its associated focus space (not shown in the figure), because T3 is the last subtask of T1; that is, FS1 is closed as well as FS3. If the next statement is "The next step is to attach the lid to the box," a new focus space, FS4, is created as shown

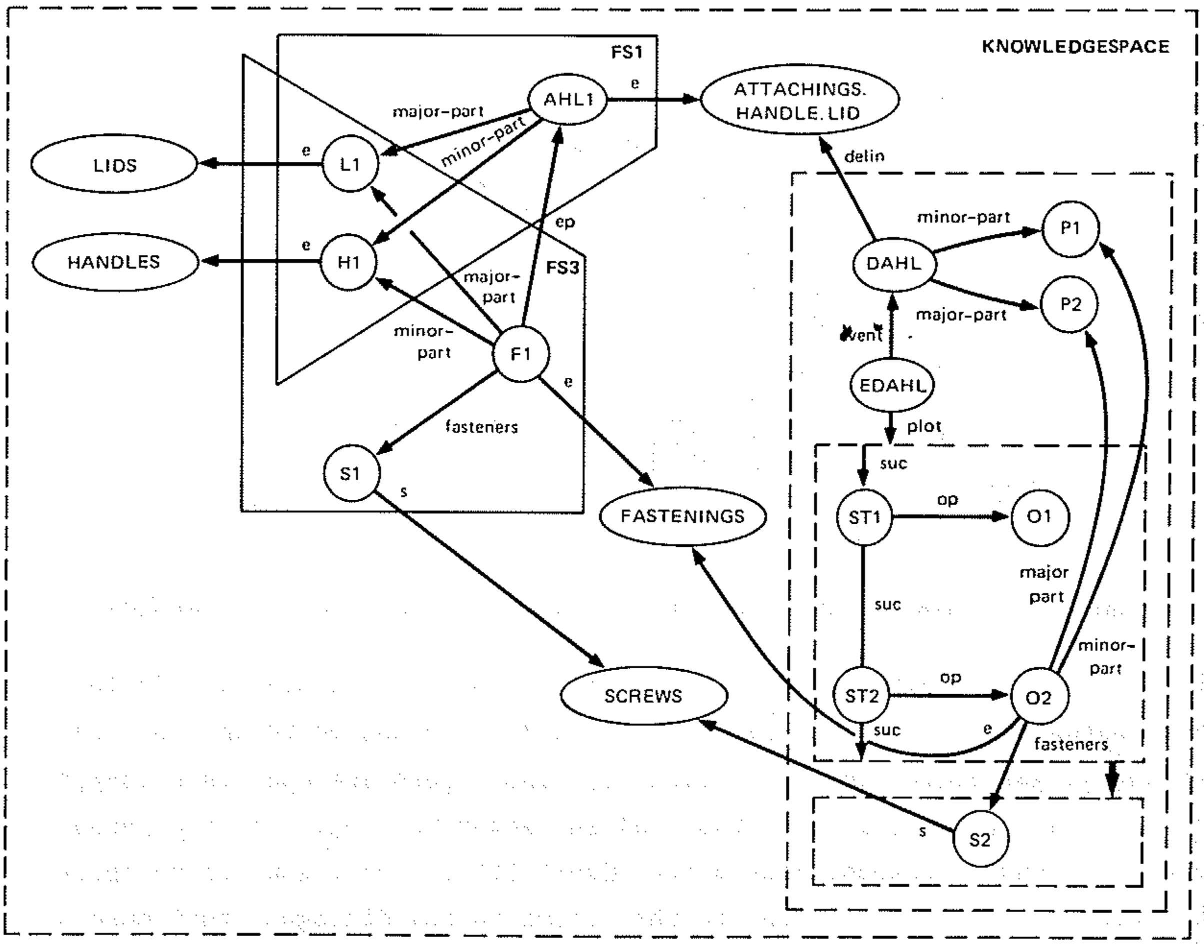


FIGURE V-7 A SAMPLE FOCUS ENVIRONMENT

in Figure V-8. In this focus, the noun phrase "the lid" will be resolved to L1 because that lid is in focus. If the phrase, "the screws" appears, it will be tentatively resolved to the set of screws implicitly focused by the lid attaching operation. The resolution is marked as implicit in the DEFNP table so that it can be checked with other resolutions that require a shift and so that a final resolution to actual screws (and not the hypothetical ones S3) is done. A tentative

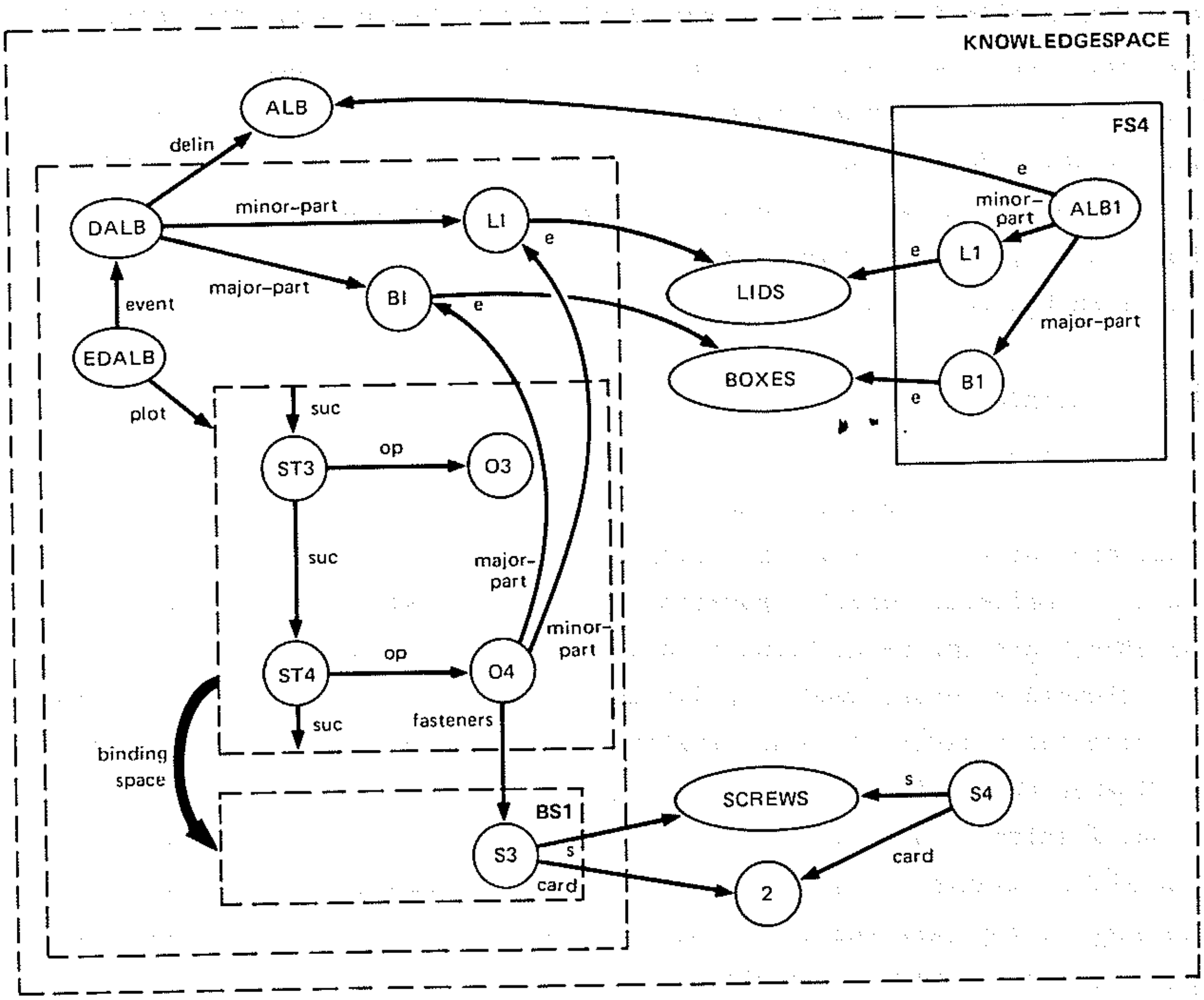


FIGURE V-8 FOCUS SPACE FOR ATTACHING LID TO BOX

shift to a new focus, corresponding to task T6, securing the lid, is recorded in the DEFNP table. If the remainder of the utterance agrees with this shift (e.g., "The screws are all in place in the lid"), then a new focus space will be created below FS4 (the focus space for T4), the set of screws S4 will be found to correspond to the hypothetical set S3 and moved into this new focus space, and focus will be shifted to this space.

As an example of the affect of a higher order predicate on the shift in focus, consider the question, "Have you ever gotten this far and then realized that the screws don't fit?" The phrase "the screws" in this utterance does not refer to the particular screws used in T4, but rather to the hypothetical screws that are part of this kind of operation. Focus shifts, but to a hypothetical world. Although partitioned networks can handle hypothetical worlds (see Hendrix, 1975a,b), no work has been done yet to incorporate this kind of shift in the focus representation.

F. LIMITATIONS AND EXTENSIONS

There are several limitations to the shift strategy discussed in this chapter. It does not take into account the influence of utterances that are not task-related (see Chapter II) and of higher level embedding (e.g., believe, want) phrases. An examination of how these constructions influence focus is needed to allow the mechanisms here to be extended to other kinds of discourse. The shift mechanism needs to provide for backup; there are instances when a subsequent utterance may influence the shift or clarify an ambiguous situation (e.g., when the exact meaning of an "ok" is unclear; see Chapter II). A more immediate issue for task-related dialogues is using information about the major substep of any task and about the time between successive utterances in deciding about shifts. Finally, extensions are needed to the focus and shift mechanisms to accommodate intrasentential references. In this section, some extensions to the preceding design to overcome these latter problems are presented.

1. INTRASENTENTIAL REFERENCES

The DEFNP table can be used to aid in the resolution of intrasentential references including (with one addition) forward pronominal reference. If an utterance is processed left to right the DEFNP table provides a kind of sentence focus. Those DEFNPs already resolved may be referred to later in the utterance (backward reference).

Most often such references are pronominal (e.g. "When Willie fired the pot, it broke"), but they may be DEFNPs (e.g. "When Sue takes the cat and the dog together for a walk, the cat gets upset"). In addition, a pronoun may be used to refer to an object not mentioned until later in the sentence (forward reference). For example, the "it" in "If it's too heavy, don't bring the ceramic bowl".

Since backward reference may also be made to indefinite NPs and pronouns, a record of any that occur in the current utterance must be accessible for reference resolution. Tables of indefinite NPs and of pronouns that occur as an utterance is parsed must augment the DEFNP table. In order to allow for forward reference, the entries in the pronoun table may be marked 'unresolved'. These tables constitute a cache for reference resolution routines.

The first place to look for both pronominal and DEFNP references is in the DEFNP and indefinites tables. If the referent of a DEFNP cannot be found in either of these tables, the focus space and implicit focus search previously described is invoked. If the referent of a pronoun cannot be found in these tables, the pronoun table is checked before any other reference finding procedures are invoked.* If this fails, the pronoun is marked as unresolved. After a complete utterance is parsed, a check is made for any unresolved references. If one is found, the entries in the DEFNP table and indefinite NP list are checked.

There is a modification of this scheme that seems important for a speech understanding system. In a speech system the input, at the signal level, is much more ambiguous than in a text system.** One of the primary roles of noun phrase resolution is to rule out interpretations:

----

* Resolving pronoun references has many problems that are not being addressed here (see Hobbs, 1976; Nash-Webber, 1976). The point of this discussion is to show how some of the mechanisms needed for DEFNP resolution can be used to help two of these problems.

** See Paxton (1977) for a discussion of some of the problems this causes.

to provide evidence that something that was 'heard' really was not said. Furthermore, as a result of the multiple parses under consideration at any time, it is difficult to maintain and use the auxiliary tables just discussed. A solution is to consider all resolutions to be temporary. Parses with unresolvable NPs will get lower priorities, but not be eliminated. Then, when an utterance is parsed, but before it is finally accepted, the whole parse must be retraversed, this time building the DEFNP and other tables. As NPs are encountered this time, a check is made for intrasentential references including the possibility of forward pronoun references. The lowering of the priority of utterances with intrasentential (especially, forward) references fits with our dialogue data. These references are much rarer than intersentential references.

2. <u>TIME AND MAJOR STEP INFORMATION</u>

Some of the subtasks of a task are more important than others. In many cases, one subtask is distinguished as comprising the key operation of the task. Questions of the tools or parts involved in doing the task most often entail the objects and actions of this <u>major subtask</u>.* Scragg (1975) points out the computational inefficiency of searching all lower level subtasks in order to decide whether some object takes part in a task. The search of implicit focus needs to take this major step into account. It is straightforward to augment the process description of Chapter III to include an indication of what the major step of a task is. The remaining problem is to decide how much the search of the task representation should proceed depth first through major subtasks and how much breadth first.

One further piece of information interacts with the major task information in providing evidence about a shift to a new subtask.

----

* See Werner (1966) for a discussion of how many verbs used to describe tasks have layers of specificity of meaning corresponding to the levels of the hierarchy of the task they denote. For example, sewing a garment can mean the whole operation of selecting a pattern, buying material, etc. or only the more minute operation of moving needle and thread through material.

Information about the time that had elapsed since the beginning of a subtask (or since the last verbal communication) was used by the experts in our dialogue experiments to help determine whether a subtask was completed. In particular, if no communication was received after an amount of time sufficient for completing some task, the expert could (and often did) ask about the completion of its subtasks. The major subtask was often asked about first (used as a reference point for further questioning). The major task information needs to be coordinated with information about the time that has elapsed since beginning a subtask. This requires a more elaborate use of time than provided for in current systems.

G. RELATED WORK

The work most closely related to the focus representation presented in this report is the work on conceptual overlays (Rieger, 1975). Conceptual overlays are a mechanism designed to address the problem of interpreting an action in context. The major emphasis in designing conceptual overlays was on providing a means of determining the inferences that result from a given input in the context of preceding input(s). For example, the statement that "They ran into the den" has different implications following "Susan and John smelled smoke" and "The fox cubs heard a strange noise." Rieger does not address any problems that result from the ambiguities that arise from the input language itself; the system he describes assumes an unambiguous input in some formal representation. For instance, the system would not be concerned about choosing the correct sense of "run" or identifying "the den" in the above example. It would assume this had been done and look at questions like why Susan and John ran into the den. Although conceptual overlays are directed at a slightly different aspect of context and its influence on understanding, the structures used are quite similar to those constructed for the focus representation.

In essence, conceptual overlays are an extension of the task model idea to more general kinds of activities. In Rieger's system, every action is represented by a set of temporally sequenced hierarchical collection of subgoals, called a commonsense algorithm. Conceptual overlays group together with a particular action (e.g., smelling smoke), a set of possible next actions (e.g., doing something about the source of the smoke), and a set of functions that select from among these actions the ones that are most likely to follow. The set of possible next actions constitutes a set of expectations about subsequent inputs. The interpretation of a new input entails identifying how the action it conveys fits into one of these expected actions (or some part of one of these actions).

This representation has direct analogues in the focus representation. Commonsense algorithms provide an implicit focus on subactions of an action similar to the implicit focus provided by the plot space of an event (and other process information reachable from nodes in that space). Conceptual overlays provide an implicit focus like that of the process representation for tasks above the current task in the task hierarchy. Hence, the process of finding where a new input fits into an active overlay is similar to the problem of deciding where a new input fits into the currently active task.

There are two major differences between Rieger's approach and ours. First, commonsense algorithms are designed to deal with actions that have a larger set of possible next actions than the tasks of the task-related dialogues; i.e., actions that are less constraining than the tasks of the task-oriented dialogues. (These actions also lack some of the time ordering constraints of the tasks considered in this report.) Hence, it is more important for him to restrict the amount of depth-first search (through possible next actions and their subactions) that is done in trying to determine where a new input fits. The resulting scheme depends on structuring the knowledge base so that the search from a particular action to all the algortihms it forms a part of is reasonable. (As Rieger points out, it is not clear whether this is possible in general.)

Second, the assumption of unambiguous input avoids the problem of needing context to figure out to what action a particular linguistic input refers. Some of the search that Rieger is concerned with from the perspective of determining inferences is needed to go from English into an internal representation. Although the mechanisms needed to build an interpretation of an utterance entail the same kind of task identification that is needed for inferencing about actions, the information that is available in building this interpretation (i.e., internal representation) is not as complete as the information in the internal representation. As a result, the search of focus described here is more top-down than the search through commonsense alogorithms.

Unfortunately, Rieger does not address the issue of switching overlays to any extent. He is partly able to avoid this problem because he does not look at problems that arise from building an interpretation of a natural language input. The similarity between conceptual overlays and the focus representation suggests that the switching strategies described in this chapter could be extended to other kinds of discourse.

H. SUMMARY

Determining how a particular input influences the focus of a discourse depends on both the kind of discourse and the topic of the particular discourse. This chapter presented mechanisms for determining shifts of focus for a limited kind of discourse, namely task-oriented dialogues. For these dialogues, top-down information about shifting is available from knowledge about the task. The interaction of individual utterances with this information was described. Recent work by other researchers provides some indication of how the focus mechanisms used here could be extended to other kinds of discourse, by providing analogues of task information for other less structured kinds of actions.

VI ELLIPSIS

CONTENTS:

A. INTRODUCTION

Focus not only provides the semantic framework for resolution of definite noun phrases, but also the syntactic and semantic framework for interpreting elliptical utterances. Ellipsis refers to the use of incomplete grammatical units in a discourse (the items left out are elided). Although such a unit is ill-formed by itself (in the traditional competence grammar sense), if the context in which it appears supplies the elided items, it is well-formed. For example, the utterance,

The crescent wrench

is an incomplete sentence, but if it occurs after the question,

What tool are you using to loosen the bolts?

then it is easy to construct the complete sentence it is meant to convey, namely,

I am using the crescent wrench.

"The crescent wrench" is an example of ellipsis at the sentence (or clause) level. Ellipses may occur at the noun phrase or verb phrase level as well.* The following sequence is an example of noun phrase ellipsis:

> Which box should I use for the tools?
> Only the largest will hold all the tools.

Verb phrase ellipsis is shown in the following sequence:

> Has the pump been tightened down?
> No, but the motor has been.

We limited the range of elliptical expressions we would handle in the speech-understanding system to noun phrases functioning as complete sentences, as in the initial example. To allow more extensive noun phrase and verb phrase ellipses would have meant greatly increasing the alternatives considered for these lower level constituents during the interpretation of an utterance. For example, if noun phrase ellipsis had been allowed, when any piece of a noun phrase was constructed, discourse could have been called to try interpreting the noun phrase elliptically. Expanding an elliptical phrase is a relatively expensive operation when compared, for example, with syntactic checks or semantic case checks. Doing it at the utterance level seems worth the cost since complete utterances are relatively infrequent compared with other constituents being proposed and found.** If we had been working with error-free text input rather than speech, the overhead requirements would have been less extreme. Because the words are clearly distinguishable in text input, it is easier to determine where a noun phrase ends. Extensions and modifications needed to do more complete ellipsis handling are described in Section E.

B. OVERVIEW OF ELLIPSIS

Ellipsis is a more local phenomenon than reference. The immediate

----

* Halliday and Hasan, 1976, contains a comprehensive discussion of these various forms of ellipsis.

** Paxton, 1977, contains a discussion of parsing strategies in the speech environment.

focus provided by one utterance is used to expand any elliptical phrases in the following utterance. The constituent phrases of the first utterance provide the framework for expanding and interpreting the second utterance if it is elliptical. Similarly, the constituents of these phrases are used to interpret elliptical noun phrases and verb phrases in the second utterance.

It is important to note that if the constituents missing from an elliptical phrase can be found at all, they can be found in the immediately preceding utterance. If there is a sequence of three utterances u1, u2, and u3, then the structure of u2 can be matched against u1, but u3 can only be matched against that of u2. The presence of u2 precludes matching u3 against u1. In Chapter II, several examples of long sequences of elliptical questions were presented. Although, in these sequences, it appears that u3 is patterned on u1, in fact u2 is expanded to a form similar to u1 and then u3 is patterned on this expansion of u2.

The process of building an interpretation of an elliptical phrase entails two steps once the ellipsis has been detected. First, the items missing from the phrase must be found in the preceding utterance (or, equivalently, the slot the elliptical phrase fills in the preceding utterance must be determined). Second, a complete phrase must be built using the elliptical phrase and the missing constituents found in the previous (old) utterance. In the remainder of the discussion, the first step will be referred to as <u>determining the slot</u>, the second as <u>expanding the utterance</u>.

The use of ellipsis in the task dialogues differed from that in the data base dialogues (see Chapter II, Section E). In the task dialogues, elliptical utterances appeared as responses to questions. In the data base dialogues, elliptical utterances were used in long sequences of questions. For purposes of building an interpretation of an utterance, the difference has most impact on the slot-determining phase of processing. In the question-and-answer pairs of the task dialogues, the slot filled by the elliptical answer is often marked in the question by a WH-phrase. Determining the slot filled by the ellipses in the question sequences of the data base dialogues is not so straightforward; syntactic and semantic clues must be used as explained below. Expansion of the utterance entails similar procedures in the two domains, but some preliminary transformations are required for the ellipses in the task domain (see Robinson, 1975).

The remainder of this chapter concentrates on capabilities in the discourse component of the speech understanding system for handling the elliptical utterances that occurred in the data base dialogues. The procedure for interpreting an elliptical utterance (EU) in the context of the preceding pattern utterance (PU) will be presented. In question-and-answer sequences, both the answer following a question and the next question itself may be elliptical. The PU for an elliptical answer is the preceding question. Expansion of this elliptical answer requires many of the same transformations as the elliptical utterances in the task dialogues. The PU for an elliptical question also is the preceding question, which is really two utterances back. This treatment is actually equivalent to using the immediately preceding utterance, the answer, since its structure corresponds directly to that of the question. The two utterances differ only in that one is marked as a question.

C. PARTITIONING TO REFLECT PARSE STRUCTURE

The ellipsis procedures require a combination of syntactic and semantic information. These two kinds of information are coordinated through the use of network partitioning. In essence, partitioning is used to overlay the parse structure of an utterance on the semantic interpretation. Hendrix (1976) describes the process of building this structure in detail. In this section, the representation will be described only in enough detail to elucidate its use in ellipsis handling.

As an utterance is parsed, a piece of network structure is associated with each constituent of the utterance. In some cases, new structure is built; in others, existing structures are referenced. To encode the parse structure of the utterance, the network structure corresponding to each constituent is isolated on a separate space in the logical partitioning. These spaces are related in a hierarchy that corresponds to the parse tree.

Figure VI-1 shows the network structures that would be built for the utterance, "John owns a red bike," using the simplified language definition:

```
        RULES

R1:    S => NP VP
R2:   VP => V NP
R3:   NP => DET MOD N
R4:   NP => N
        LEXICON

N:    John, bike
V:    own
MOD:  red
DET:  a, the
```

The spaces V1, MOD1, and N1 contain the network structure built from the lexical entries for "own", "red", and "bike" respectively. The node representing "John" is in the KNOWLEDGESPACE, so no new structure is built for this concept.

The space NP2S is built when the MOD "red" is combined with the N "bike" to form a noun phrase. The hierarchy of spaces NP2S, MOD1, and N1 reflect the syntactic structure of this NP. The noun "bike" is the head noun of this NP. The node 'B', which corresponds to this noun, is distinguished as the 'head node' of the structure built for the NP. The network structure visible from the space NP2S describes the concept referred to by the NP. This structure, considered without the space partition, corresponds precisely to the semantic interpretation of "red bike". The partitioning adds a means of recovering the parse structure of the NP.

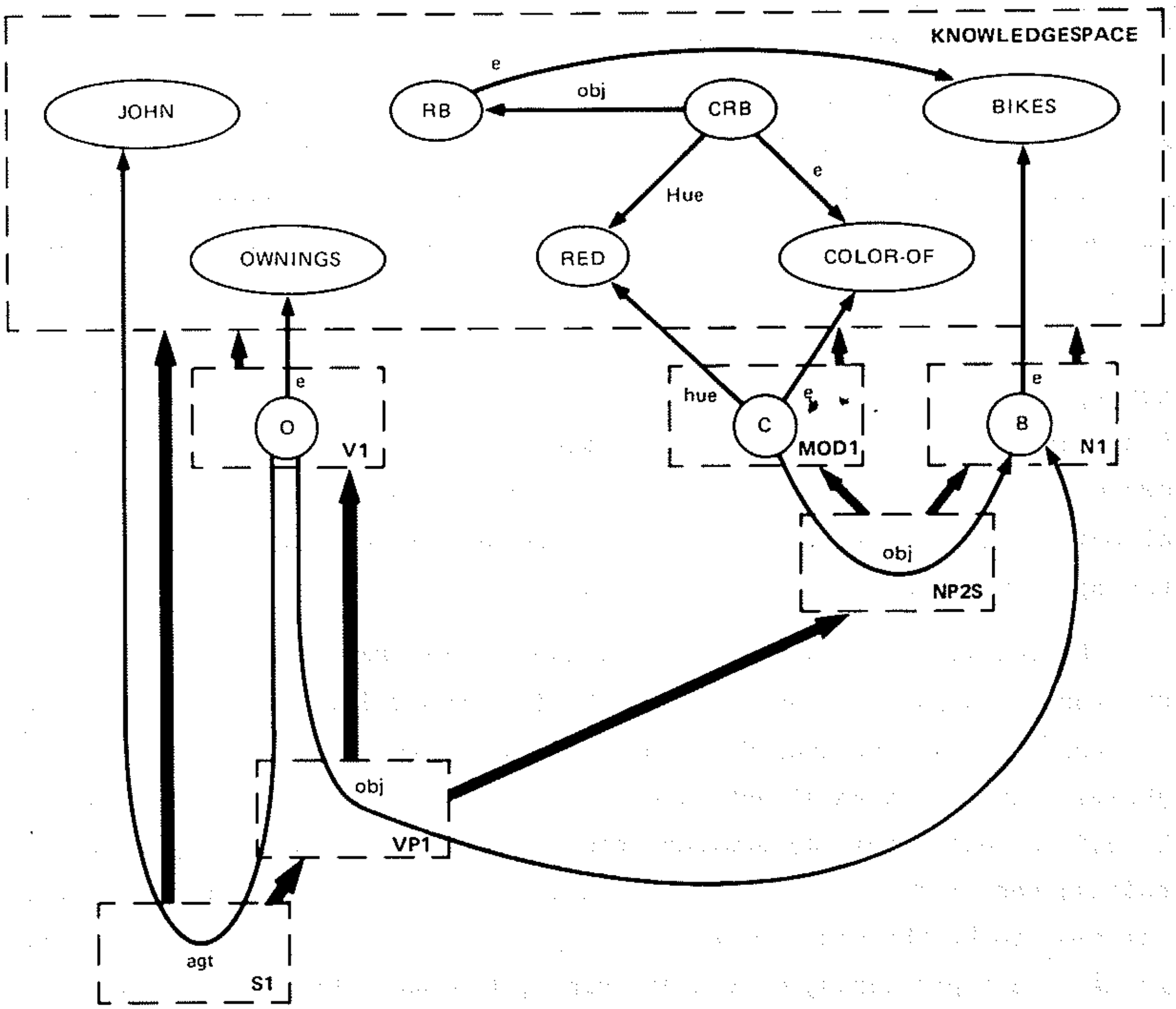


FIGURE VI-1 PARSE SPACES FOR "JOHN OWNS A RED BIKE"

In general, a noun phrase corresponds to a set of nodes and relations in the network. For each noun phrase, a single node in the network can be distinguished as central to the concept expressed in the noun phrase. This distinguished node is used by the algorithm for determining the slot an elliptical expression fills in the pattern utterance. In the example, the head node,'B', and the vista visible from NP2S contain all the information needed to combine the NP with other constituents of higher level phrases. For definitely determined NPs, the node representing the referent of the NP is used in further computations in place of the head node of the semantic representation, as the next example will illustrate.

The space VP1 results from the application of rule R2. It lies below spaces NP2S and V1 to reflect the fact that the constituents of the verb phrase are the verb "own" and the NP, "a red bike". Finally, at the sentence level, the VP is combined with the node representing JOHN.

If the spaces in the partitioning are ignored, the semantic interpretation of the utterance can be seen to be the node 'O', an instance of owning in which JOHN is the owner and C describes the object owned. The parse structure of the utterance, which is needed by the algorithm for completing an elliptical utterance, can be retrieved from the space hierarchy of S1.

Figure VI-2 shows the network structures that are built for the utterance, "John owns the red bike." The processing of this utterance differs from the previous description only in the handling of the noun phrase that serves as direct object. Because the noun phrase, "the red bike," is definitely determined, the discourse component is called to resolve the reference after space NP2S' is built. It identifies the referent to be the bike represented by the node 'RB' (see Chapter IV). In all other processing of the utterance, the node 'RB' is used in place of the node 'B' and space NP2D is used in place of NP2S'. This replacement is recorded on space NP2D so that the parse structure of the entire utterance can be retrieved for use in processing elliptical expressions.

D. SLOT DETERMINATION

The algorithm for determining the slot filled by an elliptical utterance uses a combination of syntactic and semantic filters. The following filters are applied in the order shown:

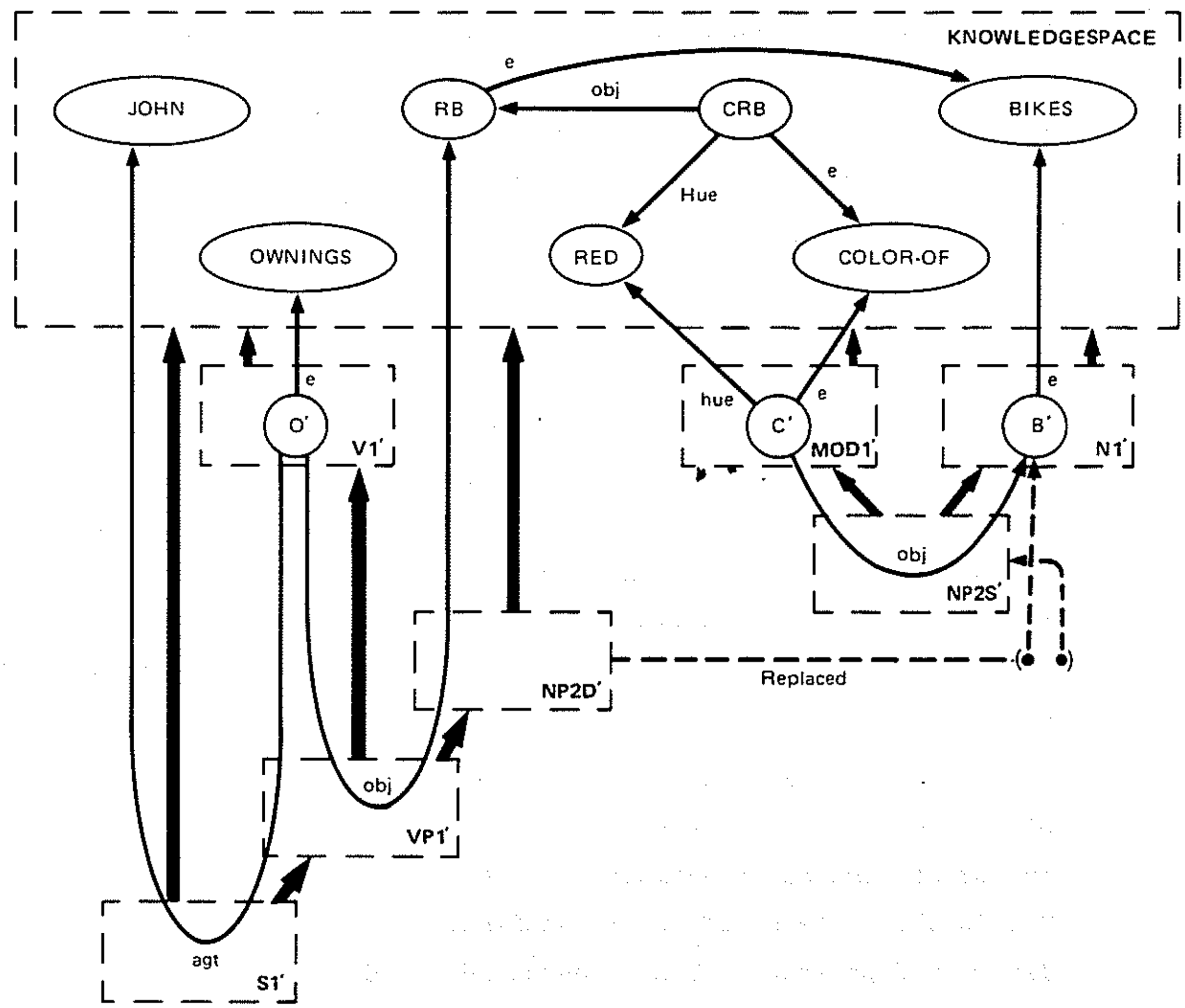


FIGURE VI-2 PARSE SPACES FOR "JOHN OWNS THE RED BIKE"

SYNTACTIC FILTERS:
1. category of phrase
2. definiteness
3. role of phrase in utterance

SEMANTIC FILTER:
4. semantic similarity

Each of the filters, and reasons for using it are described in the following sections. Syntactic filters are applied first because they are faster. If the syntactic role of the elliptical phrase is not given (an example appears below), then this filter is skipped. In this case, more than one candidate may remain, even after semantic filters are applied. A possibility, not explored in the current implementation, is to examine the syntactic roles of all of the candidates and see what keys to disambiguation might be asked of the speaker.

1. <u>SYNTAX</u>

Syntax plays a major role in determining the slot filled by an elliptical utterance (EU).* Usually, for an EU to make sense there must be a structural unit of the same type in the pattern utterance (PU). (This is not completely true: there may be an unfilled slot in the syntactic pattern for the PU that the EU fills. This case, and extensions to the algorithms for handling it, are discussed in Section E.) In addition to defining the category of phrase an EU can match, syntax also provides filters on the basis of definiteness and syntactic role.

If an EU consists solely of a noun phrase (NP), the determiner of that NP must match the determiner of the slot phrase in the PU. If the NP of the EU is definitely determined, it can match only definite NPs in the pattern; if it is indefinite, it can match only indefinitely determined phrases. The sequence PU - EU1 is fine, but PU - EU2 is awkward.

PU: Does Steven own a car?
EU1: A bike?
EU2: The bike?

The algorithm for determining the slot filled by an elliptical utterance uses the parallelism of determiners to filter out phrases to be considered as matches. The determiner of each NP in the PU is checked. If it matches the determiner of the NP constituting the EU,

----

* The systems described in Hendrix (1977), Burton (1976), and Bobrow, et al. (1976) all handle a limited range of ellipsis based only on syntactic features. Their success comes from the largely syntactic nature of ellipsis.

then the NP is a candidate for a match; the slot it fills is a candidate slot for the EU.

The parallelism of definites and indefinites is most clear when we consider utterances with two NPs that differ only in definiteness. Contrast the two sets of question sequences:

PU: Did the cat hurt a bird?
EU1: The dog?
EU2: A mouse?

PU: Did the cat hurt the bird?
EU1: The dog?
EU2: A mouse?

Without any preceding context, in the first sequence both EU1 and EU2 are unambiguous; the NPs match the correspondingly determined NPs. In the second sequence, EU1 is ambiguous; it could either be a question about the cat and the dog or one about the dog and the bird. The preference is to resolve the ambiguity on a semantic basis, but there is clearly some confusion that does not arise in the first sequence. Utterance EU2, in the second sequence, really does not make sense without some imputed context. Even then, there could be an ambiguity similar to the one for EU1.

It is possible to have a sequence of questions with indefinite NPs culminating in a definite NP, but this is an exceptional case; it occurs only when the definite NP refers to some truly unique object, or the questioner and answerer are playing a game. The following sequence showing an interchange between two people is an example of the former:

P1: Do you know what John got at the auction?
P2: Was it a document?
P1: Yes.
P2: An old one?
P1: Yes.
...
P2: The Constitution? / A copy of the Constitution?

The question-answering dialogues of the game '20 Questions' are an example of the latter. The same phenomenon happens with plurals. So the sequence PU - EU1 is fine, but PU - EU2 is not.

PU: Does England own any submarines?
EU1: Any patrol boats?

EU2: The patrol boats?

One can construct situations in which EU2 is reasonable, but again the set denoted by the NP must be unique. So, for a data base in which there was only one set of patrol boats (and these are a subset of submarines), the sequence PU - EU2 might be acceptable. This use of the definite at the end of a series of indefinites is sufficiently rare that the algorithm was not modified to handle it.

A problem arises when considering EUs consisting solely of nominals -- NPs without any determiners. Some default determiner must be chosen for the EU so that the filtering process can be done. The default currently used is definite for singular NPs and indefinite for plural NPs. This treatment is adequate for the kinds of questions in the data base domain seen in the following three examples from the data base protocols:

PU: What is the length of the Ethan Allen?
EU: Draft?

PU: Does Britain own any submarines?
EU: Patrol boats?

PU: Does the U.S. own the Ethan Allen?
EU: George Washington?

In general, however, there are cases that do not work undetermined:

PU: Did you drive the Cadillac today?
EU: Volkswagen?

"Volkswagen" alone is just not enough; "the Volkswagen" is. Other nouns require no determiner and can be matched by other undetermined nouns or by definitely determined ones:

PU: Did he write about pollution?
EU: Ecology?
EU: The environment?

The syntactic role of a noun phrase is important in choosing between candidate slots that are filled by phrases that are otherwise semantically and syntactically equivalent. Consider the sequence:

PU: Is the Ethan Allen longer than the George Washington?
EU: The Churchill?

The EU is ambiguous since "The Churchill" could replace either "the Ethan Allen" or "the George Washington". However, both "Is the Churchill" and "Than the Churchill" are unambiguous. In each case a syntactic role is assigned to "the Churchill" that can be used to eliminate one of the two candidate slots.

In summary, syntax is used to limit the candidates considered for finding slots of NPs serving as EUs. First, only NPs with matching determiners are considered. If there is more than one candidate, syntactic role is used to eliminate choices. If at either step of the process there are no candidates, there is the option of relaxing syntactic constraints. This option was not pursued in the speech understanding system because of the need to restrict, rather than increase, potential interpretations.

2. SEMANTICS

Although syntactic restrictions often eliminate all but one choice, there are cases when an appeal must be made to semantic attributes of the phrases filling candidate slots in the pattern utterance. The role of semantics in filtering out candidates may be seen by considering the sequence:

PU: Is the chicken in the cooler?
EU: The potato salad?

Syntactically, "the potato salad" matches both "the chicken" and "the cooler". Semantically, it is more similar to "the chicken": they are both foods. Therefore, the ellipsis procedures should establish the subject slot (i.e., the role filled by "the chicken") as the candidate slot.

If more than one candidate slot remains after the syntactic filters have been applied, the ellipsis procedures must determine which phrase in a candidate slot is semantically most similar to the phrase constituting the elliptical utterance. This semantic filter is applied to a set of candidate nodes each of which corresponds to a candidate slot that has passed through the syntactic constraints. Recall that each phrase in the pattern utterance is described by a piece of network structure. In particular, for each noun phrase, one node in this structure is distinguished as central to the concept expressed by the noun phrase. For each candidate slot, this distinguished node is the candidate node. The NP that constitutes the elliptical utterance can be similarly identified with a single node in the network. The candidate node that is taxonomically most similar to the distinguished node of the EU is chosen as the matching node; the slot filled by its concept is the slot the EU is taken to fill.

In a system with a semantic network knowledge representation, semantic similarity is determined from the element and superset hierarchy of the network. Given some collection of nodes N and a node m, the node n in N is most similar to m if n and m belong to a common set (in the network) that does not include any other nodes of N. In network terms, node n is most similar to m if, considering only element (e) and subset (s) arcs, n and m have the closest common ancestor. This similarity measure is a relative one. It can only be used to decide among a set of alternatives.

To find the candidate node that shares the closest common ancestor with the EU-node, the path from the EU-node to the root of the hierarchy (i.e., the node UNIVERSAL) is marked. (This path may actually split into several paths since the network is not a tree.) Paths are then grown by recursively following e and s arcs (including de and ds arcs) from each candidate node until they intersect this path (or, if the path from the EU node splits, some one of the resulting paths). The node at which two paths intersect is the least common ancestor for the two different nodes that started the paths.* The matching node is the candidate node whose least common ancestor is the smallest number of links away from the EU node. The paths traced for the sequence,

PU: Is the box-end wrench used to loosen the bolt?
EU: The socket wrench?

---

* This is not entirely true if the path from one or both of the nodes has split. In this case, there may be more than one intersection; the least common ancestor is the node at the intersection that is the fewest links away from the starting node. Such cases may be ambiguous.

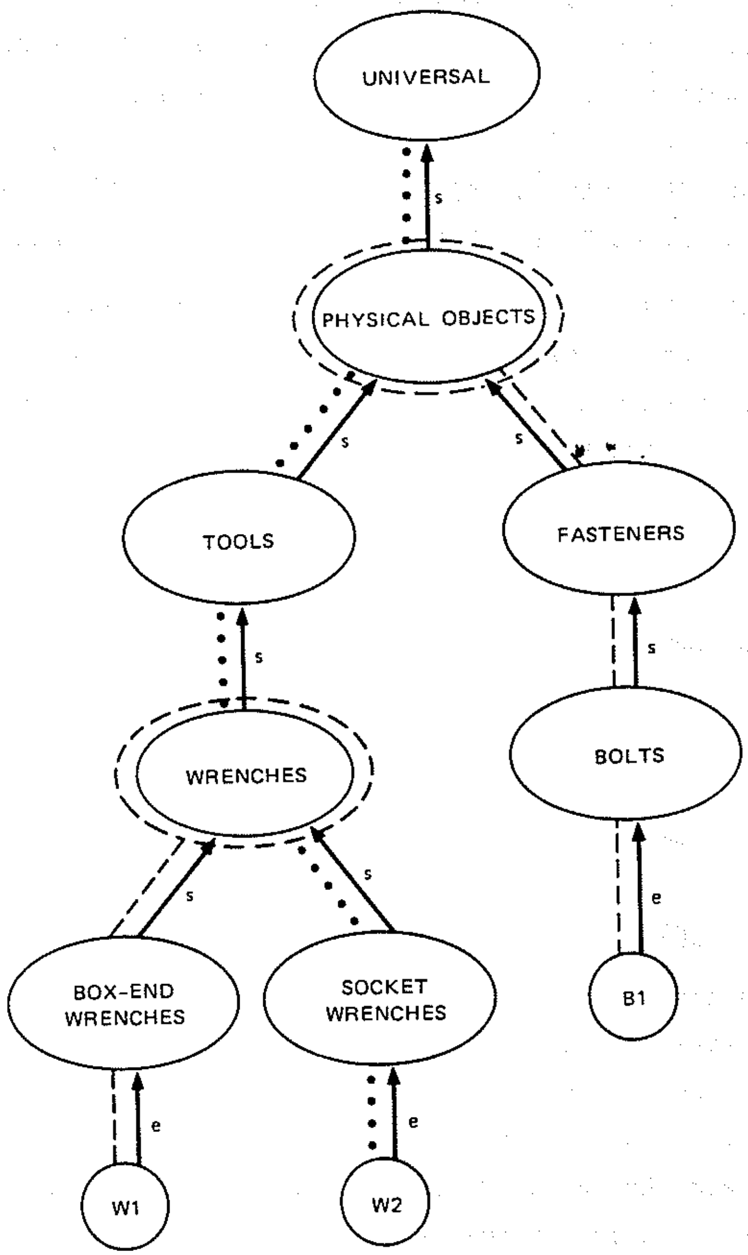


Figure VI-3. PATH-GROWING ALGORITHM

are shown in Figure VI-3. Nodes 'W1', 'W2', and 'B1' represent the items referred to by the definite noun phrases, "the box-end wrench", "the socket wrench", and "the bolt" respectively. The path from the node corresponding to the EU is shown with a dotted line. Paths from the PU candidate nodes are shown with dashed lines. The paths from 'W1' and 'W2' meet at 'WRENCHES'. The path from 'B1' intersects the path from 'W2' at 'PHYSICAL OBJECTS'. Since 'WRENCHES' is closer to 'W2' than 'PHYSICAL OBJECTS' is, 'W1' is chosen as the matching node.

When two of the phrases filling candidate slots are semantically equally similar to the elliptical phrase, the elliptical utterance is ambiguous. Such cases are detected by the path growing algorithm when paths from two (or more) candidate nodes intersect with the path from the EU node at the same node. This can happen either because the paths all intersect (for the first time) at the same node, or because the paths from the candidates have intersected at some node and the path from that node intersects with the EU node's path. Since syntactic clues have already been used as a filter, discourse has no further way of disambiguating the utterance. However, the number of candidates is usually sufficiently small at this point, that the system could resolve the ambiguity by asking a question.

E. SEMANTIC SUITABILITY CHECK

After a candidate is selected, a check must be made to determine that the EU fits semantically into the selected slot. This check is in essence the same one that is done by the semantic composition routines (Hendrix, 1976) when the original utterance (i.e., the PU) is interpreted and the matching (slot) phrase is embedded in some higher level phrase in that utterance. The need for this kind of check is especially strong in a speech-understanding environment. Even though the phrase constituting an EU syntactically and semantically matches some phrase in the PU, it may not make sense semantically to substitute the EU for this phrase. For example, in

PU: Does Britain own a sub?
EU: A commander?

the EU matches the phrase "a sub" (they are both physical objects) but the substitution does not make sense (note that it would if the PU were, "Is there a sub in Naples?").

For this reason, a semantic check on the suitability of substituting the EU in the selected slot is always done. In the above example, the phrase "own a sub" is checked by the semantics component when the original utterance is parsed. Before trying to substitute an EU, the discourse routines perform the same check with the EU. In the example, the plausibility of "own a commander" is checked and this interpretation of the utterance is rejected. The acoustic routines must find another set of words.

F. COMPLETING THE UTTERANCE

Completion of the elliptical utterance entails fitting it into the slot in the pattern utterance selected by the slot determination phase of the process. Semantic checks already have ensured that it is reasonable to substitute the EU for the NP that occupies the slot in the PU. The remaining step is to build a new structure using pieces of the PU and the EU. The use of a network partition to reflect the parse structure for an utterance is crucial to limiting the computing done in this expansion.

Elliptical expansion in an earlier version of the speech understanding system (see Walker et al., 1975) depended on having available a representation of the semantic interpretation of the complete PU in terms of the semantic representation of each of its constituents. The utterance expansion routines built a new net around the semantic representation of the EU using all of the information from the semantic interpretation of the PU not superseded by information in the EU. But, in a speech system environment, interpretations of utterances are built up from partial interpretations. Each partial interpretation has been processed by both semantics and discourse to allow assignment of scores for determining which of the competing interpretations to work on next. As a result, the final semantic interpretation of an utterance is a combination of semantic representations of some constituents, discourse representations of other constituents, and semantic processing to handle quantification. The simple surgery of the original system no longer works because there is no complete semantic template available. For example, when a definite noun phrase is resolved, the node identified with the resolution, rather than the original semantic interpretation, is used in building representations for higher (embedding) phrases. This occured with the noun phrase, "the red bike", of Figure VI-1.

It would be possible for the semantic component of the system to build dual representations, one using semantics and one using discourse results, each time phrases were merged to make a higher level phrase.* This duplication would make available a final semantic interpretation built only from semantic constituents. However, this solution would double the most expensive work done by semantics in building an interpretation. This doubling of effort would have to be done for all candidate phrases that include NPs, even the false attempts that were not part of the final interpretation. Furthermore, such an expansion algorithm requires copying all portions of the PU being used with the EU. In contrast, the algorithm described in this section overcomes both of these problems: it works using the combination of semantic and discourse representations, and it copies only those portions of an utterance that embed the slot filled by the EU.

To illustrate the basic algorithm, consider the sequence

PU: What is the speed of the submarine?
EU: The carrier?

Figure VI-4 shows the final semantic interpretation of the PU along with the semantic interpretation for each of the constituent phrases and the discourse interpretation of the NPs.

----
* The discourse and semantic components build the same kind of structures. The difference is that the structure the discourse component builds, if it is called, has considered the context in which an utterance, or one of its constituents, appears.

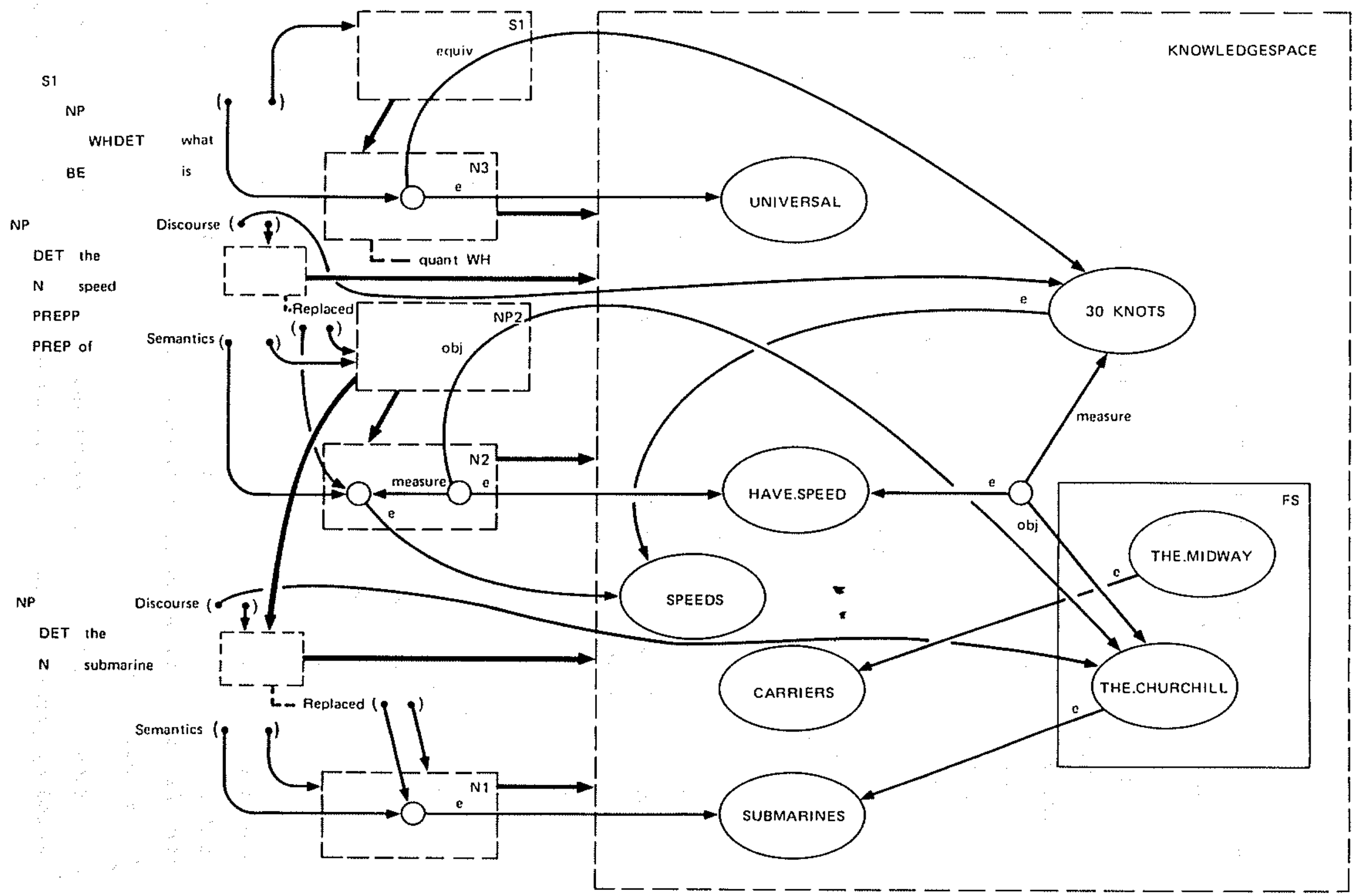


FIGURE VI-4 REPRESENTATIONS FOR "WHAT IS THE SPEED OF THE SUBMARINE?"

The structure shown is built as follows. As soon as the NP "the submarine" is encountered and semantics has built an interpretation for it, discourse is called. The submarine Churchill is found in focus and hence identified as the object referred to by the NP. Note that the node for the particular ship is used in the higher level (embedding) NP "the speed of the Churchill". Similarly, once the semantic interpretation for this NP is built, discourse is called and determines the node corresponding to the speed of the Churchill (which may or may not exist explicitly in the net; see Hendrix, 1976). This node is then used in building the semantics for the whole utterance.

Now consider what happens when the EU is encountered. The match of the phrase, "the carrier", which is first resolved to the Midway, with the slot filled by "the submarine", which was resolved to the Churchill, is found as described in the preceding section. But the node for the Churchill is nowhere to be found in the utterance level semantics, which consists solely of the nodes and arcs in the vista of Spaces S1 and N3 of Figure VI-4 (and of the knowledgespace nodes touched by those arcs). However, it is easy to find how any node was used in building a final interpretation of an utterance if enough information from the parse of that utterance is kept.

After an utterance is accepted, the discourse routines collect information about each of the NPs and VPs in history lists. In particular, for NPs the following information is recorded: (1) the semantic interpretation,(2) the discourse interpretation (which in some cases is identical to the semantic interpretation but is always different for definite NPs), (3) the phrase of which the NP is a constituent (in the accepted interpretation), or, in which it is <u>embedded</u>, and (4) syntactic factors such as number and determination. For VPs, only the semantic interpretation and the embedding phrase need to be collected.

When an EU is encountered and the candidate slot found, the embedding phrase for the EU can be constructed from the embedding phrase for the phrase filling the slot in the PU. In the example, the embedding phrase for "the carrier" is NP2. The first step of substituting the EU in the slot is to copy the space(s) created when the embedding phrase was formed from its constituents and to substitute arcs to the EU node for any arcs to the corresponding PU node. In the example, a new space NP3 corresponding to NP2 must be built with an arc to the Midway instead of the Churchill, as shown in Figure VI-5* Note that it is not necessary to copy any of the structure built for other constituents of the embedding phrase. Network partitioning, in particular the visibility restrictions it imposes, enables each of these constituents to be viewed from the perspective of the new space. The result of this step is a new constituent for some higher level embedding phrase. Again the embedding phrase can be determined easily from the history lists. The process continues recursively until the embedding phrase is the utterance. Resolution of definite noun phrases (in particular, relational NPs) is performed, if relevant, when the new constituent is built. In the example, NP3 is built as shown in Figure VI-5. Because this is a relational NP, it is passed to the resolution routines and the actual "speed of the Midway" is found. Finally, this node is embedded in a copy of the utterance level semantics as shown in Figure VI-6.

Notice that the use of network partitioning enables the copying of constituents of the PU to be limited to those phrases embedding the slot filled by the EU. Looked at another way, only those phrases on the path from the slot to the root of the parse tree were copied. This attribute of the procedure may be seen even more clearly by considering the sequence

PU: Does Britain own the carrier?
EU: The U.S.?

and examining Figure VI-7 and Figure VI-8. The phrase "the U.S." corresponds to "Britain", a top-level constituent of the sentence. Only the space S1 and the agent arc need to be copied in building the interpretation of the EU.

---

* For clarity, replacement spaces have been left out of this and following figures.

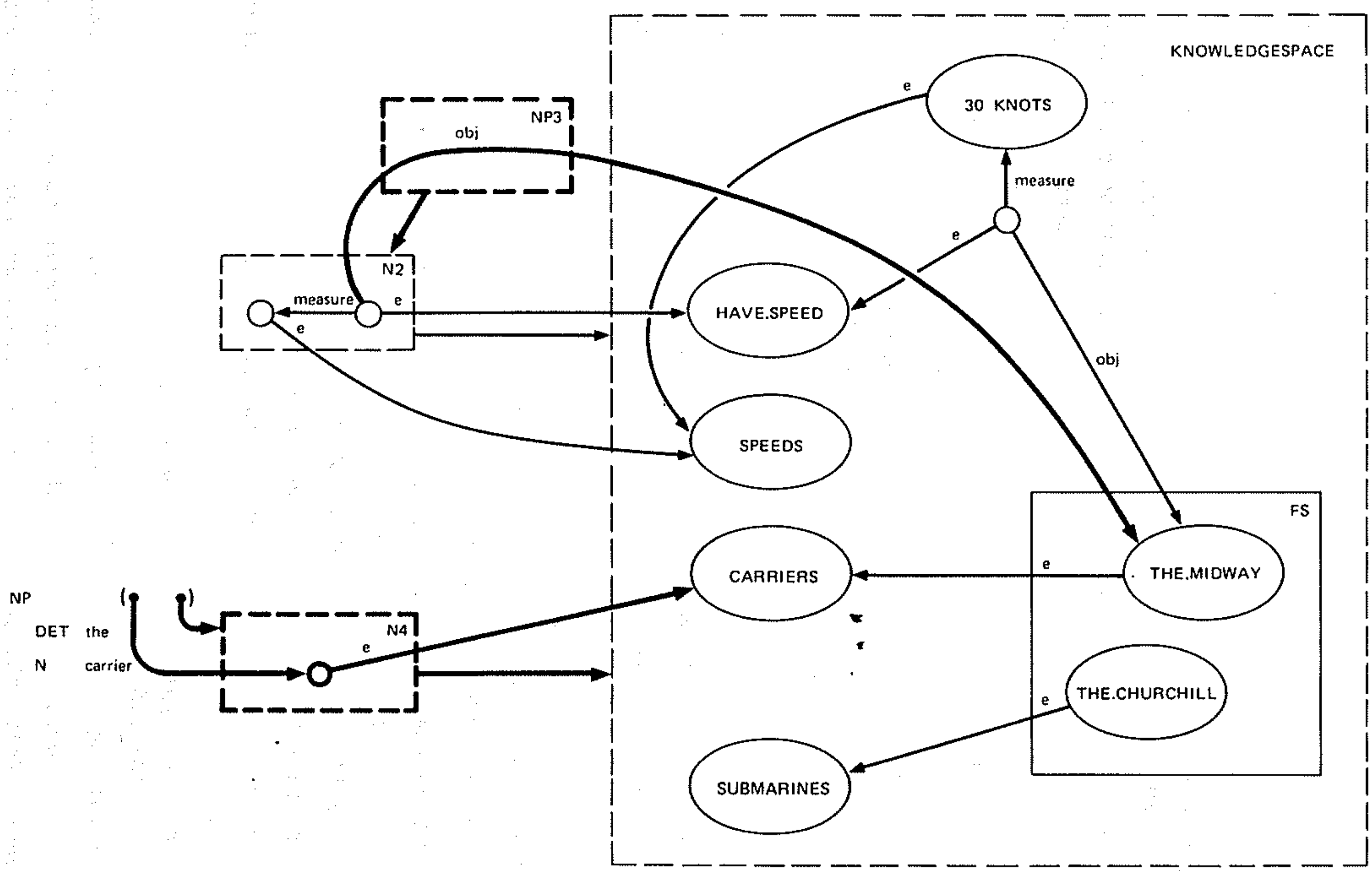


FIGURE VI-5 RESOLVING "THE CARRIER" AND FIRST LEVEL EXPANSION OF ELLIPSIS TO "THE SPEED OF THE CARRIER"

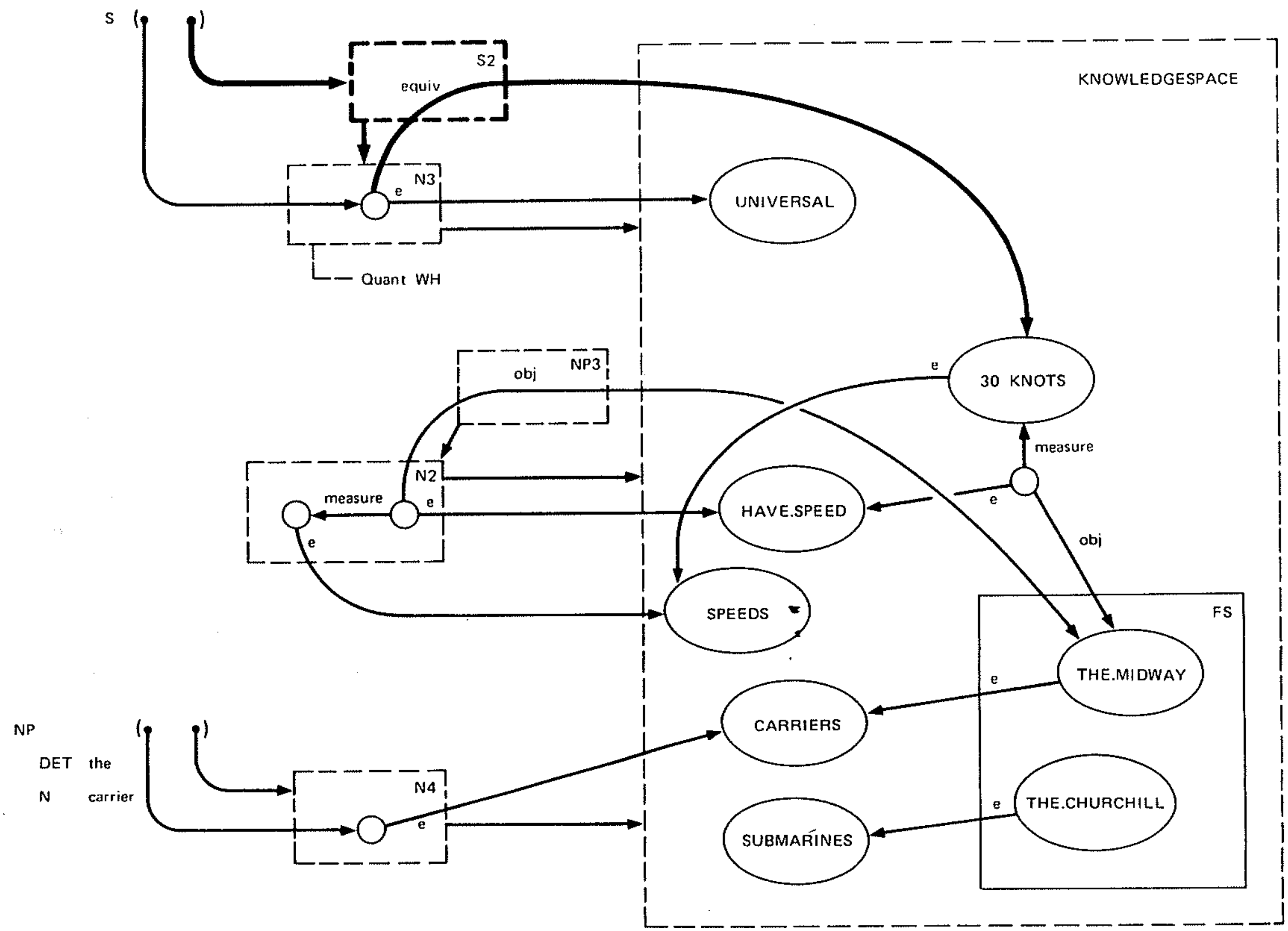


FIGURE VI-6 FINAL EXPANSION OF ELLIPTICAL UTTERANCE TO "WHAT IS THE SPEED OF THE CARRIER?"

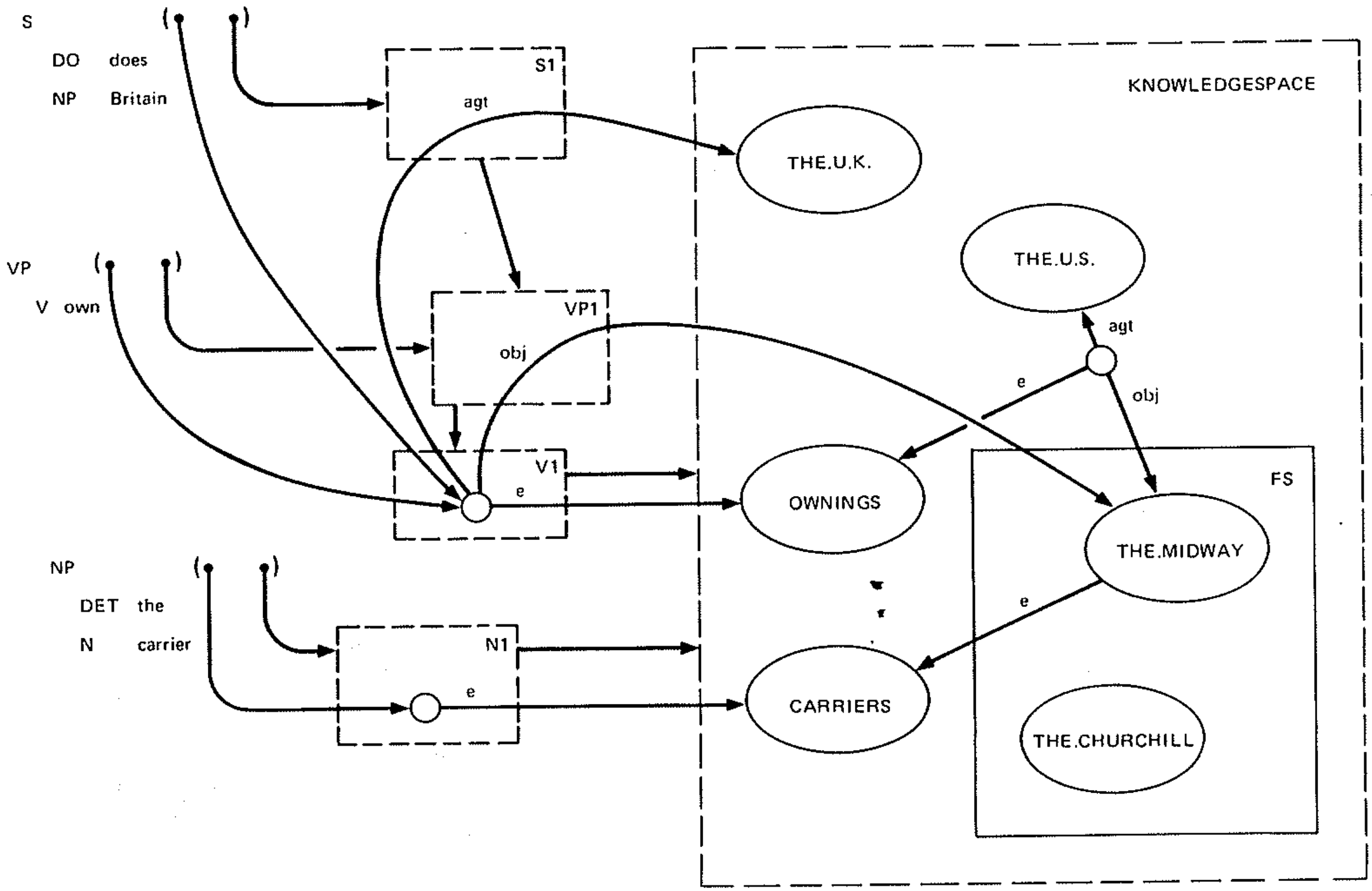


FIGURE VI-7 REPRESENTATIONS FOR "DOES BRITAIN OWN THE CARRIER?"

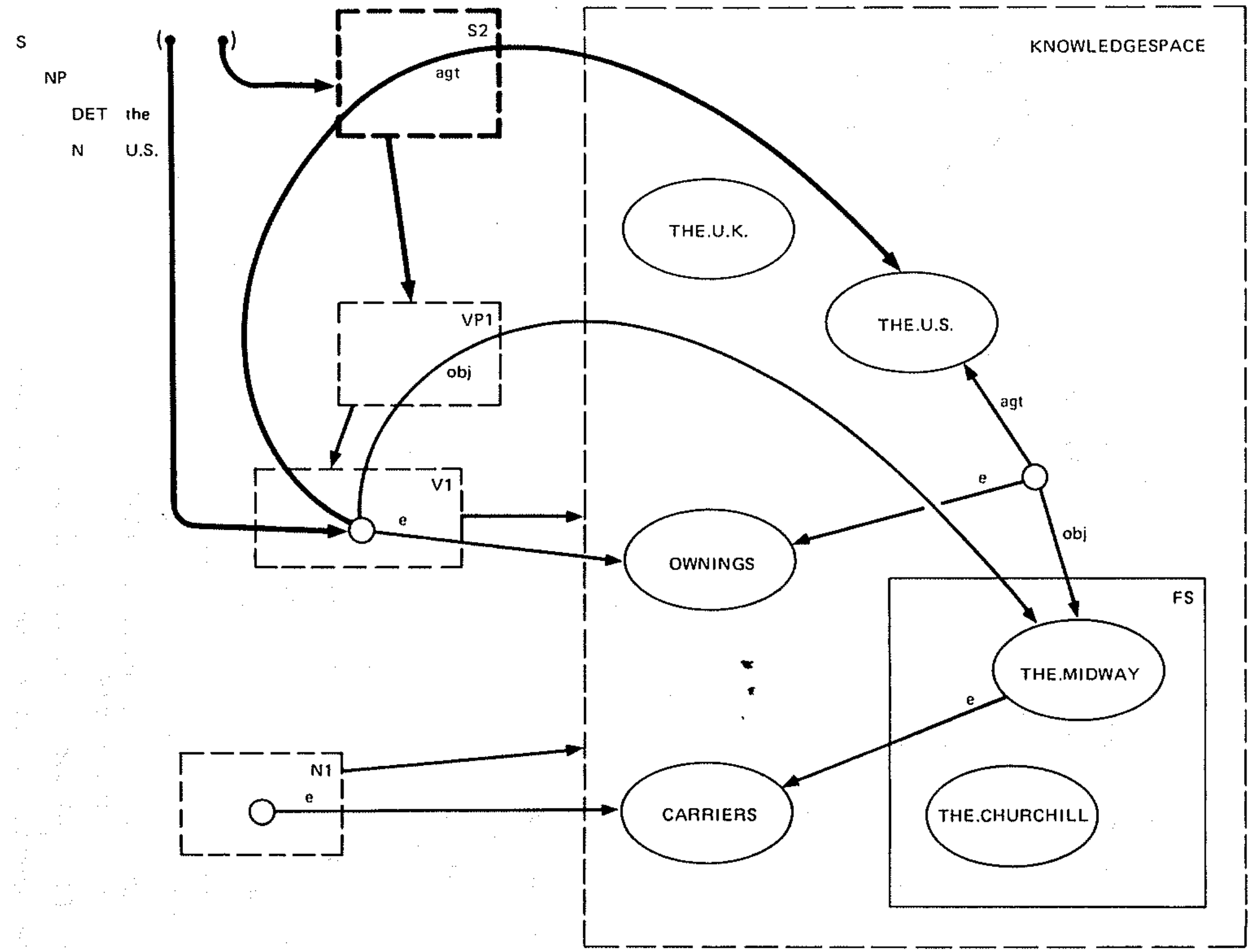


FIGURE VI-8 EXPANSION OF THE ELLIPTICAL UTTERANCE, "THE U.S.?"

In the two examples presented so far, the EU is a definite NP. The only difference in handling indefinite NPs is that the head node (and other nodes and arcs) of the NP lie on spaces below the KNOWLEDGESPACE and these spaces must be copied in the first step of the substitution. Again, network partitioning minimizes the work; the whole collection of spaces for the EU becomes visible (and the spaces for the NP that fill the slot in the PU become invisible) when the new space is created for the embedding phrase.

Another problem occurs with quantified NPs. Consider the sequence:

PU: Does Britain own all of the subs?
EU: The carriers?

The quantifier "all" operates on the NP "the carriers" in the most natural interpretation of the EU; i.e., the most natural interpretation is "Does Britain own all of the carriers?" To obtain this result, some record must be kept of what quantifier, if any, applies to a phrase. The semantic component makes precisely this record in the first step of handling quantifiers (see Hendrix, 1976). When the NP "all of the subs" is constructed from a quantifier and a prepositional phrase, the only thing that happens is the recording of the quantifier "all" on a space in the network of the NP. The actual rearrangement of the structure into one that corresponds to the network encoding of quantified statements (see Hendrix, 1976) does not occur until after the entire utterance is processed. Semantic processing must operate this way to capture the proper scoping of quantifiers. Discourse uses the tracks left at the parse structure level to transfer relevant quantifiers to elliptical utterances. In the sequence

PU: Does John own both boats?
EU: Either boat?

the EU is already quantified and the expansion process does not transfer the quantifier from the PU. The two-step process for handling quantifiers also means that an elliptical utterance when expanded may have different scoping than the PU. This difference in scoping occurs in the sequence

PU: Who owns all anthracite coal mines in the U.S.?
EU: Each natural gas pipeline?

The PU asks for the single owner of all anthracite coal mines (and carries the implicit assumption of such a single owner). The scope of "all" is inside of the scope of "who". In the EU, the scope of the "who" moves inside the scope of the quantifier, "each". For each natural gas pipeline, the particular owner of that pipeline must be identified.

G. ELLIPTICAL RELATIONAL NOUN PHRASES

The ellipses discussed so far have all been structural in the sense that some syntactic pieces of an utterance have been left out; the structure of the utterance is incomplete. As a result, syntactic clues may be used to detect the ellipsis and to guide interpretation of it. The data base dialogues also contain elliptical utterances for which there are no syntactic clues. Consider the utterance: "What is the length?"; the ellipsis here is semantic. The utterance is syntactically, but not semantically, complete. "The length" is a well-formed NP; however, semantically, "length" assumes some object for which length is a relevant measure and implicitly conveys the relation of 'having a length'. The combination of this 'relational' attribute and definiteness indicates the need for an object. (If the utterance had been "What is a length", then no object would be required. The use of the indefinite determiner distinguishes this case.)

In essence, the verb-like characteristics of the relational nouns cause a situation in which a phrase that appears to be syntactically complete is not. The object of the 'verb' is missing, but, since the verb is expressed through a noun, no syntactic indications of incompleteness occur. Case information appearing with the semantics of relational nouns can be used to detect this kind of ellipsis.

When a definitely determined relational NP (RELNP) is encountered, the discourse routines first check to see if all of the cases required by the RELNP are present. If any are missing, ellipsis handling is invoked. The preceding utterance is examined to find objects for the empty slots. The procedure for finding candidate slots in the case of structural ellipsis can be used to determine which object in the PU best fills the missing case slot. Expansion of the elliptical RELNP is straightforward: a new space is created below the space for the RELNP and the space(s) containing the slot filler(s), and the case arcs are added to this space.

A compound case of structural and RELNP ellipsis occurs in the sequence

PU: What is the draft of the submarine?
EU: The length?

In processing this EU, the RELNP ellipsis is handled at the noun phrase level, resulting in the structure of (a) in Figure VI-9 being transformed into the structure of (b). The structural ellipsis is handled at the utterance level. At this point the problem is equivalent to processing the EU, "the length of the submarine". The result appears in (c) of Figure VI-9.

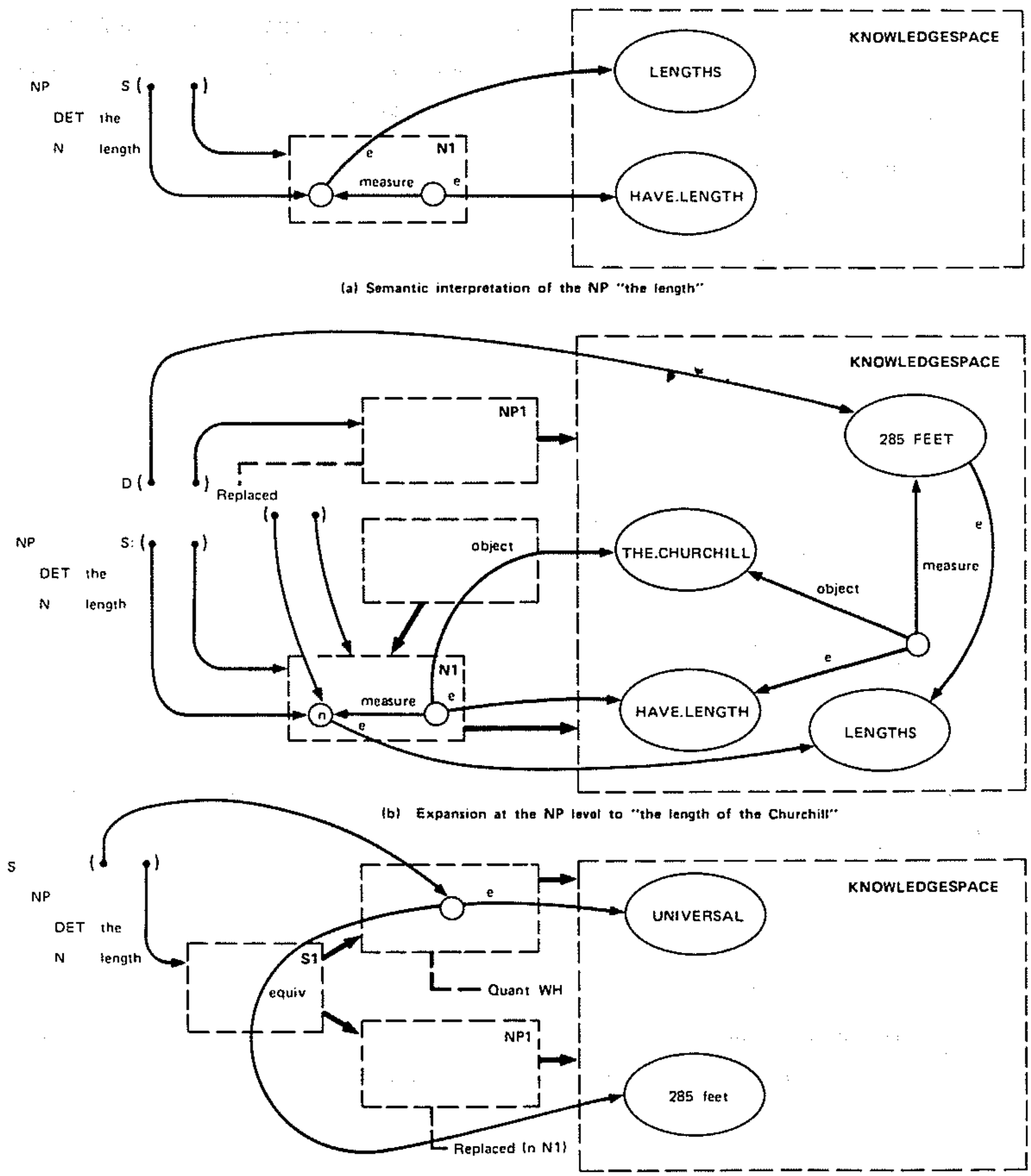
H. LIMITATIONS AND EXTENSIONS

The ellipsis-handling capabilities described in this section are limited in at least two ways. However, the algorithm for expanding an elliptical utterance is general. In the remainder of this section we discuss these limitations and present the extensions necessary for handling less restricted forms of ellipsis.

The major limitation of the current ellipsis routines stems from the assumption that the EU will fill a single slot in the PU, which is not true of ellipsis in general. At the utterance level, the general case is that any number of constituents may be present or missing in the EU. In the sequence,

PU: Did you take the coat to the cleaners?
EU: The shoes to the shoemaker?

the EU contains an object NP and an adverbial prepositional phrase. The subject NP and the VP must be retrieved from the PU. This kind of ellipsis is even more common when more complex sentences are considered.



(a) Semantic interpretation of the NP "the length"

(b) Expansion at the NP level to "the length of the Churchill"

(c) Expansion to full utterance

Figure VI-9. EXPANSION OF THE ELLIPTICAL UTTERANCE, 'THE LENGTH'

In particular, when two clauses or phrases are conjoined, the second is often elliptical; consider the above sequence joined by "and". Rather than looking for a single slot filled by the EU, the ellipsis routines should determine the constituents missing from the PU and then build the full utterance. (The latter step would be quite similar to the work done by the semantic composition routines).*

The mechanism for handling ellipsis this way would entail a closer coupling of syntax and discourse and would proceed basically as follows. The parsing routines would determine which constituents of the utterance were present in the EU and which were missing, on the basis of the context-free structural description associated with each rule in the language definition. Using this information and the parse of the PU, the discourse routines would build the complete utterance in a manner similar to the one now used for expansion. The only difference would be that several components might get replaced at once. Both semantic and syntactic checking could be done, based on the mapping between the structure of the PU and that of the completed EU.

Adopting such a strategy eliminates two major limitations of the current approach. First, the EU may consist of any number of constituents, not just a single NP (the only exception to this restriction in the current routines is with RELNP ellipsis). In particular, the EU may consist solely of a modifying phrase not present in the PU, as in the sequence:**

PU: Plot the distribution of soybeans.
EU: In the year 2000.

Second, the extension to handling NP and VP ellipsis is straightforward. The only additional step needed is to determine the NP (or VP) in the PU that matches the elliptical phrase. The PU phrase then takes the role

----

* Hendrix, 1977 describes a system that does this for a limited kind of language.

** Thanks to W. H. Paxton for this example and for a suggestion of how to handle it. The content of this section was greatly influenced by discussions with him.

of the PU and the elliptical phrase takes the role of the EU in the above description of ellipsis handling for a complete utterance. The result of the processing is a complete NP (or VP) to be used in building the rest of the utterance. For example, in the sequence:

PU: Is the Churchill the smallest sub?
EU: Is the Lafayette the largest?

the elliptical NP "the largest" gets matched with the PU phrase "the smallest sub", and is then expanded to "the largest sub". This complete NP can then be used in the (now complete) EU.

Processing the kinds of ellipsis occurring in the question answering pairs of the task dialogues also entails only one additional step. The question (PU) must be transformed before it can be used as a template. As an example, consider the sequence

PU: Which bolts did you tighten?
EU: The front bolts.

The PU must get transformed to "You did tighten which bolts", then an I/you transformation must be done. Then the EU can be placed in the slot (nicely indicated by the WH-phrase). A means of expanding the language definition to facilitate this kind of processing is currently being explored.

I. CONCLUSIONS

Ellipsis is an example of the local influence an utterance exerts on the interpretation of the following utterance. This chapter has examined the kinds of information that need to be recorded from one utterance to help in processing the following one. Syntactic information is a central part of immediate focus. The constituents of an utterance and the roles each plays in the utterance influence the immediate focus of that utterance. The use of network partitioning to overlay the parse structure of an utterance on the semantic interpretation provides a means of coordinating syntactic and semantic information. This coordination facilitates constructing an interpretation of an elliptical utterance from the interpretation of the preceding utterance.

VII RECAPITULATION AND A LOOK AHEAD

CONTENTS:

A. SUMMARY OF REPORT

The influence of focus on the interpretation of utterances in a dialogue and representations of focus for a computer system for understanding dialogue have been examined in the preceding chapters and the effects of two ranges of focus, global and immediate, have been demonstrated. To recapitulate, briefly: the linguistic form of an utterance -- the syntactic structure of the utterance and even the particular words that appear in it -- constitutes an immediate focus that constrains the linguistic form and hence the interpretation of the following utterance. Global focus is more long lasting; it derives not just from an individual utterance but from the total discourse context. The global focus in which an utterance is interpreted is determined by a combination of elements: the topic of the discourse, the particular form of the discourse, the setting in which it occurs, and the place in the discourse that the utterance occupies. Global focus influences what gets talked about, how different concepts get introduced, and how concepts are referenced.

To use immediate focus in the interpretation of an utterance, syntactic and semantic information about the preceding utterance must be coordinated so that the focus information conveyed syntactically is directly available in the underlying knowledge representation. The interpretation of elliptical sentence fragments illustrates the use of immediate focus. The coordination of syntactic and semantic information is used to minimize the work done in expanding the fragment into a complete utterance. In the approach described here, coordination is achieved by recording the relationship between syntactic units and their semantic translations in the knowledge representation, a semantic network. A partitioning of the network is used to superimpose the parse structure of an utterance on its semantic interpretation. The expansion of an elliptical utterance then reduces to the replacement of spaces in the structure built for the preceding utterance with spaces built for the new utterance.

The representation of global focus is based on partitioning the knowledge base into focus spaces. Each focus space contains those items that are in the focus of attention of the dialogue participants at a given point in the dialogue. The need for mechanisms to change what is highlighted as new utterances are encountered is as important as the separation of the knowledge base so that certain elements are highlighted. In general, the problem of deciding when and how to shift focus is extremely difficult. Shifts in focus depend both on the particular kind of discourse being interpreted and on the topic of discourse. The mechanism for shifting focus developed in this report was designed specifically for task-oriented dialogues. For these dialogues, the structure of the task provides a guide for detecting shifts of focus and a framework for structuring the focus spaces.

The problem of identifying the referents of definite noun phrases illustrates the role of focus in the interpretation of utterances. The method developed in this report for resolving definite noun phrases is new. It takes discourse structure into account and uses the distinction between highlighted items and those that are not highlighted to constrain the search made by the deduction and retrieval component when used to identify the object referred to by a definite noun phrase. With a representation of focus, the process of identifying the referent of a noun phrase looks quite different than in systems that search sequentially back through a discourse to look for a referent. The important questions are what items are relevant at a given point in a discourse and when an item ceases to be relevant; not how many sentences have occurred since the item was last mentioned.

B. EXTENSIONS

The work reported here can be quite naturally extended in several directions. First, the focus representation can be generalized by integrating global and immediate focus, coordinating the focus representation with a user model, and extending the mechanisms for shifting focus to other kinds of discourse. Second, the focus representation can be used in generating descriptions. Third, the use of the focus representation for interpreting definite noun phrases can be extended so that exact matches are not required.

1. INTERACTION BETWEEN IMMEDIATE AND GLOBAL FOCUS

Although two ranges of focus have been examined in the preceding chapters, no mechanisms for coordinating them have been developed. It is clearly important to provide for the interaction between global and immediate focus both for language understanding and for language generation. Although each of the items that are mentioned in an utterance enters the focus of attention of the discourse, some of them are more focused than others. For example, the item in the subject position of an utterance is more salient than an item that is the object of a prepositional phrase (Fillmore, to appear). This differentiation is part of the immediate focus an utterance provides for the following utterance. If the difference in strength of focus is recorded in the representation of global focus, it can be used both to determine shifts in focus and, when generating an utterance, for ordering its constituents.

In essence, what is needed is a means of discriminating among degrees of focus for the items that are entered in a focus space. The partitioning of the net that encodes the correspondence between syntactic units and their meanings can be used here as well as in the handling of ellipsis. Since each syntactic unit of an utterance is isolated on a space (or, set of spaces) in the network, the network structures that correspond to syntactic units that are focused to a higher degree can be easily identified. These items can then be separated from other items that are in focus. The remaining problem is to determine a metric for ordering the syntactic units of an utterance.* Work in linguistics (Fillmore, to appear) suggests that the subject and direct object should be more sharply focused than the other items in the sentence.**

2. USER MODEL

The use of the focus representation presented here has been based on the simplifying assumption that the speaker and the hearer have a common model of the world. The system assumes that what is in its focus of attention is identical to what is in the user's focus of attention. This assumption does not hold in general. In fact, one of the purposes of dialogue is to communicate beliefs about the world, including what is in focus, to another party. To participate in a dialogue with a user, a computer system needs to distinguish between its own beliefs and what it believes the user knows. By coordinating the focus representation with a model of the user, the system can distinguish between its focus of attention and what it believes to be the user's. This distinction is important both for language generation and for language understanding. For example, if I know that Bob owns two cars but that you only know about one of them, then I can identify the car you mean by the phrase "Bob's car."

---

* Hendrix (1976) defines and orders these spaces to determine quantifier scope.

** The problem is more complex when considering spoken language since intonation affects the metric.

Cohen and Perrault (1976) have developed a representation of a user model in terms of network partitioning for use in generating dialogue. This representation can be coordinated with the focus representation by augmenting the procedures that use focus to consider the intersection of certain spaces in the user model partitioning with spaces in the focus partitioning. A remaining problem is how to decide dynamically what should be changed both in the space that describes the user's beliefs and in the space that describes the system's beliefs about the user (Chapter II, section G briefly discusses the problem of dynamically forming a model of an apprentice's skill level).

3. <u>FOCUS IN LESS STRUCTURED DISCOURSE</u>

The major problem in adapting the focus representation to kinds of discourse other than task-oriented dialogues is to augment the mechanisms for shifting focus. Major indicators of shifts in focus (e.g., chapter headings) are easy to accommodate. The difficulty lies in identifying local indications of shifts in focus. These can be more subtle and tenuous when the structure of a discourse is not tied to anything as strong as a task model. For such discourses, shifts in focus are often more gradual than in the task dialogues, and structural indications of shifts (segmentation) occur less often.

The identification of a local shift in focus involves both detecting signals of a possible shift and ascertaining when such signals indicate an actual shift in focus. The focus representation described here can detect two kinds of change that indicate possible shifts in focus; it can detect a change in the perspective from which an event or object is discussed, and it can detect references to items that are only implicitly in focus.

A change in perspective can be signaled by a reference to an unfocused relationship in which a focused object participates, as in switching from talking about the owner of a house to discussing the location of the house. This kind of change can be detected by the retrieval component; it knows when it is forced outside of focus to find a match for a relationship. A change in perspective also occurs when the event participants that are highlighted in an utterance differ from those that are highlighted in the previous utterance. If an item is referred to in a prepositional phrase in one utterance and becomes the subject of the next utterance, this change in syntactic status may indicate a shift of focus to that item. Detecting this kind of change requires integrating immediate and global focus.

Implicitly focused items are those concepts not explicitly mentioned in a discourse, but closely related to items that have been. For example, when talking about a particular tree, the trunk, although not mentioned, is implicitly focused. Implicitly focused concepts can be referred to by a definite noun phrase. In some instances such references indicate a shift in focus. In the task dialogues, references to items implicitly focused through a process (e.g., the bolts involved in a substep of an attaching operation) indicate a shift in focus. In other forms of discourse such as descriptive narration, reference to items implicitly focused as parts of objects in focus may be indicators of shifts in focus. To extend the focus representation to these other forms of discourse, the representation of implicit focus needs to be extended to handle relationships between objects as well as events. This requires a consideration of the different kinds of relationships that objects can enter into, since the implicit focus for an object includes associations like ownership and location as well as subparts. Hayes (forthcoming) investigates this problem.

The search of items in implicit focus (e.g., to identify referents of definite noun phrases or interpret action descriptions) needs to be modified in two ways to accommodate other forms of discourse. First, the current strategy only considers event subpart relationships. The inclusion of associations other than subpart relationships into implicit focus influences the question of depth vs. breadth in the search of implicit focus. The depth of search for the different kinds of associations is different. For example, subparts of subparts may be implicitly focused, but not locations of owners. Second, the strictly top-down search described in Chapter V must be modified. It works for the task dialogues because the task structure sets up strong expectations about what will be discussed next. For other forms of discourse, a combination of this top-down processing with more bottom-up processing is probably needed (cf. Rieger, 1975; Hobbs, 1976).

4. GENERATION OF DESCRIPTIONS

A representation of focus is as important for language generation as it is for language understanding. Although focus affects many aspects of generation, its influence is perhaps clearest when considering the generation of descriptions of objects. The role of the focus representation here, as in the interpretation of definite references, is to circumscribe those items from which the object to be identified must be distinguished. In addition, implicit focus can be used to provide a metric for deciding when definite reference can be made to an item not explicitly in focus, but closely related to an item that is.

The focus representation needs to be augmented by several features to enable natural descriptions to be produced; merely distinguishing an object from the others in focus is not sufficient (see Chapter II, Section D.5). There is a trade-off between the time taken to understand more detailed (and hence precise) descriptions and the time taken to locate an object from a minimal description. A generation procedure needs to include some planning mechanisms that consider the actual problem of locating the object being identified. If an object is in focus, this means considering what attributes will make it easiest to distinguish, a different problem from deciding what attributes are sufficient to distinguish it! If the object is not explicitly in focus, the generation procedure needs to consider the problem of locating the object (from the hearer's perspective), i.e., how to minimize the search needed to bring this object into focus. Finally, the coordination of the focus representation with a user model is needed so that the user's own model of the world is taken into account. The user model influences both the complexity of the semantic description and the particular words used in expressing that description.

5. AMBIGUITY AND INEXACT MATCHES

The procedures described in Chapter IV for identifying the referent of a definite noun phrase required that the retrieval component locate exactly one object that matches the description of the noun phrase (or, in the case of plurals, a set). The retrieval component can fail to find such a match even though for most people the noun phrase suffices to identify an object. For example, suppose two men, one wearing a beret and the other hatless, are having a conversation and I want to tell you something about one of them. Although the phrase, "the man wearing the ski cap," does not describe either one, it is clear that the man wearing the beret is being identified. There are two ways such failures can occur. As in the example, there may be no object that exactly matches the description in the noun phrase. Alternatively, more than one object may match, but the ambiguity may not matter for the purposes of the utterance. The problem in either case is to determine the nature of the mismatch and whether it matters.

Although the general question of inexact matches is a semantic issue and has to do with how the general deduction component works, the question of when an inexact match is sufficient is context dependent. The difference between yellow and green may not matter when a yellow-green shirt is being distinguished from a red one; it does matter when picking lemons. The focus representation provides one crucial element for deciding about inexact matches. It separates those items that are in the focus of attention from all other known items. If an exact match cannot be found in focus, it is reasonable to ask if any of the items in focus come close to matching the description of the noun phrase (the question of what is close is the other crucial element in such decisions) and if so which is closest. This illustrates another use of the distinction of items on the basis of relevance that is provided by the focus representation. By making clear those items a reference must distinguish among and by eliminating those items that can match a description but that are not relevant to the discourse, it provides a set of items that can be examined to determine if an inexact match is possible.

> Ambiguity seems to be an essential, indispensable element for the transfer of information from one place to another by words, where matters of real importance are concerned.
>
> Lewis Thomas, The Lives of a Cell

VIII REFERENCES

Bobrow, Daniel G., Kaplan, Ronald M., Kay, Martin, Norman, Donald A., Thompson, Henry, Winograd, Terry. GUS, A Frame Driven Dialog System. Artificial Intelligence 1977, V8:2.

Bobrow, Daniel G. and Winograd, Terry. An Overview of KRL, A Knowledge Representation Language. Cognitive Science 1977, 1.

Burton, Richard R. Semantic Grammar: An Engineering Technique for Constructing Natural Language Systems. BBN Report 3453, ICAI Report 3, Bolt Beranek and Newman, Cambridge, Massachusetts, December 1976.

Chafe, Wallace L. Discourse Structure and Human Knowledge. In: Language Comprehension and the Acquisition of Knowledge. Edited by Roy O. Freedle and John B. Carroll. Winston, Washington, D.C., 1972. Pp. 41-69.

Chafe, Wallace L. Language and Consciousness. Language, 1974, 50, 111-133.

Chafe, Wallace L. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In: Subject and Topic. Edited by Charles N. Li. Academic Press, New York, 1976. Pp. 25-55.

Chapanis, Alphonse. The Communication of Factual Information Through Various Channels. Information Storage and Retrieval, 1973, 9, 215-231.

Chapanis, Alphonse. Interactive Human Communication. Scientific American, March 1975, 36-42.

Charniak, Eugene. Toward a Model of Children's Story Comprehension. AI TR-266, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1972.

Cohen, Philip R. and Perrault, C. Raymond. Preliminaries for a Computer Model of Conversation. In Proceedings of the First CSCSI/SCEIO National Conference, Vancouver, British Columbia, Canada, 25-27 August 1976.

Deutsch[Grosz], Barbara G. Typescripts of Task Oriented Dialogs. SUR Note 146, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, August 20, 1974.

Evans, Thomas G. A Heuristic Program to Solve Geometric-Analogy Problems. Ph. D. Thesis, Department of Mathematics, MIT. Cambridge, Mass. May, 1963.

Fikes, Richard E. The Deduction Component. In: Speech Understanding Research. Edited by Donald E. Walker. Final Technical Report. Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, October 1976.

Fillmore, Charles J. An Alternative to Checklist Theories of Meaning. Proceedings of the First Annual Berkeley Linguistics Society. Institute of Human Learning, Berkeley, California, 1975.

Fillmore, Charles J. The Case for Case Reopened. To appear in: Syntax and Semantics. Edited by John P. Kimball. Academic Press. New York.

Freedle, Roy O. Language Users as Fallible Information-Processors: Implications for Measuring and Modeling Comprehension. In: Language Comprehension and the Acquisition of Knowledge. Edited by John B. Carroll and Roy O. Freedle. Winston, Washington, D.C., 1972. Pp. 169-209.

Garvey, Thomas D. Perceptual Strategies for Purposive Vision. Technical Note 117, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, September 1976.

Gregory, R. L. Eye and Brain: The Psychology of Seeing. McGraw Hill, New York, 1966.

Grimes, Joseph E. The Thread of Discourse. Technical Report 1. Department of Modern Languages and Linguistics, Cornell University. 1972.

Halliday, Michael A. Notes on Transitivity and Theme in English. Part 2. Journal of Linguistics, 1967, 31, 177-274.

Halliday, Michael A., and Hasan, Ruqaiya. Cohesion in English. London, Longman, 1976.

Hankamer, Jorge, and Sag, Ivan. Deep and Surface Anaphora. Linguistic Inquiry, 1976, 7, 390-428.

Hart, Peter E. Progress on a Computer Based Consultant. Technical Note 99, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, January 1975.

Haviland, Susan E. and Clark, Herbert H. What's New? Acquiring New Information as a Process in Comprehension. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 512-521.

Hayes, Philip J. Some Association-Based Techniques for Lexical Disambiguation by Machine. Doctoral Thesis, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, forthcoming.

Hendrix, Gary G. Expanding the Utility of Semantic Networks Through Partitioning. Advance Papers, International Joint Conference on Artificial Intelligence, Tbilisi, Georgian SSR, 3-8 September 1975, 115-121. (a)

Hendrix, Gary G. Partitioned Networks for the Mathematical Modeling of

Natural Language Semantics. Technical Report NL-28, Department of Computer Sciences, University of Texas, Austin, Texas, 1975. (b)

Hendrix, Gary G. The Representation of Semantic Knowledge. In: Speech Understanding Research. Edited by Donald E. Walker. Final Technical Report. Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, October 1976.

Hendrix, Gary G. Human Engineering for Applied Natural Language Processing. Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, February 1977.

Hobbs, Jerry R. Pronoun Resolution. Research Report 76-1, Department of Computer Sciences, City University of New York, New York, August 1976.

Karttunen, Lauri What Makes Definite Noun Phrases Definite. Paper P-3871, The RAND Corporation, Santa Monica, California, 1968.

Levin, James A. Proteus: An Activation Framework for Cognitive Process Models. Working Paper WP-2. USC/Information Sciences Institute, Marina del Rey, California, 1976.

Linde, Charlotte and Labov, William. Spatial Networks as a Site for the Study of Language and Thought. Language, 1975, 51, 924-939.

Malhotra, Ashok. Design Requirements for a Knowledge-Based English Language System for Management: An Experimental Analysis. Ph.D. Thesis, Sloan School of Management, Massachusetts Institute of Technology, February 1975.

Maratsos, Michael P. The Use of Definite and Indefinite Reference in Young Children. Cambridge University Press, Cambridge, 1976. Pp. 133-136.

Minsky, Marvin. A Framework for Representing Knowledge. AI Memo 306, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1974.

Nash-Webber, Bonnie L. Semantic Interpretation Revisited. BBN Report 3335, AI Report 48, Bolt Beranek and Newman, Cambridge, Massachusetts, July 1976.

Newell, Allen, et al. Speech Understanding Systems: Report of a Steering Committee. North-Holland Publishing Company, Amsterdam, 1973.

Norman, Donald A., Rumelhart, David E., and the LNR Research Group. Explorations in Cognition. W.H. Freeman, San Francisco, 1975.

Olson, David R. Language and Thought: Aspects of a Cognitive Theory of Semantics. Psychological Review, 1970, 77, 257-273.

Paxton, William H. A Framework for Speech Understanding. Ph. D. Thesis, Department of Computer Science, Stanford University, June, 1977.

Quillian, M.R. Semantic Memory. In: Semantic Information Processing.

Edited by Marvin Minsky. The MIT Press, Cambridge, Massachusetts, 1968. Pp. 227-270.

Rieger, Charles J., III. Conceptual Overlays: A Mechanism for the Interpretation of Sentence Meaning in Context. Technical Report TR-354. Computer Science Department, University of Maryland, College Park, Maryland. February 1975.

Robinson, Jane J. Performance Grammars. In: Speech Recognition: Invited Papers of the 1974 IEEE Symposium. Edited by D. Raj Reddy. Academic Press, New York, 1975. Pp. 401-427.

Rumelhart, David E. Notes on a Schema for Stories. In: Representation and Understanding: Studies in Cognitive Science. Edited by Daniel R. Bobrow and Alan Collins. Academic Press, New York, 1975.

Sacerdoti, Earl D. A Structure for Plans and Behavior. Technical Note 109, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, February 1977.

Sacks, Harvey, Schegloff, Emanuel, A. and Jefferson, Gail. A Simplest Systematics for the Organization of Turn-taking for Conversation. Language, 50, 1974, 696-735.

Schank, Roger C. and the Yale A.I. Project. SAM -- A Story Understander. Research Report 43, Computer Science Department, Yale University, New Haven, Connecticut, 1975.

Schegloff, Emanuel. Notes on a Conversational Practice: Formulating Place. In: Studies in Social Interaction. Edited by David Sudnow. The Free Press. N.Y., 1972.

Scragg, Greg W. Answering Questions About Processes. In: Explorations in Cognition. Edited by Donald A. Norman and David E. Rumelhart. W.H. Freeman, San Francisco, 1975. Pp. 349-375.

Searle, John R. Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge, 1969.

Silva, Georgette. SDC-SRI Protocol Gathering Experiments and Computer Analysis of Dialog. SUR Note 141, System Development Corporation, Santa Monica, California, October 1975.

Simmons, Robert F. Semantic Networks: Their Computation and Use for Understanding English Sentences. In: Computer Models of Thought and Language. Edited by Roger C. Shank and Kenneth M. Colby. Freeman, San Francisco, 1973. Pp. 63-113.

Tenenbaum, Jay M. On Locating Objects by Their Distinguishing Features in Multisensory Images. Computer Graphics and Image Processing, December 1973, 2, 308-320.

Van Dijk, Teun A. Some Aspects of Text Grammars: A Study in Theoretical Linguistics and Poetics. Mouton, The Hague, 1972

Walker, Donald E., et al. Speech Understanding Research. Annual Technical Report, Project 3804, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, June 1975.

Walker, Donald E. (Ed.). Speech Understanding Research. Final Technical Report, Project 4762. Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, October 1976.

Werner, Oswald. Pragmatics and Ethnoscience. Anthropological Linguistics, 1966, 8.8, 42-65.

Winograd, Terry. Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. MAC TR-84. M.I.T. Artificial Intelligence Laboratory, 1971.

Winograd, Terry. Frame Representations and the Declarative/Procedural Controversy. In: Representation and Understanding. Edited by Daniel G. Bobrow and Allen M. Collins. Academic Press, New York, 1975. Pp. 185-210.

Winston, Patrick H. Learning Structural Descriptions From Examples. MAC TR-76. M.I.T. Artificial Intelligence Laboratory, 1970.

Appendix A

SUMMARY OF RELATED RESEARCH

Appendix A
SUMMARY OF RELATED RESEARCH

Research on various aspects of discourse is being carried on in many different fields. The following references are not complete but rather are meant to provide some indication of the range of material that addresses issues related to the problems discussed in this report.

Related work on cohesion in discourse and discourse structure includes that of Grimes (1972), Halliday and Hasan (1976), and VanDijk (1972). Work in sociolinguistics that has investigated the structure present in spontaneous conversation (e.g., Schegloff, 1972; Sacks et al., 1974) and narrative (e.g., Linde, 1975) is especially relevant.

The importance of focus for interpreting utterances in a discourse is closely related to issues of givenness and definiteness in noun phrases (e.g., Chafe, 1976; Haviland and Clark, 1974; Halliday, 1967) and to the problem of interpreting actions in context (Rieger, 1975; this work is discussed in some detail in Chapter 5). Recent work by Fillmore on case frames and perspective (Fillmore, 1975, to appear) is particularly relevant to the problem of immediate focus but also provides insights into some facets of global focus and to the interface between these two.

An extensive amount of research has been done on various aspects of definite noun phrases. This includes investigations of when definite noun phrases are referential (Karttunen, 1968) and what properties an object must have to be referred to definitely (Chafe, 1976; Karttunen, 1968). Several computer systems have incorporated procedures for resolving definite noun phrases in unstructured discourse (e.g., Winograd, 1971; Norman et al., 1975). These procedures have been ad hoc, to some extent because they have not taken discourse structure into account. Designs for more general reference resolution strategies are presented in Hobbs (1976) and Levin (1976).

The need to associate groups of items in the knowledge base is closely related to general issues in the representation of knowledge (see Bobrow and Winograd, 1977). Scripts (Schank, et al., 1975) and frames (Minsky, 1974; Winograd, 1975) are two representation schemes that address problems of a more global and static structuring of knowledge than the dynamic grouping for focus investigated in this report. Work on partitioned semantic networks (Hendrix, 1975a,b) has had a direct influence on the development of the focus representation.