INTERPRETING NATURAL-LANGUAGE UTTERANCES IN DIALOGS ABOUT TASKS

Technical Note 210

March 15, 1980

By: Ann E. Robinson, Douglas E. Appelt
Barbara J. Grosz, Gary G. Hendrix,
and Jane J. Robinson

Artificial Intelligence Center
Computer Science and Technology Division

SRI Project 5844

The work reported herein was supported by the National Science Foundation under Grant No. MCS76-22004.

INTERPRETING NATURAL-LANGUAGE UTTERANCES IN DIALOGS ABOUT TASKS

Ann E. Robinson, Douglas E. Appelt, Barbara J. Grosz, Gary G. Hendrix, and Jane J. Robinson*

ABSTRACT

This paper describes the results of a three-year research effort investigating the knowledge and processes needed for participation in natural-language dialogs about ongoing mechanical-assembly tasks. Major concerns were the ability to interpret and respond to utterances within the dynamic environment effected by progress in the task, as well as by the concomitant shifting dialog context.

The research strategy followed was to determine the kinds of knowledge needed, to define formalisms for encoding them and procedures for reasoning with them, to implement those formalisms and procedures in a computer system called TDUS, and then to test them by exercising the system.

Principal accomplishments include: development of a framework for encoding knowledge about linguistic processes; encoding of a grammar for recognizing many of the syntactic structures of English; development of the concept of "focusing," which clarifies a major role of context; development of a formalism for representing knowledge about processes, and procedures for reasoning about them; development of an overall framework for describing how different types of knowledge interact in the communication process; development of a computer system that not only demonstrates the feasibility of the various formalisms and procedures, but also provides a research tool for testing new hypotheses about the communication process.

CONTENT INDICATORS: 3.60, 3.69, 3.42

KEY WORDS: Natural-language understanding, Task-oriented dialogs

----

* Authors' current address: Artificial Intelligence Center, SRI International, Menlo Park, California 94025.

I INTRODUCTION

A major research effort at SRI International has been devoted to investigating the knowledge and processes necessary for participation in natural-language dialogs about ongoing mechanical-assembly tasks.* One primary concern has been the ability to follow both the progress of the task and the shifting context of the dialog; another has been the ability to interpret and respond to written utterances within that dynamic environment. Such sensitivity to the dynamics of the dialog and the task environment is central to the communication process.

Dialog participation requires the coordinated use of knowledge of many kinds. Knowledge about the domain of discourse comprises information as to what entities (objects and actions) are present, what processes are needed for reasoning about them, how a specific task is performed, and what goals the dialog participants are attempting to achieve. Also necessary is knowledge about language--i.e., what words are used and how they combine into phrases, how words and phrases are related to entities in the domain, what procedures may derive the relationships between phrases and domain entities, how utterances interact with the dialog and domain context.

The principal research issues we have addressed include:

* How to represent knowledge about words and phrases--and how to use it in interpreting utterances.
* How to represent and use knowledge about the relationship between phrases and domain entities.
* How to represent and use knowledge about context.

----

* This research has been funded under a three-year National Science Foundation Continuing Research grant No. MCS76-22004. We wish to thank many of our colleagues for significant contributions to this work, including Jerry Hobbs, Robert Moore, William Paxton, Earl Sacerdoti, Candace Sidner, and Donald Walker.

* How to represent knowledge about processes.
* How to reason about domain processes.
* How to discern the goals of the dialog participants from their utterances.
* How to coordinate the use of linguistic and domain knowledge in interpreting utterances, and, in particular, how to identify the entities referred to by noun phrases and verb phrases.

Our research strategy has been to determine the kinds of knowledge needed, to define formalisms for encoding both the knowledge and the procedures for reasoning with it, to implement the formalisms and procedures in a computer system called TDUS, and then to test them by exercising the system.

Detailed descriptions of particular aspects of our research are available in other publications [2], [7], [8], [9], [10], [11], [12], [15], [16] [26], [27], [28], [29], [30], [18], [19]. To make the discussion more concrete, we will explicate several of our ideas within the context of the TDUS system.

Figure 1 shows a dialog segment in which one participant is the TDUS system. TDUS acts as an expert, guiding an apprentice through an assembly task. TDUS interpreted the utterances preceded by a '#' and produced the succeeding sentences in reply. The task under discussion is a subtask in the assembly of an air compressor--installation of the pump.

The dialog clearly illustrates the ability of TDUS to deal with a changing environment. The system follows both the task as it progresses and the changes that occur in the apprentice's goals and in focusing--the highlighting of entities that are most relevant at any particular time during the dialog.

Utterances (1) and (2) of Figure 1 include a computer-generated commentary describing the system's knowledge about the progress of the task. Only the apprentice's utterances and the system's replies are given for Utterances (3)-(15). Section III contains an expanded version of this dialog with computer-generated commentary for all the

```
#I AM ATTACHING THE PUMP                                        (1)
OK

        The following has been assumed:

        Focus has shifted to:
          Primary focus:
             PL - a platform.
             PU - a pump.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
                  COMP - a compressor.

        Expected immediate focus:
        PU - a pump.

        New goal-step is:
        Attaching the pump to the platform.

#WHICH WRENCH SHOULD I USE TO BOLT IT                           (2)
Use the box-end wrench.

        Focus has shifted to:
          Primary focus:
             You are using the box-end wrench.
             W4 - a box-end wrench.
             The pump is being bolted.
             PL - a platform.
             PU - a pump.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
                  COMP - a compressor.

        Immediate focus:
        PU - a pump.

        New goal-step is:
        Knowing the wrench to use.

        .
        .
        .
#WHERE IS THE WRENCH                                            (5)
The box-end wrench is on the table.
```

#I FOUND IT (6)
OK

.
.
.

#WHAT SHOULD I DO NOW (12)
Install the aftercooler elbow on the pump.

#I'VE DONE IT (13)
OK

.
.
.

Figure 1 Sample Dialog with the Computer

utterances, including focusing and goal information and indications as to the steps TDUS assumes the apprentice has performed or is performing. In this dialog segment, Utterance (1) shows the person reporting the start of a new step in the task. The commentary indicates the entities currently focused on by the dialog participants, primarily the pump (PU) and platform (PL), and the goal the system assumes the apprentice is attempting to achieve, i.e., that the pump be attached. Utterance (2) shows how focusing shifts when the apprentice mentions new entities and a new task step--in this case the bolting substep--and how the system assumes the apprentice's most immediate goal has become that of knowing which wrench to use.*

The use and interpretation of "it" in this dialog presents one example of how the context of an utterance affects its interpretation. "It" occurs in Utterance (2) and refers to the pump. "It" is also used in Utterances (6) and (13), in each case referring to a different entity. In Utterance (6) "it" refers to the wrench W4. In Utterance (13) "it" refers not to an object, but to an action. Without knowledge of the current situation, determining what "it" refers to in each of these utterances is impossible.

----

* See Grosz [7] and [8] for a discussion of focusing and shifting focus.

The interpretation of each utterance includes analyzing its grammatical structure, determining its propositional content (loosely, the literal meaning), deriving a representation of the propositional content in terms of the system's representation of the domain, inferring from the utterances any changes in the domain, the speaker's goal, or focus of attention, and deciding on an appropriate reply. In Section II we discuss the knowledge and procedures necessary for performing these tasks. In Section III we show how they enter into the interpretation of the utterances in the sample dialog. In Section IV we summarize some of the research results and their relationship to other research.

II KNOWLEDGE ABOUT LANGUAGE AND THE DOMAIN

Dialog participation requires knowledge about both the subject area and the language in which the dialog occurs. Development of a computer system capable of participating in a dialog requires encoding that knowledge, as well as developing effective procedures for using it in the interpretation and production of utterances. In the following pages we describe the knowledge encoded in the TDUS system, the formalisms developed for encoding it, and the ways in which it is used for interpreting and responding to utterances, and for following changes in the domain and dialog contexts.

A. The Subject Domain

We have developed a general representational formalism for encoding information about a dynamic environment and have used it in TDUS to embody knowledge about the objects, actors, relationships, and events relevant to the task of assembling an air compressor. The resulting domain model provides a semantic framework for interpreting utterances against the background of a changing environment and provides the subject-dependent information needed for responding to domain-specific questions.

The representational formalism is a partitioned network data structure, as described in Hendrix [14], [16]. This network formalism has been augmented during the past three years with another formalism, process models [2], for encoding information about processes. The incorporation of the formalism for process models significantly extends our representational capabilities. Its use in the implementation of TDUS provides a means for maintaining an up-to-date record of task progress. This record forms an essential part of the domain context within which utterances are interpreted and questions answered. At any given moment the domain context indicates what assembly actions have already occurred (and in what order), what actions are in progress, and what actions can be initiated next.

This formalism provides the ability to specify the hierarchical decomposition of events into subevents, as well as to describe individual types of events. The description of each event type includes information about its participating actors and objects, the preconditions for its enactment, its effects, and the alternative sequences of substeps that may be performed to accomplish it. The sequence of substeps can be partially ordered. This decomposition of events builds on earlier research regarding the hierarchical decomposition of the planning process [31] and on the work by Hendrix [13], [14] dealing with modeling actions and processes.

Figure 2 illustrates a process model for the pump-attaching process in the assembly domain. The network node ATTACH PUMP represents the set of pump-attaching events. The large box depicts a separate space in the network in which the schema of the ATTACH PUMP event is represented. This schema specifies the participants in the attaching operation, marked by the MAJORPART, MINORPART, and AGENT arcs, and describes the event of attaching. The event description includes the PRECONDITIONS that must be true for the action to be performed, the EFFECTS of performing the action, and the PLOT or steps by which the action is performed. The plot is encoded on a separate space. Each step in the plot is in turn further described by a process model. In this example, the substeps of attaching are positioning and bolting the pump (SECURE WITH BOLTS).

As an action is being performed, a record of progress is kept by instantiating a schema for that action and any of its subactions. The new piece of network so created is incorporated into the domain model. Records of actions are linked both temporally by a time lattice and through their taxonomic relationships with other events and objects in the task. Internally each event represented has associated with it an interval bounded by two times: a start time and an end time. For events

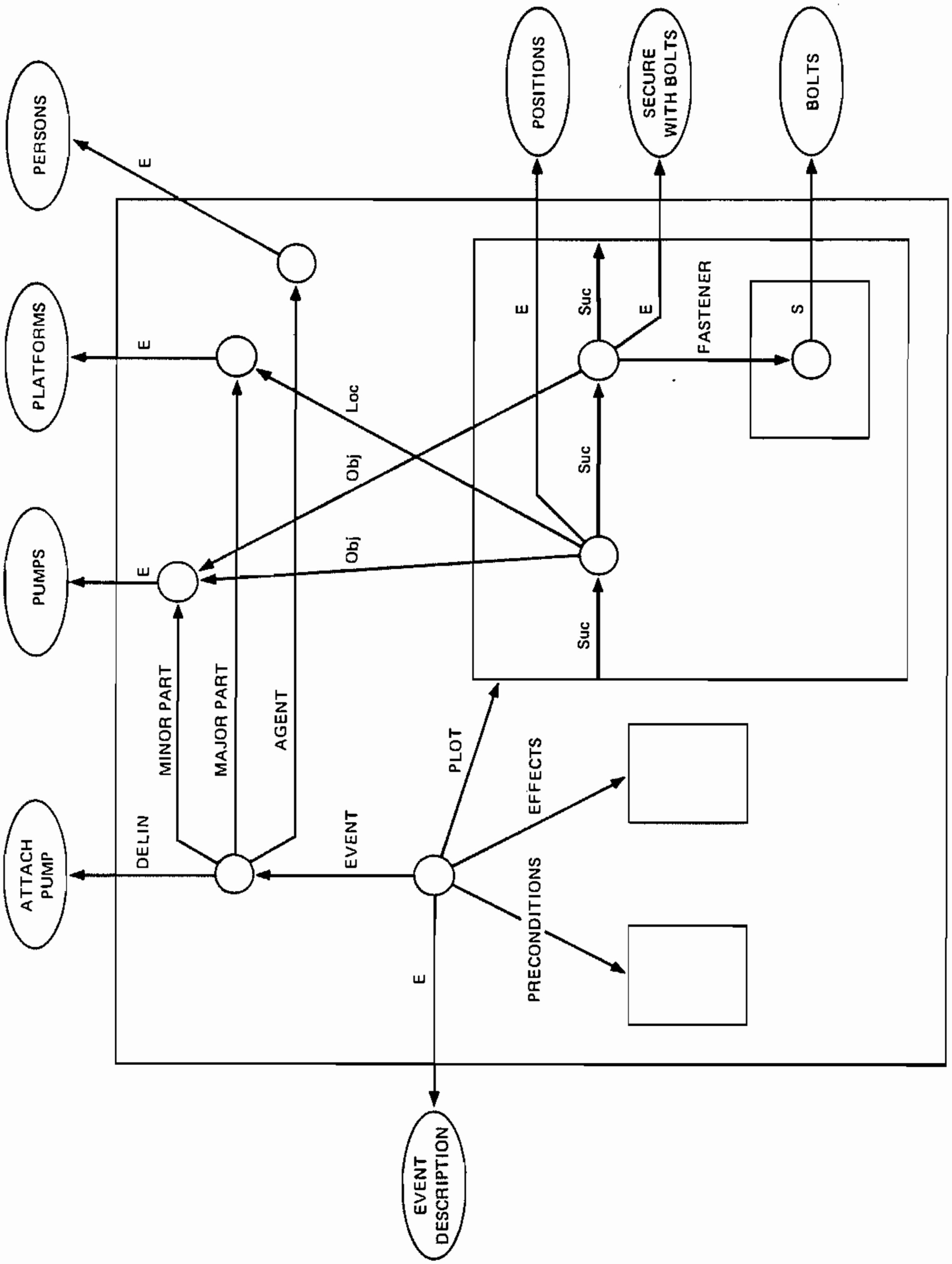

Figure 2 An Attaching Process

treated as points, the start and end times are identical. For events whose start and/or end time is not precisely known, the values may be left unspecified or represented by parameters that are bounded above and/or below by known points in the time lattice.

Once an instance of an action is recorded in the domain model, it can be used in subsequent deductions and is available for answering questions about past events.

In conjunction with the development of the process model formalism, we have extended our deductive system [5] by adding specific knowledge about how to manipulate process models. These procedures are used to retrieve or deduce answers to questions (which might be posed either by some process during the interpretation of an utterance or by the other participant in a dialog) and to assimilate into the model new domain information communicated by the user.

Examples of questions about the task domain posed by the system and answered during the interpretation process include: "Can one object be in a certain relation to another object?" ("Can a wrench be on a table?") and "Can an object participate in a given type of event in a certain role?" ("Can the wrench be used to tighten bolts?"). These questions are asked during interpretation to help decide if particular words or phrases combine meaningfully within the domain (e.g., "red pump" versus "red location").

Questions posed by the user include: "Does an object have a certain property?" ("Is the pump on the table?"), "What is the value of a certain property for a specific object?" ("Where is the wrench?"), "What are the substeps of a certain type of event?", and "Does there exist a record of an occurrence of a certain type of event?" ("Have the bolts been tightened?"). New information can be communicated by the user in statements like: "I am tightening the bolts," "I attached the pulley," and "I found it." The situations described by such utterances are assimilated into the model along with certain implied situations, such as the completion of substeps or other actions that are prerequisites of the situation reported.

B. Knowledge About the Context

Perhaps the most important achievement of our research has been the establishment of a framework within which to characterize language as it is used in a changing context. Utterances are influenced both by the domain context provided by the ongoing task and by the linguistic or dialog context provided by the preceding utterances.

Knowledge of the domain context is essential, for example, in determining the referents of such phrases as "the tool used in the last step" and "that's done." The importance of the prior discourse, which we call the dialog context, is illustrated by the pair of utterances

Speaker 1: Why did John take the pump apart?
Speaker 2: He did it to fix it.

in which the interpretation of the second utterance must take into account the context established by the first. In particular, "he" refers to John, "did it" refers to the disassembly task, and the second "it" refers to the pump.

Two major aspects of discourse context for which we have identified and developed a representation are focusing [7], [8], [9] and the goals of the apprentice [33], [28].*

Focusing is an active process. As a dialog progresses, the participants continually shift their focus of attention and thus form an evolving context against which utterances are produced and interpreted. A speaker provides a hearer with clues of what to look at and how to look at it--what to focus upon, how to focus upon it, and how wide or narrow the focusing should be. We have developed a representation for discourse focusing, procedures for using it in identifying objects referred to by noun phrases, and procedures for detecting and representing shifts in focusing [7], [8], [9], [11].

A speaker can have multiple goals of various types. These include "domain goals" related to the subject domain being discussed,

----

* The current implementation of goals in TDUS is an extension and partial revision of one developed at SRI by Sidner and described in [33].

"knowledge-state goals" concerned with changing the knowledge of one or more of the dialog participants, and "social goals" arising from both the social context in which a dialog takes place and from the interpersonal relationships of the dialog participants. For a particular speaker in a dialog, a single utterance can be viewed as having multiple facets corresponding to the multiplicity of goals the speaker is trying to achieve [10]. For example, an utterance can simultaneously request information (a knowledge-state goal) and convey politeness.

Each dialog participant can be viewed as having individual domain, knowledge-state, and social goals. Many of the goals may be shared by the participants, particularly in cooperative dialogs, but they can also be conflicting or complementary. TDUS provides an explicit representation of the system's view of the apprentice's goals, including domain goals and certain knowledge-state goals. Other interesting insights into, and methods for, recognizing and expressing separate goals and for taking this separation into account in interpreting and generating utterances have been discussed elsewhere [4], [1], [3], [36].

Figure 3 illustrates the relationship between actions and goals. Showing the same portion of the task as was illustrated in Figure 2, it is a simplification of the assembly task hierarchy currently encoded in TDUS.* Each node in Figure 3 represents an action and its associated goal. The hierarchy encodes the substep relationships: child nodes represent substeps of their parent nodes. The top-level node in the tree, node (1), is the action of attaching a pump whose associated goal is that the pump be attached.** Nodes (2) and (3) are substeps of this attaching process, i.e., the actions of positioning the pump and

----

* Although the assembly task currently encoded in TDUS provides strong structuring of actions and goals, the representations and procedures we have developed are also applicable to less structured domains.

** Our current assumption is that each action has one main goal, or purpose; we have not addressed the problem of determining which of a set of effects is the apprentice's goal.

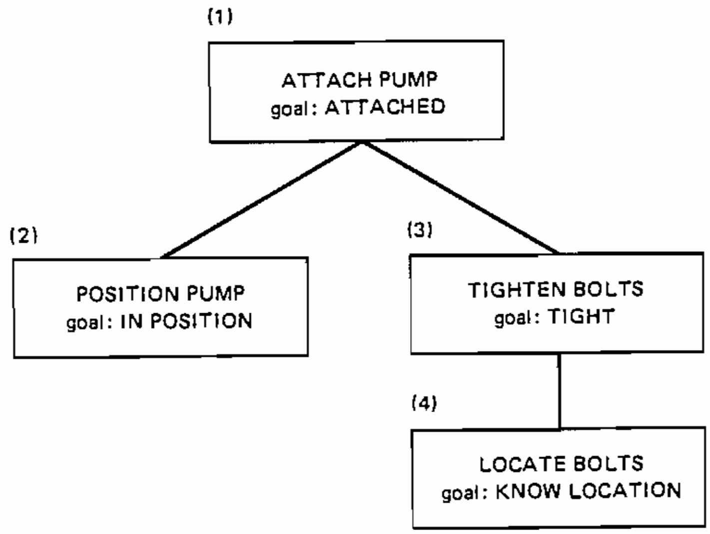

Figure 3 Goal/Action Tree

tightening the bolts, with the associated goals that the pump be in position and the bolts be tight. These substeps correspond to the PLOT steps in Figure 2. The action of locating bolts encoded in node (4) is not an action that is an explicit step in the task model, but it is nevertheless essential to its performance. It has an associated knowledge-state goal "know the location of the bolts."

We distinguish two classes of goals: "direct goals" achieved by actions the apprentice has explicitly or implicitly said <u>are currently being performed</u> or <u>have been performed</u> and "potential goals," mentioned by either participant, that <u>have not been acted upon</u> but <u>might possibly be</u>. In the task context of Figure 3, "I am attaching the pump" states that the speaker is performing an instance of the attaching action represented by node (1). Thus the utterance establishes having the pump attached as a direct goal. "Should I tighten the bolts?" indicates that the speaker _might_ perform the tightening action represented by node (3)--and thus establishes the bolts being tight as a _potential_ goal.

Knowledge of the goals of dialog participants interacts with the interpretation of verbs in at least two ways: (1) the interpretation of verbs comprises recognition of those goals of the speaker that are expressed or implied by the verb; (2) current goals are a part of the context that provides possible events that may serve as the referents of verbs.

Our discussion of the dialog sample in Section III, below, illustrates the role of focusing in identifying referents of definite noun phrases, as well as the role of focusing and goals in identifying the referents of verb phrases.

C. Knowledge about Language

Linguistic knowledge includes information not only about words and how they combine to form phrases, but also about the ways in which words and phrases relate to concepts in the subject domain. Domain, contextual, and linguistic knowledge interacts in many ways during the interpretation process. Before describing the specific linguistic knowledge we have represented we shall describe the framework we have developed for specifying the interactions of the different kinds of knowledge used during the interpretation process.

In this section, we shall describe our use of the framework for participating in dialogs about assembly tasks. In Section IV we shall examine the way it can function as a general-purpose natural-language interface.

1. The Framework for Interpretation

The basic framework of TDUS is embodied in a 'language definition system' called DIAMOND. DIAMOND, which was designed and implemented by William Paxton, may be viewed in broad outline as both a sophisticated programming language and an associated execution system. As a programming language, DIAMOND is used to specify how to interpret utterances. The interpretation process is performed in multiple stages under control of the DIAMOND executive and in accordance with the specifications of the language definition. DIAMOND's flexibility allows experimentation with different numbers of stages and with alternative sequences for applying various types of knowledge. We are now working with three basic processing stages:

The first stage interprets the input 'bottom-up' (i.e., words -> phrases -> larger phrases -> sentences). Phrases are constructed in isolation, without reference to the context in which they might be embedded. This stage is used for relatively simple tests that do not depend on surrounding context. The result of this stage is a complex data structure (a generalization of a syntax tree) that reflects a decomposition of the input utterance into component phrases and associates a number of attributes with each phrase. This structure is expanded and refined during subsequent stages of processing.

In the second stage of processing, each phrase is further interpreted within the context of the preliminary interpretation of the entire utterance. Phrases processed during this stage are generally subsets of those initially considered and are thus more likely (but not guaranteed) to form a correct interpretation of the utterance. Consequently, for the sake of computational efficiency, many of the more extensive processing tasks are delayed until this stage. For example, a relationship is established in this stage between phrases identified during syntactic analysis and descriptions of (sets of propositions about) objects in the domain model.

In the third stage, phrases can be further interpreted within the context of the entire utterance by building on the results of the previous two stages. Major tasks of this stage include delimiting the scopes of quantifiers and, by taking into account the overall dialog and domain context, identifying the objects and actions referred to by noun phrases and verb phrases, and identifying any goals implied by an utterance.

The separation of processing into stages has allowed us to examine more easily the question of when certain types of knowledge can or should be used during interpretation. For example, identification of a pronoun's referent requires knowledge of the role that pronoun plays in the utterance, as well as knowledge about the dialog context [33], [28]. Thus identification is not possible until after the utterance has been interpreted with respect to concepts in the domain. By contrast, simple structural tests (e.g., number agreement) of phrase constituents can be made during the first stage of processing.

A language definition encoded in DIAMOND consists of (1) a lexicon in which words are separated into categories with associated attributes and (2) phrase structure rules augmented with procedures to be evaluated during successive stages of processing. Figure 4 presents a noun-phrase rule in the language definition (simplified for illustrative purposes). It is discussed in some detail in the following section to show how its different constituents identify linguistic structure, relate linguistic form to domain concepts, and relate the utterance as a whole to its context. In Section III the results of each stage of processing are described for Utterance (1).

2. Linguistic Structure

The sample noun phrase rule has four constituents. The first part, the phrase structure, indicates the phrase to be formed. Phrases are generally standard linguistic units, such as noun phrases and verb phrases. The rule specification allows for optional and alternative elements. Thus, as in this example, one rule for interpreting a noun phrase may allow many alternatives. In this example, the DET (determiner) and QUANT (quantifier) constituents are alternatives, indicated by the braces, and the ADJ and PP constituents are optional, indicated by the parentheses. The NOUN constituent is required.

The second part of the rule, the CONSTRUCTOR, is a procedure to be evaluated when a phrase is formed using this rule. Generally a constructor assigns attributes to the phrase and tests for consistency

```
NP = {DET/QUANT} (ADJ) NOUN (PP);                                    |
                                                                     |
   CONSTRUCTOR                                                       |
     (PROGN (@FROM NOUN NUMBER)                                      | Stage 1
            (@FROM DET DEF)                                          |
            (COND ((@ ADJ)(OR (AGREE TYPE ADJ NOUN)                  |
                              (F.REJECT 'NO-AGREEMENT)))))           |

   TRANSLATOR                                                        |
     (@SET SEMANTICS (COMBINE (@ SEMANTICS ADJ)                      | Stage 2
                              (@ SEMANTICS NOUN)))                   |

   INTEGRATOR                                                        |
     (@SET D.IDENT (RESOLVE (@ SEMANTICS)))                          | Stage 3
```

Figure 4 Sample Noun Phrase Rule

among the attributes of its constituents. For example, (@FROM NOUN NUMBER) copies the value of the NUMBER attribute from the NOUN constituent to the NP being built. Procedures can rate an interpretation of a phrase, based on an assessment of its likelihood, and reject unlikely ones. The F.REJECT statement at the end of the CONSTRUCTOR will reject a proposed phrase if the ADJ and NOUN are not of the same type (i.e., the type of adjective in the phrase cannot modify the type of noun present). Unlikely phrases need not be rejected, but can be marked as "less-than-good" (e.g., FAIR or POOR); they can be used, however, if no better interpretation is found.

Much of what is commonly called "syntactic" information (information about words and phrases and how they combine independently of their relationship to the domain) is encoded in the phrase-structure and constructor parts of the language definition.

Using the DIAMOND formalism, we have encoded and tested rules for syntactic analysis that cover a wide range of English constructions. J. Robinson [30] presents an extended discussion of these rules, collectively called the DIAGRAM grammar, and discusses their grounding in current linguistic theory.

Sample utterances covered by DIAGRAM's syntax include:

What size wrench should I use to tighten the bolts?
The wrench that I used before is not on the table.
I've installed the brace, pulley and Woodruff key.
How many bolts of each size should I be using?
Could there have been any bolts removed from the box?
After I took the pulley off, I found two setscrews in the back of the shaft.
If I install the pump on the platform, will the pressure be too much for it?

3. Relation of Linguistic Form to Domain Concepts

The third part of the rule, applied during the second stage of processing, is the TRANSLATOR procedure. A translator for a phrase is evaluated after the phrase has been combined with others to provide a syntactic analysis for an entire utterance. Thus, unlike the constructor, the translator has information available about how the phrase fits into the utterance as a whole. In Figure 4 the operation is simply one of combining the semantics for the ADJ with that for the NOUN.

Within the translator procedures we have formalized our knowledge of the relationship between words and syntactic structures in the language, on the one hand. and concepts in the domain model, on the other. These procedures are extensions of our earlier "semantic composition" routines (see Hendrix in Walker [34]). They can map common nouns, such as "wheel," into restricted free variables, such as "X where WHEEL(X)" and pronouns, such as "he," into restricted free variables, such as "X where MALE(X)." They can map verb-like concepts, such as "removed" or "green-colored," into propositions in the system's representation of knowledge about the world. In addition, these procedures can use knowledge about the domain to reject as meaningless such phrases as "the wheel removed John" (although the same string of words can be used in such meaningful constructions as "with the wheel removed John could fix it"). They can create network structures corresponding to meaningful phrases, such as "John removed the wheel," by combining information about the individual words, the syntax that relates them, and the relevant relationship in the assembly domain. The association between a phrase in the utterance and its corresponding network representation is encoded as an attribute of the phrase.

4. Relating Utterances to the Context

The fourth rule component, the INTEGRATOR procedure, is applied during the third stage of processing. An integrator for a rule specifies how to relate the concepts mentioned in a phrase to specific domain entities. During this stage, the specific objects referred to by definite noun phrases and pronouns are identified, along with the specific actions referred to in verb phrases.

To perform this process, a relatively complex control strategy is used that takes into account focusing, goals, and the domain and dialog contexts in deciding when and where to look for referents of phrases. This strategy governs when noun phrases, pronouns, and verb phrase referents are identified [28]. For example, for the phrase "the wrench," the result of the second stage would be a network structure that describes a wrench. In the third stage, focusing information would be used to help identify the wrench of current interest in the dialog. The model entity representing that particular wrench would be associated with the interpretation of the phrase, as occurred in Utterance (5) of the Figure 1 dialog.

Other examples of integrating an utterance into the system's model involve the occurrence of a pronoun or pro-verb (e.g., "do"). A pronoun or pro-verb is normally used to refer to some focused entity, or to an entity closely related to a focused entity.

D. Responding to Utterances

Participation in the communication process requires determining the appropriate action to take, once an utterance has been interpreted. This means much more than simply answering questions that are posed by the other party. In fact, no clear boundary exists between the process of interpreting an utterance and that of responding to it. Many of the processes we have described above, such as finding indications of changes in the user's goals and focus of attention, could be considered as responding to utterances as well as interpreting them.

The most explicit response is answering questions. We have developed procedures that answer certain types of questions by making deductions from the information in the domain model [5]. A network-to-English translation system developed in a previous project (see Slocum in Walker [34]) produces English replies to questions, using simple syntactic constructions.

III CAPABILITIES OF TDUS -- A SAMPLE DIALOG

A significant aspect of the TDUS system is that it operates in and is responsive to a dynamically changing environment. As utterances indicate that progress has been made in the task and/or that changes have occurred in the apprentice's focus of attention or goals, the change in context is reflected in the system's model of the situation. Changes in the context influence the way in which subsequent utterances are interpreted. Changes can be indicated either by statements that directly specify a change ("I have attached the pump") or less directly by utterances about particular objects or substeps. For example, "Where is the Allen wrench?" suggests that the apprentice may be ready to perform a task requiring the Allen wrench.

In addition to interpreting statements that indicate changes in the task state, TDUS can interpret and respond to questions about the current situation ("Where is the box-end wrench?"), about past situations ("Was the pump on the table before it was attached?"), about how to perform a step in the task ("How do I install the pulley?"), and about what step should be performed next ("What should I do now?"). Similarly, TDUS can interpret and respond to questions about participants in and results of previous steps ("Who attached the pump?", "Have the bolts been tightened?").

TDUS maintains a history of previous events and situations in the domain. This history is used to derive responses to queries about the past and to help interpret such phrases as "the wrench that was used to tighten the motor bolts." To process questions about the assembly task, the steps necessary for accomplishing it are encoded in process models that are also used to deduce the effect of performing an action. The same encoding is used to follow the dynamics of the dialog. As a task is performed, the system infers the resulting situation, which includes the consequences of intermediate steps. The computer representation of focusing and the apprentice's goals are then updated to reflect the revised situation so that references to specific entities and actions can be correctly identified in subsequent inputs.

We shall now look more closely at the sample dialog excerpted in Figure 1. The computer-generated commentary at the right, following each utterance, reflects the system's perception of task progress. The initial focus includes the user (you), the compressor (COMP), the pump (PU), and the table (T1). As entities are mentioned, they are focused upon, becoming part of what we call here "primary focus." The first step of the installation is to attach the pump. Attachment decomposes into smaller steps (see Figure 2). In the initial context for this dialog, the next step to be performed is to install the pump, the first substep of which is to attach the pump.

```
#I AM ATTACHING THE PUMP                                    (1)
 OK

        The following has been assumed:

        Focus has shifted to:
          Primary focus:
             PL - a platform.
             PU - a pump.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
                  COMP - a compressor.

        Expected immediate focus:
        PU - a pump.

        New goal-step is:
        Attaching the pump to the platform.
```

In Utterance (1), the apprentice indicates the start of the pump-attaching action. This is assumed to be the current "goal-step" (i.e., the action associated with the current goal). The utterance also shifts focus to the pump (PU) and the platform (PL), with the pump (PU) as the expected immediate focus*. Note that the platform (PL) became focused without being explicitly mentioned either by the apprentice or by the system, because it is a participant in the attaching step--the pump is attached to it. A subsequent reference to "the platform" will be interpreted as a reference to the platform (PL). The apprentice was already focused and although s/he is a participant in the action, s/he is not part of the new focus because referents of indexicals (e.g., "I" and "you") are treated separately by the focusing mechanism. The previously focused entities (T1, PU, and COMP) are still focused, but to a lesser degree. If or when PU and PL are no longer focused, focusing will shift back to T1, PU and COMP as primary. Meanwhile, they are still accessible as focused. Grosz [7] discusses the structure of focusing and the mechanisms of focus shifting.

The constituents of Utterance (1) are the subject pronoun "I," the verb phrase "attach the pump," and the auxiliaries "am -ing".** The constituents of the verb phrase are the verb "attach" and the noun phrase "the pump." During the first stage of interpretation, a tree indicating these constituents is constructed. A simplified form of that tree is illustrated in Figure 5. During this stage the CONSTRUCTORs of the DIAGRAM grammar are also evaluated. Following this stage, nodes in the tree will have attributes indicating, for example, the tense and aspect (present and imperfective) of the verb and the number of each noun phrase (singular).

During the second stage of interpretation, when the TRANSLATORs are evaluated, a representation of the propositional content of the utterance is constructed, and associated with appropriate nodes in the tree. This representation is independent of the current context. Figure 6 shows the representation for Utterance (1). The nodes PERSONS, ATTACHING, and PUMPS were part of the initial representation of

----

* See Sidner [33] for a discussion of expected and potential immediate focus.

** The "ing" has been stripped from the verb and shifted ahead of it to facilitate analysis [30].

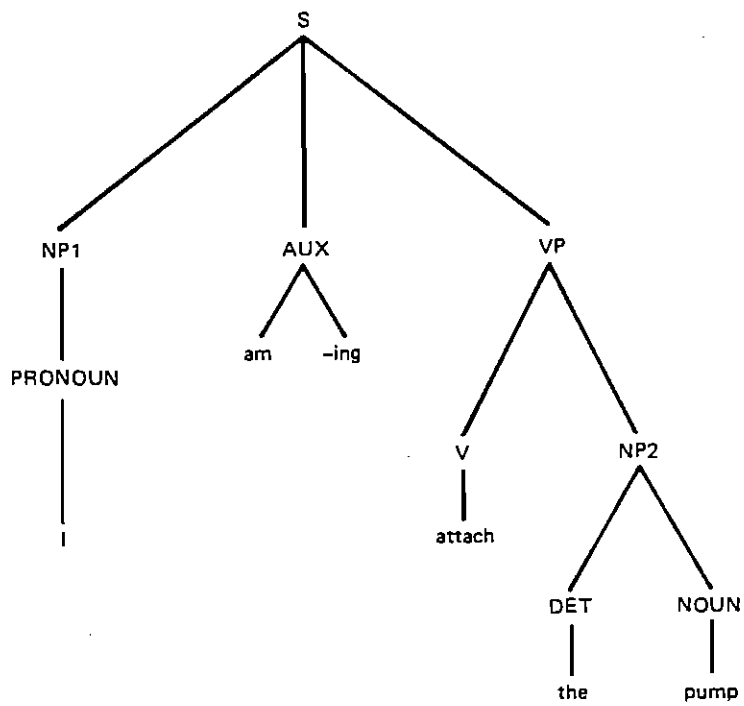

Figure 5 Tree for Utterance (1)

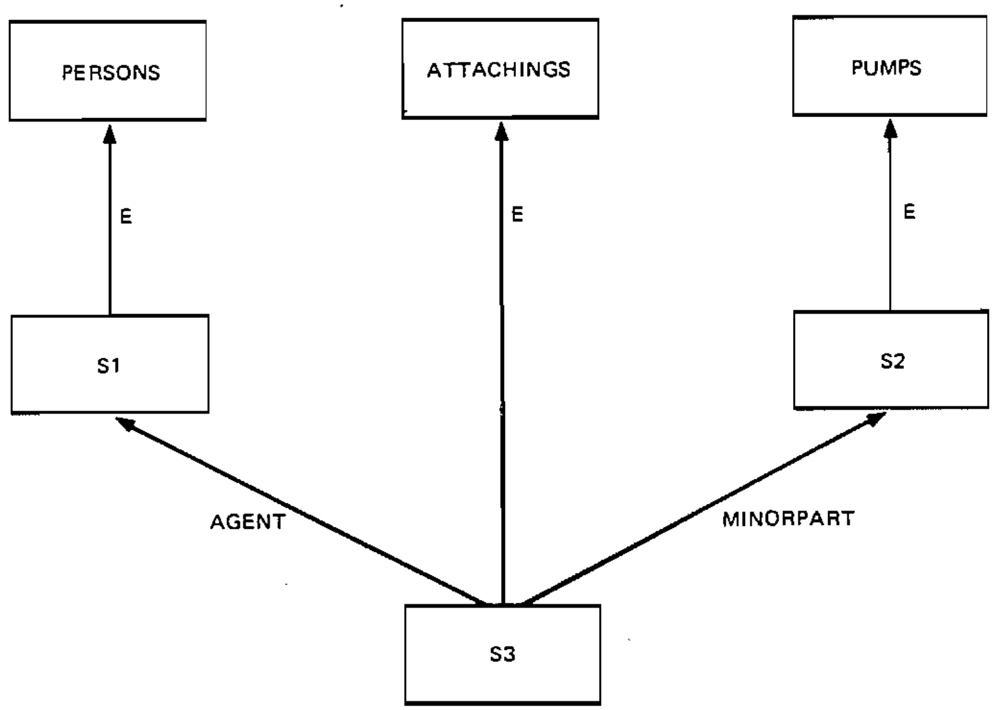

Figure 6 Representation of Propositional Content of Utterance (1)

the domain; S1, S2, and S3 are newly created during the interpretation process and are kept separate from the domain model via the network partitioning mechanism [34]. Node S3 (and its outgoing arcs) is associated with the "S" node of the tree in Figure 5 as the value of the SEMANTICS attribute. The node labeled S1 is associated with the NP1 node as the value of the SEMANTICS attribute and S2 is associated with NP2. The actual representation also indicates that the action is currently in progress. Appelt, et al. [2] discusses the representation of progressive actions.

During the third stage of interpretation, when the INTEGRATORs are evaluated, entities referred to by definite noun phrases and verbs are identified in the domain model. Figure 7 shows the representation that results from this third stage (tense and aspect markings are not shown). "I" is identified as the apprentice, "the pump" as pump (PU), and "attach" as an instance of the attach-pump step in the task model. Node (I1) represents the newly created instance of an attach-pump event. Each of the nodes PU, APPRENTICE, and I1 is associated with the corresponding node in the tree as the value of the D.IDENT attribute. During this third interpretation stage, any new goal and possible focusing shift are also recognized. The implied changes in the context are recorded following this stage as part of the process of responding to the utterance.

Since this utterance is a statement, TDUS's response consists of incorporating into the model the information asserted and updating the task model to reflect the start of the "attach pump" step. TDUS also changes the representation of the context to reflect both the new goal (that the pump be attached) and the focusing shift to the pump and platform.

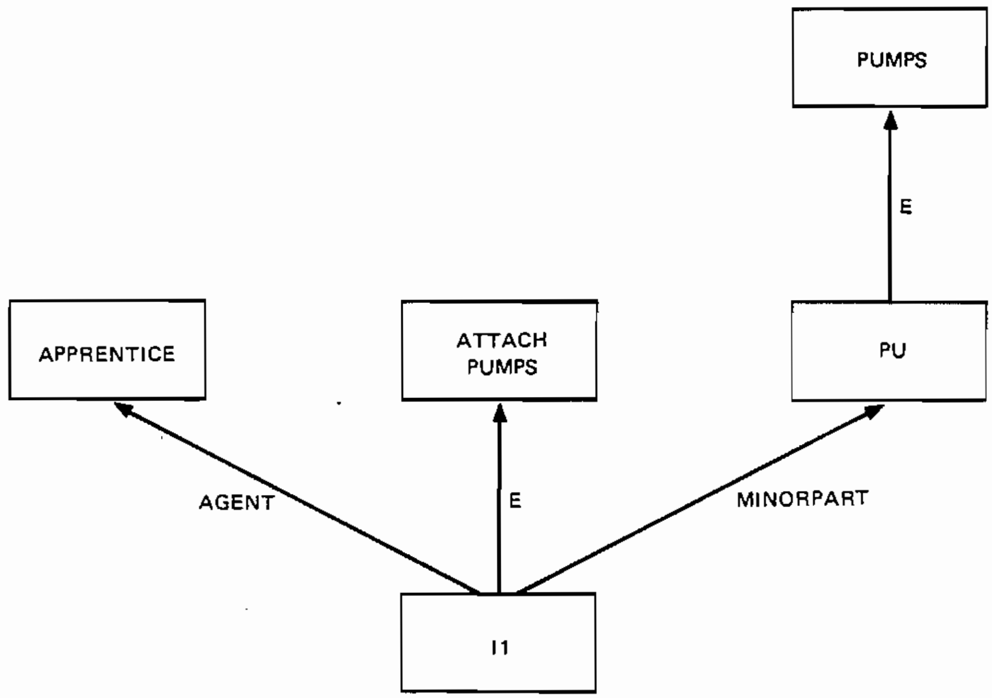

Figure 7   Interpretation of Utterance (1) After Stage 3

```
#WHICH WRENCH SHOULD I USE TO BOLT IT                          (2)
 Use the box-end wrench.

        Focus has shifted to:
          Primary focus:
             You are using the box-end wrench.
             W4 - a box-end wrench.
             The pump is being bolted.
             PL - a platform.
             PU - a pump.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
                  COMP - a compressor.

        Immediate focus:
        PU - a pump.

        New goal-step is:
        Knowing the wrench to use.
```

Utterance (2) is a question about a substep of the pump-attaching action mentioned in Utterance (1). The goal of Utterance (2) is interpreted to be a knowledge-state goal (knowing what wrench to use). Responding to the question entails examining the model to find the tool that participates in a bolting action. In the model there is no tool directly associated with bolting. Instead there is a box-end wrench associated with a substep of bolting, tightening the bolts. The procedures that reason about actions and, in this case, the procedures that reason about "using," examine substeps of an action to find an occurrence of a tool to be used--in this instance a box-end wrench. Since there is only one box-end wrench in the model, it can be unambiguously referred to as <u>the</u> box-end wrench, which TDUS does in responding to the question.

This utterance also affects focusing. Since the bolting step is mentioned, it is focused along with the box-end wrench and the event of using the wrench. These are focused at the same level as the pump and platform.

```
#WHERE ARE THE BOLTS                                        (3)
 I don't know.

        Focus has shifted to:
          Primary focus·
             bolts
             You are using the box-end wrench.
             W4 - a box-end wrench.
             The pump is being blted.
             PL - a platform.
             PU - a pump.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
                  COMP - a compressor.

        Immediate focus:
        bolts

        New goal-step is:
        Knowing a location.

        Potential goal-step is:
        The pump is bolted to the platform with the bolts.
```

Utterance (3) is another question about a substep of attaching the pump, in this case concerning the location of the bolts used to bolt it. The direct goal expressed in this utterance is a knowledge-state goal, namely, knowing the location of the bolts. The potential goal, a domain goal, is that the pump be bolted with the bolts, a condition associated with the bolting substep in which the bolts are used. The mention of the bolts causes them to be focused.

```
#I FOUND THEM                                                   (4)
 OK


       Immediate focus:
       bolts

       Completed goal-step was:
       Knowing the location.

       Potential goal-step is:
       The pump is bolted to the platform with the bolts.
```

Utterance (4) shows fulfillment of the current knowledge-state goal: knowing the location of the bolts. Focusing remains unchanged and bolting the bolts remains a potential goal of the apprentice.

```
#WHERE IS THE WRENCH                                            (5)
 The box-end wrench is on the table.

       Focus has shifted to:
         Primary focus:
            The box-end wrench is on the table.
            T1 - a table.
            bolts
            You are using the box-end wrench.
            W4 - a box-end wrench.
            The pump is being bolted.
            PL - a platform.
            PU - a pump.
               then
                 T1 - a table.
                 PU - a pump.
                 You - a person.
                 COMP - a compressor.

       Immediate focus:
       bolts
```

Potential immediate focus:
W4 - a box-end wrench.

New goal-step is:
Knowing a location.

Potential goal-step is:
The pump is bolted to the platform with the bolts.

#I FOUND IT (6)
OK

Immediate focus:
W4 - a box end wrench.

Completed goal-step was:
Knowing the location.

Potential goal-step is:
The pump is bolted to the platform with the bolts.

In Utterance (5) the apprentice asks about the location of "the wrench." This utterance illustrates how focusing information helps disambiguate noun phrase referents. There are several wrenches in the model, so the phrase "the wrench" might be considered ambiguous. However, in Utterance (2) a particular wrench (W4) was focused through TDUS's reply and has remained focused, so the phrase "the wrench" can be interpreted as referring to a unique wrench--the box-end wrench previously mentioned and identified. Since the location of the wrench is mentioned in TDUS's reply, it also becomes focused.

The goal inferred from this utterance is "knowing the location of the wrench." In both this utterance and Utterance (2), TDUS has apparently satisfied the apprentice's knowledge-state goal by supplying the relevant information, but the knowledge-state goal is not assumed to be satisfied unless the apprentice confirms it. This is a design decision that could be changed by assuming the reply satisfied the goal or by identifying the goal as a "potentially satisfied" goal. Different choices reflect different assumptions about the coparticipant. In one case, the coparticipant is assumed to always understand; in the other, understanding is not assumed but must be explicitly confirmed.

In Utterance (6) the apprentice explicitly confirms that the wrench has been located. Thus, knowing the location of the wrench becomes an achieved goal and bolting the pump remains a potential domain goal.

In Utterance (7) the apprentice explicitly indicates completion of the attaching step, from which the system infers that all the substeps have been performed.

```
#I ATTACHED THE PUMP                                              (7)
 OK

        The following has been assumed:
        You put the pump at the platform.
        You bolted the pump to the platform with the bolts.

        Immediate focus:
        PU - a pump.

        Potential immediate focus:
        PU - a pump.

        Completed goal-step was:
        Attaching the pump to the platform.

#SHOULD I INSTALL THE PULLEY NOW                                  (8)
 No
 The next step is:

 Install the aftercooler elbow on the pump.
      or
 Install the brace on the pump.

        Focus has shifted to:
          Primary focus:
             ACE - a aftercooler elbow.
             BR - a brace.
             You attached the pump.
             The box-end wrench was on the table.
             T1 - a table.
             bolts
             You used the box-end wrench.
             W4 - a box-end wrench.
             The pump has been bolted.
             PL - a platform.
             PU - a pump.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
```

COMP - a compressor.

Immediate focus:
PU - a pump.

Potential immediate focus:
a pulley.

The question in Utterance (8) is about a possible next step--in this case the wrong one. The algorithm for identifying the action indicated by the verb phrase recognizes that the step mentioned is not a possible next step. The apprentice is so informed and is told what the possible next steps actually are.

Recognition of the impossibility of an action illustrates how the identifications of referents of noun phrases and verbs are interrelated. Since the utterance contains no pronouns, but does include a definite noun phrase, "the pulley," the first step in interpreting the action is an attempt to identify the pulley. Failing to find a pulley among focused entities, the interpretation process then scrutinizes the actions associated with the potential goal and, if necessary, the possible next steps in the task to determine if any of those actions can be referents of the verb. In this case no action matching the input is found, so the input action is not identified in the model and no referent is identified for "the pulley". Rather, the phrase "the pulley" is interpreted only as referring to some unidentifiable pulley and "install" is interpreted to mean some general type of installing action. The representation of this utterance remains essentially the one produced after Stage 2 of the interpretation, as is shown in Figure 6.

At this point many entities are focused. The ones that were previously focused remain so because no new step has been started, while the newly focused entities come from the system's reply.

#I'M DOING THE BRACE NOW (9)
OK

The following has been assumed:

```
DO is assumed to mean:
You are installing the brace.

Focus has shifted to:
  Primary focus:
     PU - a pump.
     BR - a brace.
        then
          T1 - a table.
          PU - a pump.
          You - a person.
          COMP - a compressor.

Immediate focus:
PU - a pump.

Potential immediate focus:
BR - a brace.

New goal-step is:
Installing the brace on the pump.
```

Utterance (9) shows a use of the verb "do," in this case to mean "install," referring back to TDUS's reply in which "install" was used. Since installing the brace is the action that has "the brace" as a participant, the verb phrase interpretation algorithm selects it as the referent.

Focusing now shifts to the installation step and away from the previously focused entities.

```
#I'VE DONE IT                                                    (10)
 OK

        The following has been assumed:
        You attached the installed brace to the pump.
        You put the installed brace at the pump.
        You secured the installed brace.

        DO is assumed to mean:
        You installed the brace.

        Immediate focus:
        BR - a brace.

        Completed goal-step was:
        Installing the brace on the pump.
```

In Utterance (10) "do" is used differently -- as "perform," with "it" referring to the action of installing the brace. The utterance's tense and aspect (past and perfective) imply that the action has been completed. From this TDUS infers that the substeps of positioning and securing the brace have also been completed.

Utterances (11) through (15) illustrate not only other references to steps started and completed, but also the corresponding inferences, shifts in focusing, and changes in goals.

```
#I'VE INSTALLED THE PULLEY                                    (11)
 OK

        The following has been assumed:
        You installed the woodruff key on the pump.
        You attached the woodruff key to the pump.
        You put the woodruff key at the pump.
        You secured the woodruff key.
        You attached the installed pulley to the pump.
        You put the installed pulley at the pump.
        You fastened the installed pulley to the pump with
            the screws.

        Focus has shifted to:
          Primary focus:
             You installed the installed pulley.
             PU - a pump.
             PUL - a pulley.
                then
                  T1 - a table.
                  PU - a pump.
                  You - a person.
                  COMP - a compressor.

        Potential immediate focus:
        PUL - a pulley.

#WHAT SHOULD I DO NOW                                         (12)
 Install the aftercooler elbow on the pump.

        Focus has shifted to:
          Primary focus:
             ACE - a aftercooler elbow.
             You installed the installed pulley.
             PU - a pump.
             PUL - a pulley.
                then
                  T1 - a table.
                  PU - a pump.
```

You - a person.
COMP - a compressor.

Potential immediate focus:
PUL - a pulley.

New goal-step is:
Knowing the actions.

Potential goal-step is:
Install the aftercooler elbow on the pump.

#I'VE DONE IT (13)
OK

The following has been assumed:
You attached the installed aftercooler elbow to the pump.
You put the installed aftercooler elbow at the pump.
You secured the installed aftercooler elbow.

DO is assumed to mean:
You installed the installed aftercooler elbow on the pump.

Focus has shifted to:
Primary focus:
You installed the installed aftercooler elbow on the pump.
PU - a pump.
ACE - a aftercooler elbow.
then
T1 - a table.
PU - a pump.
You - a person.
COMP - a compressor.

#SHOULD I INSTALL THE AFTERCOOLER (14)
Yes

Potential goal-step is:
Installing the aftercooler.

#I'VE DONE IT (15)
OK

The following has been assumed:
You attached the installed aftercooler to the pump.
You put the installed aftercooler at the pump.
You secured the installed aftercooler.

DO is assumed to mean:
You installed the installed aftercooler.

Focus has shifted to:
  Primary focus:
    You installed the installed aftercooler.
    PU - a pump.
    AC - a aftercooler.
      then
        T1 - a table.
        PU - a pump.
        You - a person.
        COMP - a compressor.

IV RESEARCH RESULTS

Drawing on the preceding discussion, this section examines some of the accomplishments of this research.

A. A Framework for Integrating Knowledge

A significant aspect of our research has been a continuing emphasis on the development and integration of several kinds of knowledge about language and the subject domain, including knowledge about how the domain changes. Much natural-language research to date has been concerned primarily with individual aspects of language (syntax, semantics, etc.), not with their overall integration. Furthermore, the domains for natural-language research have often been static (e.g., retrieval from databases), rather than dynamic.

The DIAMOND framework for representing knowledge about how to interpret utterances was developed at the outset of this work. It is based on an earlier framework for interpreting spoken utterances [34], [26]. Within DIAMOND, knowledge of various kinds can be easily integrated. In particular, the CONSTRUCTOR, TRANSLATOR, and INTEGRATOR procedures in DIAMOND (as described in Section II-C) provide a simple and flexible means of specifying what knowledge is to be used in interpreting particular phrase types and when during the interpretation process each kind of knowledge is to be used.

One of the first efforts integrating multiple kinds of knowledge was Winograd's work on SHRDLU [37]. This system had, in some form, many of the kinds of knowledge represented in TDUS, although the specifics (such as grammar and domain entities) were different. The most notable differences are the absence in SHRDLU of any task context provided by a model of the process to be performed and a weaker model of discourse context (Winograd relied on a linear history of past events rather than on a more structured representation that takes into account focusing in the dialog). Components of SHRDLU were closely integrated, with semantic representations being produced for each phrase as it was constructed and, when possible, referents of phrases being identified in the domain model. However, the knowledge was so closely integrated that it was hard to ascertain or specify what knowledge was most useful at particular points during processing. Our definition of interpretation stages allows us to study these interactions and evaluate the effects of alternative procedures. Konolige [20] discusses an evaluation of the roles of syntax, domain-independent semantics, and domain-dependent semantics in a system that uses DIAMOND and DIAGRAM in interpreting natural-language queries of a large data base.

The structure of the domain and of the task being performed is taken into account in such systems as SAM and PAM [32], [36]. Scripts, plans, goals, and themes provide them with an overall structure within which inputs are interpreted. However, knowledge about linguistic form and structure does not play a significant role in these systems. Here too, as in Winograd's system, the form of the integration is such that the contributions of individual kinds of knowledge are not easily discernible.

Other question-answering systems--such as LUNAR [38], TQA [24], and PHLIQA [21]--combine syntactic and semantic knowledge to form queries, but do not incorporate knowledge about the overall task and dialog structure.

B. DIAMOND/DIAGRAM a General-Purpose Natural-Language Interface

It is important to note that the DIAMOND system and much of the knowledge encoded in it, particularly the DIAGRAM grammar, constitutes a general-purpose natural-language interface that is readily adaptable to other domains, other representations of domain knowledge, and other applications.* The DIAMOND/DIAGRAM system is being used both at SRI

----

* The "domain" is the subject discussed (e.g., equipment, ships, personnel records, or an assembly task). The "application" is the way a system is used (e.g., answering questions, giving advice, or solving problems).

[22], [20]* and at USC-ISI by W. Mann, J. Moore, et al.

The development of DIAMOND/DIAGRAM into a general facility has been accomplished by clearly separating the domain- and representation-dependent information from the domain- and representation-independent information by means of basic semantic operators. DIAMOND, as a programming language, is clearly domain-independent. DIAGRAM, because it captures many generalities about linguistic structures, is also for the most part application-independent [30]. Although particular lexical items may be necessary for special domains or applications, these lexical items can be easily entered without extensive knowledge by the user about the rest of the system.

To separate domain- and representation-independent from domain- and representation-dependent information, we have developed a set of basic semantic operators [20]. These operators, written as TRANSLATORs in the DIAMOND framework and designed to work with the DIAGRAM grammar, are a set of functions that, when evaluated, build representations of concepts in the domain. The semantic operators provide a flexible, uniform interface between the linguistic analysis of an utterance and the representation of the domain's semantics. Sentences with similar meaning but differing phrase structures are mapped into a similar set of semantic operators. The translation from surface structure to calls on semantic operators thus encodes general linguistic knowledge about the way surface structure relates to meaning. This knowledge is domain-independent. To change domains or representations, one need only implement a small, well-defined set of functions.

C. Knowledge about Language and the Domain

We have identified and formalized a broad range of knowledge about language and the domain. A principal aspect of this knowledge is contextual knowledge, including knowledge about focusing, the changing goals of the participants, and the processes in the domain.

----

* Supported in part by the Defense Advanced Research Projects Agency under Contract No. N00039-78-C-0060, with the Naval Electronics Systems Command.

Knowledge about focusing, particularly how it is signaled and how it shifts in a dialog, has proven to be important in interpreting both noun phrases and verb phrases. Focusing information has been used for identifying entities referred to by noun phrases, for which purpose it was originally developed, and has played an important role in strategies for identifying actions referred to by verbs. The concept of focusing has been extended and further explored elsewhere [33], [25], [6].

The role of the speaker's plans and goals in interpreting his or her utterances has been examined in several research projects. Two recent studies are those of Wilensky [36], who has investigated the use of plans and goals in a computer system that interprets stories (see also Schank and Abelson [32]), and Allen [1], who has examined the role of the speaker's plans and goals in interpreting indirect speech acts. The importance of goals in interpreting referents of pronouns has been demonstrated by Sidner [33]. While our representation of goals is still preliminary, we have begun separating different kinds of speaker goals and developing procedures for recognizing them. We have extended research in this direction by developing mechanisms for interpreting references made by verb phrases. These mechanisms use knowledge about goals coordinated with other aspects of the context (e.g., focusing and the task model) [28].

D. Benefits Resulting from Implementation

Our research strategy has been to proceed from an examination of actual dialogs between a human apprentice and an expert to characterizing and formalizing the knowledge and procedures employed by the dialog participants, and then to encoding and testing our formalizations in a computer system. Given the current understanding of the communication process (i.e., the existence of partial theories, but of no complete theory as to how utterances are interpreted and produced), system implementations are useful for at least the following reasons: (1) they allow experimentation with the integration of many kinds of knowledge to determine the role of each and the trade-offs that become possible when different kinds of knowledge can be used for essentially the same purpose; (2) for a particular kind of knowledge, they allow the completeness and accuracy of complex interconnected sets of rules and processes to be tested; (3) in addition to revealing gaps and inaccuracies, the discipline of implementation may suggest patterns and clarify the multiple roles of particular kinds of knowledge.

The TDUS implementation yielded benefits in each of these categories. In the following paragraphs we discuss briefly three items that illustrate the usefulness of this implementation.

The development of the DIAGRAM grammar, a description of the syntax of a large portion of English [30], benefited from the implementation of TDUS. As J. Robinson discusses, DIAGRAM was influenced by the fact that it was written within the context of a larger system for interpreting English utterances. This larger context simplified separating and encoding the knowledge about language that is generally considered to be syntactic, because within the larger context it was easier to determine what knowledge should be encoded as "syntactic" and what should be encoded in other ways. Although the actual writing of the rules did not require a system, both the size of the grammar and its interconnectedness meant that the system played a critical role in testing for consistency and accuracy.

The basic semantic operators described above also resulted from experience with the system and from interaction with other projects that use DIAMOND and DIAGRAM for interpreting natural-language utterances. The need to simplify and standardize the interactions between syntactic and semantic processes led to a specification of some of the basic structural building blocks of semantic interpretation.

Finally, the TDUS environment, within which algorithms for recognizing and changing contextual elements can be developed and tested, played an important part in clarifying the roles of the domain and dialog contexts in natural-language communication. In particular, the development of the goal mechanisms, the refinement of algorithms for focus shifting, and the separation of the task and dialog context were facilitated by the implementation and subsequent testing. The critical role of goals in even the most basic aspects of interpretation became apparent during the development of algorithms for interpreting pronouns [33]. The goal mechanisms were then refined and extended as part of our research on the interpretation of verb phrases. Similarly, focus-shifting algorithms were tested with various dialog segments, which helped determine when focusing should or should not shift.

V CONCLUSION

The research described in this paper is part of a continuing research effort whose goals are to identify the types and structures of knowledge required for participation in natural-language, problem-solving dialogs, to create efficient and flexible computer-based representations of such knowledge, to develop procedural mechanisms for applying it to support linguistic interaction between humans and machines, and to articulate a high-level computational theory of natural-language communication. As part of this effort, we developed a system that participates in a natural-language dialog. This has demonstrated the feasibility of our approach and has created a powerful research tool for testing new hypotheses about the communication process.

This system and the research supporting it comprise a solid basis for further research and experimentation. Building from it, our next steps will be to incorporate the linguistic and domain knowledge required to enable consideration of underlying intentions and communication strategies. One major extension, drawing on research performed elsewhere [4], [23], will be the explicit separation of the system's knowledge and beliefs from those of the user. This separation will allow utterances to be interpreted and generated in relation to what the user is believed to know. For example, replies can be generated in terms the user is expected to understand [18], [19], [29]. Furthermore, determining the presuppositions of an utterance is of value in inferring the user's beliefs. A second major extension will be a generalization of the notion of goals and planning in communication. Included in this generalization will be three principal components: the integration of language processing with a general planning mechanism; the incorporation of plan recognition into the interpretation process;

and the generation of utterances planned so as to achieve multiple goals simultaneously. Such research is essential both for understanding human communication and for building computer systems that communicate with humans in natural language.

REFERENCES

1. Allen, J. "A plan-based approach to speech act recognition". Department of Computer Science. Ph. D. Th. University of Toronto, Toronto, Canada, 1979.

2. Appelt, D. E., Grosz, B. J., Hendrix, G. G., and Robinson, A. E. "The representation and use of process knowledge." Technical Note 207. Artificial Intelligence Center, SRI International, Menlo Park, California (1980).

3. Clark, H. H., and Marshall, C. "Definite Reference and Mutual Knowledge." In Elements of Discourse Understanding: Proc. of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting. A. K. Joshi, I. A. Sag, and B. L. Webber, eds. Cambridge University Press, Cambridge, England, 1980.

4. Cohen , P. R. "On knowing what to say: planning speech acts." Technical Report No. 118, Department of Computer Science, University of Toronto, Toronto, Canada, 1978.

5. Fikes, R. E. and Hendrix, G. G. "A network-based knowledge representation and its natural deduction system." Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, (August 22-25, 1977), pp. 235-246.

6. Grimes, J. G. "Context structure patterns." To be presented at Nobel Symposium on Text Processing, Stockholm, Sweden (August 11-15, 1980).

7. Grosz, B. J. The representation and use of focus in dialogue understanding. Technical Note No. 151, SRI International, Menlo Park, California (1977).

8. Grosz, B. J. "The representation and use of focus in a system for understanding dialogues". Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts (August 22-25, 1977) pp. 67-76.

9. Grosz, B. J. "Focusing in dialog." Proceedings of TINLAP-2, Urbana, Illinois (July 24-26, 1978).

10. Grosz, B. J. Utterance and objective: issues in natural language communication. Proceedings of the Sixth International Joint Conference on Artificial Intelligence, Tokyo, Japan (August 20-23, 1979) pp. 1067-1076.

11. Grosz, Barbara J. Focusing and description in natural language dialogues. In Elements of Discourse Understanding: Proc. of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting, A. K. Joshi, I. A. Sag, and B. L. Webber, eds. Cambridge University Press, Cambridge, England, 1980.

12. Grosz, B. J., Hendrix, G. G., and Robinson, A. E. "Using process knowledge in understanding task-oriented dialogs." Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts (August 22-25, 1977) p. 90.

13. Hendrix, G. G. "Modeling simultaneous actions and continuous processes." Artificial Intelligence, 4, (1973) pp. 145-180.

14. Hendrix, G. 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).

15. Hendrix, G. G. "Some general comments on semantic networks." Panel on Knowledge Representation, Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts (August 22-25, 1977) pp. 984-985.

16. Hendrix, G. G. "Encoding knowledge in partitioned networks." In Associative Networks--The Representation and Use of Knowledge in Computers, N. V. Findler, ed. Academic Press, New York, New York, 1979.

17. Hobbs, J. R. "Coherence in english discourse." Proceeding of the Seventh International Conference on Computational Linguistics, Bergen, Norway (August 14-18, 1978).

18. Hobbs, J. R. and Robinson, J. J. "Why do you ask?" Proceedings of Linguistic Society of America Summer Meeting, University of Illinois, Urbana-Champaign, Illinois (July 28-30, 1978).

19. Hobbs, J. R. and Robinson, J. J. "Why ask?" Discourse Processes Vol. 3. (1979) pp. 311-318.

20. Konolige, K. "A framework for a portable natural-language interface to large data bases." Technical Note No. 197, SRI International, Menlo Park, California (1979).

21. Landsbergen, S. P. J. "Syntax and formal semantics of english in PHLIQA1." In Coling 76, Preprints of the 6th International Conference on Computational Linguistics, Ottawa, Ontario, Canada (June 28 - July 2, 1976).

22. Moore, R. C. "Mechanical intelligence: research and applications." Final Report, Artificial Intelligence Center, SRI International, Menlo Park, California (1979).

23. Moore, R. C. "Reasoning about knowledge and actions." Ph.D. Th. Massachusetts Institute of Technology, Cambridge, Massachusetts. 1979.

24. Petrick, S. R. "Automatic syntactic and semantic analysis." In Proceedings of the Interdisciplinary Conference on Automatized Text Processing (Bielefeld, German Federal Republic, November 8-12, 1976). J. Petofi and S. Allen (Eds.). Reidel, Dordrecht, Holland, 1979.

25. Reichman, R. "Conversational coherency." Cognitive Science. 2:4, (1978) pp. 283-327.

26. Robinson, A. E. "Speech, language and artificial intelligence." Proceedings of the Seventh International Conference on Computational Linguistics, Bergen, Norway (August 14-18, 1978).

27. Robinson, A. E. Natural language communication. SISTM Quarterly incorporating the Brain Theory Newsletter, Vol II, No. 4, (Summer 1979).

28. Robinson, A. E. "The interpretation of verb phrases in dialogs." Technical Note 206. Artificial Intelligence Center, SRI International, Menlo Park, California (January 1980).

29. Robinson, J. J. "Theoretical foundations of linguistics and automatic text processing." Technical Note 199. Artificial Intelligence Center, SRI International, Menlo Park, California (1979). To appear in: Methodological Problems in Automatic Text Processing. J. Petofi, ed. De Gruyter, Berlin, West Germany, 1979.

30. Robinson, J. J. DIAGRAM. Technical Note No. 205. SRI International, Menlo Park, California (1980).

31. Sacerdoti, E. D. A Structure for Plans and Behavior. Elsevier North-Holland, New York, 1977.

32. Schank, R. C. and Abelson, R. P. Scripts, Plans, Goals and Understanding. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.

33. Sidner, C. . "Towards a computational theory of definite anaphora comprehension in English discourse." Ph.D Th. Massachusetts Institute of Technology, Cambridge, Massachusetts, 1979.

34. Walker, D. E., (Ed.) Understanding Spoken Language. Elsevier North-Holland, New York, 1978.

35. Webber, B. L. "A formal approach to discourse anaphora," BBN Report No. 3761. Bolt Beranek and Newman Inc. Cambridge, Massachusetts (May, 1978).

36. Wilensky, R. "Understanding goal-based stories." Ph.D. Th. Yale University, New Haven, Connecticut. 1978.

37. Winograd, T. Understanding Natural Language. Academic Press, New York, 1972.

38. Woods, W. A., Kaplan, R. M., and Nash-Webber, B. The lunar sciences natural language information system. Final Report. BBN Report 2378, Bolt Beranek and Newman, Cambridge, Massachusetts 1972.