SRI International

THE INTERPRETATION OF VERB PHRASES IN DIALOGS

Technical Note 206

January 31, 1980

By: Ann E. Robinson, Computer Scientist

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.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

ABSTRACT

This paper discusses two problems central to the interpretation of utterances: determining the relationship between actions described in an utterance and events in the world, and inferring the "state of the world" from utterances. Knowledge of the language, knowledge about the general subject being discussed, and knowledge about the current situation are all necessary for this. The problem of determining an action referred to by a verb phrase is analogous to the problem of determining the object referred to by a noun phrase.

This paper presents an approach to the problems of verb phrase resolution in which knowledge about language, the problem domain, and the dialog itself is combined to interpret such references. Presented and discussed are the kinds of knowledge necessary for interpreting references to actions, as well as algorithms for using that knowledge in interpreting dialog utterances about ongoing tasks and for drawing inferences about the task situation that are based on a given interpretation.

I INTRODUCTION

This paper discusses two problems central to the interpretation of utterances: determining the relationship between actions described in an utterance and events in the world, and inferring the "state of the world" from utterances.* Knowledge of the language, knowledge about the general subject being discussed, and knowledge about the current situation are all necessary for this. The problem of determining an action referred to by a verb phrase is analogous to the problem of determining the object referred to by a noun phrase. Although considerable attention has been given to the latter (Donellan, 1977; Grosz, 1977a, 1977b; Sidner, 1979; Webber, 1979), little has been done with the former.**

The need to identify an action is obvious in utterances containing verbs like "do", "have", and "use", as in "I've done it", "what tool should I use?", or "I have it". In these utterances the verb does not name the action, but rather refers to it more generally, much as pronouns or "nonspecific" nouns (e.g., "thing") refer to objects. Even when more specific verbs are used, complex reasoning may be required to ascertain the particular action being referred to. For example, the utterance "I've glued the pieces together" can refer to different steps in a task--depending on what objects "the pieces" refers to, because each gluing action is a different step in the task. Similarly, the verb "cut" refers to different types of cutting actions when used with

---

* This research has been funded under three-year NSF Continuing Research Grant No. MCS76-22004. This paper and the research reported in it has benefited from interactions with all the members of the natural language research group at SRI. Barbara Grosz, Jerry Hobbs, Gary Hendrix, and Jane Robinson have been particularly helpful in the preparation of the paper.

** A problem related to identifying verb phrase referents--interpreting verb phrase ellipsis--has been investigated by Webber (1979).

different objects, as in "cut grass", "cut wood", or "cut cake" (Searle, 1978).

A variant of this problem is deciding whether a verb is intended to refer to a general or a specific action. For example, "cutting wood" can refer to the general activity of cutting many pieces of wood or it can refer to the action of cutting a particular piece. (Werner, 1966)

This paper presents an approach to these problems in which knowledge about language, the problem domain, and the dialog itself is combined to interpret references by verbs. Presented and discussed are the kinds of knowledge necessary for interpreting references to actions, as well as algorithms for using that knowledge in interpreting dialog utterances about ongoing tasks and for drawing inferences about the task situation that are based on a given interpretation. The algorithms have been implemented and tested in a computer system (TDUS) that participates in a dialog about the assembly of an air compressor (Robinson, 1979). The system acts as an expert, guiding an apprentice through the steps of the task. The knowledge available will be described first, followed by a discussion of a sample dialog in which the system participated, then by a detailed description of the algorithms for verb interpretation.

II KNOWLEDGE NEEDED

Interpreting any utterance and relating it to a task requires knowledge about the language and the task, as well as the relationships between them. This paper will concentrate on knowledge needed to identify actions. It builds directly on the concepts of global and immediate focusing, through which certain entities are highlighted (Grosz, 1977a, 1977b, 1978; Sidner, 1979). General familiarity with that research will be assumed. More detailed descriptions of other aspects of the knowledge needed for interpreting utterances can be found elsewhere (Grosz, 1977a; Hendrix, 1977, 1979; Robinson, 1979; J. Robinson, 1980).

A. Actions and Events

Interpreting verbs requires knowing about events that have occurred, are occurring, or can occur in the domain. Such knowledge typically includes the steps necessary to perform the actions associated with the events, the possible participants, the conditions that must be true before the actions can be performed, and their results (e.g., the goals they achieve or their possible side effects). Knowledge about actions and events includes both general knowledge about possible actions and events and more specific knowledge about those that occur during a particular task.

We have developed a formalism, process models, for encoding information about actions (Grosz et al., 1977; Appelt et al., 1980). This formalism enables the specification of a hierarchical decomposition of actions into subactions, as well as the description of individual types of actions. It is an extension of the network formalism used for representing other knowledge about objects and relationships, as described by Hendrix (1979). The description of each action 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 followed to accomplish it. A sequence of substeps may be partially ordered. This decomposition of actions builds upon earlier research on the hierarchical decomposition of the planning process (Sacerdoti, 1977) and upon the work by Hendrix (1973, 1975) on modeling actions and processes. Many of the actions in the assembly task domain have been encoded in this formalism for use in the TDUS system.

Figure 1 illustrates a process model for the pump-attaching process in the assembly domain. The network node ATTACH PUMP represents the set of pump-attaching actions. The large box depicts a separate space in the network in which the schema of the ATTACH PUMP action is represented. This schema specifies the participants in the attach operation, marked by the MAJORPART, MINORPART, and AGENT arcs, and describes the action of attaching. The action 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. Each step in the plot (encoded on a separate space) is in turn further described by a process model. In this example, the substeps of attaching are positioning and bolting the pump.

During a task, a record of progress is kept by filling in, or instantiating, the schema for an action as that action is performed and then incorporating the newly created piece of network 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 instantiated action that is represented has associated with it a time interval. The interval can be past, present, or future, and it can be bounded by two times: a start time and an end time. For events 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.

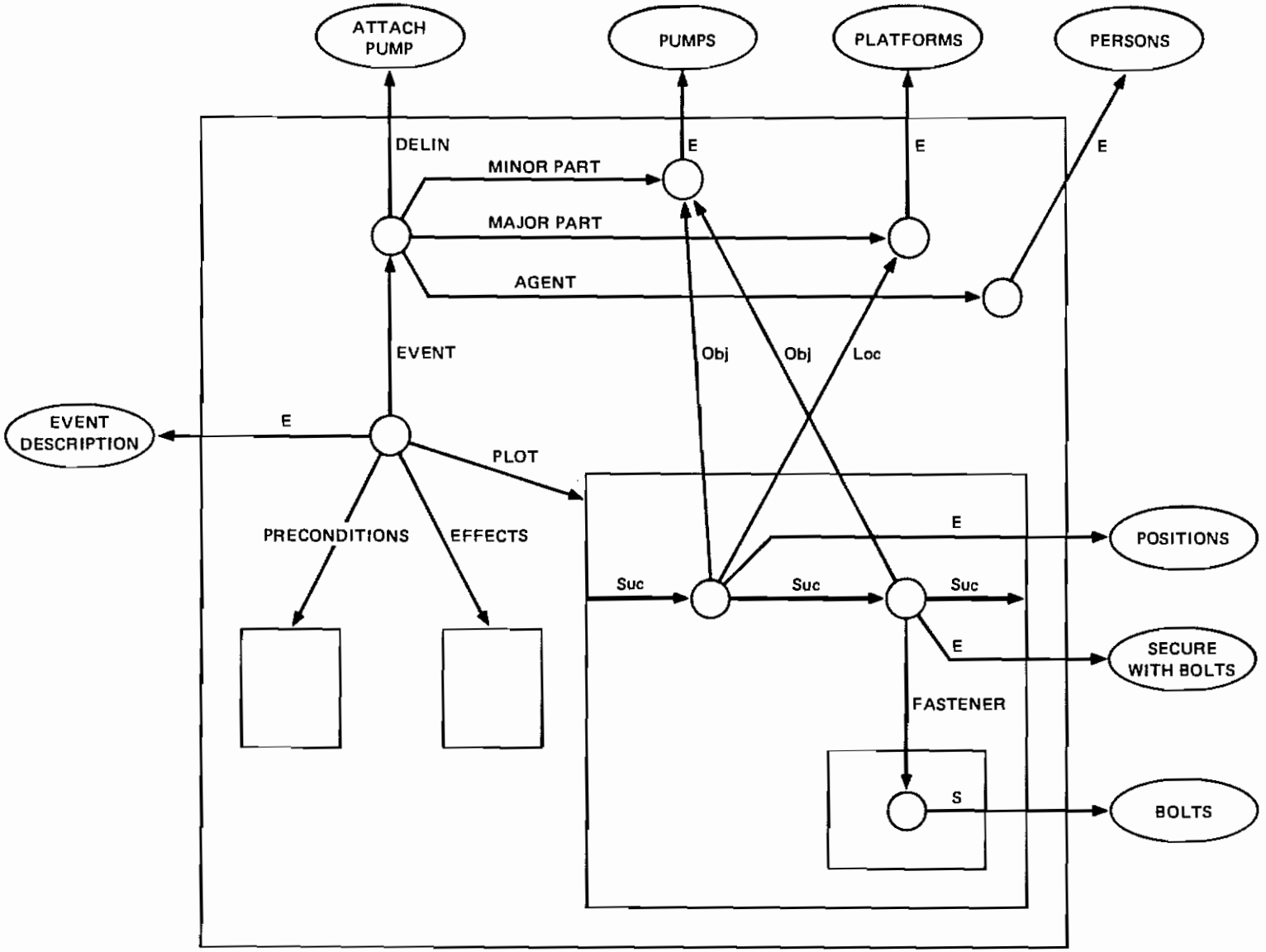

Figure 1 An Attaching Process

Once an instance of an event is recorded in the domain model, it can be used in subsequent deductions and is available for answering questions about past events. This provides a means of maintaining an up-to-date record of assembly progress. Such a record comprises 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.

We have developed procedures encoding specific knowledge about how to manipulate process models. These are reasoning procedures supplementing and dependent upon those that embody general knowledge about logical deduction (Fikes and Hendrix, 1977). These new procedures for reasoning about processes are used to retrieve or deduce answers to questions (that might be posed either by some process during the interpretation of an utterance or by a coparticipant in a dialog) and to assimilate into the model new domain information communicated by the user (e.g., the user may indicate that certain actions have been performed or that new conditions have arisen). Such new facts must be assimilated in a way that enables derivation of their logical consequences.

B. Goals

Related to knowledge about current actions is knowledge about the goals of the dialog participants. The goals that are expressed or implied by an utterance can be of many types. These include "domain goals" related to the subject domain; "knowledge-state goals" related to 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 the interpersonal relationships of the dialog participants.*

---

* The distinction between domain and knowledge-state goals was drawn by Appelt (1979).

Each dialog participant can be viewed as having individual domain, knowledge-state, and social goals. Many of these goals may be shared by the participants, particularly in cooperative dialogs, but the participants' goals can also be conflicting or complementary. Interesting insights into, and methods for, recognizing and expressing separate goals, as well as for taking this separation into account in interpreting and generating utterances, have been discussed elsewhere (Cohen, 1978; Allen, 1979; Clark & Marshall, 1980; Wilensky, 1978).

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 (Grosz, 1979). For example, an utterance can simultaneously request information (a knowledge-state goal) and convey politeness.

The goals of dialog participants affect the interpretation of verbs in at least two ways: (1) interpreting verbs entails recognizing the speaker's goals as expressed or implied by the utterance; (2) current goals are a part of the context within which verbs are interpreted. In this section we will discuss the goals currently recognized in TDUS, how to recognize whether a goal is current or achieved, and how goals are represented.* In the following sections, we will see how these goals are used for interpreting verbs.

1. Recognizing Goals in TDUS

The TDUS system handles two kinds of goals: domain goals and certain knowledge-state goals. Domain goals concern states to be achieved by task-related actions in the domain, while knowledge-state goals concern states to be achieved by acquiring a specific piece of information.

Figure 2 illustrates the relationship between actions and goals. The hierarchy shown is a simplification of a portion of the

---

* The current implementation of goals in TDUS is an extension and partial revision of one by Sidner described in her dissertation (1979).

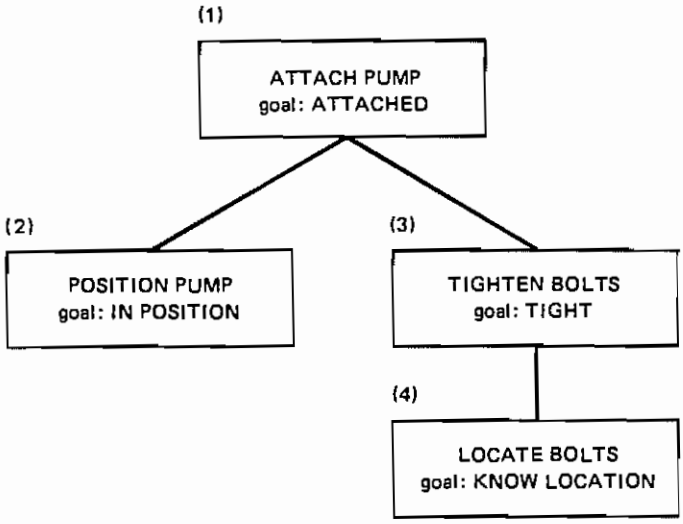

Figure 2 Goal/Action Tree

assembly task hierarchy currently encoded in TDUS.* Each node 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), represents the action of attaching a pump whose associated goal is that the pump be attached. Nodes (2) and (3) represent substeps of this attaching process--the actions of positioning the pump and tightening the bolts, with the associated goals that the pump be positioned and that the bolts be tight. The action of locating bolts represented by node (4) is not an explicit step in the task, but is necessary for its performance. Node (4) has an associated knowledge-state goal: "know the location of the

---

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

bolts". All these goals have associated actions that, in the process model formalism, are specific instantiations of actions, not action schemata.

We distinguish two classes of goals: "direct goals" achieved by actions the apprentice has explicitly or implicitly said are being performed now or have been performed and "potential goals" mentioned by either participant that have not been acted upon but might possibly be. In the context of the task steps shown in Figure 2, "I am attaching the pump" states that the speaker is performing an attaching action, as represented by node (1). Thus the utterance establishes that the pump be 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 that the bolts be tight as a potential goal.

A knowledge-state goal can also be a direct goal. For example, the goal associated with node (4), "know the location of the bolts", can be introduced by utterances like "where are the bolts?" In the current implementation, knowledge-state goals cannot be potential goals. This limitation arises primarily because knowledge-state goals have not been as fully explored as task goals. Clearly, potential knowledge-state goals can occur and there is no reason they cannot be recognized. For example, the utterance "I'd like to read more Plato" implies the knowledge-state goal of knowing more about the philosophy of Plato.

Direct and potential goals are distinguished from one another because of the different roles they play in the interpretation of verbs. Basically, direct goals are those that are known as existing or former goals associated with actions that are being or have been performed. Potential goals are possible near-term goals associated with possible future actions. Depending on the type of utterance, one or the other class of goal might be considered first. The different roles of the two goal classes will be illustrated when the interpretation of verbs is discussed in detail in Section IV.

In the TDUS system, a potential goal can be introduced either by the apprentice who is performing the task or by the system which is acting as an expert advisor. Such a goal can be introduced in at least three different ways. (1) The apprentice can introduce a potential goal by mentioning a possible future action, while not explicitly stating that it will be performed. This distinguishes between "I am going to take the lid off now" and "should I take the lid off now?" The former expresses a direct goal because the speaker explicitly says s/he is planning to perform the action. The latter expresses a potential goal because the speaker has not made a commitment to performing the action, but implies that s/he might. When a potential action is mentioned in this way, if it is an appropriate next step in the task the system will establish the associated goal as a potential goal. For example,

"Should I tighten the bolts now?"

will cause establishment of the potential goal "that the bolts be tight" if the appropriate reply is "yes".

(2) The expert can introduce a potential goal by telling the apprentice what actions to perform. The goal is potential and not direct, because the expert cannot, on the basis of the utterance alone, infer that the apprentice will perform the action. For example, the expert's reply to

"What should I do now?"

will cause establishment of the potential goal -- or goals if there are multiple possibilities -- associated with the action in the reply.

(3) The apprentice can also introduce a potential goal by indirectly mentioning an action in the task. For example, if the apprentice says

"I found the pulley."

in a situation in which one of the next steps is to install the pulley, but neither the installation nor the pulley has been mentioned before, the potential goal "that the pulley be installed" will be inferred from the reference to the pulley and the knowledge that it is a possible next step. This forward reference to an object implicitly focuses the object

and the step it is associated with. Previously, algorithms for shifting focus caused a shift to the step associated with the object (Grosz, 1977b). However, this is problematic because the speaker may not intend to perform the step or even discuss it, but rather intends to talk about the object. Establishing the step in which the object participates as a potential goal highlights the step but does not force a shift of focus to it. This change has proved to be important, as will be seen during discussion of the algorithm.

Utterances can introduce direct and potential goals simultaneously. In the examples above showing potential goals, direct knowledge-state goals are also being introduced in cases (1) and (2). In particular, the knowledge-state goals are 'knowing whether tightening the bolts is the next step' and 'knowing the action to perform'.

2. Recognizing the State of a Goal

Besides recognizing a goal, it is necessary to recognize whether the goal is the current one, one that has already been achieved, or one that has been abandoned. It is also necessary to recognize when goals are no longer potential.

A direct goal is assumed to be current when an utterance states that an action that will achieve the goal is in progress. A goal is assumed to have been achieved either

(1) when an explicit statement such as "I have attached it" or "I'm done" or "OK"* indicates the completion of the action achieving the goal;

(2) when an explicit statement indicates an action intended to achieve the goal is finished; or

(3) when the start of a new action implies completion of the current one and thus achievement of the associated goal.**

An utterance such as "never mind" is interpreted as indicating that a goal has been abandoned.

* See the discussion in Grosz(1977a) of the roles of OK.

** As Sidner(1979) points out, in the first two cases the information comes from the utterance, while in the third case it is from the task model.

Potential goals are not achievable as such. Rather, they can either become direct goals through the mechanisms for establishing direct goals or they can disappear when a new potential goal is recognized.

3. Representing Goals in TDUS

The structure of goals in a dialog about a task is related both to the structure of the task and to the structure of the dialog. The structure of tasks and the structure of dialogs have been discussed elsewhere (Grimes, 1980; Grosz, 1977a, 1977b, 1978; Hobbs, 1978; Reichman, 1978; Sacerdoti, 1977; Sidner, 1979; Wilensky, 1978). Open questions remain about the structure of the goals that arise and how they should be represented.

In TDUS direct goals are represented in a single list, acting like a first-in-last-out stack. Both knowledge-state and task goals are entered on the same list. This simplification has proved adequate for current purposes.

In general, there can be only one potential goal at a time. The exception is when two possible actions are introduced at once, as in "install the aftercooler or install the brace". Because it is simplest to view a potential goal as a single item, hereafter references to the potential goal can be read as referring to the possible conjunction or disjunction of potential goals when appropriate.

C. Knowledge about Language

To interpret verbs and infer the current task and dialog situation, the domain knowledge outlined above must be combined with knowledge about the language including what is generally characterized as syntactic, semantic, and discourse knowledge.

1. Syntactic Knowledge

One of the most important elements of syntactic knowledge necessary for interpreting verbs--and the one discussed here--is knowledge about tense and aspect. Tense and aspect are used to indicate the relative time of an event and whether it is or was occurring or completed.

Tense and aspect are indicated syntactically by auxiliaries and/or certain verb forms. In TDUS utterances are analyzed and marked for tense (past, present, or future) and for imperfective (event in progress) and perfective (event completed) aspect.

Figure 3 gives examples of some of the verb forms TDUS can interpret along with their tense and aspect markings.

I am going.	present, progressive
I had gone.	past, perfective
I had been going.	past, progressive, perfect
I will be going.	future, progressive

Figure 3 Examples of Tense and Aspect

So far, we have considered primarily verbs that refer to events rather than states, and to the usage that is most common in dialogs about tasks, such as references to single occurrences of actions. However, the analysis and representation are compatible with analyses that consider other kinds of usage (Leech, 1976). With suitable extensions we should be able to account for such things as repeated actions ("he walks to work daily") and actions that occur over a period of time ("she is learning Spanish").

2. Semantic Knowledge

The interpretation of references to actions and events requires knowledge of the relationship between words for actions or events and the internal representation of that class of action or event;* it also requires knowledge of the relationship between nouns and

* Note that all that can be known when a dialog starts is the relation between words and classes of concepts. The problem addressed in this paper is that of identifying the particular action or event referenced in a particular utterance.

entities in the domain. For example, the 'SELLING' action is an action whose participants include a buyer, seller, some object being sold, and some money. In an utterance whose main verb is "sell" in the active voice, the syntactic subject of the utterance is the "seller" in a selling event, the syntactic object is the item sold, the indirect object is the one to whom the item is sold, and the object of the 'for' preposition is the selling price. The information necessary to make this mapping and to build the appropriate representation is encoded with the verb. (Hendrix in Walker, 1978; Konolige, 1979).

3. Discourse Knowledge

Discourse knowledge is knowledge about how the domain and dialog contexts in which an utterance occurs contribute to and are influenced by the interpretation of the utterance.

a. Focusing

During a dialog, the participants focus their attention on only a small portion of what each of them knows or believes. Both what is said and how it is interpreted depend on a shared understanding of this narrowing of attention to a small highlighted portion of what is known.

Focusing is an active process. As a dialog progresses, the participants continually shift their focus and thus form an evolving context within 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 on, how to focus on it, and how wide or narrow the focusing should be. We have developed a representation for discourse focusing (or global focusing), procedures for using it in identifying objects referred to by noun phrases, and procedures for detecting and representing shifts in focusing (Grosz, 1977a, 1977b, 1978, 1980).

In addition to global focusing, we have incorporated the concept of immediate focus (Sidner, 1979) through which a single discourse entity is singled out. This is a more localized focusing

phenomenon that is closely related to the use and recognition of anaphora, as well as to changes in global focusing.

b. Shared and Joint Knowledge

In our framework, the dialog participants are assumed to share knowledge about processes in the task model* and the history of the task performance to date, along with knowledge about direct and potential goals and focused entities. We view this shared knowledge as composed of at least two parts: (1) the processes in the task model and the history of its performance comprise knowledge about the world that is assumed to be shared by the participants independently of the dialog; (2) knowledge about the goals and focusing, which is assumed to be shared as a result of the dialog. We will distinguish these two types of shared knowledge and their roles in the interpretation of utterances and use the terms shared and joint to refer to them.

We use shared knowledge to refer to what is known by both participants because of their common background and experiences, and is assumed to be shared, but has not been explicitly discussed by them. This includes knowledge of the language and the domain.

We use joint knowledge to refer to what has been explicitly communicated between the dialog participants. The steps of the task that are explicitly mentioned are joint knowledge, as are other focused entities that have been mentioned. Since we are considering dialogs in which the only mode of communication is verbal (there is, for example, no shared visual context) only what actually is said is assumed to be known jointly.

This analysis distinguishes as "joint knowledge" essentially what Clark and Marshall (1980) distinguish as the mutual knowledge that results from "linguistic co-presence." Our use of the term "shared knowledge" covers the mutual knowledge they describe as

---

* Note that the apprentice knows neither all the steps in the task nor their ordering--otherwise there would be no need for the expert. However, the apprentice does know how to perform most of the basic actions, such as bolting and tightening.

resulting from "cultural co-presence" and a limited form of "physical co-presence".*

To help clarify our distinction between shared and joint knowledge, consider a dialog about building a particular wooden structure. The dialog participants share knowledge about actions that are used in building (cutting wood, hammering nails, measuring), about building materials (lumber, nails, screws), about tools (saws, hammers, chisels), and about terminology for talking about them. All this is shared at the beginning of the dialog. During the dialog some things become joint knowledge. Consider the following exchange between an expert (E) and an apprentice (A):

E: First, cut the planks for the base.
A: How many and how long?
E: Cut 4 planks, 5 feet long.
A: OK.
A: They're cut.

Shared knowledge here includes knowing about cutting boards, knowing what planks are, and how to measure lumber. Following the expert's first utterance it has become joint knowledge that the first step is to cut some planks and that cutting them is a potential goal of the apprentice. The expert's second utterance makes joint the fact that 4 planks should be cut to the length of 5 feet each. The apprentice's response then makes joint the knowledge that the cutting has taken place. The fact that the boards were found and measured by the apprentice can be assumed by the expert, drawing on knowledge of the task. Since these actions were not mentioned, they are shared knowledge but not joint.

Assumptions about things that are jointly known play a critical role in the interpretation and production of utterances (Clark and Marshall, 1980), as the use of anaphora illustrates. Pronouns and pro-verbs (when used felicitiously) always refer to jointly known concepts, so that any utterance containing a pronoun or pro-verb must

---

* Physical copresence is limited by the sensory constraints of the computer system. The system can assume that both it and the apprentice are aware of the physical situation, but it can verify its assumptions only on the basis of the apprentice's actual utterances in the dialog.

drawn upon joint knowledge. In the example above, if the apprentice's second utterance had been "I'm cutting them now" followed by "I've done it", the "it" could have referred only to the cutting step, which is jointly known, not to any substep, such as measuring the board, which is not joint knowledge.

A similar observation about the use of anaphora has been made by Hankamer and Sag (1976). They differentiate the linguistic and nonlinguistic components of joint knowledge, using the term 'pragmatic environment' to refer to the nonlinguistic environment--which, as we pointed out, is limited in our situation. Hankamer and Sag state that "the conditions on insertion (and interpretation) are that the speaker presumes the content of the anaphor to be recoverable, either from linguistic context (in which case the anaphor has an 'antecedent' in linguistic structure, a fully specified linguistic form with the same semantic content) or from the pragmatic environment." (Pg. 422). The algorithms we have developed for interpreting verbs draw on these observations and distinguish between utterances containing and not containing anaphora, relying more heavily on joint knowledge when anaphora is present.

Jointly known entities can be referred to by anaphora, but they are not always, as is demonstrated by the use of definite noun phrases to refer to focused objects. In the foregoing example, the planks are focused and are thus part of joint knowledge after the expert's first utterance--but when the expert refers to them the second time, a noun phrase is used instead of a pronoun. The 'degree' of focusing, which influences the choice of anaphora or a definite noun phrase to refer to some jointly known entity, has been discussed elsewhere (Sidner, 1979; Grosz, 1977b; Reichman, 1978).

When referring to something not assumed to be jointly known, a speaker not only cannot use anaphora, but must draw on other shared knowledge and supply enough information to enable the hearer to interpret the reference correctly. In our example, if the apprentice had asked where to find the planks, the expert could have said "beside

the toolshed", assuming the apprentice was generally familiar with the surroundings and knew where the toolshed was. The expert could not have said "beside it" unless the toolshed had already been mentioned and comprised a highly focused part of joint knowledge.

III SAMPLE DIALOG

Before discussing the interpretation algorithm, we present a sample dialog in which the TDUS system was one of the participants. This dialog illustrates some utterances that can be interpreted and responded to, the goals that are inferred, and the inferences that are drawn about the task. The apprentice's utterances are preceded by the symbol '#' and numbered for purposes of discussion. The rest of the dialog was generated by the system acting as an expert.

In the initial context for this dialog, the next step to be performed is to install the pump. The first step in installing the pump is the pump-attaching step illustrated in Figure 1. At the outset, the table (T1), the pump (PU), the apprentice (you), and the compressor (COMP) are focused.

```
#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.
```

In utterance (1), the apprentice indicates the start of the pump-attaching action. This is assumed to be the current goal-step, as the commentary indicates. The utterance also causes focusing to shift to the pump (PU) and the platform (PL), with the pump (PU) as the expected immediate focus.* The hierarchy of focused entities (Grosz, 1977a, 1977b) is illustrated in the two levels of focusing shown here. "Primary focus" indicates the most highly focused entities, with the other objects T1, PU, You, and COMP at the next level in the hierarchy. Because the pump is explicitly mentioned in utterance (1) it appears in both places in the focusing hierarchy.

Utterance (2) is a question about a substep of the attaching action. The goal is interpreted as a knowledge-state goal--knowing what wrench to use. This goal is added to the stack of direct goals as the most recent goal.

---

* See Sidner(1979) for a discussion of expected and potential immediate focus.

```
#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 bolted.
             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:
        Bolting the pump to the platform with the bolts.

#I FOUND THEM                                                 (4)
OK

        Immediate focus:
        bolts

        Completed goal-step was:
        Knowing the location.

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

Utterance (3) is another question about a substep, in this case the location of the bolts used for bolting the pump. The direct goal is a knowledge-state goal, to know the location of the bolts; it is placed on the stack atop the goal from utterance (2). The potential goal, a domain goal, is that the bolts be bolted; this is the goal associated with the bolting substep in which the bolts are used. It replaces the previous potential goal.

Utterance (4) shows satisfaction of the goal of knowing the location of the bolts, which is removed from the stack of direct goals.

```
#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:
        Bolting the pump 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:
        Bolting the pump 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 was focused by the expert's reply and has remained focused, so the phrase "the wrench" can be interpreted as referring to a unique wrench -- the particular box-end wrench previously mentioned and identified. The goal inferred from utterance (5) 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 TDUS does not assume that the knowledge-state goal will 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 distinguishing the goal as one that has been potentially-satisfied. Different choices reflect different assumptions about the other participant. In one case, it is assumed that the coparticipant understands, whereas in the other case, such understanding is not assumed, but must be explicitly confirmed.

In utterance (6) the apprentice explicitly indicates the completion of the attaching step, from which the system infers that the substeps shown in Figure 1 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.

#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.

#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.
```

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 told this and, in addition, is told the possible next steps.

Utterance (9) shows the verb "do" used in this case to mean "install"--referring back to the expert's reply, in which "install" was used.

In utterance (10) "do" is used differently--as "perform" with "it" referring to the brace-installing action, which is inferred to have been completed. TDUS also infers completion of the substeps of installing the brace, putting it on the pump, and securing it.

Utterances (11) through (15) illustrate other references to steps started and/or completed and the corresponding inferences, focusing shifts, and goal changes.

```
#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 INTERPRETING VERBS

In this section we address issues that arise in applying domain and linguistic knowledge to interpret verbs and to infer the current situation on the basis of the interpretation.

The possible referents of a verb phrase are constrained by both the context and the utterance itself. Coordination of the constraints is necessary for interpreting verbs in a computer system.

Contextual constraints are derived from two sources: the domain and the dialog. Knowledge about the domain and, in particular, the task being performed, is part of the knowledge shared by the participants at the beginning of the dialog, including knowledge as to which actions can be performed, how to perform them, and when. The dialog provides knowledge about the actual progress of the task; it causes certain entities to be focused, as well as providing information about the goals of the participants. This knowledge is the joint knowledge we described previously.

Utterance constraints include tense and aspect information and the type of action denoted by the verb. The tense and aspect of the utterance restrict the alternatives within the task model and limit the goals that might be considered as referents. Generally, present tense and progressive aspect are used when referring to a new action, indicating that it has been started. Only if the utterance is somehow marked, as in "I'm still tightening the bolts", will the reference be to an action that already has been mentioned as in progress. Consequently, when TDUS is interpreting a present and progressive utterance, the actions considered in the task model are those closely related to the most recent action performed. The only goal considered is the potential goal since a direct goal is associated with an action already under way.

Past tense and/or perfect aspect indicate that an action has been finished. However, the hearer may or may not have known that the action was in progress. Consequently, the actions known to have been in progress and those that can be next steps are possible referents, as are actions associated with all direct goals and the potential goal.

The search for the referent of a verb can be conducted either top-down or bottom-up. The top-down search uses contextual constraints to find the place in the task that the utterance fits and utterance constraints to limit alternatives. The bottom-up mode uses information from the utterance, such as verb type, to find its relationship to the task. If the top-down search is successful, the action and its place in the task are identified simultaneously.

In the current domain, in which all the utterances are directly related to the task and in which the system has already encoded all the relevant steps to be performed, top-down constraints are strong enough to allow a top-down search to be conducted first--and only if that fails is a bottom-up search conducted. In a domain where there is less structure provided by the task, a bottom-up search will clearly play a more central role. This search can be improved by doing more extensive reasoning based on the verb in the utterance. For such domains, we have been examining what other linkage between actions should be introduced.

One of the major limitations of previous natural-language systems has been a lack of coordination of the strategies for identifying referents of noun phrases and pronouns with one another or with the interpretation of the verb. In fact, except for the pronoun resolution procedure that used a very simple goal recognition algorithm (Sidner, 1979), the verb phrase was not even taken into account. However, since the interpretation of each of these utterance elements cannot be carried out in isolation, the previous strategies have been modified and now the procedures for identifying noun phrase and pronoun referents are coordinated with the search for the verb phrase referent. Details of this modified strategy will be discussed in conjunction with the details of the verb phrase strategy.

A. The Top-down Algorithm

Different types of utterances can draw upon different contextual constraints. Three major factors are considered by the interpretation algorithm in determining which contextual constraints to draw upon. The factors are (1) whether or not a pronoun is present in the utterance; (2) whether or not all the noun phrases in the utterance refer to focused entities; (3) whether or not the main verb is "do". The presence of a pronoun indicates that joint knowledge, particularly goals and immediate focus, is being drawn upon. If no pronoun is present, other factors weigh more heavily in determining constraints. When all the definite noun phrases refer to focused entities, focusing information is also key in interpreting the verb. If the referents are not focused, knowledge about the task and its structure must be used. When "do" appears as the main verb, joint knowledge plays a more central role than when other verbs are used. The particular usage of "do", as signalled by the other constituents, indicates which aspects of joint knowledge are most important.

We will discuss the interpretation algorithm by examining the interpretation of utterances resulting from various combinations of these factors. The utterances we will discuss are those containing the verb "do", those containing verbs other than "do" and pronouns, and those containing verbs other than "do" and definite noun phrases.

Within the first type of utterances, those containing "do", we further distinguish utterances like "I've done it" from utterances like "I've done the screws." In the former, "do" refers to the general action of performing an action and "it" refers to the action. In the latter, "do" refers to a particular action, such as remove. Our discussion will first cover these two types of utterances containing "do", then utterances with other verbs and pronouns, then utterances with other verbs and definite noun phrases.

1. Do and Pronouns

In interpreting verb phrases such as "do it", knowledge about the context is used first to determine possible referents. If "it" has been used felicitously, it must refer to an action jointly known by the dialog participants. As we have discussed, joint knowledge in TDUS is represented by goals and focusing. Goals are a subset of all focused entities and, by definition, those actions that could possibly be performed by the apprentice. Consequently, possible referents are contained in the subset of joint knowledge represented by the most current direct goals and by the potential goal.

The main utterance constraints are derived from the tense and aspect, which, as we have observed, limit the goals whose associated actions could be referents. The three cases we distinguish are past tense, present tense and progressive aspect, and future tense.

As we have discussed, direct and potential goals can be referred to in a past-tense utterance. For such utterances, the algorithm examines the most recent direct goal first. If it is associated with a domain action (i.e., not a knowledge-state goal), the action is taken to be the referent of "it" because it is the action known to be in progress. Utterance (10) (repeated below) illustrates such a reference to a task goal.

A: I'm doing the brace now. (9)
E: OK
A: I've done it. (10)

Here "it" refers to the action of installing the brace, the action associated with the current goal.

Because of restrictions in our current implementation, the most recent direct goal is not considered as a referent if it is a knowledge-state goal. Instead, the action associated with the potential goal is taken to be the one referred to since it is always a domain action.* Clearly, if potential goals were extended to include knowledge-

---

* This is clearly a limitation that should be removed as linguistic and representational capabilities improve. An example of "it" referring to a knowledge-state goal would be "I wanted to learn Spanish and I've done it", where the goal was a knowledge-state goal of 'KNOWING SPANISH'.

state goals, a more sophisticated test would be required.

Utterances (12) through (15) illustrate reference to a potential goal.

A: What should I do now (12)
E: Install the aftercooler elbow on the pump.
A: I've done it (13)
E: OK
A: Should I install the aftercooler (14)
E: yes
A: I've done it (15)

The apprentice's utterance (12) establishes a direct knowledge-state goal of knowing what action to perform, while the expert's reply establishes a potential goal that the aftercooler elbow be installed. Utterance (13) refers to the potential goal. Utterance (14) similarly establishes a direct knowledge-state goal of knowing about the action--in this case, whether the action is installing the aftercooler; here the apprentice's utterance establishes the potential goal that the aftercooler be installed. Utterance (15) refers again to the potential goal.

An utterance that is present-tense and progressive (e.g., "I'm doing it") refers to an action that has been previously mentioned but only just started. As we have seen, a potential goal is associated with such an action, so that the latter is taken as the referent. For example, utterance (15) could have been "I'm doing it", referring to the action of installing the aftercooler.

For a question referring to a future or a hypothetical action (e.g., "What should I do now?"), no attempt is made to identify the action as part of the interpretation. Instead, the reasoning process makes use of the task model to identify the appropriate reply.

2. Do and Definite Noun Phrases

For the other use of "do", (e.g., "I'm doing the screws") where "do" refers to an action, an action of that type must be part of joint knowledge. However, only the action type may be jointly known and not the specific action referred to. For example in the sequence:

"I've attached the pump."
"I'm doing the pulley now."

the first utterance makes joint the attaching action for the pump. In the second utterance, "do" refers to another attaching action, but this one is attaching the pulley, a separate action. "Do" is not referring to the same specific action, but rather to the same type of action, "attaching".

To interpret such utterances, the contextual knowledge used is joint knowledge and knowledge about the task. The joint knowledge used is focusing information, because an action of the same type as the one referred to should be focused.* The interpretation algorithm examines focused actions for one that is of a type capable of having the newly mentioned participating objects. For example, the algorithm might find "attach pump" as a focused action, determine that it is an "attach" and then that a pulley can also participate in an "attach" action. If an action is found, task knowledge is used to determine if an action of that type with the participants indicated is an appropriate action in the current situation. Thus, if attach + pulley is an appropriate action, "attach pulley" is taken as the referent of "do".

Tense and aspect information from the utterance help determine which actions in the task model are appropriate. As we noted, a present-progressive utterance indicates initiation of a new step, whereas the past tense could be used either with a new step or with one in progress.

Utterances (8)-(9) illustrate a related situation. Here two steps have been mentioned and are essentially equally focused and both potential goals, so "do it" could not refer unambiguously to one of the actions. However, both actions are "install" actions, so "do" can refer to an "install" type action. The interpretation algorithm outlined above works for this case as well.

---

* Goal information could be used by examining the types of the actions associated with domain goals. However, access to the action type is more direct through focusing information.

3. Pronouns with Verbs Other Than Do

For utterances containing verbs other than "do" and pronouns, contextual constraints also stem from joint knowledge, since the object or objects referred to by the pronoun must be joint knowledge--in our case, mentioned in the dialog. The way the referent of the pronoun was introduced into the dialog affects the interpretation of utterances with pronouns. The distinction we make is whether the object was mentioned as a participant in an action that is part of the task, (e.g., "I attached the pump.") or was not mentioned as a participant in an action (e.g., "Where is the pump?"). In the first case, if the object has been mentioned as participating in an action, the action will be recognized as a direct or potential goal and all its participating objects will be focused. In the second case, if no action has been mentioned but the object is a participant in some task action, the action will be inferred through the potential-goal recognition mechanism and become a potential goal. However, in this case only the object mentioned will be focused and not the other participants in the action. An example of the second case is:

Where are the bolts?
[Immediate focus = bolts]
[Potential goal = THE BOLTS ARE BOLTED]

I've tightened them with the wrench.
[with the wrench not in focus]

In this situation, the first reference to the bolts has established the potential goal that the bolts be bolted.

In both these situations the object mentioned is focused and, when appropriate, an action it participates in is established as a goal. The difference between the two is whether the actions and the other participating objects are also focused. This difference affects the interpretation of successive utterances containing pronouns.

Three cases are distinguished in the algorithm: (1) If there is a pronoun and there are no definite noun phrases, the actions associated with the most recent direct goal and the potential goals are considered as possible referents of the verb, since either of the two

cases described above could obtain. (2) If there are definite noun phrases, all of which refer to focused entities, then the actions associated with the most recent direct goal and the potential goal are the most likely referents. Since all the objects are focused, the actions was presumably mentioned as in the first case described above. (3) If there are a pronoun and also definite noun phrases, but not all the definite noun phrases refer to focused entities, then only an action associated with a potential goal is a possible referent. Since a direct goal associated with this object could not have been established only the second case described above could obtain.

In all three cases, utterance information about tense and aspect and about action type (from the verb) is used either to verify that the action associated with the goal is a possible referent or to choose a matching action type among possible referents.

4. No Pronoun or Do

When there is no anaphora in the utterance, the contextual knowledge used for interpretation is focusing and the task model. Focusing is used to determine the relationship between the utterance and focused entities, including the current action. The task model, including the record of task progress, is used to determine which actions can reasonably be talked about in the context. First, focusing information is used to determine if the referents of any definite noun phrases associated with the verb are currently focused.

a. All Noun Phrases in Current Focus

If the noun phrase referents are focused, it indicates that the action involves objects currently being discussed by discourse participants and that the action is related to the current step (because it involves the same objects). The task model provides information about actions the apprentice can perform and has performed. Tense and aspect information from the utterance and the verb type restrict alternatives within the task model.

As we discussed earlier, present and progressive utterances generally refer to newly started actions. Thus the actions in the task model considered are those closely related to the most recent action performed and that involve objects referred to in the utterance. Possible actions might be: a substep of the last step started but not completed; the potential goal; a step not involving any different objects that is closely linked in the plan to the last step started or completed (i.e., a step that is a substep of or successor to the last step, or succeeds a parent of the last step).

Utterance (1) in the sample dialog ("I am attaching the pump") illustrates a present-progressive utterance with a noun phrase referring to a focused object. In this instance, the pump-attaching step is a substep of the last step started--installing the pump.

For utterances that are past tense and/or perfective aspect, actions in the task model known to have been in progress and those that could be next steps are possible referents. The alternatives considered during interpretation are: a step in progress; the potential goal; a substep of the last step started; a substep of any step in progress; and a step closely linked to the last step started or completed. Utterance (7) ("I attached the pump") shows a reference to a completed action that was a step in progress--attach pump. The verb in utterance (11) ("I've installed the pulley") refers to a completed action which was the next step to perform, but was not explicitly mentioned as having been started--install pulley.

b. Not all Noun Phrases in Current Focus

If the referents of the noun phrases are not currently focused, the focusing hierarchy is searched because it indicates previously focused objects that might become focused again. If the noun phrase referents are identified somewhere in the focusing hierarchy, the action named in the utterance is matched against any action occurring at that place in the hierarchy.

If the utterance contains noun phrases referring to objects participating in the action and those objects cannot be identified among focused entities, the actions associated with direct goals are eliminated as possible referents of the verb. This happens because all actions associated with direct goals have been mentioned, which has caused all their participants to be focused.

Possible referents of such verbs include: the action associated with the potential goal; a substep of the current step in progress; a substep of all the steps in progress (if the utterance is past and/or perfective); any action which can achieve some current goal (e.g., knowing a location -> found the object). Since the objects described in the noun phrases and the action both have to be tested when examining the substeps, the algorithm first checks the objects described by the noun phrases to see if they are participants in any of the substeps (by looking at the binding space) and if so, it then examines the actions to ascertain whether one of them matches the input action.

B. Bottom-Up Search

Currently the bottom-up algorithm consists of a search for the most specific occurrence of an event in the model whose participants are compatible with those in the utterance. This strategy is being expanded to include a search for a more general event that can then be found in the task. This can be either the most specific event type that is compatible with all the elements in the utterance, or a more general or 'similar' event type that is compatible and can be found in the task . An example of the first is an utterance containing "tighten the bolt". The verb "tighten" refers to a general tightening action, that can have more specific uses--such as tighten screws, tighten bolts, etc. From the knowledge that one kind of tightening is bolt tightening and from the concomitance of "bolts" in the utterance, it can be inferred that the "tighten bolts" action is intended. In the second case, a more specific verb might have been used (e.g., bolt the pump) to mean securing the bolts. The verb "bolt" might be initially interpreted as

referring to a specific action of tightening bolts. However, that might not be an explicit step in the task, but rather, perhaps, only some general securing step. From the bolting action and knowledge of the more general actions of which it is a subset (e.g., securing), its relation to the task model can be found.

C. Setting Limits to a Search

Knowing when to stop searching for a referent of a verb is another important part of interpreting it. In general, the extent to which a verb reference is interpreted depends on the type of utterance. For example, a verb may refer to an action that does not fit into the current task context, such as one that could not or should not be performed at that time. If the verb is contained in a question (e.g., "Should I cut the end off now?"), a reasonable assumption may be as follows: if the action cannot be identified it is not the appropriate one to take, as illustrated in Utterance (8). On the other hand, if the verb is contained in a statement (e.g., "I have cut off the end."), identifying the specific action performed is more important, since a model of the current situation could not otherwise be maintained. Thus, any process for identifying a verb referent should be able to determine the amount of resources it should expend in each situation.

Another factor to be considered when determining how much effort to expend in identifying the referent is the extent to which the speaker can be assumed to be cooperative, and, consequently his or her utterances to be relevant. If some fairly direct connection between the utterance, the task, and/or dialog context can be postulated, devoting more effort to the search for a connection is more reasonable than in a less task-oriented dialog, in which such a connection may not even exist. In the TDUS system it is assumed that the user is cooperative and that all his or her utterances are relevant. Thus, considerable effort is expended when necessary to relate a statement about an executed step to the task of which it is a part.

D. Effect of Automatic Planning

The strategy described here has been developed in a system in which the plan for accomplishing the task has already been determined. The incorporation of an automatic planning facility should not cause it to be modified substantially. With automatic planning, the search forward to next possible steps would generally require planning 'next steps' to see if the action in the utterance would fit, and bottom-up searching could include plan recognition to see how the action might be part of a plan.

V FUTURE DIRECTIONS

In this paper, we have discussed the problem of identifying the actions and events referred to by verbs. In particular, we have considered dialogs about an ongoing task. We have examined some of the knowledge needed for identifying the actions and have presented a strategy for finding them. This problem is of interest both because it is an important part of interpreting utterances and because it illustrates the need for combining knowledge of many types when interpreting utterances.

The research discussed here shows how the knowledge about language and about the domain that is currently identified and represented in a computer system can be used when interpreting verbs. Important extensions of this research include determining: (1) how top-down and bottom-up searching can be combined more effectively; (2) on what basis decisions can be made to stop looking for a connection between an action and a plan; (3) what extensions of this algorithm are necessary for handling dialogs in which the lack of a strong model of the task being performed results in weaker top-down constraints. Further research on finding referents of verb phrases, building on the algorithm presented here, should contribute to solving the more general natural-language processing problems of determining what other knowledge is needed for interpreting utterances and how that knowledge can be used most effectively.

VI REFERENCES

Allen, J. 1979. "A Plan-Based Approach to Speech Act Recognition". Department of Computer Science. University of Toronto, Toronto, Canada. PhD Dissertation.

Appelt, D. E., Grosz, B. J., Hendrix, G. G., and Robinson, A. E., 1980. "The Representation and Use of Process Knowledge". Technical Note 207. Artificial Intelligence Center, SRI International, Menlo Park, California.

Clark, H. H., and Marshall, C. 1980. "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.

Cohen , P. R., 1978. "On Knowing What to Say: Planning Speech Acts". Technical Report No. 118, Department of Computer Science, University of Toronto, Toronto, Canada.

Donellan, K. S., 1977. "Reference and Definite Description." In S. P. Schwartz (Ed.) Naming, Necessity, and Natural Kind. Ithaca, New York: Cornell University Press, pp. 42-65.

Fikes, R. E. and Hendrix, G. G., 1977. "A Network-Based Knowledge Representation and its Natural Deduction System." Proceedings of the Fifth International Joint Conference on Artificial Intelligence, pp. 235-246 Cambridge, Massachusetts, 22-25 August 1977.

Grimes, J. G., 1980. "Context Structure Patterns." To be presented at Nobel Symposium on Text Processing, Stockholm, Sweden, 11-15 August, 1980.

Grosz, B. J., 1977a. The Representation and Use of Focus in Dialogue Understanding. Technical Note No. 151, SRI International, Menlo Park, California.

Grosz, B. J., 1977b. "The Representation and Use of Focus in a System for Understanding Dialogues." Proceedings of the Fifth International Joint Conference on Artificial Intelligence, pp. 67-76, Cambridge, Massachusetts, 22-25 August 1977.

Grosz, B. J., 1978. "Focusing in Dialogue." Proceedings of TINLAP-2, Urbana, Illinois, 24-26 July, 1978.

Grosz, B. J., 1979. Utterance and Objective: Issues in Natural Language Communication. Proceedings of the Sixth International Joint Conference on Artificial Intelligence, pp. 1067-1076. Tokyo, Japan. 20-23 August 1979.

Grosz, Barbara J. 1980. 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.

Grosz, B. J., Hendrix, G. G., and Robinson, A. E., 1977. "Using Process Knowledge in Understanding Task-Oriented Dialogs." Proceedings of the Fifth International Joint Conference on Artificial Intelligence, pp. 90. Cambridge, Massachusetts, 22-25 August 1977.

Hankamer, J. and Sag, I. 1976. Deep and Surface Anaphora. Linguistic Inquiry Volume 7 Number 3 (Summer, 1976), Pp. 391-426.

Hendrix, G. G., 1973. "Modeling Simultaneous Actions and Continuous Processes." Artificial Intelligence, 4, pp. 145-180.

Hendrix, G. G., 1975. "Partitioned Networks for the Mathematical Modeling of Natural Language Semantics." Technical Report NL-28, Department of Computer Sciences, University of Texas, Austin, Texas.

Hendrix, G. G., 1977. " Some General Comments on Semantic Networks." Panel on Knowledge Representation, Proceedings of the Fifth International Joint Conference on Artificial Intelligence, pp. 984-985, Cambridge, Massachusetts, 22-25 August 1977.

Hendrix, G. G., 1979. "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.

Hobbs, J. R., 1978. Why is Discourse Coherent? Technical Note No. 176, SRI International, Menlo Park, California.

Konolige, K., 1979. "A Framework for a Portable Natural-Language Interface to Large Data Bases". Technical Note No. 197, SRI International, Menlo Park, California.

Leech, G. N. 1976. Meaning and the English Verb. Longman Group Ltd. London, England.

Reichman, R. 1978. "Conversational Coherency." Cognitive Science. 2:4.283-327.

Robinson, A. E., 1979. Natural Language Communication. SISTM Quarterly incorporating the Brain Theory Newsletter, Vol II, Number 4, Summer 1979.

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

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

Searle, J. R., 1978. Literal Meaning. Erkenntnis, 13, pp. 207-224.

Sidner, C. , 1979. "Towards a Computational Theoy of Definite Anaphora Comprehension in English Discourse." Massachusetts Institute of Technology, Cambridge, Massachusetts, PhD dissertation.

Walker, D. E., ed., 1978. Understanding Spoken Language. Elsevier North-Holland, New York.

Webber, B. L., 1978. "A Formal Approach to Discourse Anaphora," BBN Report No. 3761, Bolt, Beranek, and Newman Inc. Cambridge, Massachusetts, May, 1978.

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

Wilensky, R., 1978. "Understanding Goal-Based Stories". Yale University, New Haven, Connecticut, PhD dissertation.