THE REPRESENTATION AND USE OF FOCUS IN A SYSTEM FOR UNDERSTANDING DIALOGS

Barbara J. Grosz
Artificial Intelligence Center
SRI International, Menlo Park, California 94025

ABSTRACT

As a dialog progresses the objects and actions that are most relevant to the conversation, and hence in the focus of attention of the dialog participants, change. This paper describes a representation of focus for language understanding systems, emphasizing its use in understanding task-oriented dialogs. The representation highlights that part of the knowledge base relevant at a given point in a dialog. A model of the task is used both to structure the focus representation and to provide an index into potentially relevant concepts in the knowledge base The use of the focus representation to make retrieval of items from the knowledge base more efficient is described.

I INTRODUCTION

To understand the sentences in a discourse, a computer system, like a person, must have knowledge about the domain of the discourse. However, the knowledge required to understand even simple, real-life domains is so extensive that it will overwhelm a system that does not apply it selectively. This means that the ability to focus on the subset of knowledge relevant to a particular situation is crucial. This paper addresses the problem of focus from the perspective of building a computer system that can participate in a task-oriented dialog. A representation for focus is presented; its use is illustrated by showing how the referents of definite noun phrases are identified.

A combination of contextual factors influences the interpretation of an utterance. In fact, what is usually meant by "the context of an utterance" is precisely that set of constraints which together direct attention to the concepts of interest in the discourse in which the utterance occurs. Both the preceding discourse context -■- the utterances that have already occurred — and the situational context — the environment in which an utterance occurs — affect the interpretation of the utterance. For a dialog, the situational context includes the physical environment, the social setting, and the relationship between the participants in the dialog. This paper shows how the task and dialog contexts combine to provide a focus on those concepts relevant to the interpretation of utterances in task-oriented dialogs.

This research is currently supported by the National Science Foundation under Grant No. MCS76-22004; it was previously funded under ARPA contract DAAG29-76-C-0011, administered through the Army Research Office.

The following two dialog fragments illustrate the role of focus in interpreting utterances.

SI: The lid is attached to the container with four 1/4-inch bolts.
R1: Where are the bolts?

S2: Attach the lid to the container.
R2: Where are the bolts?

In the first sequence, statement S1 explicitly points out a set of bolts that are then referred to in response R1. The dialog context provides the focus for understanding the phrase In the second sequence, no such explicit mention occurs. Instead, the attaching referred to in S2 implicitly focuses on the fasteners involved. In a particular task context, knowledge about the process of attaching a specific lid to a specific container focuses on a specific set of bolts Hence, the noun phrase in R2 is unambiguous.

A. REQUISITES FOR A FOU REPRESENTATUION

There are three requisite properties of focus representation. The most crucial is that it differentiate among the items in the knowledge base (i.e., the encoding of that portion of the world the system knows about) on the basis of relevance. By highlighting those items in the knowledge base that are most relevant to the current discourse, the focus representation enables the system to access more important information first during its retrieval and deduction operations.

Second, the focus representation must account for implicitly focused items Specific mention of an object brings into focus not only the object itself, but also certain associated items. For example, mention of "the house" brings into focus such associated objects as "the living room", "the roof", "the yard", and "the owner". Parts of actions as well as objects may enter focus in this way. For example, "sewing a dress" brings into focus "cutting out the skirt".

Third, the focus representation must include mechanisms for shifting focus. As successive utterances in a discourse are interpreted, the items in focus change. Shifts of focus occur both gradually with time, and more abruptly with change of topic. Not only the objects in focus, but also the particular way of viewing them, can change. For example, a doctor can be viewed as a member of the medical profession, or aA having a role in a family.

B. OVERVIEW OF PAPER

The next section describes a focus representation that satisfies these requirements. The representation was developed in the context of a system for understanding language (Walker 1976). The knowledge base of this system is embodied in a semantic network. In such a system, identifying a network structure is the analog of the human process of identifying and retrieving an item from memory. Hence, the matching of network structures is a crucial process in interpreting utterances. A major use of the focus representation is to make this matching process more efficient. Once the focus representation has been described, and its use illustrated, mechanisms for shifting focus in task-oriented dialogs are discussed. These mechanisms are the only part of the focus representation that is specific to task-oriented dialogs. Finally, a description is given of how the focus representation can be extended to help solve two essential problems in natural language understanding: focusing on different attributes of the same object under different circumstances, and forgetting information no longer relevant to a discourse.

The focus representation was implemented in the SRI speech understanding system (Walker, 1976) and used by the discourse component to resolve definite noun phrases. The process representation used to encode the task model, which is needed both for implicit focusing and to guide shifts of focus, is designed and currently being implemented. (The speech understanding system implementation used a simpler shift strategy than the one described here.)

II THE FOCUS REPRESENTATION

This section describes a two-part focus representation. One part corresponds to explicit focus, the other to implicit focus. The explicit focus data structure contains those items that are relevant to the interpretation of an utterance because they have participated explicitly in the preceding discourse. Implicit focus consists of those items that are relevant because they are closely connected to items in explicit focus. For instance, in the dialog fragments given in the Introduction, both S1 and S2 result in the lid, the container, and the attaching operation being in explicit focus. The bolts involved in the operation are in explicit focus following S1, but implicit focus following S2.

Concepts that are implicitly focused are separated from those that are explicitly focused (i.e., they are not added to the explicit focus data structure) for two reasons. First, there are numerous implicitly focused items, many of which are never referenced in a dialog. Including these items in the explicit focus data structure would clutter it, weakening its highlighting function. Second, references to implicitly focused items are considered indications of shifts of focus.

EPRESENTATIQN QF EXPLICIT FOCUS

The representation of explicit focus is achieved by partitioning a semantic network. A semantic network is a directed graph: a set of nodes, and a set of (labelled, directed) arcs, connecting pairs of those nodes. Networks have been used in several previous language understanding systems (e.g., Quillian. 1968; Simmons, 1973). Conventions for the use and meaning of nodes and arcs vary. The networks described here use the conventions of Hendrix (1975)—nodes are used to represent "objects," where "objects" include such things as physical objects, events, relationships, and sets. Arcs are used only to encode those binary relationships that do not change over time.

Partitioning adds to the structure of a semantic network by segmenting the nodes and arcs of the network into units called spaces. Hendrix (1975) introduced the notion of network partitioning, and described its use in encoding quantification, abstraction, and hypothetical worlds. In addition to separating the nodes of a network into spaces, partitioning provides for grouping the spaces into ordered sets called vistas. Vistas typically are used to restrict the network entities seen by procedures that reference the network (i.e., vistas Impose visibility constraints). When given a vista, a procedure can operate as though the only nodes and arcs in the network are those contained in some space in the vista. Although any set of spaces may be collected into a vista, vistas typically are used to group spaces hierarchically.

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

An example of a partitioned semantic network is displayed in Figure 1. The network is divided into four spaces: SO, S1, S2, and S3- The conventions adopted for figures in this paper are that a node lies on the space(s) inside of which It is drawn, and an arc lies on the space(s) inside of which its label appears. If the boxes representing two spaces overlap, but neither contains the other, then the nodes and arcs in the overlap lie on both spaces. In Figure 1, space SO groups the nodes representing EXCHANGES (the set of all exchange situations), ATTACHINGS, BOLTS, PUMPS, and

PLATFORMS (the sets of all attach operations, bolts, pumps, and platforms, respectively). Space S1 contains the specific exchange, represented by the node 'EX1' (single quotes denote node names), in which one object, the set of bolts B1. repressnted by the node B1', is exchanged for another, the amount of money represented by the node '$1». The e arc from 'EX1 ' to 'EXCHANGES' indicates that EX1 is an element of the set EXCHANGES. The s arc from 'B1» to 'BOLTS' indicates that the set B1 is a subset of BOLTS. Space S2 contains the representation of a specific attaching operation, A1, of the minor part, PU1, and the major part, PL1. Space S3 also contains the specific attaching operation A1 but shows that this operation involves the specific set of bolts, B1.

The hierarchy of spaces in Figure 1 is shown by the heavy arrows between spaces. Each space is associated with a particular vista that is the "orthodox" vista for that space. In Figure 1, the orthodox vista associated with each space S is composed of the space S itself and all spaces that can be reached from S by following the heavy arrows. For instance, the orthodox vista of SO is (SO) and the orthodox vista of S3 is (S3 S2 SO).

The visibility constraints that result from this partitioning may be seen by considering different views of the bolts B1 and the attaching operation A1. B1 is shown as taking part in two different events. A1 is a 3ingle operation shown at two different levels of detail. From the vista (S1 SO), the set of bolts B1 is seen only to be involved in the exchange EX1. However, from the vista (S3 S2 SO), B1 is seen as the fasteners in the operation of attaching PU1 to PL1. The two vistas give two alternative views of B1. A similar situation occurs with A1. From the vista (S2 SO), A1 is seen only as an attaching between two parts, with the fasteners left unspecified. When S3 is added as the bottom space in the vista, A1 is seen to involve the specific fasteners B1.

The focus partitioning makes it possible to highlight the particular way of looking at a concept that is germane to a given point in a dialog. When the same object enters the dialog twice, in two different subdialogs (e.g., a tool used in two distinct subtasks), the node corresponding to that object will appear in two distinct focus spaces. If different aspects of the object are focused on in the two subdialogs, different relationships in which the object participates will be in the two focus spaces. For example, in Figure 1, B1 is focused on in S1 as a part of an exchange. In contrast, in S3 it is focused on as part of an attaching operation.

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

There are two principles governing what is contained in a focus space. First, if a concept is in focus, type information about that concept must also be in focus. (The type information both indicates the aspect of the concept being focused on and provides the key index to additional knowledge about the concept.) Therefore, in the network representation, every node in focus must have one outgoing element or subset arc also in focus. Second. if a concept's participation in some situation (e.g., a book's being the object of a specific owning relationship) is in focus, then the situation itself (i.e., the specific owning relationship) also must be in focus. Therefore, the node from which any focused case arc emanates must also be in focus.

New focus spaces are created as the focus of a discourse shifts. At any point in a dialog, only one focus space is "active," but several may be considered "open." The active focus space reflects the focus of attention at the current point in the dialog. The open focus spaces reflect previous active spaces that contain some unfinished topics and hence may become active again; they are areas to which the dialog may return. The relationship between focus spaces Is determined by (and hence reflects) the structure of the particular discourse being processed For task dialogs, the task hierarchy provides a framework for this structure (Grosz, 1977, discusses the structure of these dialogs).

B. <u>IMPLICIT FOCUSING</u> <u>THROUGH</u> A <u>TASK REPRESENTATION</u>

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

For physical objects, the subparts of the object are among the concepts that must be implicitly focused when the object is in focus. For events, the situation is somewhat more complicated. The direct analogy of subparts of an object is subevents of an event. However, the participants in the subevents of an event are also implicitly focused. The dialog sequence S2-R2 presented in the introduction illustrates this point. S2 implicitly focuses on the bolts involved in the attaching as well as the subevent of fastening the lid down.

To enable implicit focusing on both the subevents and the participants involved in them, the representation of an event indicates both its subevents and the participants in its subevents. Figure 2 shows a network representation that accomplishes this fOr the task step of attaching a pump to a platform. The logical space KNOWLEDGESPACE, only part of which is shown here, contains representations for all items in the knowledge base. The set of ATTACHINGS.PUMP.PLATFORM is shown to be a subset of all ATTACHINGS. The delin arc from 'APP' to 'ATTACHINGS.PUMP.PLATFORM' indicates that APP is the prototypical element of the set of such attachings (see Hendrix, 1975, for a discussion of delineations). The two nodes 'APP' and 'APPD' together with the other structures inside the delineation space, DS, describe the nature of events in which a pump is attached to a platform. APP relates the participants in the event. The outgoing arcs from 'APP' indicate that these attachings involve a minor part, which is an element of the set PUMPS, and a major part, which is an element of the set PLATFORMS.

APPD is the event descriptor for APP. It relates the preconditions, effects, and substeps of the event. The two constituents of APPD that are most relevant here are the plot space and the binding space. The plot space, PS, contains the breakdown of APP into two substeps, S1 and S2, specifying a POSITION operation, 0P1, and a SECURE operation 0P2. The sue arcs indicate successor links between substeps. (Although not shown here, the representation allows for partial ordering of substeps, as in Sacerdoti, 1975.) The binding space, BS, contains a set of four bolts that take part in the securing substep. When the task step of attaching a particular pump to a particular platform is in (explicit) focus, then the corresponding substeps for S1 and S2, and the set of bolts in the binding space are considered implicitly in focus.

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

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

III USING FOCUS FOR NETWORK STRUCTURE MATCHING

The retrieval of items from memory is one of the most frequent operations any knowledge-based system must do. In a system with a semantic network knowledge base, the central process involved in retrieval is matching a network fragment containing variables with the knowledge base. This matching process typically entails considerable search that is guided only by local constraints. A major use of the focus representation is to constrain the search on the basis of discourse information. In this paper, the system component that performs this matching process will be called the matcher. Fikes and Hendrix (1977) describes in detail how this component works. Only enough detail will be given here to elucidate the need for and the role of the focus representation in this process.

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

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

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

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

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

Optimally, for both kinds of decision, the matcher will choose the most constraining element. In an unfocused match, the choice can be made only on the basis of local structural information.

A. MATCHING IN FOCUS

The focus representation is used to order the candidates considered for binding by the matcher. The term "focused match" is used to denote matches that are constrained by focu3. Focusing on certain concepts (both nodes and arcs) constrains the matcher to consider only objects germane to the dialog. Since arcs provide indices from focused items into general network (KVISTA) information, focusing on an arc also guides the matcher in establishing properties about nodes being matched. That is, focused arcs provide a means of differential access to unfocused information. Using the arcs in focus for differential access does not eliminate consideration of a concept from a different perspective. Instead, it orders the way in which aspects of the concept are to be examined in looking for new (to the dialog) information about the concept.

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

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

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

Figure 3 illustrates the use of focus to reduce the number of candidates considered for binding by the matcher. Consider the KVISTA of Figure 3 and the QVISTA (q.wl) of Figure 4. The KVISTA contains several wrenches: W1 is a box-end wrench that is in focus FS1; W2 is a box-end wrench in focus FS2; W3 is an open-end wrench also in focus FS2; W4 is another open-end wrench not in focus at all. There is another object, 01, with a box-end. The QVISTA represents the noun phrase "the box-end wrench". In an unconstrained match, the matcher would consider all the nodes with e arcs to 'WRENCHES', or all of the nodes with endtype arcs to 'BOX-END' (depending on which set is smaller) as candidates for binding to QW1. Eventually, it would try 'W1' or 'W2' and obtain a successful match. In the worst case, this would entail one node and two arc bindings for each of the candidate nodes that fails as a complete match. In general, there may be many such unsuccessful candidates (e.g., many wrenches that are not box-end wrenches, but are considered by the matcher before it selects 'W1 or 'W2').

The focused match is able to avoid all this searching. If focus space FS1 is used, only nodes 'H1' and *W1' are considered. 'H1' will be rejected immediately because the e arc to 'HAMMERS' is incompatible with the e arc from 'QW1' to 'WRENCHES' (the matcher knows that the sets HAMMERS and WRENCHES have no intersection from the ds arcs from 'WRENCHES' and 'HAMMERS' to 'TOOLS'). With focus space FS2 as the constraint, both 'W3' and 'W2' are considered, but 'W3' is eliminated because of its incompatible endtype. In the worst case, one set (one e arc and one node) of unnecessary bindings is made. Even greater savings are obtained when deduction is necessary to achieve a match, that is, when general rules — chunks of information stored in the net as applicable to whole sets of concepts — must be applied (see Grosz, 1977). In such cases, focus constrains the application of such rules, avoiding a combinatorial explosion of trial bindings.

IV SHIFTING FOCUS AND THE RESOLUTION OF DEFINITE NOUN PHRASES

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

In task dialogs, a shift in focus takes place whenever a new task is entered or an old one completed. A narrowing of focus takes place whenever a subtask of the active task is opened for discussion. The focus shifts back up to the higher level task when that subtask is completed. Hence, when a subtask of the current task is referenced, a new active focus space is created below the current active focus space. When the subtask is completed, the new focus space is closed and the old space (i.e., the higher space) becomes the active focus space again. The top of the focus space hierarchy is the focus of the overall task.

A shift in focus may be cued by any part of an utterance: a noun phrase, a verb phrase, or modifying phrases. Although an individual constituent (e.g., a noun phrase) may indicate a shift in focus, the constituent alone cannot be used to determine the shift, because the remainder of the utterance influences the decision. For instance, an isolated noun phrase may seem to indicate a shift to some task but, when considered with its embedding verb phrase, may indicate a shift to a different task. The following discussion examines the relationship between identifying the referent of a definite noun phrase and shifting focus. Grosz (1977) describes the interaction of various constituents of an utterance in determining a shift.

To illustrate how a noun phrase may indicate a shift in focus, consider the task hierarchy in Figure 5 and the focus environment portrayed in Figure 6. Figure 5 is only for the reader's benefit; this information is actually encoded in structures like those in Figure 2 and Figure 6. The dotted lines show the task hierarchy and the solid lines show time sequencing. Suppose that task T2, installing the aftercooler, is the current task. The focus spaces FSO, FS1, and FS2 in Figure 6 correspond to subtasks TO, T1, and T2 respectively. FS2 is the active focus space; the vista (FS1 FSO) is the hierarchy of open focus spaces.

A reference to an object in either the active focus space or one of the open focus spaces does not cause a shift in focus. Those items in the active focus space are considered first when resolving a reference because the currently active task is more in focus than its embedding tasks. The phrases "the aftercooler", "the wrench", and "the crescent wrench" all refer to objects in FS2, the active focus space. Hence, the use of any of these phrases does not affect the focus of attention. The referent can be retrieved immediately. The use of either "the air compressor", "the pump", or "the ratchet wrench" also does not cause a change in focus. Since these objects are in open focus spaces, they are also in focus, but are accessed only after considering the objects in FS2. Note that the noun phrase, "the wrench," is not ambiguous because of the distinction between the active focus and the open focus spaces (this distinction is evident in references occurring in actual task dialogs between two people; see Grosz, 1977).

References to either a new subtask, or a new parallel or higher task, *or* to subtasks of any of these, do change focus. In the example, space IADS contains the delineation of the process for installing the aftercooler. The plot space of this delineation is the implicit focus for node 'IACI'. It shows that this kind of installation has two substeps (corresponding to T3 and TU in Figure 5). The first substep involves a connection operation between the aftercooler and one of its subparts, an aftercooler elbow. In this focus environment, since there is no aftercooler elbow in explicit focus, the phrase, "the aftercooler elbow", indicates a possible shift in focus to task T3- If the remainder of the utterance concurs with such a shift, a new focus space will be created below FS2 in the hierarchy. The utterance, "Attach the aftercooler elbow", indicates such a shift, but the utterance, "The aftercooler elbow is broken", does not. Note that a shift in focus may entail instantiating new entities *or* identifying real entities corresponding to hypothetical entities in implicit focus. If focus is shifted to task T3, the aftercooler elbow ACE1 is brought into focus and the noun phrase "the aftercooler elbow" is identified with it; i.e., the new focus space will contain the node 'ACEV.

With a representation of focus, the process of identifying the referent of a noun phrase looks quite different than in systems that search sequentially back through a discourse for a referent. The important question is when and how to shift focus, not how far back (i.e., how many sentences) to look for the referent. The search for the referent of a definite noun phrase takes into account the difference between those items which do and those which do not shift focus. Items in explicit focus, which do not indicate a shift in focus, are checked before items in implicit focus, which do.

V EXTENSIONS

This section explores the use of the focus space representation in the solution of two other problems that arise in building knowledge based systems. First, there is a space-time tradeoff between storing derived information in the knowledge base and recomputing the information. Ideally, the information would be stored only as long as needed and then erased from the knowledge base. This issue is closely related to the general issue of forgetting in a knowledge based system. Second, any given object may be viewed from several different perspectives. Highlighting a particular view may be used to capture the information conveyed by the specific way an object is described in a given utterance.

A. DERIVED INFORMATION M.D FORGETTING

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

As an example of this use of focus spaces, consider the situation portrayed in Figure 7. The (logical) space oew.dese represents the fact that all elements of the set 0-E have endtype OPEN-END. Bew.desc represents a similar rule. Suppose that initially the nodes 'W1' and W2' were in focus as elements of the sets B-E and 0-E respectively (e.g., the wrenches were selected from two boxes each containing one type of wrench). If the matcher is given the structure for "box-end wrench" (see Figure 4) to match, it will create two new arcs, an endtype arc from 'W1 ' to 'BOX-END', and an explicit e arc from 'W1' to 'WRENCHES'. These new arcs are added to the focu3 space, FS, as shown in Figure 7. Any further matches sought for "the box-end wrench" while the focus is FS will be able to take advantage of this explicitly stored information. When FS ceases to be open, the arcs will be erased. If the deduction had resulted in new nodes being created, they too could be erased. Using focus spaces in this way creates the double advantage of having the information available when it is relevant, and allowing it to be "garbage collected" or "forgotten" after it ceases to be relevant.

B. DIFFERENTIAL ACCESS and DESCRIPTION

The representation of some concept C may include descriptions of C as an Instance of several different kinds of other concepts. Focusing allows the particular way of looking at C germane to a given point in a dialog to be highlighted (cf. the use of multiple perspectives in Bobrow and Winograd, 1977). The arcs from focused items to unfocused items provide the matcher with preferential access to information that is most likely to become relevant to a discourse.

This use of focusing addresses one part of the "mayor of San Diego" problem posed by Norman et al. (1975). Consider the situation portrayed in Figure 8. The person represented by the node 'MNMSD' is shown to be both D's neighbor and the mayor of San Diego. If MNMSD is referred to by D either as "the mayor of San Diego" or "D's neighbor", then node 'MNMSD' represents the individual referred to. The problem is that looking only at that node provides no reflection of the differences between the two references to MNMSD, even though the surface noun phrases do express this difference. Focus spaces provide a means of representing this difference. Even though node 'MNMSD' will be in focus no matter which reference is used, arcs from 'MNMSD' that are in focus in the two cases will differ. Focus spaces FS1 and FS2 illustrate this difference.

VI SVMMARX

The focus representation groups items relevant to a particular point in a discourse, providing a small subset of the knowledge base on which the understanding system can concentrate. In particular, the focus representation may be used to guide the retrieval of information from the knowledge base. It reduces the size of the search space that the retrieval mechanism must traverse. The representation of explicit focus in focus spaces also appears to be useful for related understanding system problems, such as describing objects and forgetting information. Although the representation presented is in terms of a semantic network, partitioning a memory representation for the purpose of reflecting focus of attention is a general mechanism which may be used in other representation schemes as well.

REFERENCES

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

Fikes, Richard E. and Hendrix, Gary G. A Network-Based Knowledge Representation and its Natural Deduction System. Fifth International Joint Conference on Artificial Intelligence, 1977.

Grosz, Barbara J. The Representation and Use of Focus in Dialogue Understanding. Ph.D Thesis, University of California, Berkeley, California, June 1977.

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

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

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

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

Quillian, M.R. Semantic Memory. In: Semantic Information Processing. Edited by Marvin Minsky. The MIT Press, Cambridge, Massachusetts, 1968.

Rieger, C. Conceptual Overlays: A Mechanism for the Interpretation of Sentence Meaning in Context. Technical Report TR-35M. Computer Science Department, University of Maryland, College Park, Maryland. February 1975.

Sacerdoti, Earl D. A Structure for Plans and Behavior. Technical Note 109, Artificial Intelligence Center, Standford Research Institute, Menlo Park, California, 1975.

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

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

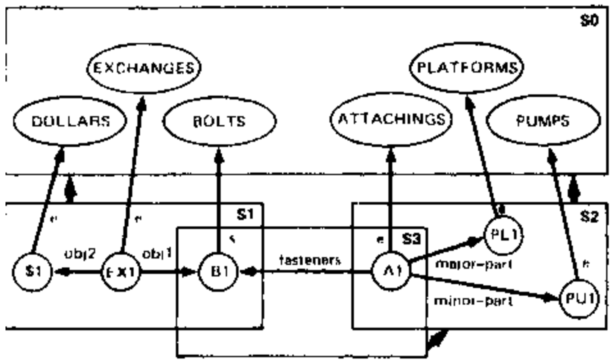


FIGURE 1 A SAMPLE PARTITIONED SEMANTIC NETWORK

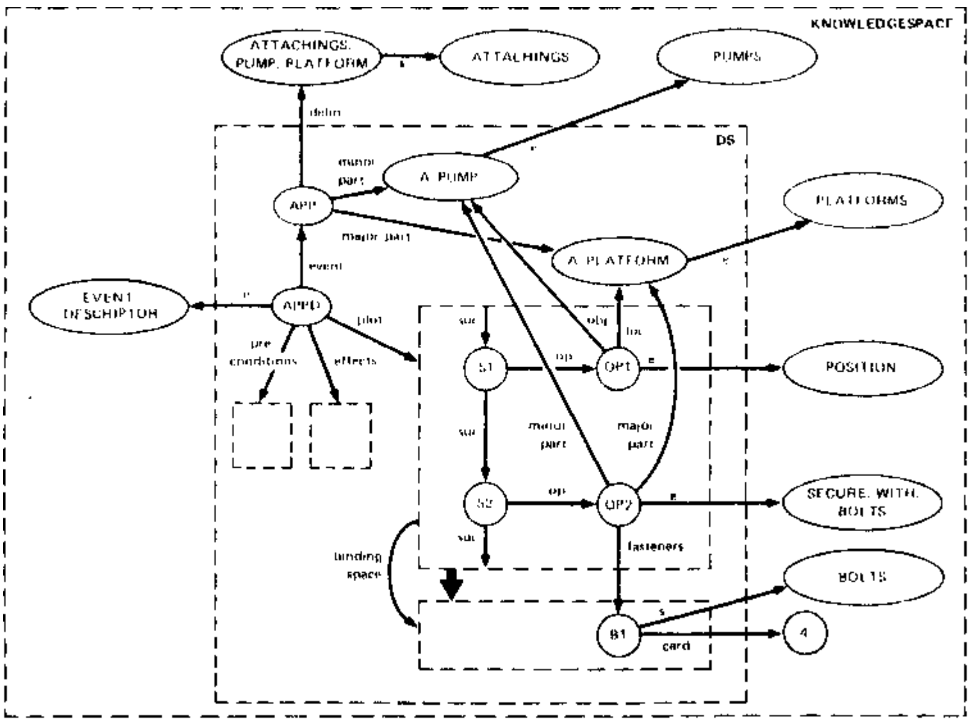


FIGURE 2 EVENT ENCODING SHOWING IMPLICIT FOCUS

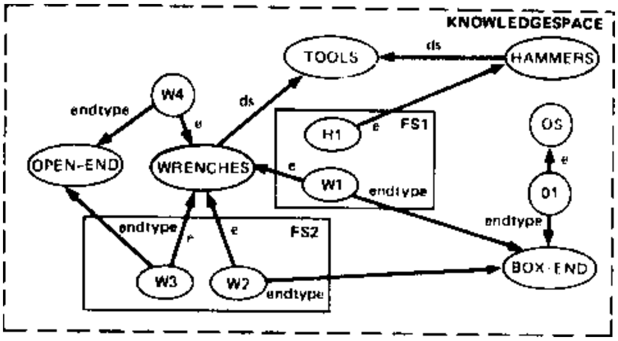


FIGURE 3 A SIMPLE KVISTA WITH TWO FOCUS SPACES

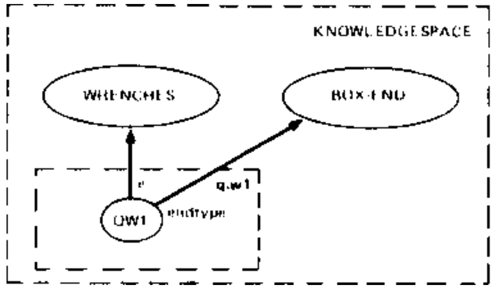


FIGURE 4 QVISTA FOR "THE BOX-END WRENCH"

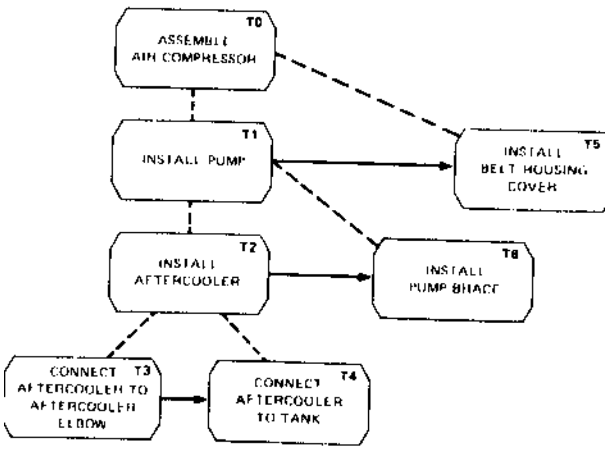
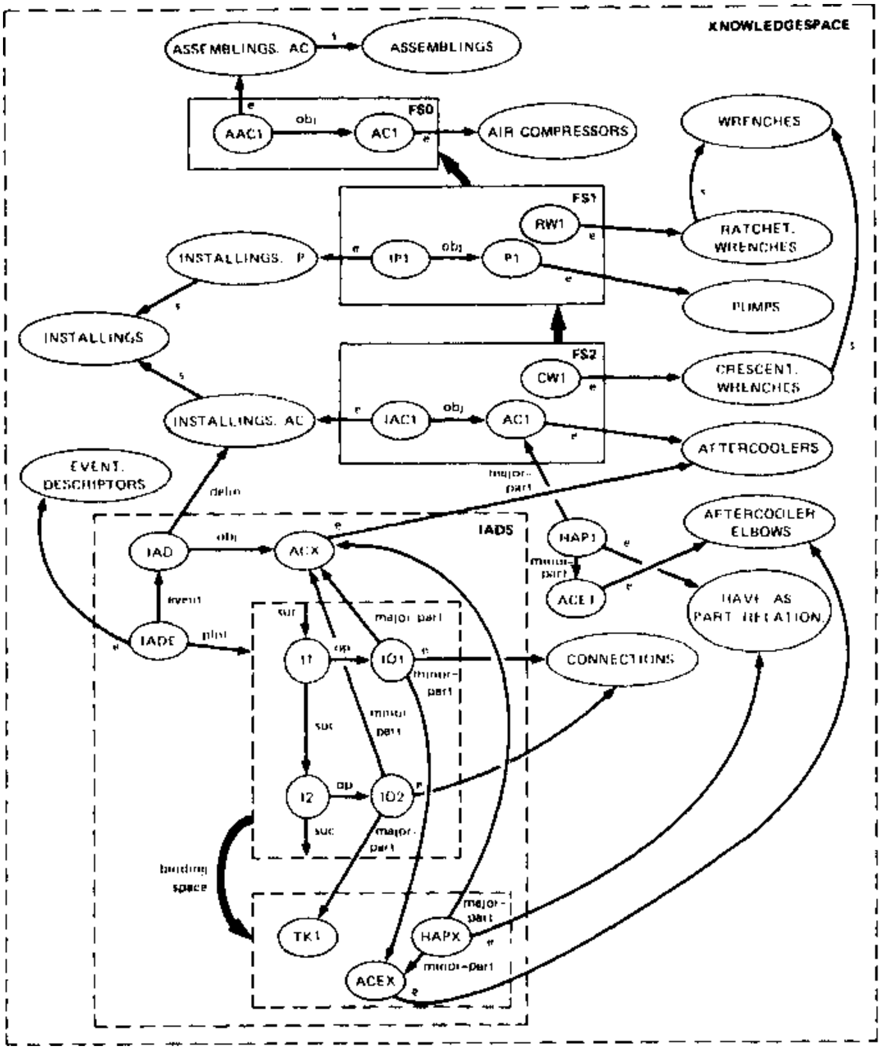


FIGURE 5 PARTIAL TASK HIERARCHY FOR ASSEMBLING AIR COMPRESSOR



FIGURF 6 FOCUS SPACFS AND IMPLICIT FOCUS FRAGMENT FOR SHIFTING FOCUS

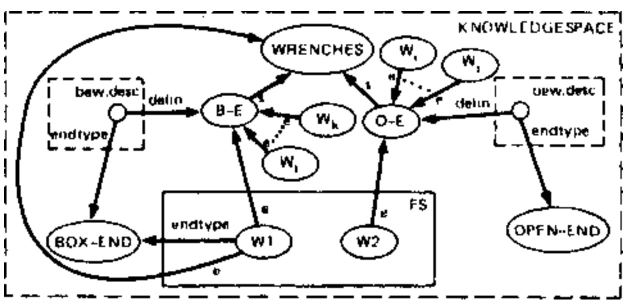


FIGURE 7 THE WRENCHES KVISTA WITH FOCUS ADDED

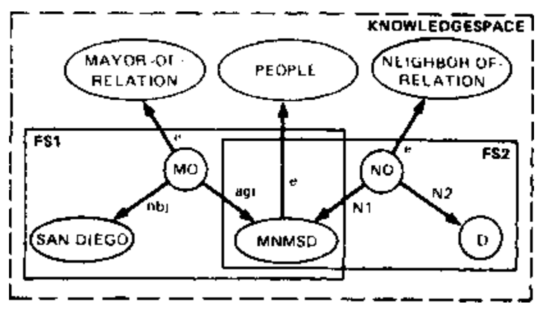


FIGURE 8 MY NEIGHBOR THE MAYOR OF SAN DIEGO