FOCUSING IN DIALOG

Technical Note 166

July 6, 1978

By: Barbara J. Grosz
Artificial Intelligence Center

SRI Projects 5844 and 6891

To appear in Proceedings of Theoretical Issues in Natural Language Processing - 2: An Interdisciplinary Workshop, University of Illinois at Urbana Champaign, July 25-27, 1978.

The work reported herin was supported by the National Science Foundation under Grant No. MCS 76-22004 and by the Advanced Research Projects Agency of the Department of Defense under Contract No. N00039-78-C-0060.

ABSTRACT

When two people talk, they focus their attention on only a small portion of what each of them knows or believes. Both what gets said and how it gets interpreted depend on a shared understanding of this narrowing of attention to a small highlighted portion of what is known. One of the effects of understanding an utterance is to be focused on certain entities (both relationships and objects) from a particular perspective. A speaker provides a hearer with clues to 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. These clues may be linguistic or they may come from knowledge about the relationships between entities in the domain. Linguistic clues may be either explicit, deriving directly from certan words, or implicit, deriving from sentential structure and from rhetorical relationships between sentences.

This paper examines focusing in dialog, discusses focusing mechanisms based on domain structure clues, and, from this perspective, indicates future research problems entailed in modeling the focusing process more generally. The importance of focusing is illustrated by considering the problem of generating and understanding definite descriptions.

I INTRODUCTION

When two people talk, they focus their attention on only a small portion of what each of them knows or believes. Not only do they concentrate on particular entities (objects or relationships), but they do so using particular perspectives on those entities. In choosing a particular set of words with which to describe an entity, a speaker indicates a perspective on that entity. The hearer is led, then, to see the entity more as one kind of thing than as another. For example, a single building may be viewed as an architectural wonder, a house, or a home, and a single event may be viewed at one time as a selling, another as a buying, and still another as a trading. Some entities are central to the dialog at a certain point and hence are focused on more sharply than others. More importantly, much of what each participant knows is not clearly in view at all; it is not considered by the speaker in choosing what to say or how to say it, or by the hearer in interpreting an utterance.

Focusing is an active process.[1] As a dialog progresses, the participants shift their focus to new entities or to new perspectives on entities previously highlighted by the dialog. Furthermore, an actor is involved in focusing (as the term is used in this paper): if an entity is in focus, it is the object of someone's focusing; it cannot be impersonally in focus. When I use the constructions "highlighted", "focused on", or "in focus", there is always an implicit actor doing the highlighting or focusing. Finally, the entities that the speaker and hearer focus on are entities in their (external) shared reality. Focusing, then, is the active process, engaged in by the participants in a dialog, of concentrating attention on, or highlighting, a subset of their shared reality.

The relationship between language and focusing is two-way: what is said influences focusing; what is focused on influences what is said.

---

[1] This is the reason the verb "focusing" rather than the noun "focus" is used most often in this paper.

The speaker provides clues for the hearer both to what s/he is currently focused on and to what s/he wants to focus on next. These clues may be linguistic or may derive from shared linguistic or nonlinguistic knowledge. The hearer depends on shared beliefs about what entities are highlighted to interpret such things as the appropriate sense of a particular word and the object or event corresponding to a definite description. The link between the entities discussed in an utterance and the entities focused on when the utterance is spoken is thus an important aspect both of producing and of interpreting that utterance.

The use and interpretation of definite descriptions in dialog demonstrate the importance of focusing to dialog participants.[2] This paper examines the relationship between focusing and definite description and the implications of this relationship for computer systems for dialog understanding. Section B presents an example that illustrates this relationship. Section C discusses definite descriptions from both the speaker's and the hearer's perspectives and presents problems that arise for both participants whose solutions are influenced by how the participants are focused. Section D addresses some problems that arise in computationally capturing the notion of focusing and discusses other aspects of dialog with which focusing mechanisms must be coordinated in a natural language processing system, in order to handle the problems introduced in the preceding sections.

II AN EXAMPLE

To begin, I want to examine a sample dialog between two people, an expert and an apprentice, cooperating to complete a task. It illustrates several important aspects of the role of focusing in communication. The sample comes from a corpus of task-oriented dialogs collected in situations simulating direct interaction between a person

---

[2] Although I will concentrate on dialog, much of what I have to say carries over to other forms of discourse.

and a computer (Grosz, 1977; Deutsch, 1974).[3] The particular task being performed is disassembly of an air compressor.

(1) E: First you have to remove the flywheel.
(2) A: How do I remove the flywheel?
(3) E: First, loosen the two allen head setscrews holding it to the shaft, then pull it off.
(4) A: OK.
(5) I can only find one screw. Where's the other one?
(6) E: On the hub of the flywheel.
(7) A: That's the one I found. Where's the other one?
(8) E: About ninety degrees around the hub from the first one.
(9) A: I don't understand. I can only find one. Oh wait, yes I think I was on the wrong wheel.
(10) E: Show me what you are doing.
(11) A: I was on the wrong wheel and I can find them both now.
(12) The tool I have is awkward. Is there another tool that I could use instead?
(13) E: Show me the tool you are using.
(14) A: OK.
(15) E: Are you sure you are using the right size key?
(16) A: I'll try some others.
(17) I found an angle I can get at it.
(18) The two screws are loose, but I'm having trouble getting the wheel off.
(19) E: Use the wheelpuller. Do you know how to use it?
(20) A: No.
(21) E: Do you know what it looks like?
(22) A: Yes.
(23) E: Show it to me please.
(24) A: OK
(25) E: Good. Loosen the screw in the center and place the jaws around the hub of the wheel, then tighten the screw onto the center of the shaft. The wheel should slide off.

First, consider the use of the phrase "the two screws" in (18) to refer to the two setscrews holding the pulley on its shaft and the use of the phrases "the screw in the center" and "the screw" in (25) to refer to a part of the wheelpuller.[4] Since most objects do not have proper names, definite descriptions are a primary means of identifying

---

[3] For most of these dialogs the expert and apprentice had only limited visual contact.

[4] The modifying phrase "in the center" does not distinguish the main wheelpuller screw from the setscrews, but from other screws that are part of the wheelpuller.

objects. However, as in this dialog, the same description may be used to identify different objects at different times. When (25) was uttered, the two screws mentioned in (3) through (18) were the most recently mentioned objects that could be referred to by a phrase such as "the screw", but they were no longer focused on by the dialog participants -- they were no longer relevant to either the dialog or the task -- and hence were not considered as possible referents for either "the screw in the center" or "the screw" in (25).

One can see in this example that the most recently mentioned object that satisfies a description may not be the object identified by that description. What entities a speaker and hearer are focused on influences both the kinds of descriptions they use and how their descriptions are interpreted. In utterance (3), the expert indicates that he is focused on, and concurrently gets the apprentice to focus on, the two subtasks involved in removing the pulley. In particular, the two allen head setscrews involved in the first task are brought into focus; they continue to be in focus through the first part of (18). The initial clause of (18) indicates the completion of the task involving the screws and hence suggests that the apprentice will shift her attention to some new task (she might not -- she could still say something more about the screws). She does make such a shift in the second clause of (18) ("but I'm having trouble getting the wheel off"). In (19), the expert indicates that he has followed this shift (note that he might have asked a question about the screws -- e.g., "How loose are they?" -- and thereby continued to focus on them and the associated task) and narrows focusing from the task of removing the flywheel to a particular tool involved in that task. In this context, it is clear that the phrase "the screw" cannot refer to either of the setscrews, but must refer to something else.[5]

---

[5] It is interesting that some people who are not familiar with the compressor or wheelpuller find this sequence confusing: (18) seems to end any concern with screws and hence (25) is unintelligible. One must know -- or infer -- that the wheelpuller has a screw for the statement to make sense.

This dialog also indicates some of the ways in which focusing is manipulated in a dialog. In particular, it illustrates how the structure of the entities being discussed (the 'domain') influences focusing and hence the structure of the discourse. The dialog concerns the performance of a task; its topic is that task. As a result, the way in which the apprentice and expert focus, and hence the structure of the dialog,[6] are closely linked to the structure of the task. Information about the structure of entities in the domain provides one kind of clue to how focusing can change. What about general linguistic clues to focusing? What information in words themselves or in sentence structure can influence focusing? The use of "but" in (18) illustrates one kind of linguistic clue to focus. The indication of contrast suggests a shifting of focus to the entities described in the clause following the "but". In fact, this shift does occur and the remainder of the fragment concerns things involved with "getting the wheel off".[7]

The final point I want to make with respect to this fragment concerns the relationship between how the speaker and hearer are focused and how differences in focusing affect understanding. It is clearly crucial for speaker and hearer to be able to distinguish their own beliefs from each other's. What about focus? I am concerned here not with the consistent difference in focusing that results from the speaker being one step ahead of the hearer (closing this gap is one goal of an utterance), but rather with whether speaker and hearer purposely maintain differences in focusing over several interactions (as they do with beliefs). An analysis of the dialogs we collected indicates that,

---

[6] The concept of structure used here is similar to that in Levy (1977), but different from that in work on story and text grammars (cf. vanDijk 1972; Rumelhart 1975). In particular, I am not interested in such things as generating or recognizing a valid dialog (the analogy to sentence grammars), but rather in those dynamic aspects of intersentential relationships such as focusing that influence the interpretation and generation of utterances in a dialog.

[7] One of the key open problems for incorporating focusing mechanisms in natural language processing systems is identifying the different kinds of clues to focusing and how they interact. Some aspects of this problem are discussed in Section D.

in most cases, whether or not a speaker and hearer are focused similarly, they speak as though they were. Speaker and hearer assume a common focus; they usually do not have distinct models of each other's focus. That is, the speaker assumes that the hearer in understanding an utterance has followed any shift in focus indicated by that utterance and is, to the extent it matters, focused on the entities the speaker intended (from the perspective the speaker intended). It is only when a difference in focusing results in some fairly major incompatibility that a problem is detected. The interchange in (5) through (11) illustrates what happens when the two participants in a dialog believe erroneously that they are focused on the same entity. Initially, the apprentice is focused on the motor pulley, which she thinks is the flywheel. Because the expert is not aware of this (he probably doesn't even consider the possibility), his responses are not very helpful.

III DESCRIPTIONS

One of the key ways in which the influence of focusing on dialog is manifest is in the definite descriptions used. There is a two-way interaction between definite descriptions and focusing: what entities a speaker and hearer concentrate on (and from what perspectives) influences how they describe entities, and how entities are described influences how the speaker and hearer continue to focus their attention. Two specific problems relating to descriptions are strongly influenced by focusing. From the speaker's perspective, there is the problem of what to include in a description. From the hearer's perspective, there is the problem of what to do when a description doesn't correspond to any known entity, when it doesn't "match" anything.

A. Generating Descriptions

Three factors that influence the production of a description are: the information speaker and hearer share about the entity being described, the perspectives they have on it, and the use of redundancy. The following fragment of dialog illustrates the first two of these factors.[8]

> E: OK. Now we need to attach the conduit to the motor. The conduit is the covering around the wires that you . . . were working with earlier. There is a small part . . . oh brother
> A: Now wait a s . . . the conduit is the cover to the wires?
> E: Yes and . . .
> A: Oh I see, there's a part that . . .a part that's supposed to go over it.
> E: Yes.
> A: I see . . . it looks just the right shape too. Ah hah! Yes.
> E: Wonderful, since I did not know how to describe the part.

The problem that arises here is that there is no simple shape-based description for the object the expert needs to identify, so he must find some other shared information on which to base his description (cf. Downing, 1977; Chafe, 1977). The problem is complicated because the expert and apprentice do not share a visual field. If they did, the expert could point (if they and the object being pointed at were all in the same location) or use relative location (e.g., "it's next to the red-handled screwdriver").[9] The expert's solution in this case is to anchor the description on the basis of a past action the apprentice performed and then to describe the object functionally (i.e., to describe its function rather than its shape). Functional descriptions often enable bypassing other more complex descriptions. The statement

---

[8] This segment also illustrates the cooperative nature of task-oriented dialogs: the two participants work together to achieve a shared goal of identifying the object the expert wants the apprentice to locate.

[9] Rubin (1978) describes spatial and temporal commonality between speaker and hearer as two dimensions along which language experiences may differ and considers how these dimensions affect the interpretation of deictic expressions.

"it is used for doing x" or "it has the right shape for doing x" may be used to communicate complex shapes and structures. As always, the success of such descriptions depends on the hearer's ability to determine what such an object is like, or to pick out the object from a set.

The fragment also illustrates the problems that arise when two participants in a dialog have different perspectives on what is being described. The expert's orientation is basically functional; he has a model of what is going on, of how the compressor works, and of how it goes together. His descriptions are based on this model. The apprentice's orientation is basically visual or shape-based. He can see the parts and can tell by trying whether they fit. This discrepancy is even clearer in the following fragment, where from the functional perspective of the expert we get the descriptions "pump" and "cooling fins", while from the shape-based perspective of the apprentice, the same objects are described as "thing with flanges" and "little ribby things":

E: Remove the <u>pump</u> and the belt.
A: Is this <u>thing with flanges on it</u> the pump?
E: Point at "the thing with flanges on it" please.
A: I'm pointing at the thing with flanges on it. These <u>little ribby things</u> are flanges.
E: Yes, the thing you are pointing at is the pump. The little ribby things are <u>cooling fins</u>.

In this fragment, one can see the expert and apprentice working toward a shared view, trying to establish, or check that they have established, a common referent and hence a common focus.[10] An implicit goal in a dialog is to establish this commonality -- the effort this requires is very clear here. One of the ways in which misunderstandings arise is when the participants in a dialog fail to establish this commonality but <u>think</u> they have (this happened with the flywheel and motor pulley in the

---

[10] There is a clear indication at the end of the previous fragment that the expert realizes the importance of shape in the apprentice's orientation: he says he didn't know how to describe the part, apparently meaning that he didn't have a description of its shape (he did describe it functionally and in fact that seems to have worked just fine).

initial dialog fragment). Not only do such mismatches occur, they are difficult to detect and often go unnoticed until a fairly major problem arises.

A further problem that arises in producing a description is deciding how much information to include in it. The linguistic description of an object must distinguish it from all others currently focused on by the speaker and hearer.[11] But the situation is more complicated than this. It is clear from an analysis of the task-oriented dialogs and from other data (Freedle, 1972) that the description of an object seldom contains only the minimal amount of information necessary to distinguish it. Descriptions, like the rest of language, are often redundant.[12] What appears to be the case for physical objects is that the speaker describes an object not in the minimum number of 'bits' of information, but rather in a manner that will enable the hearer to locate the object as quickly as possible. Clear distinguishing features (e.g., color, size, and shape) are part of a description precisely because they eliminate large numbers of wrong objects and hence help the hearer to isolate the correct object more quickly.

The use of redundant information (and not just distinguishing information) to speed up the search for a referent can be seen easily from an example. If someone asks "What tool should I use?" the response "The red-handled one." may not be satisfactory even if there is only one red-handled tool, because processing such a description

---

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

[12] Olson, 1970, p.266, comments on this phenomenon and on the need for further investigation of it.

requires considering too many alternatives. The phrase "the red-handled screwdriver" is more helpful, because it limits the search to screwdrivers. In giving a description that minimizes the time it takes the hearer to identify the referent of a referring expression, a balance must be reached. Too much information is as harmful as too little, since all parts of the description must be processed to make sure the object is the correct one. Furthermore, the hearer may wonder whether he is mistaken if he thinks he has determined the referent but there is more description to process (cf. Grice, 1975). Using the phrase, "the red-handled screwdriver with the small chip on the bottom and a loose handle" to identify the only red-handled screwdriver will probably both increase the hearer's search time and confuse him. Rather than minimize either the communication time (including processing of the description) or the search time alone, the combination of communication time and search time must be minimized. A speaker should be redundant only to the degree that redundancy reduces the total time involved in identifying the referent.

B. Matching a Description

As the preceding discussion illustrates, a major role of descriptions is to point; the speaker is directing the hearer's attention to some entity. For the hearer, focusing is crucial in providing a small set of items from which to choose that entity. Being able to so restrict attention is necessary both for identifying the correct referent (as the interpretation of the phrase "the screw" in the initial dialog fragment illustrates) and constraining search time (see Grosz 1977).

One problem that arises for a hearer, especially a computer system in the role of hearer, is what to do when a reference does not correspond to (or match) any known entity. If the description suffices to distinguish the entity being pointed at from others that are currently focused on, then the mismatch does not matter. But, what does "suffice to distinguish" mean? The question of what kind of mismatch is

significant depends on more than the entities in focus. For example, the difference between yellow and green may not matter when a yellow-green shirt is being distinguished from a red one; it does matter when picking lemons.

In addition, the hearer must decide whether or not an inexact match should even be considered. In the usual use of definite descriptions, to identify some entity in the domain of discourse, inexact matches are always acceptable. Donellan (1966) distinguishes this referential use from an attributive use for which an inexact match is not possible: "In the attributive use, the attribute of being the so-and-so is all important, while it is not in the referential use" (p.102). But the distinction in the terms that Donnellan makes it poses a problem for a hearer, since it is the speaker's intent and not the speaker's beliefs[13] that distinguishes attributive from referential uses of a description. This means that the hearer (whether a person or a computer system) must be able to detect this intent. In certain cases (for example, descriptions of entities that do not yet exist), the attributive use is usually clear. In using the phrase, "the winner of the 1979 Nobel Peace Prize", a speaker is describing a person whose identity is not yet known; there is no other way to describe that person (yet).[14] There are other instances in which the distinction relies on knowledge outside the dialog in which the reference occurs (in particular, what the hearer believes the speaker wants). It seems that for this problem the dialog participants must rely on the potential for clarification available in further dialog. If a hearer misinterprets an attributive use of a description, the speaker can explicitly indicate the need for an exact

---

[13] "A definite description can be used attributively even when the speaker believes that some particular person fits the description, and it can be used referentially in the absence of this belief."(p.111)

[14] There is, of course, the possibility that the speaker meant to say 1977, in which case s/he is referring (wrongly) to an existing entity, but then we are back with the referential case.

match.[15]

To summarize, the importance of focusing to both the interpretation and the generation of definite descriptions comes from the highlighting function it serves. By separating those items currently highlighted from those that aren't, focusing provides a boundary around the entities from which the entity being either described or identified must be distinguished. For generation purposes, this boundary circumscribes those items from which the entity being described must be distinguished, and thus provides some means of determining when a description is complete enough. It is useful for interpretation in providing a small set of items from which to choose. If an exact match cannot be found in focus, it is reasonable to ask if any of the items in focus comes close to matching the definite description and if so, which is the closest.

IV FOCUS IN DISCOURSE: PROSPECTS AND PROBLEMS

The major implication of the role of focusing in dialog for a natural language processing system is that such a system needs mechanisms for focusing. In particular, suppose the system has a knowledge base which encodes the portion of the world the system knows about, and that this knowledge base contains formal elements which stand for entities in that world. Then the system needs a means of highlighting those elements in its knowledge base that correspond to the entities currently focused on and must be able both to use this highlighting (for example, to interpret and generate descriptions) and to change it appropriately as the dialog progresses. This section presents several issues that arise in constructing such a computational model and for each discusses what structures and procedures are needed and what research issues must be resolved.

---

[15] I have ignored a third issue that arises when considering a computer system for natural language processing: the formalism used for encoding knowledge in the system must be adequate for handling attributive descriptions. For a discussion of this issue, see Cohen, 1978 and Webber, 1978.

Grosz (1977) describes focusing mechanisms incorporated in a computer system for understanding task-oriented dialogs. These include structures for highlighting elements of a knowledge base, operations on those structures, procedures that use them for interpreting definite noun phrases, and procedures for updating them. The implementation provides for two kinds of highlighting, explicit and implicit, and uses task information to determine shifts in focus. An explicit focus data structure contains those elements that are relevant to the interpretation of an utterance because they have been discussed in the preceding discourse. In addition, the focusing mechanisms provide for differential access to certain information associated with these elements. In particular, the subactions and objects involved in a task are implicitly highlighted whenever that task is highlighted. That is, implicit focus consists of those elements that are relevant to the interpretation of an utterance because they are closely connected to task-related elements in explicit focus.[16]

There are several directions in which these mechanisms must be extended for a system to be able to handle the general problems posed by focusing and definite descriptions in dialog. First, the only clues to how focusing changes that have been incorporated in the system are clues based on shared knowledge about the structure of entities in the domain (in particular, the structure of the task); linguistic clues and the interaction between different kinds of clues remain to be examined. Second, the highlighting of explicit and implicit focus are used in interpreting definite descriptions, but an exact match is required; the question of what constitutes an inexact match has not yet been faced. Third, although the highlighting structures provide for focusing on different aspects of an entity, the deduction routines do not use this

---

[16] Elements in implicit focus are separated from those in explicit focus for two reasons. First, there are numerous entities implicitly focused on in a dialog, many of which are never referenced. Including the elements corresponding to such entities in the explicit focus data structure would clutter it, weakening its highlighting function. Second, references to implicitly focused entities may indicate a shift of focus to those entities, making it useful to distinguish such references from others.

information in accessing information about an entity in focus. Finally, the question of how the focusing mechanisms interact with representations of belief has not been addressed. The following sections examine the problems posed by each of these extensions in more detail.

A. Ranges of Focusing and Clues to Shifts in Focus

The term focus (as well as theme) is sometimes used (e.g., Halliday, 1967) to refer to prominence in a sentence, a more local phenomenon than focus as discussed here. It is clear that a speaker and hearer are focused not only globally on some set of entities but also more locally, and that this more local focusing affects the way in which a particular idea is expressed in an utterance. This raises the question of how sentential focusing interacts with the more global focusing discussed in this paper. When does the way in which an utterance is phrased not only highlight certain entities, but also change the global focusing of the dialog participants? An answer to this question requires looking more closely at what kinds of clues a speaker can use to shift focus.[17]

A speaker's clues on how to focus may be linguistic or may come from knowledge about the relationships among entities being discussed. Linguistic clues may be either explicit, given directly by certain words, or implicit, deriving from sentential structure or from rhetorical relationships between sentences. In the model described in Grosz (1977), both implicit focus and the procedures for shifting focus are based on clues that derive from knowledge a speaker and hearer share about the structure of the entities being discussed; they use a

---

[17] It is important to note that shifting and focusing are not separable tasks. Focusing is an ongoing process that both influences and is influenced by the interpretation of an utterance. This dynamic aspect of focusing is clear in the interpretation of the phrase "one screw" in utterance (5) of the initial dialog fragment. The focusing established by the expert in utterance (3) highlights a set of screws from which the one screw can be chosen. The reference to one screw shifts focus to the particular subtask of loosening those screws.

representation of the task to decide when and how to shift focus.[18] For the focus mechanisms to be useful for discourse in general, they must be extended to handle the linguistic clues that a speaker may use. In particular, two kinds of implicit linguistic clues must be understood and their use for shifting formalized.

First, there are the global linguistic clues that come from patterns of relationships between sentences, such as paraphrase and elaboration (Grimes, 1975; Halliday and Hasan, 1976). For example, by elaborating on some element of a sentence, a speaker shifts focus to that element (really the entity expressed by that element). A major question here is how to recognize when such patterns occur (cf. Hobbs 1976). Perhaps more important, there is the question of whether recognizing the patterns requires knowing how the focus of attention in the two sentences is related. It may be that such global patterns are more useful in setting expectations about where focus may be in the following utterances than in determining the focus in a particular utterance.

The second kind of implicit clue comes from the syntactic form of an utterance. Sidner (1978) presents rules for determining focus, based on syntactic structure. A particularly important aspect of her work involves the recognition that focusing is only <u>predicted</u> by a single utterance and that the "expected focus" must be confirmed by succeeding utterances. That is, the question of whether an utterance changes global focus cannot be answered on the basis of the individual utterance. Rather, an utterance can only suggest a global shift in focus. This expectation may then be confirmed in a following utterance (if the speaker continues; if the hearer speaks next s/he may choose to accept or reject this shift).

---

[18] The structure need not be that of a task. For example, in describing a house, focus can move from the total house to one of the rooms of the house.

B. Inexact Matches: The Problems that Remain

Before the focusing mechanisms can be extended to handle inexact matches two major problems must be addressed: determining how to decide whether an inexact match is close enough and determining how to decide between accepting an inexact match and considering a shift in focus. For the first problem, focusing makes it possible to determine the closest match, but not to decide whether that match is close enough. For example, if a red ball and a green ball are in focus, then the red ball comes closest to matching the description "the red block" but not close enough to be considered the referent of that phrase. For the second problem, if no exact match can be found in explicit focus the matching procedures must decide whether to accept a referent that inexactly matches a description or to consider the possibility that the speaker wants to focus on some new entity. For example, should a hearer confronted with the phrase "the red spot" in the situation just described look for a red spot on one of the balls? Answers to these questions require research on some fundamental issues in semantics and on speech errors.

C. Focusing and Perspective

Focusing involves not only highlighting certain entities, but also highlighting certain ways of viewing those entities. For example, a doctor may be viewed as a member of the medical profession or as having a role in a family. In the process of focusing on some entity, the speaker also chooses a certain perspective on that entity and, as a result, focuses on that entity from that perspective

---

[19] Fillmore says,

> The point is that whenever we pick a word or phrase, we automatically drag along with it the larger context or framework in terms of which the word or phrase we have chosen has an interpretation. It is as if descriptions of the meanings of elements must identify simultaneously "figure" and "ground".
>
> To say it again, whenever we understand a linguistic expression of whatever sort, we have simultaneously a background scene and a perspective on that scene.

(Fillmore, 1977[19]; Halliday, 1977).

The perspective from which an entity is viewed influences how further information about that entity is accessed. The representation of focus presented in Grosz (1977) allows for differential access to properties of an entity, but this addresses only one part of the problem.[20] Using the initial perspective from which an entity is viewed for differential access does not rule out considering a concept differently from the way it has already been portrayed. Instead, it orders the way in which aspects of the concept are to be examined. One of the problems this raises is deciding when to consider a switch in perspective, when to abandon deriving properties or searching items implicitly focused by an initial perspective and examine other aspects of the entity.

Another problem that relates to perspective is how perspective influences the particular description a speaker chooses. Does global focus give an indication to a speaker of which properties to choose? The preceding fragments of dialog contained several examples that illustrated the effect of differences in how a speaker and hearer were focused on communication. This suggests that focusing, though often quite useful, can cause problems for people; similar problems may be unavoidable in a natural language processing system.

D. Focusing and Beliefs

An additional aspect of focus that has not yet been addressed is its interaction with a representation of beliefs. The dialog fragments in the section on description pointed out some of the problems that arise when the two participants know different things about the entity being described. It is important, then, for a speaker to be able to separate his own beliefs from what he believes his hearer knows or believes. It seems equally clear from the dialogs, however, that focusing is not one of the things that is separate for the two

---

[20] Consequently, the reference resolution mechanisms did not use this feature.

participants. There is a pervasive assumption by speaker and hearer that they share a common focus (this is, in fact, an important part of how and why focusing works). The extension that seems to be needed here is to have the focusing mechanisms interact with an encoding of knowledge that distinguishes beliefs (e.g., Cohen 1978) rather than, as is now the case, with some uniform encoding of knowledge that does not distinguish between speaker and hearer.

V SUMMARY

Focusing is the active process, engaged in by the participants in a dialog, of concentrating attention on, or highlighting, a subset of their shared reality. Not only does it make communication more efficient, it makes communication possible. Speaker and hearer can concentrate on a small portion of what they know and ignore the rest. The importance of focusing to communication is clearly demonstrated by the definite descriptions that are used in dialog. For a natural language processing system to carry on a dialog with a person it must include mechanisms that computationally capture this focusing process. This paper has examined the requirements definite descriptions impose on such mechanisms, discussed focusing mechanisms included in a computer system for understanding task-oriented dialog, and indicated future research problems entailed in modeling the focusing process more generally.

VI ACKNOWLEDGMENTS

The work reported herein was supported by the National Science Foundation under Grant No. MCS 76-22004 and by the Advanced Research Projects Agency of the Department of Defense under Contract No. N00039-78-C-0060. I would like to thank Jerry Hobbs, David Levy, Ann Robinson, Jane Robinson, Candy sidner, and Brian Smith for discussing the ideas in this paper and commenting on various drafts of it.

REFERENCES

Chafe, Wallace L. The Flow of Thought and the Flow of Language. In Proceedings of the Symposium on Discourse and Syntax, Los Angeles, California, November, 1977. In press.

Cohen, Philip R. On Knowing What to Say: Planning Speech Acts. Ph. D. thesis, University of Toronto, Canada. 1978.

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

Donnellan, Keith. Reference and Definite Description. The Philosophical Review, vol. 75, 1966. Reprinted in: Semantics, Danny P. Steinberg and Leon A. Jakobovits, Eds. pp. 100-114. The University Press, Cambridge. 1971.

Downing Pamela A. On "Basic Levels" and the Categorization of Objects in English Discourse. Proceedings of the Third Annual Meeting of the Berkeley Linguistics Society, Berkeley, California, February 1977.

Evans, Thomas G. A Heuristic Program to Solve Geometric-Analogy Problems. Ph. D. thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, May, 1963.

Fillmore, Charles J. The Case for Case Reopened. In: Syntax and Semantics, John P. Kimball, Ed. Academic Press. New York. In press.

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

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

Grice, H. Logic and Conversation. In: Syntax and Semantics, P. Cole and J. Morgan, Eds. Vol. 3, pp. 41-58. Academic Press, New York, 1975.

Grimes, Joseph E. The Thread of Discourse. The Hague, Mouton, 1975.

Grosz, Barbara J. The Representation and Use of Focus in Dialogue Understanding. Ph. D. thesis, University of California, Berkeley, California; also Technical Note No. 151, SRI International, Menlo Park, California, 1977.

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

Halliday, Michael A. Language as Code and Language as Behaviour: A Systemic-functional interpretation of the nature and ontogenesis of dialogue. In: Semiotics of Culture and Language, Sydney M. Lamb and Adam Makkai, Eds. 1977. In press.

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

Hobbs Jerry R. A Computational Approach to Discourse Analysis. Research Report 76-2, Department of Computer Sciences, City College, CUNY, December 1976.

Levy, David M. Communicative Goals and Strategies: Between Discourse and Syntax. In Proceedings of the Symposium on Discourse and Syntax, Los Angeles, California, November, 1977. In press.

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

Rubin, A.D. A Theoretical Taxonomy of the Differences Between Oral and Written Language, In: Theoretical Issues in Reading Comprehension, R. Sprio, B. Bruce and W. Brewer, Eds., Lawrence Erlbaum, Hillsdale, N.J., 1978. Also as Center for the Study of Reading Technical Report No. 35, January 1978.

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

Sidner, Candace L. A Computational Model of Co-reference Comprehension in English. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, forthcoming.

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

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

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

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