

REFERENCE IDENTIFICATION AND REFERENCE IDENTIFICATION FAILURES

Bradley A. Goodman

BBN Laboratories
10 Moulton Street
Cambridge, Massachusetts 02238

The goal of this work is the enrichment of human-machine interactions in a natural language environment.¹ Because a speaker and listener cannot be assured to have the same beliefs, contexts, perceptions, backgrounds, or goals at each point in a conversation, difficulties and mistakes arise when a listener interprets a speaker's utterance. These mistakes can lead to various kinds of misunderstandings between speaker and listener, including reference failures or failure to understand the speaker's intention. We call these misunderstandings miscommunication. Such mistakes can slow, and possibly break down, communication. Our goal is to recognize and isolate such miscommunications and circumvent them. This paper highlights a particular class of miscommunication – reference problems – by describing a case study and techniques for avoiding failures of reference. We want to illustrate a framework less restrictive than earlier ones by allowing a speaker leeway in forming an utterance about a task and in determining the conversational vehicle to deliver it. The paper also promotes a new view for extensional reference.

1 INTRODUCTION

Cohen, Perrault, and Allen (1981) argued that "... users of question-answering systems expect them to do more than just answer isolated questions – they expect systems to engage in conversation. In doing so, the system is expected to allow users to be less than meticulously literal in conveying their intentions, and it is expected to make linguistic and pragmatic use of the previous discourse." Following in their footsteps, we want to build robust natural language processing systems that can detect and recover from miscommunication. The development of such systems requires a study on how people communicate and how they recover from miscommunication. The study of miscommunication is a necessary task for building natural language understanding systems since any computer capable of communicating with humans in natural language must be tolerant of the complex, imprecise, or ill-devised utterances that people often use. This paper summarizes the results of a dissertation (Goodman 1984) that investigated the kinds of miscommunication that occur in human communication, with a special emphasis on **reference problems**: problems a listener has in determining about whom or what a speaker is talking. To cope with such problems, we proposed an algorithm for extending the reference paradigm. We have also implemented computer programs that

demonstrate how one could solve these problems in a natural language understanding system.

Our current research (Sidner et al. 1981, 1983) assumes most dialogue as being cooperative and goal directed: we assume that a speaker and listener are working together to achieve a common goal. In order for the listener to interpret utterances, he must identify the underlying plan or goal that the utterances reflect (Cohen 1978, Allen 1979, Sidner and Israel 1981, Sidner 1985, Carberry 1985, Litman 1985, Pollack 1986). This plan, however, is rarely obvious at the surface sentence level. A central process, therefore, in the interpretation of utterances is the transformation of sequences of complex, imprecise, or ill-devised utterances into *well-specified* plans that might be carried out by dialogue participants. Within this process, miscommunication can occur. In this paper, we are particularly concerned with cases of miscommunication from the hearer's viewpoint, such as when the hearer is inattentive to, confused about, or misled about the intentions of the speaker.

In ordinary exchanges speakers usually make assumptions regarding what their listeners know about a topic of discussion. They will leave out details thought to be superfluous (Appelt 1981, McKeown 1983). Since the speaker really does not know exactly what the listener knows about a topic, it is easy to make statements that can be misinterpreted or not understood by the listener

Copyright 1986 by the Association for Computational Linguistics. Permission to copy without fee all or part of this material is granted provided that the copies are not made for direct commercial advantage and the CL reference and this copyright notice are included on the first page. To copy otherwise, or to republish, requires a fee and/or specific permission.

0362-613X/86/040273-305\$03.00

because not enough details were presented. One principal source of trouble is the descriptions constructed by the speaker to refer to actual objects in the world. A description can be, for a given listener, either imprecise, confused, ambiguous, or overly specific. In addition, it might be interpreted under the wrong context (which can cause one of the problems with the description to occur or can cause the description to successfully refer when it should not have). As a result, reference identification errors² occur; the listener cannot determine what object is being described. The descriptions that cause reference identification failure are a type of “ill-formed” input. The blame for ill-formedness may lie partly with the speaker and partly with the listener. The speaker may have been sloppy or may not have taken the hearer into consideration. The listener may be remiss, unwilling to admit he can’t understand the speaker, or unwilling to ask the speaker for clarification. It may even be the case that the listener does not know that he has misunderstood the speaker.

The interactions that can occur among the speaker’s description, the context of the communication, and the listener’s view of the world, as well as the listener’s own abilities, especially in a task-oriented environment, all contribute to make the reference task more complicated. Our work provides a new way to look at reference that involves a more active, introspective approach to repairing communication. It redefines the notion of finding a referent since previous paradigms have proven inappropriate in the real world (see Section 4 for a detailed discussion).

We introduce a new process to reference called **negotiation** that is used during the reference task to take into account all the language and perceptual knowledge people have about the world, especially when reference fails. We illustrate this process by introducing a new computational model for the reference process called **FWIM**, for “Find What I Mean”. In addition, we develop a theory of the use of extensional descriptions that will help explain how people successfully use imperfect descriptions. This theory is called the theory of extensional reference miscommunication.

The last part of this section provides an introduction to the domain of our work and outlines the methodology used. We also present a description of other relevant research in this domain. Section 2 of this paper briefly highlights some aspects of normal communication and then provides a general discussion on the types of miscommunication that occur in conversation, concentrating primarily on reference problems and motivating many of them with illustrative protocols. The protocols demonstrate the complexity of the reference process and help illuminate the kinds of knowledge sources people consult when performing reference. Section 3 describes those knowledge sources in more detail, providing information about the language and physical knowledge that

people use to perform reference identification and to recover from reference failure. Section 4 presents initial solutions to some of the problems of miscommunication in reference. Motivated there is a partial implementation of a reference mechanism that attempts to overcome many reference problems. Finally, we conclude in Section 5 with a summary and suggestions for future research.

1.1 THE DOMAIN AND METHODOLOGY

We are following the task-oriented paradigm of Grosz (1977) since

- it is easy to study (through videotapes),
- it places the world in front of you (a primarily extensional world), and
- it limits the discussion while still providing a rich environment for complex descriptions.

The task chosen as the target for the system is the assembly of a toy water pump. The water pump is reasonably complex, containing four subassemblies built from plastic tubes, nozzles, valves, plungers, and caps that can be screwed or pushed together. A large corpus of dialogues concerning this task was collected by Cohen (see Cohen 1981, 1984; Cohen, Fertig, and Starr 1982). These dialogues contained instructions from an “expert” to an “apprentice” that explain the assembly of the toy water pump. Both participants were working to achieve a common goal – the successful assembly of the pump. This domain is rich in perceptual information, allowing for complex descriptions of its elements. The data provide examples of imprecision, confusion, and ambiguity as well as attempts to correct these problems.

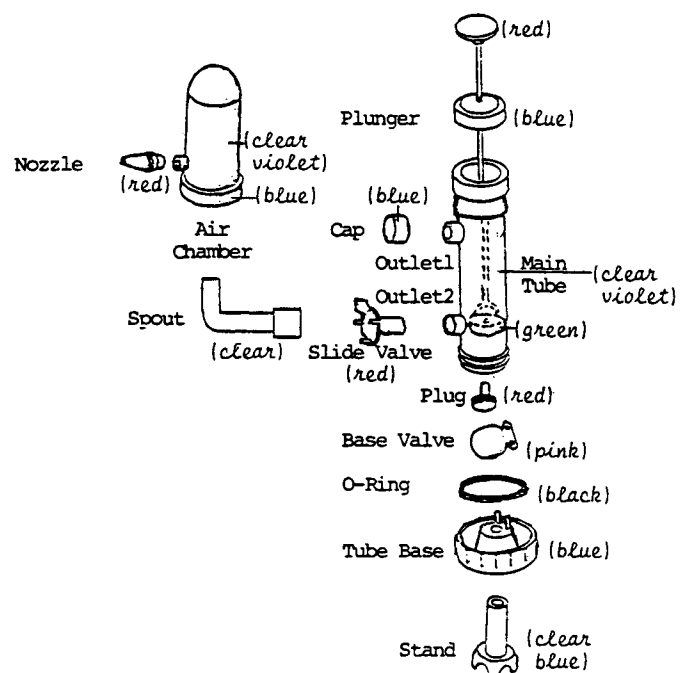


Figure 1a. The toy water pump.

The following exchange exemplifies one such situation. In it, E is instructing A to assemble part of the water pump. E and A are communicating verbally but neither can see the other. (The bracketed text in the excerpt tells what was actually occurring while each utterance was spoken.) Notice the complexity of the speaker's descriptions and the resultant processing required by the listener. This dialogue illustrates that listeners

- repair the speaker's description in order to find a referent,
- repair their initial reference choice once they are given more information, and
- can fail to choose a proper referent.

In Line 7, E describes the two holes on the BASEVALVE as "the little hole". A must repair the description, realizing that E doesn't really mean "one" hole but is referring to the "two" holes. A apparently does this since he doesn't complain about E's description and correctly attaches the BASEVALVE to the TUBEBASE.

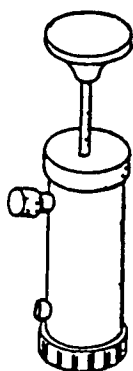


Figure 1b. Configuration of the pump after the TUBEBASE is attached to the MAINTUBE (Line 10).

In Line 13, A interprets "a red plastic piece" to refer to the NOZZLE. When E adds the relative clause "that has four gizmos on it," A is forced to drop the NOZZLE as the referent and to select the SLIDEVALVE. In Lines 17 and 18, E's description "the other – the open part of the main tube, the lower valve" is ambiguous, and A selects the wrong site, namely the TUBEBASE, in which to insert the SLIDEVALVE. Since the SLIDEVALVE fits, A doesn't detect any trouble. Lines 20 and 21 keep A from thinking that something is wrong because the part fits loosely. In Lines 27 and 28, A indicates that E has not given him enough information to perform the requested action. In Line 30, A further compounds the error in Line 18 by putting the SPOUT on the TUBEBASE.

Excerpt 1 (Telephone)

- E: 1. Now there's a blue cap
[A grabs the TUBEBASE]
2. that has two little teeth sticking

3. out of the bottom of it.

A: 4. Yeah.

E: 5. Okay. On that take the

6. bright shocking pink piece of plastic
[A takes BASEVALVE]

7. and stick the little hole over the teeth.

[A starts to install the BASEVALVE, backs off, looks at it again and then goes ahead and installs it]

A: 8. Okay.

E: 9. Now screw that blue cap onto

10. the bottom of the main tube.

[A screws TUBEBASE onto MAINTUBE]

A: 11. Okay.

E: 12. Now, there's a–

13. a red plastic piece

[A starts for NOZZLE]

14. that has four gizmos on it.

[A switches to SLIDEVALVE]

A: 15. Yes.

E: 16. Okay. Put the ungizmoed end in the uh

17. the other–the open

18. part of the main tube, the lower valve.

[A puts SLIDEVALVE into hole in TUBEBASE, but E meant OUTLET2 of MAINTUBE]

A: 19. All right.

E: 20. It just fits loosely. It doesn't

21. have to fit right. Okay, then take

22. the clear plastic elbow joint.

[A takes SPOUT]

A: 23. All right.

E: 24. And put it over the bottom opening, too.

[A tries installing SPOUT on TUBEBASE]

A: 25. Okay.

E: 26. Okay. Now, take the–

A: 27. Which end am I supposed to put it over?

28. Do you know?

E: 29. Put the–put the–the big end–

30. the big end over it.

[A pushes big end of SPOUT on TUBEBASE, twisting it to force it on]

The example illustrates the complexity of reference identification in a task-oriented domain. It shows that people do not always give up when a speaker's description isn't perfect (or isn't readily assimilable for them), but that they try to plow ahead anyway. The rest of this paper formalizes the kinds of problems that occur during reference and then extends the reference paradigm to get around many of the problems.

1.2 RELATED WORK IN REFERENCE AND MISCOMMUNICATION

There are two major pieces of work in AI literature that laid the foundation for our research: those in reference and those in miscommunication.

Cohen (1981, 1984) presents a detailed analysis of the pragmatics of reference and the effects of different modalities of communication. His work was a major starting point of this research. It showed that it was reasonable to consider reference identification as separate from the whole process of language understanding instead of being too intimately tangled to consider on its own. There is evidence presented by Cohen (1981, 1984) that a speaker attempts as a separate step in his overall plan of communication to get a hearer to identify a referent. He provided grounds for an IDENTIFY action by illustrating particular requests to identify from his water pump protocols. For example, utterances like "Notice the two side outlets on the tube end" or "Find the rubber ring shaped like an O" showed that the speaker wanted the hearer to perform some kind of action. That action is the IDENTIFY act, which is to search the world for a referent for the speaker's description (and thus identify it). Cohen also showed that the hearer's response to a request to identify provided further evidence. He pointed out excerpts in the protocols where hearers responded to a request to identify with a confirmation that the identification had actually occurred (e.g., "Got it."). Cohen went on to show how reference fits into a plan-based theory of communication.

The reference paradigm we followed was closest to that developed by Grosz (1977). Her basic reference identification paradigm was similar to that of many others in the past (e.g., Winograd 1971, Woods 1972): put the speaker's description into a searchable form (i.e., parse and semantically interpret the speaker's description) and then use that form as a pattern that can be compared against objects (i.e., the possible referents) in the world. A referent is found when a match occurs between the pattern and one or more of the objects. The pattern and a target referent match each other if *all* the attributes specified in the pattern exactly fit the corresponding attributes in the target. There is variability in each of the past reference schemes in what pattern is generated, how the world is represented, and how the actual search progresses, but the general scheme remains the same. Success in all cases occurs if and only if a perfect match exists between all the pattern's attributes and the corresponding attributes on a target. Grosz's reference mechanism departed from past works by introducing the notion of focus. Focus provides a better way to resolve referents by constraining the search space. For definite noun phrases, the choice of possible referent candidates is guided by the focus mechanism. The information provided in the definite noun phrase itself (i.e., by the head noun and modifiers) is used to distinguish the referent from other objects in focus. Grosz showed how

both the surrounding non-linguistic environment and the global context of preceding discourse are part of focus and how it is used to resolve definite noun phrases. Grosz (1977:161) also proposed the need for inexact matching in the reference process should something go wrong:

The retrieval component can fail to find such a match even though for most people the noun phrase suffices to identify an object. ... Alternatively, more than one object may match, but the ambiguity may not matter for the purposes of the utterance. The problem in either case is to determine the nature of the mismatch and whether it matters. ... The focus mechanism provides one crucial element for deciding about inexact matches. It separates those items that are in the focus of attention from all other known items. If an exact match cannot be found in focus, it is reasonable to ask if any of the items in focus come close to matching the description of the noun phrase (the question of what is close is the other crucial element in such decisions) and if so which is closest.

Ringle and Bruce (1981) present a survey of numerous types of miscommunication in conversation. They point out problems across a wide spectrum of dialogue types and situations. Two primary ways that conversation fail are described by them. The first one, **input failure**, occurs when the listener is unable to form a complete or at least coherent interpretation for an utterance. Input failure can occur due to such causes as misinterpretation of a single word, incorrect resolution of a referential term, and misplacement of a negation. Such failures cause the listener to misunderstand without weakening the listener's comprehension of the overall context of the communication (making the failures local in nature). The second way that Ringle and Bruce say that people fail, **model failure**, happens when the listener cannot incorporate the inputs into a coherent belief model as intended by the speaker. The problem could be due to an input failure when information is lost that is needed to assimilate the speaker's utterances into the belief model. It can also occur when a listener does not have sufficient background knowledge, has a different thematic emphasis than the speaker, or fails to make the proper inference (or any at all) from the speaker's input. Ringle and Bruce describe repair techniques that often occur between the listener and speaker when a failure occurs. Such repairs are usually initiated by the listener providing a failure cue (e.g., recapitulating the speaker's important points) to the speaker to indicate possible trouble. The repairs primarily require action by both the listener and the speaker. Sometimes the dialogue situation affects the ability of the listener to provide such cues. For example, in a teacher-student relationship, it is hard for the student to interrupt the teacher's lecture/conversation to initiate a repair due to a mistake the student feels has occurred. In other conversational settings, such interruptions are easier.

McCoy (1985a, 1985b) focuses on a particular class of communication problems. She considers misconceptions about the objects modelled by a system in its

knowledge base. She is concerned with discrepancies between the beliefs of the system and that of the user as seen in a system/user dialogue. Her work concentrates on two kinds of misconceptions about the properties of an object: misclassification and misattribution.

Misclassification occurs when one classifies an object incorrectly. For example, a person may think that whales are fish when in fact they are mammals (McCoy 1985b:17). McCoy called the way to correct this problem the **like-super** strategy since an expert may believe that the user misclassified the misconception object (*whale*) because it is similar to the posited superordinate (*fish*). She defines two other kinds of misclassifications that her system can detect: **Like-Some-Super** and **No-Support**. Like-Some-Super occurs when the expert believes a user wrongly classified an object because it is like some subclass of the posited superordinate. For example, a whale may be viewed by someone as a fish because they think that a whale is like a shark, and a shark is a fish (McCoy 1985b:24). No-Support occurs when the system can find no support in the user model for the misclassification. McCoy's system simply denies the incorrect information in that case and provides the correct information.

Misattribution is the second class of misconceptions with which McCoy deals. They occur when the user wrongly attributes a property to an object that the object doesn't have. One reason that misattribution can occur is that the user either has confused the object with one the user thinks is a similar object or has made a bad analogy from a similar object (the "Wrong Object" strategy). McCoy presents an example where the user attributes the "high liquidity" property of a money market fund to a money market certificate. Another reason that misattribution can occur is that the user attributes to an object a related property instead of the actual one (the "Wrong Attribute" strategy). An example that McCoy presents for this strategy occurs when the user talked about the "interest" on the stock but really meant the "dividend." The correction in that case is the substitution of the proper property for the incorrect one. The last case of misattribution that McCoy considers is No-Support. It occurs when the expert can find no support for the misattribution in his model of the user. In that case, McCoy's system denies the incorrect information and asserts the correct information.

McCoy's work demonstrates the power of representing objects using a taxonomic knowledge base that indicates an object's superordinates and subtypes, and its attributes and their values. That paradigm allows her to notice several classes of user's misconceptions and to correct them. Her solutions blend in nicely with the relaxation mechanism motivated and described in this paper.

2 MISCOMMUNICATION

People must and do manage to resolve lots of (potential) miscommunication in everyday conversation. Much of it seems to be resolved subconsciously – with the listener unconcerned that anything is wrong. Other miscommunication is resolved with the listener *actively* deleting or replacing information in the speaker's utterance until it fits the current context. Sometimes this resolution is postponed until the questionable part of the utterance is actually needed. Still, when all these fail, the listener can ask the speaker to clarify what was said.³

In this section we present evidence that people do miscommunicate and yet they often manage to repair reference failures. We look at specific forms of miscommunication and describe ways to detect them. We highlight relationships between different miscommunication problems and demonstrate ways for resolving some of them. The different kinds of miscommunication we present directly motivate the need by listeners for many of the types of knowledge we describe in Section 3.

2.1 CAUSES OF MISCOMMUNICATION

This section motivates a paradigm for the kinds of conversation that we studied and points out places in the paradigm that leave room for miscommunication.

EFFECTS OF THE STRUCTURE OF TASK-ORIENTED DIALOGUES

Task-oriented conversations have a specific goal to be achieved: the performance of a task (e.g., the air compressor assembly in Grosz (1977)). The participants in the dialogue can have the same skill level and they can work together to accomplish the task; or one of them, the expert, could know more and could direct the other, the apprentice, to perform the task. We have concentrated primarily on the latter case – due to the protocols that we examined – but many of our observations can be generalized to the former case.

The viewpoints of the expert and apprentice differ greatly in apprentice-expert exchanges. The expert, having an understanding of the functionality of the elements in the task, has more of a feel for how the elements work together, how they go together, and how the individual elements can be used. The apprentice normally has no such knowledge and must base his decisions on perceptual features such as shape (Grosz 1981). These differences can lead to problems.

The structure of the task affects the structure of the dialogue (Grosz 1977), particularly through the center of attention of the expert and apprentice during the accomplishment of each step of the task. The common center of attention of the dialogue participants is called the focus (Grosz 1977, Reichman 1978, Sidner 1979). Shifts in focus correspond to shifts between the tasks and subtasks; e.g., the objects in a task and the subpieces of each object. Focus and focus shifts are governed by many rules (Grosz 1977, Reichman 1978, Sidner 1979).

Confusion may result when expected shifts do not take place. For example, if the expert changes focus to some object but never bothers to talk about the object reasonably soon after its introduction (i.e., between the time of its introduction and its use, without digressing in a well-structured way in between (see Reichman 1978)), or never discusses its subpieces (such as an obvious attachment surface), then the apprentice may become confused, leaving him likely to misunderstand further utterances. The reverse influence between focus and objects can lead to trouble, too. A shift in focus by the expert that does not have a manifestation in the apprentice's world will also perplex the apprentice.

Focus also influences how descriptions are formed (Grosz 1981, Appelt 1981). The level of detail required in a description depends directly on the elements currently in focus. If the object to be described is similar to other elements in focus, the expert must be more specific in the formulation of the description or may consider shifting focus away from the confusing objects.

2.2 INSTANCES OF MISCOMMUNICATION

Figure 2 outlines some of the ways people get confused during a conversation. These instances were derived from analyzing the water pump protocols. We only discuss referent confusion in this paper. The other forms of confusion – Action, Goal, and Cognitive Load – are described in Goodman (1982, 1984). The confusions themselves, coupled with the description at the end of this section on how to recognize when one of them is occurring and the knowledge people use to perform reference described in Section 3, provide motivation for the use of the algorithm outlined in Section 4 as a means for repairing communication problems.

We illustrate here many of the confusions in the taxonomy through numerous excerpts. Each excerpt has marked in parentheses the modality of communication

that was used in the excerpt (face-to-face, over the telephone, and so forth). A description about the collection of these excerpts can be found in Cohen (1984). Each bracketed portion of the excerpt explains what was occurring at that point in the dialogue.

ERRONEOUS SPECIFICITY

A speaker can be over- or underspecific in his descriptions (which violates Grice's (1975) maxim of quantity). Such descriptions are a form of erroneous specificity that can lead to mistakes on the part of the listener even though, technically, nothing is wrong with the description.

A request is overspecific if extra details are given that seem obvious to the listener (Grosz 1978). Since the listener would not expect the speaker to provide him with obvious details, the listener might become confused; thinking that he had done something incorrectly as the task seemed easier than the one apparently described by the speaker.⁴ For example, in Excerpt 2, E's description of the bubbled piece (i.e., the AIRCHAMBER) is overspecific because it supplies many more features than needed to identify the piece. The extra description in Lines 15 to 17 confused the listener, who appeared to have correctly identified the piece by Line 13 but ended up taking the wrong one when the expert kept adding more details. See Excerpt 10 in the section on bad analogies for other related examples of overspecificity.

Excerpt 2 (Telephone)

- E: 1. Okay?
 2. Now you have two devices that
 3. are clear plastic.
 [A picks up MAINTUBE and SPOUT]
- A: 4. Okay.
- E: 5. One of them has two openings
 6. on the outside with threads on

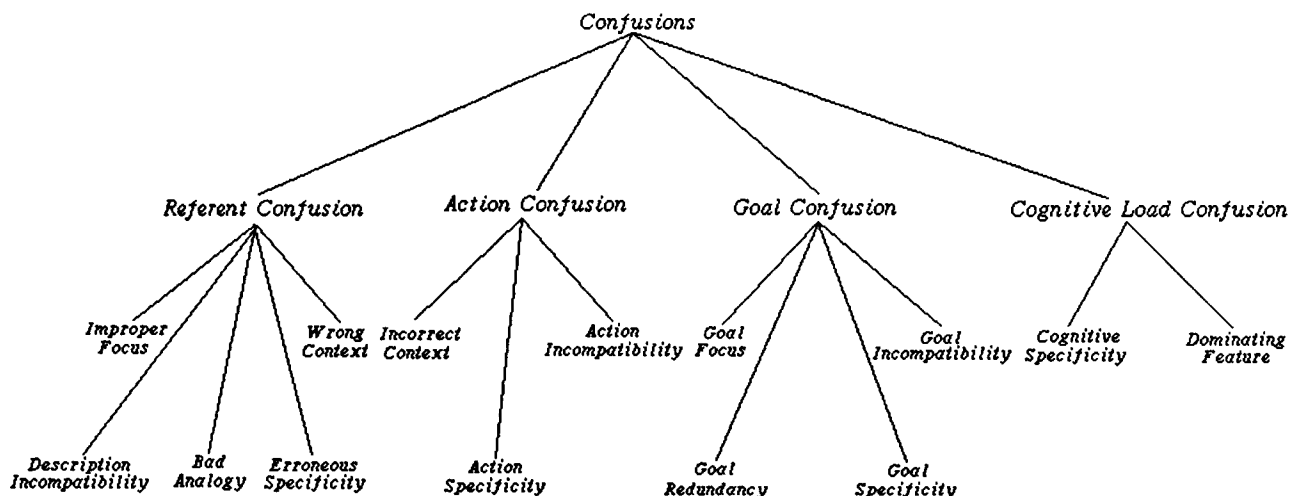


Figure 2. A taxonomy of confusions.

7. the end, and its about five
8. inches long.

[A rotates MAINTUBE confirming E's description]

9. Do you see that?

A: 10. Yeah.

E: 11. Okay,

12. the other one is a bubbled
13. piece with a blue base on it
14. with one spout.

[A looks at AIRCHAMBER]

15. Do you see it?

16. About two inches long.

[A picks up STAND and drops MAINTUBE]

17. Both of these are tubular.

[A puts down SPOUT]

A: 18. Okay.

19. not the bent one.

Ambiguous descriptions are underspecified and can cause confusion about the referent. Excerpt 3 below illustrates a case where the speaker's description is underspecified – it does not provide enough detail to prune the set of possible referents to one.

Excerpt 3 (Face-to-Face)

- E: 1. And now take the little red
2. peg,

[A takes PLUG]

3. Yes,
4. and place it in the hole at the
5. green end,

[A starts to put PLUG into OUTLET2 of MAINTUBE]

6. no

7. the—in the green thing

[A puts PLUG into green part of PLUNGER]

A: 8. Okay.

In Line 4 and 5, E describes the location to place a peg into a hole by giving spatial information. Since the location is given relative to another location by “in the hole *at the green end*,” it defines a region where the peg might go instead of a specific location. In this particular case, there are three possible holes to choose from that are near the green end. The listener chooses one – the wrong one – and inserts the peg into it. Because this dialogue took place face to face, E is able to correct the ambiguity in Lines 6 and 7.

An underspecified description can be imprecise in many possible ways.

- A description may consist of features that do not readily apply or that are inappropriate in the domain. In Line 3, Excerpt 4, the feature “funny” has no meaning to the listener here. It is not until E provides a fuller

description in Lines 5 to 8 that A is able to select the proper piece.

- It may use imprecise feature values. For example, one could use an imprecise head noun coupled with few or no feature values (and context alone does not necessarily suffice to distinguish the object). In Excerpt 5, Lines 8 and 9, “attachment” is imprecise because all objects in the domain are attachable parts. The expert's use of “attachment” was most likely to signal the action the apprentice can expect to take next. The use of the feature value “clear” provides little benefit either because three clear, unused parts exist. The size descriptor “little” prunes this set of possible referents to two contenders.

Another use of imprecise feature values occurs when enough feature values are provided but at least one value is too imprecise. In Excerpt 6, Line 3, the use of the attribute value “rounded” to describe the shape does not sufficiently reduce the set of four possible referents (though, in this particular instance, A correctly identifies it) because the term is applicable to numerous parts in the domain.⁵ A more precise shape descriptor such as “bell-shaped” or “cylindrical” would have been more beneficial to the listener.

Excerpt 4 (Telephone)

E: 1. All right.

2. Now.

3. There's another funny little

4. red thing, a

[A is confused, examines both NOZZLE and SLIDEVALVE]

5. little teeny red thing that's

6. some—should be somewhere on

7. the desk, that has um—there's

8. like teeth on one end.

[A takes SLIDEVALVE]

A: 9. Okay.

E: 10. It's a funny-loo-hollow,

11. hollow projection on one end

12. and then teeth on the other.

Excerpt 5 (Teletype)

E: 1. take the red thing with the

2. prongs on it

3. and fit it onto the other hole

4. of the cylinder

5. so that the prongs are

6. sticking out

A: 7. ok

E: 8. now take the clear little

9. attachment

10. and put on the hole where you

11. just put the red cap on

12. make sure it points

13. upward
A: 14. ok

Excerpt 6 (Teletype)

- E: 1. Ok,
2. put the red nozzle on the outlet
3. of the rounded clear chamber
4. ok?
A: 5. got it.

IMPROPER FOCUS

Earlier we talked about focus and problems that occur due to it. In this section, we discuss how misfocus can cause misreference. Focus confusion can occur when the speaker sets up one focus and then proceeds with another one without letting the listener know of the switch (i.e., a focus shift occurs without any indication). The opposite phenomenon can also happen – the listener may feel that a focus shift has taken place when the speaker actually never intended one. These really are very similar – one is viewed more strongly from the perspective of the speaker and the other from that of the listener.

Excerpt 7 illustrates an instance of the first type of focus confusion. In the excerpt, the speaker (E) shifts focus without notifying the listener (A) of the switch. As the excerpt begins, A is holding the TUBEBASE. E provides in Lines 1 to 16 instructions for A to attach the CAP and the SPOUT to OUTLET1 and OUTLET2, respectively, on the MAINTUBE. Upon A's successful completion of these attachments, E switches focus in Lines 17 to 20 to the TUBEBASE assembly and requests A to screw it on to the bottom of the MAINTUBE. While A completes the task, E realizes (Line 22) she left out a step in the assembly – the placement of the SLIDEVALVE into OUTLET2 of the MAINTUBE before the SPOUT is placed over the same outlet. E attempts to correct her mistake by requesting (Line 23) A to remove "the plas"⁶ piece. Since E never indicated a shift in focus from the TUBEBASE back to the SPOUT, A interprets "the plas" to refer to the TUBEBASE.

Excerpt 7 (Face-to-Face)

- E: 1. And place
2. the blue cap that's left
 [A takes CAP]
3. on the side holes that are
4. on the cylinder,
 [A lays down TUBEBASE]
5. the side hole that is farthest
6. from the green end.
 [A puts CAP on OUTLET1 of
 MAINTUBE]
A: 7. Okay.
E: 8. And take the nozzle-looking
9. piece,
 [A grabs NOZZLE]
10. no

11. I mean the clear plastic one,
 [A takes SPOUT]
12. and place it on the other hole
 [A identifies OUTLET2 of
 MAINTUBE]
13. that's left,
14. so that nozzle points away
15. from the
 [A installs SPOUT on OUTLET2 of
 MAINTUBE]
16. right.
A: 17. Okay.
E: 18. Now
19. take the
20. cap base thing
 [A takes TUBEBASE]
21. and screw it onto the bottom,
 [A screws TUBEBASE on
 MAINTUBE]
22. ooops,
 [E realizes she has forgotten to have
 A put SLIDEVALVE into
 OUTLET2 of MAINTUBE]
23. un-undo the plas
 [A starts to take TUBEBASE off
 MAINTUBE]
24. no
25. the clear plastic thing that I
26. told you to put on
 [A removes SPOUT]
27. sorry.
28. And place the little red thing
 [A takes SLIDEVALVE]
29. in there first,
 [A inserts SLIDEVALVE into
 OUTLET2 of MAINTUBE]
30. it fits loosely in there.

Excerpt 8 demonstrates the latter type of focus confusion that occurs when the speaker (E) sets up one focus – the MAINTUBE, which is the correct focus in this case – but then proceeds in such a manner that the listener (A) thinks a focus shift to another piece, the TUBEBASE, has occurred. Thus, Line 15, "a bottom hole," refers to "the lower side hole in the MAINTUBE" for E and "the hole in the TUBEBASE" for A. A has no way of realizing that he has focused incorrectly unless the description as he interprets it doesn't have a real world correlate (here something does satisfy the description so A doesn't sense any problem) or if, later in the exchange, a conflict arises due to the mistake (e.g., a requested action can not be performed). In Line 31, A inserts a piece into the wrong hole because of the misunderstanding in Line 15. Line 31 hints that A may have become suspicious that an ambiguity existed somewhere

in the previous conversation, but since the task appeared to be successfully completed (i.e., the red piece fit into the hole in the base), and since E did not provide any clarification, he assumed he was correct.

Excerpt 8 (Telephone)

- E: 1. Um now.
 2. Now we're getting a little
 3. more difficult.
- A: 4. (laughs)
- E: 5. Pick out the large air tube
 [A picks up STAND]
 6. that has the plunger in it.
 [A puts down STAND, takes
 PLUNGER/MAINTUBE assembly]
- A: 7. Okay.
- E: 8. And set it on its base,
 [A puts down MAINTUBE, standing
 vertically, on the TABLE]
 9. which is blue now,
 10. right?
 [A has shifted focus to the
 TUBEBASE]
- A: 11. Yeah.
- E: 12. Base is blue.
 13. Okay,
 14. Now
 15. You've got a bottom hole still
 16. to be filled,
 17. correct?
- A: 18. Yeah.
 [A answers this with MAINTUBE still
 sitting on the TABLE; he shows no
 indication of what hole he thinks is
 meant – the one on the
 MAINTUBE, OUTLET2, or the one
 in the TUBEBASE]
- E: 19. Okay.
 20. You have one red piece
 21. remaining?
 [A picks up MAINTUBE assembly and
 looks at TUBEBASE, rotating the
 MAINTUBE so that TUBEBASE is
 pointed up, and sees the hole in it;
 he then looks at the SLIDEVALVE]
- A: 22. Yeah.
- E: 23. Okay.
 24. Take that red piece.
 [A takes SLIDEVALVE]
 25. It's got four little feet on
 26. it?
- A: 27. Yeah.
- E: 28. And put the small end into
 29. that hole on the air tube–

30. on the big tube.

A: 31. On the very bottom?

[A starts to put it into the bottom
 hole of TUBEBASE – though he
 indicates he is unsure of himself]

E: 32. On the bottom,
 33. Yes.

Misfocus can also occur when the speaker inadvertently fails to distinguish the proper focus because he did not notice a possible ambiguity; or when, through no fault of the speaker, the listener just fails to recognize a switch in focus indicated by the speaker. Excerpt 8 is an example of the first type because E failed to notice that an ambiguity existed since he never explicitly brought the TUBEBASE either into or out of focus. He just assumed that A had the same perspective as he had – a perspective in which no ambiguity occurred.

WRONG CONTEXT

Context differs from focus. The context of a portion of a conversation is concerned with the intention of the discussion in that fragment and with the set of objects relevant to that discussion, though not attended to currently. Focus pertains to the elements currently being attended to in the context. For example, two people can share the same context but have different focus assignments within it – we're both talking about the water pump, but you're describing the MAINTUBE and I'm describing the AIRCHAMBER. Alternatively, we could just be using different contexts – I think you're talking about taking the pump apart, but you're talking about replacing the pump with new parts – in both cases we may be sharing the same focus – the pump – but our contexts are totally off from one another.⁷ The kinds of misunderstandings that can occur because of context inconsistencies are similar to those for focus problems:

- the speaker might set up or use one context for a discussion and then proceed in another one without effectively letting the listener know of the change,
- the listener may feel a change in context has taken place when in fact the speaker never intended one, or
- the listener fails to recognize an indicated context switch by the speaker.

Context affects reference identification because it helps define the set of available objects that are possible contenders for the referent of the speaker's descriptions. If the contexts of the speaker and listener differ, then misreference might result.

BAD ANALOGY

An analogy (see Gentner 1980) for a discussion on analogies) is a useful way to help describe an object by attempting to be *more* precise by using shared past experience and knowledge – especially shape and functional information. If that past experience or knowledge doesn't contain the information the speaker assumes it does, then trouble occurs. Thus, one more way referent confusion

can occur is by describing an object using a poor analogy.

An analogy can be improper for several reasons. It might not be specific enough – confusing the listener because several potential referents might conform to the analogy. Alternatively, the analogy might fail because discovering a mapping between the analogous object and something in the environment is too difficult. In Excerpt 9, A at first has trouble correctly satisfying E's functional analogy "stopper" in "the big blue stopper," but finally selects what he considers to be the closest match to "stopper." The problem for A was that E's functional analogy was not specific enough. It would have been better to use *cap* instead of *stopper*.

Excerpt 9 (Telephone)

- E: 1. Okay. Now,
 2. take the big blue
 3. stopper that's laying around
 [A grabs AIRCHAMBER]
 4. ... and take the black
 5. ring–
- A: 6. The big blue stopper?
 [A is confused and tries to communicate it to E; he is holding the AIRCHAMBER here]
- E: 7. Yeah,
 8. the big blue stopper
 9. and the black ring.
 [E drops AIRCHAMBER and takes the O-RING and the TUBEBASE]

In other cases the analogy might be too specific – confusing the listener because none of the available referents appear to fit it. In Line 8 of Excerpt 7, "nozzle-looking" forms a poor shape analogy because the object being referred to actually is an elbow-shaped spout and not a nozzle. The "nozzle-looking" part of the description convinced the listener that what he was looking for was something identified by the typical properties of a nozzle (which is a small tube used as an outlet). However, sometimes when an object is a clear representative of a specified analogy class, the apprentice will not tend to select it as the intended referent. He would assume that, to refer to that object, the expert would not bother to form an analogy instead of just directly describing the object as a member of the class. Hence, the apprentice may very well ignore the best representative of the class for some less obvious exemplar. Given the case just mentioned, it is therefore better to say *nozzle* instead of *nozzle-looking*. In Excerpt 10, the description "hippopotamus face shape" (a shape analogy) in Lines 2 and 3, and "champagne top" (a shape analogy) in Line 9, are too specific and the listener is unable to easily find something close enough to match either of them. He can't discover a mapping between the object in the analogy and one in the real world (a

discussion on discovering such mappings can be found in Gentner (1980)). In fact, when this excerpt was played back to one listener, he was so overwhelmed by E's descriptions that he exclaimed "What!" when he heard them and was unable to correctly proceed.

Excerpt 10 (Audiotape)

- E: 1. take the bright pink flat
 2. piece of hippopotamus face
 3. shape piece of plastic
 4. and you notice that the two
 5. holes on it
 [E is trying to refer to BASEVALVE]
 6. match
 7. along with the two
 8. peg holes on the
 9. champagne top sort of
 10. looking bottom that had
 11. threads on it
 [E is trying to refer to TUBEBASE]

2.3 DETECTING REFERENCE MISCOMMUNICATION

The previous section illustrated some of the ways reference miscommunication occurs. Part of our research, however, has been to examine how a listener discovers the need for a repair of a description during communication. The incompatibility of a description or action with the scene is the strongest signal of possible trouble.

DESCRIPTION INCOMPATIBILITY

The strongest hint that there is a description incompatibility occurs when the listener finds no real world object to correspond to the speaker's description (i.e., referent identification fails). This can occur when the description does not agree with the current state of the world:

- when one or more of the specified feature values in the description are not satisfied by *any* of the pieces (e.g., saying "the orange cap" when none of the objects are orange);
- when one or more specified constraints do not hold (e.g., saying "the red plug that fits *loosely*" when all the red plugs attach tightly); or
- if no *one* object satisfies *all* of the features specified in the description (i.e., there is, for each feature, an object that exhibits the specified feature value, but no one object exhibits all of the values). In Lines 7 and 8 of Excerpt 10 above, E's description of "the two peg holes" leads to bewilderment for the listener because the "champagne top sort of looking bottom that had threads on it" (i.e., the TUBEBASE) has no holes in it. E actually meant "two pegs".

An impossible reference might not only suggest a mistake in the speaker's description but it could instead indicate an earlier action error (e.g., two parts were put together improperly or never had been intended to be assembled together).

With respect to actual reference mechanisms, description incompatibility means that a referent could

not be found. The reference mechanism was not able to find a match between its representation of the speaker's description and the representations of the objects in the world (i.e., the possible referents). Section 4.2.1 provides details on how our reference mechanism attempts such a match.

ACTION INCOMPATIBILITY

An action incompatibility problem is likely if

- the listener cannot perform the action specified by the speaker because of some obstacle;
- the listener performs the action but does not arrive at its intended effect (i.e., a specified or default constraint isn't satisfied); or
- the current action affects a previous action in an adverse way, yet the speaker has given no sign of any importance to this side-effect.

Such action incompatibility might indicate an earlier misreference (e.g., the wrong part was chosen and used in an earlier action).

The detection of most misreferences isn't so hard – the difficult part is determining *why* there is a problem so that the problem can be repaired. The problem could be one of the many illustrated in this section. The knowledge sources described in the next section help provide a better handle for determining the problem with the speaker's description.

3 KNOWLEDGE FOR REFERENCE

This section describes the language and physical knowledge that people use to perform reference identification and to recover from reference failure. The classification of knowledge sources and the observations on how to perform reference and to recover from reference failures were motivated from the analysis of the excerpts in the previous section. Those observations have been formalized as a set of metarules (which we call **relaxation rules**) that are used both to guide the reference process and to determine when to delete or modify portions of a speaker's description. Section 4 presents those rules in the context of the reference and miscommunication recovery mechanism. We feel that the knowledge sources motivated in this section carry across different people and domains. However, we recognize that the particulars described within each knowledge source are not universal and can vary across people and domains. For example, we would expect a difference in the knowledge used by two experts communicating as opposed to that employed by a novice and an expert.

3.1 KNOWLEDGE FOR REPAIRING DESCRIPTIONS

When things go wrong during a conversation, people have many sources of knowledge that they bring to bear to get around the problem (e.g., see Ringle and Bruce 1981). Much of the time the repairs are so natural that we aren't conscious that they have taken place. At other times, we must make an effort to correct what we have

heard, or determine that we need clarification from the speaker. Either repair process involves the use of knowledge about conversation, social conventions, and the world around us.

In this work, we chose to consider the repair of descriptions rather than complete utterances. The most relevant knowledge for repairing descriptions is the conversation itself and the real world described therein (as illustrated by the excerpts in Section 2.2). This knowledge can be broken down into numerous forms.

- **Linguistic knowledge** is the knowledge that expresses the use of the structure and meaning of a description.
- **Perceptual knowledge** is composed of information about a person's abilities to distinguish feature values, his preferences in features and feature values (i.e., what features are most important to him in this domain), and his extraction of information from the internal representation of his perception of an object.
- **Discourse knowledge** is concerned with how a person interprets the flow of conversation and its effects on highlighting relevant parts of the world.
- **Hierarchical knowledge** is concerned with the use of knowledge about generality and specificity of descriptions to decide if a description is either too vague or overly specific.
- **Trial and error knowledge** is information gained when a listener attempts a requested action on requested objects and then compares the result of the action with his expectations.

Other knowledge sources will not be covered here. For example, **Pragmatic knowledge** about mutual belief and actions (Cohen 1978, Allen 1979, Perrault and Cohen 1981, Appelt 1981) is missing because we restricted our work to noun phrases instead of complete utterances. **Domain knowledge** (including functional information) isn't covered because it is treated well elsewhere (Grosz 1977).

These knowledge sources can be used to guide the repair of the speaker's description when no referent is found. They are part of a "relaxation" process. Relaxation would typically mean in the reference identification paradigm that the system drops features in the speaker's description one at a time until a referent is found or none are left. We have something different in mind. First, relaxation means more than simply dropping a feature value. It also means replacing the feature value with another one the knowledge sources consider reasonable. Second, we want an order to be chosen to drop the features. The interesting part is that this ordering comes from a negotiation among the knowledge sources. The actual negotiation, which is a control problem, is discussed in Section 4.

3.1.1 LINGUISTIC KNOWLEDGE IN REFERENCE

Speakers can utilize many different kinds of linguistic structures to describe objects in the extensional world. This section outlines some of these structures and their

meanings and shows how they can be used to guide repairs in the description.

A description of an object in the extensional world usually includes enough information about physical features of the object so that listeners can use their perceptual abilities to identify the object.⁸ Those physical features are normally specified as modifiers of nouns and pronouns. The typical modifiers are adjectives, relative clauses, and prepositional phrases. They are often interchangeable; that is, one could specify a feature using any of the modifier forms. One modifier form, however, may be better suited for expressing some particular feature than another.

Relative clauses are well suited for expressing complicated information since they are separate from the main part of the noun phrase and can be arbitrarily complex themselves. They can restrict the word or phrase they modify. They function in the following ways in extensional reference:

- Complex relationships such as spatial relations (e.g., *the blue cap that is on the main tube*), and function information (e.g., *the thing with the wire that acts like a plunger*).
- Assertions of *extra* (usually restrictive) information, information possibly outside the domain knowledge and not useful for finding the referent at this time (e.g., *an L-shaped tube of clear plastic that is defined as a spout*).
- Material useful for confirming that the proper referent was found (e.g., *the long blue tube that has two outlets on the side*).
- A respecification of the initial description in more detail. For example, in the case of the descriptions *the thing that is flared at the top* and *the main tube which is the biggest tube*, the relative clauses are needed because the initial descriptions are too general to distinguish any one object.

Prepositional phrases are better fitted for simpler pieces of information. They are often part of expressions of predicative relationships.

- A comparative or superlative relation (e.g., *the smallest of the red pieces*).
- A subpart specification – used to access the subpart of the object under consideration (e.g., *the top end of the little elbow joint, that water chamber with the blue bottom and the globe top*).
- Most perceptual features (e.g., *with a clear tint, with a red color*).

Just like relative clauses, prepositional phrases can also provide confirmation information.

Adjectives are used to express almost any perceptual feature – though complex relations can be awkward. Usually they modify the noun phrase directly, but sometimes they are expressed as a predicate complement. In those situations, the complement describes the subject of the linking verb (e.g., *the tube is large*). As with some of the relative clauses above, predicate complements have

an assertional nature to them because they are normally used to state something about the subject of a sentence.

Sometimes the head noun carries feature information. For example, one can use *the bell* to refer to a bell-shaped object (though it does *not* necessarily have the function of a bell), or can say *the cube* instead of saying *the block* to refer to an object.

It is implicitly clear that the structure of a noun phrase can affect its meaning in many ways (such as the ones mentioned above under relative clauses). Since there is no one-to-one mapping between a noun phrase's structure and its meaning, it is the hearer's job to determine how the structural information is being used.

3.1.2 RELAXING A DESCRIPTION USING LINGUISTIC KNOWLEDGE

We examined the water pump protocols and noted where and when the modifiers of a noun phrase come into play during reference resolution (e.g., we saw that people would often commence their search for a referent immediately, using each piece of the description as it is heard). Adjectives and prepositional phrases play a more central role during referent identification, because they are heard first, while relative clauses usually play a secondary role, because they normally come at the end of a description, often after a pause. However, relative clauses and predicate complements exhibit an assertional nature that, while reducing their usefulness for resolving the current reference, provides useful information that can be expressed in subsequent (anaphoric) references. For example, a speaker can describe the MAINTUBE by saying *the long violet tube that has two outlets on the side* versus the shorter *the long violet tube with two outlets on the side*. Our claim is that the speaker would use the relative clause version to *emphasize* the information in the relative clause. Thus, relative clauses promote their contents (especially linguistically since they provide separation from the main clause) to an almost independent status. We feel this independent status stresses that the speaker took care in formulating the relative clause and that the information it conveys is *less* likely to be in error than if it had been expressed in a prepositional phrase or as an adjective; the water pump protocols tend to back up this claim (e.g., listeners would often use the information in a relative clause to confirm that their referent choice was correct). The head noun of the description can also be relaxed. It normally is relaxed last but could be relaxed prior to a relative clause (especially in the instances where the relative clause expresses confirmational information). Hence, our relaxation process attempts to weaken or remove features in a description in this order: adjectives, then prepositional phrases and finally relative clauses and predicate complements.

For example, consider the description *the blue cap that is on the main tube*. Here, the features "color" and "function" are described in the adjective and head noun of the description, and the "position" in the relative clause. Following the rule suggested above, the relaxation

of function and color should be attempted before position. The relaxation order proposed here is not meant to be the only way to relax the description. The order, in fact, may be modified by other knowledge sources.

There are many other kinds of linguistic constituents that can be examined to see if there are principled ways to relax them, too. These include premodifier and post-modifier forms, nominals, participles, and genitives. While we didn't consider any of them in detail, there is no reason why they should not fit into the relaxation framework.

3.1.3 PERCEPTUAL KNOWLEDGE IN REFERENCE

Our system must take into account how people perceive objects in the world and how their perceptions can be represented. To do so, each object in the world has two representations in our system: a spatial (3-D) representation and a cognitive/linguistic representation that shows how the system could actually talk about the object. The spatial description is a physical description of the object in terms of its dimensions, the basic 3-D shapes composing it, and its physical features (along the lines developed in Agin (1979) and Goodman (1981)). It represents the result of human perceptual skill. The cognitive/linguistic form is a representation of the parts and features of the object in linguistic terms. In many ways this representation encodes the human capacity to extract information from our perceptual system and turn physical representations into words. It overlaps the spatial form – which holds relatively constant across people – in many respects, but it is more suggestive of the listener's own perceptions. The cognitive/linguistic form often describes aspects of an object, such as its subparts, by its position on the object (“top”, “bottom”) and its functionality (“outlets”, “places for attachment”). More than one cognitive/linguistic form can refer to the same physical description. Some properties of an object differ in how they are expressed in the two forms. In the 3-D form, there are primarily properties such as numerical dimensions (e.g., *3 feet by 5 feet*) and basic shapes (e.g., generalized cylinders), while, in the cognitive/linguistic form, there are relative dimensions (e.g., *large*) and analogical shapes (e.g., *the L-shaped tube or the champagne top sort of looking bottom*).

Perceived objects, when spoken about, must be interpreted. This can lead to discrepancies between individuals. People usually agree on the spatial representation but not necessarily on the cognitive/linguistic description. This disagreement can lead to reference problems. For example, misjudgements by the speaker in calling an object “large” can cause the hearer to fail to find an object in the visual world that has dimensions that are perceptually “large” to the listener.

To avoid confusing the listener, a speaker must distinguish the objects in the environment from each other using perceptually useful features because these perceptual features provide people with a way to discriminate one object from another. A speaker must take care when

selecting from these features since the hearer can become confused about the values of a feature irrespective of the actual object being described. Perceptual features may be inherently confusing because a feature's values are difficult to differentiate (e.g., is the tube a cylinder or a slightly tapering cone?). They may also be confusing because the speaker and listener may have differing sets of values for a feature (e.g., what may be blue for someone may be turquoise for another). These characteristics affect the salience of a feature (see McDonald and Conklin (1982) for a description of feature salience) which in turn determines the feature's usefulness in a description. A feature that is common in everyday usage (e.g., color, shape, or size) is salient because the listener assumes that he can readily distinguish the feature's possible values from one another. Of course, very unusual values of a feature can stand out, making it even easier to discriminate a unique object from all other objects (McDonald and Conklin 1982).

The objects in the world may exhibit a feature whose possible values are difficult to distinguish. This occurs when a perceived feature does not have much variability in its range of values: all or subsets of the values are clustered closely together making it hard to tell the difference between one value and the next.⁹ This increases the likelihood of confusion because the usefulness of specifying the feature to a non-expert is diminished (especially if the speaker is more expert than the listener in distinguishing feature values). Hence, if one of these difficult feature values appears in the speaker's description, the listener, if he isn't an expert, will often relax the feature value to any of the members of the set of feature values. For example, if the speaker knows many shades of the color “red” (such as scarlet, crimson, cherry, maroon, or magenta), the average listener may not be able to distinguish them from each other and may be just as happy to pick up the maroon plug for *the magenta plug*.

When the number of features available for describing an object is small, one could expect to have trouble discerning one object from the next depending on the quality of the features themselves. If the environment is full of objects whose perceived features (e.g., color, size or shape) are similar, one would expect more miscommunication the larger the similarities. In those cases where perceptual information can only group objects instead of highlighting a unique one, the members of the group might become distinguishable when functional information is added.¹⁰ In other words, one may only know about the appearance of an object, but once one knows the function, the object and other potential contenders (might) become dissimilar (Grosz 1981). Of course, poor functional descriptions, like the ones illustrated in Section 2.2 for Bad Analogies, can lead to even more trouble.

3.1.4 RELAXING A DESCRIPTION USING PERCEPTUAL KNOWLEDGE

When examining the features presented in a speaker's description, one can consider perceptual aspects to deter-

mine which features are most likely in error. Such an inspection generates a partial ordering of features for use during the repair process to determine which feature in a description to relax. The relaxation ordering suggested by the inspection of features interacts with ordering proposals from other knowledge sources.

Active features are ones that require a listener to do more than simply recognize that a particular feature value belongs to a set of possible values – the listener must perform some kind of evaluation. They include the use of relative dimensions (e.g., *large*), comparatives (e.g., *larger*), or superlatives (e.g., *largest*). When considering the water pump domain, we found that listeners were better at judging less active feature values (e.g., color values). Speakers, however, seem to be casual with less active features (possibly because they feel listeners are better with them) while the active ones require their full attention. Hence, in a reference failure, the source of the problem is often the less active ones. This suggests that one should first relax those features that require less active consideration such as color (though it is easier to relax red to orange than red to blue; we will ignore such facts until a later stage of the relaxation process), composition, transparency, shape, and function *because* we would expect a speaker to be more serious about his use of active features. Only after them should one relax those features that require active consideration of the object under discussion and its surroundings (such as superlatives, comparatives, and relative values of size, length, height, thickness, position, distance, and weight).

The water pump dialogues provided some evidence for this. For example, many speakers described the MAINTUBE using a relative size adjective such as *big* or *large*. One of the descriptions of the tube was *the large blue tube*. The MAINTUBE, which was the largest object, actually was violet but there was a smaller blue tube, the STAND. Subjects still tended to select the MAINTUBE over the STAND, even with the color discrepancy, hinting that they preferred relaxing color (a less active feature) before relative size (an active feature).

3.1.5 DISCOURSE KNOWLEDGE IN PREFERENCE

Discourse knowledge concerns discourse structure, the flow of discourse, and the use of discourse to highlight parts of the real world (see Grosz (1977), Reichman (1978, 1981), Sidner (1979), Allen, Frisch, and Litman (1982), Litman (1983), and Polanyi and Scha (1984) for detailed treatments on discourse). There are several mechanisms that can highlight objects in discourse (see work on focus by Grosz (1977), Reichman (1978) and Sidner (1979)). They provide a partition of the real world that prunes the set of objects to consider during referent identification. Discourse knowledge also helps highlight what knowledge a speaker and listener have in common at any point in a dialogue. Conversants share knowledge about past actions and objects and general knowledge about the world (e.g., how to fit objects together or the functions of common objects). Focusing

can demarcate which of several perspectives of world knowledge conversants should be using to interpret each other's utterances. This simplifies the amount of information that must be packaged in each utterance, reducing places for error. For example, deictics can be used to anchor descriptions to current or past context. The description *the yellow polka-dotted motor* requires a listener to look to see how the description hooks up to the current discourse situation. However, the description *the yellow polka-dotted motor I showed you yesterday* is anchored by the deictic *yesterday* and is more easily searchable.

3.1.6 RELAXING A DESCRIPTION USING DISCOURSE KNOWLEDGE

Discourse knowledge helps the listener determine whether or not the problem is in the speaker's description or resides elsewhere. When normal reference fails (i.e., no referent corresponds to a description) and recovery is attempted, discourse knowledge can be used to determine whether the problem resides not in the description itself but possibly at the discourse level. For example, midstream corrections in an utterance by a speaker could cause a listener to either miss a shift in focus or to shift focus when no shift was intended. This was exemplified in Excerpt 7 in Section 2.2 when the speaker attempted to undo an earlier request and did not properly demark the shift of focus. The work of Grosz (1977, 1981), Reichman (1978, 1981), Webber (1978), and Sidner (1979) provided rules on deictics, anaphoric definite noun phrases, the use of pronominals versus nonpronominals, and so forth, that can be used to zero in on discourse problems. So, for example, if a self-correction of the use of a pronominal occurs (e.g., "...it – the X"), then a rule might state that focus could have shifted to X. Relaxation is then achieved by trying the hypothesized focus to see if a referent can now be found. In general, discourse knowledge can suggest when the problem may be due to the listener focussing on the wrong set of objects. Correction can be attempted by shifting to another set and testing whether or not the description better fits one of the objects in the new set.

3.1.7 HIERARCHICAL KNOWLEDGE IN REFERENCE

Imprecision (i.e., being overly general) in a speaker's description can lead to confusion. Being too specific can lead to similar results. Hierarchical knowledge – that is, knowledge about a hierarchy of taxonomic information about our world – can be used by a listener to determine the degree of imprecision or specificity of a description. We can model this behavior by consulting a prestored generic/specific hierarchy of world elements, using the current context to guide the comparison of the speaker's current description to elements in the hierarchy, and deciding on the basis of the comparison if the description was imprecise. This comparison can isolate two types of imprecision: imprecision of the whole description or imprecision of a particular feature value.

An imprecise description, missing details needed to fully distinguish a unique real world object, should point out numerous candidates that exhibit the general features in the description rather than none at all. Imprecise descriptions can, however, lead to confusion that blocks the listener from finding any referent. If a particular feature specified in a description is difficult to apply because it isn't specific or well-defined, then it may be necessary to ignore it (e.g., the use of a value like "funny" such as in *that funny red thing*). If a feature is ambiguous with respect to how it should be applied, then it may either require relaxation or further restriction (e.g., for the use of a feature value like "rounded," we must ask whether we mean "2-D" or "3-D" rounded, "cylindrical" or "bell-shaped", and so on). The determination that a feature is too imprecise might be possible *before* a search for a referent is commenced. An examination of how high in the hierarchy the feature value appears could signal when a more detailed value is needed. Each of these problems was reflected in the water pump protocols by listeners (e.g., see Excerpts 4 and 6). They often avoided searching for a referent because the speaker's description was just too imprecise, causing them confusion from the onset.

The condition of being too specific is more difficult to detect. In a task-oriented environment, one would not easily notice that something was too specific since normally being very specific is a wise goal for a speaker. The drawback of being too specific occurs not so much because of the specificity itself but because of its adverse side-effects. These side-effects include the use of feature values that are too difficult for a non-expert to determine, leading to confusion. A description can also be overspecific if it contains *too* many feature values or contains a feature that is overpowering (e.g., see, respectively, Excerpts 2 and 10).

3.1.8 RELAXING A DESCRIPTION USING HIERARCHICAL KNOWLEDGE

Hierarchical knowledge can resolve certain ambiguities by climbing or descending the hierarchy. Such a hierarchy search requires looking at a description at two levels:

- the description's placement in the generic/specific hierarchy and
- the placement of the filler of each feature of the description in the generic/specific hierarchy.

Hierarchical knowledge also interacts with perceptual knowledge. The hearer can become confused when a feature value in the speaker's description is too hard to judge. For example, it is difficult to determine whether a particular feature value applies when it is too specific. If a more imprecise value is used (and it applies only to one object), it might be easier to find the described object (e.g., *hippopotamus face shaped valve* would be better stated as *rounded valve*, as seen in Excerpt 10). Hence, in cases where a feature value is too specific, more imprecise values could be tried to see if a referent can then be found. These more imprecise values are found by looking

higher in the hierarchy above the current feature value for more general terms.

The use of hierarchical knowledge isn't always the most appropriate way to repair a description. Consider descriptions introduced only by head nouns (i.e., category descriptions) such as *the plunger*. In such instances, when no clear representative of the category object is present, it is not necessarily best to check the generic/specific hierarchy to see what is "above" the concept representing the category (e.g., finding *device* above *plunger*). It might be better to examine the attributes relevant to an average member of the category set since it may be the standard values of those attributes that the speaker is trying to get across in his or her description. For example, in the description *the man drinking the martini*, the speaker may be trying to get the listener to look for someone drinking a clear liquid from a certain shaped glass with an olive in it. The speaker isn't particularly concerned with the fact that the drink contains gin and vermouth. If we just consulted the generic/specific taxonomy, however, we might simply relax *martini* to *alcoholic beverage* (such as to *the man drinking the beer*) to *liquid* (such as to *the man drinking water*) and miss the descriptors that the speaker really intended us to use.¹¹

3.1.9 TRIAL AND ERROR KNOWLEDGE IN REFERENCE

Trial and error knowledge has to do with performance feedback. Its primary use is to determine whether a referent was properly identified (including ones found with the relaxation process). Performance of a requested action is the strongest determining factor of whether or not the listener correctly interpreted a speaker's description.¹² Successful completion of an action will be likely to build confidence in the listener that he correctly interpreted a description. Failure to find an object after relaxation leads the listener to ask the speaker to clarify; failure to successfully perform the requested action on the object found during referent identification causes the listener to ask himself what is wrong. The trouble might be due to:

- the object identified from the speaker's description,
 - the action attempted, or
 - some prior (probably unnoticed) mistake that occurred.
- Failure may come not only from the inability to perform an action but also from an action's postcondition failing.¹³ Determination of how badly a postcondition must fail before the listener asks for clarification – instead of reconsidering the description – is unclear from the current protocols; further analysis collected from different protocols might resolve this matter.

4 REPAIRING REFERENCE FAILURES

4.1 INTRODUCTION

The previous sections illustrated how task-oriented natural language interactions in the real world can induce contextually poor utterances and the kinds of knowledge

people use to reason about them. Given all the possibilities for confusion, when confusions do occur, they must be resolved if the task is to be performed. This section explores the problem of fixing reference failures.

Reference identification is a search process where a listener looks for something in the world that satisfies a speaker's uttered description. A computational scheme for performing such reference identifications has evolved from work by other artificial intelligence researchers (e.g., see Grosz 1977, Hoepfner et al. 1983). That traditional approach succeeds if a referent is found, or fails if no referent is found (see Figure 3a). However, a reference identification component must be more versatile than those previously constructed. The excerpts provided in Section 2.2 show the traditional approach is inadequate because people's real behavior is much more elaborate. In particular, listeners often find the correct referent even when the speaker's description does not describe any object in the world. For example, a speaker could describe a turquoise block as the *blue block*. Most listeners would go ahead and assume the turquoise block was the one the speaker meant since turquoise and blue are similar colors.

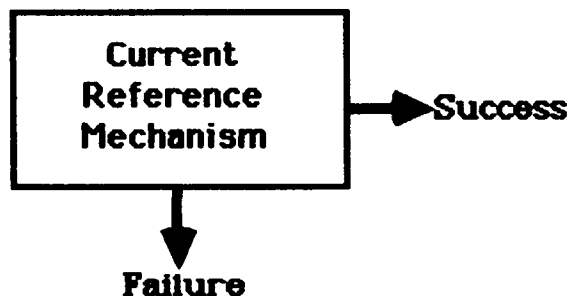


Figure 3a. The Traditional approach to reference identification.

A key feature to reference identification is **negotiation**. Negotiation in reference identification comes in two forms. First, it can occur between the listener and the speaker. The listener can step back, expand greatly on the speaker's description of a plausible referent, and ask for confirmation that he has indeed found the correct referent. For example, a listener could initiate negotiation with *I'm confused. Are you talking about the thing that is kind of flared at the top? Couple inches long. It's kind of blue.* Second, negotiation can be with oneself. This *self-negotiation* is the one we are most concerned with in this research. The listener considers aspects of the speaker's description, the context of the communication, the listener's own abilities, and other relevant sources of knowledge. He then applies that deliberation to determine whether one referent candidate is better than another or, if no candidate is found, what are the most likely places for error or confusion. Such negotiation can result in the listener testing whether or not a particular referent works. For example, linguistic descriptions can influence

a listener's perception of the world. The listener must ask himself whether he can perceive one of the objects in the world the way the speaker described it. In some cases, the listener's perception may *override* parts of the description because the listener can't perceive it the way the speaker described it.

To repair the traditional approach we have developed an algorithm that captures for certain cases the listener's ability to negotiate with himself for a referent. It can search for a referent and, if it doesn't find one, it can try to find possible referent candidates that might work, and then loosen the speaker's description using knowledge about the speaker, the conversation, and the listener himself. Thus, the reference process becomes multi-step and resumable. This computational model, which we call **FWIM** for "Find What I Mean", is more faithful to the data than the traditional model (see Figure 3b).

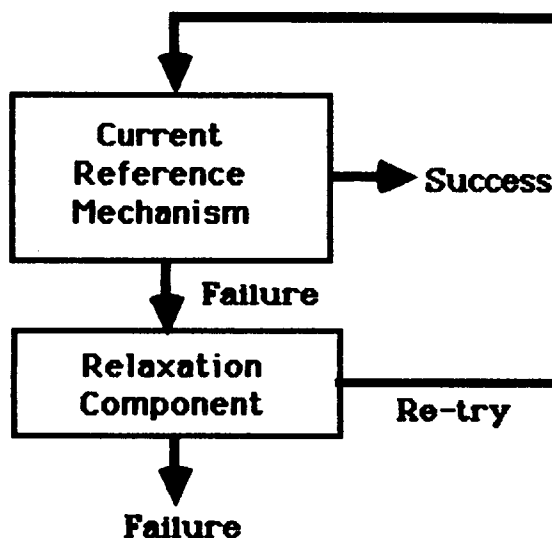


Figure 3b. FWIM approach to reference identification.

One means of making sense of a failed description is to delete or replace the portions that cause it not to match objects in the hearer's world. In our program we are using "relaxation" techniques to capture this behavior. Our reference identification module treats descriptions as approximate. It relaxes a description in order to find a referent when the literal content of the description fails to provide the needed information. Relaxation, however, is not performed blindly on the description. We try to model a person's behavior by drawing on sources of knowledge used by people. We have developed a computational model that can relax aspects of a description using many of these sources of knowledge. Relaxation then becomes a form of communication repair (in the style of the work on repair theory found in Brown and VanLehn (1980)). A goal in our model is to use the knowledge sources to reduce the number of referent candidates that must be considered while making sure that a particular relaxation makes sense.

4.2 THE REFERENT IDENTIFIER AND RELAXATION COMPONENT

This section describes the overall relaxation component in the context of the referent identifier. We explain how the relaxation component draws on knowledge sources about descriptions and the real world as it tries to relax an errorful description to one for which a referent can be identified.

4.2.1 FIND A REFERENT USING A REFERENCE MECHANISM

Identifying the referent of a description requires finding an element in the world that corresponds to the speaker's description (where every feature specified in the description is present in the element in the world but not necessarily vice versa). This process corresponds to the technique employed in the traditional reference mechanism. The initial task of our reference mechanism is to determine whether or not a search of the (taxonomic) knowledge base that we use to model the world is necessary. For example, in the water pump domain, the reference component should not bother searching – unless specifically requested to do so – for a referent for indefinite noun phrases (which usually describe new or hypothetical objects) or extremely vague descriptions (which are ambiguous because they do not clearly describe an object since they are composed of imprecise feature values). A number of aspects of discourse pragmatics can be used in that determination. For example, the use of a deictic in a definite noun phrase, such as *this X* or *the last X*, hints that the object was either mentioned previously or that it probably was evoked by some previous reference, and that it is searchable. We will not examine such aspects any further in this paper.

The knowledge base contains linguistic descriptions and a description of the listener's visual scene itself. In our implementation and algorithms, we assume it is represented in KL-One (Brachman 1977), a system for describing taxonomic knowledge. KL-One is composed of CONCEPTS, ROLES on concepts, and links between them. A CONCEPT denotes a set, representing those elements described by it. A SUPERC link (" \supseteq ") is used between concepts to show set inclusion. It defines a relation called **subsumption** that specifies that the set denoted by one concept is included in the other. For example, consider Figure 4. The SUPERC from Concept B to Concept A is like stating $B \subseteq A$ for two sets A and B. An INDIVIDUAL CONCEPT is used to guarantee that the set specified by a concept denotes a singleton set. The Individual Concept D shown in the figure is defined to be a unique member of the set specified by Concept C. ROLES on concepts are like attributes or slots in other knowledge representation languages. They define a functional relationship between the concept and other concepts that specifies a restriction on what can fill a particular slot.

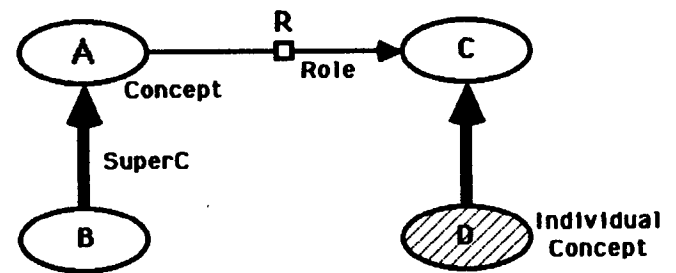


Figure 4. A KL-One taxonomy.

Once a search of the knowledge base is considered necessary, a reference search mechanism is invoked. The search mechanism uses the KL-One Classifier (Lipkis 1982) to search the knowledge base taxonomy. This search is constrained by a focus mechanism based on the one developed by Grosz (1977). The Classifier's purpose is to discover all appropriate subsumption relationships between a newly-formed description and all other concepts in a given taxonomy. With respect to reference, this means that descriptions of all possible referents of the description will be subsumed by the description after it has been classified into the knowledge base taxonomy. If more than one candidate referent is below (when a concept A is subsumed by B, we say A is *below* B) the classified description, then, unless a quantifier in the description specified more than one element, the speaker's description is ambiguous. If exactly one concept is below it, then the intended referent is assumed to have been found. Finally, if no referent is found below the classified description, the relaxation component can be invoked. Prior to actually using the relaxation component, FWIM checks to see if the problem resides not with the description but with pragmatic issues. We will only consider the no reference case in the rest of the paper.

4.2.2 COLLECT VOTES FOR OR AGAINST RELAXING THE DESCRIPTION

If the referent search fails, then it is necessary to determine whether the lack of a referent for a description has to do with the description itself (i.e., reference failure) or with outside forces that are causing reference confusion. For example, an external problem due to outside forces may be with the flow of the conversation and the speaker's and listener's perspectives on it; it may be due to incorrect attachment of a modifier; it may be due to the action requested; and so on. Pragmatic rules are invoked to decide whether or not the description should be relaxed. For example, aspects on focus, metonymy and synecdoche are considered to see if they affected the referent search.¹⁴ These rules will not be discussed here; we will assume that the problem lies in the speaker's description.

4.2.3 PERFORM THE RELAXATION OF THE DESCRIPTION

If relaxation is demanded, then the system must

- find potential referent candidates,
- determine which features in the speaker's description to relax and in what order, and use those ordered features to order the potential candidates with respect to the preferred ordering of features, and
- determine the proper relaxation techniques to use and apply them to the description.

FIND POTENTIAL REFERENT CANDIDATES

Before relaxation takes place, the algorithm looks for potential candidates for referents (which denote elements in the listener's visual scene). These candidates are discovered by performing a "walk" in the knowledge base taxonomy in the general vicinity of the speaker's classified description as partitioned by the focusing mechanism. The walk is performed by moving up and down the SuperC links, checking each candidate. A KL-One partial matcher is used to determine how close the candidate descriptions found during the walk are to the speaker's description. The partial matcher generates a numerical score to represent how well the descriptions match (after first generating scores at the feature level to help determine how the features are to be aligned and how well they match). This score is based on information about KL-One (e.g., the subsumption relationship between or the equality of two feature values) and does not take into account any information about the task domain. The set of best descriptions returned by the matcher (as determined by some cutoff score) is selected as the set of referent candidates. The ordering of features and candidates for relaxation described below takes into account the task domain.

ORDER THE FEATURES AND CANDIDATES FOR RELAXATION

At this point the reference system inspects the speaker's description and the candidates, decides which features to relax and in what order,¹⁵ and generates a master ordering of features for relaxation. That ordering is important since relaxing in different orders could yield matches to different objects. Once the feature order is created, the reference system uses that ordering to determine the order in which to try relaxing the candidates.

We draw primarily on sources of linguistic knowledge, pragmatic knowledge, discourse knowledge, domain knowledge, perceptual knowledge, hierarchical knowledge, and trial and error knowledge during this repair process. A detailed treatment of many of them was presented in Section 3. These knowledge sources are consulted to determine the feature ordering for relaxation. We represent information from each knowledge source as a set of relaxation rules. Most of the rules were motivated by the problems illustrated in the protocols. They are written in a PROLOG-like language. Figure 5 illustrates one such linguistic knowledge relaxation rule.

This rule is motivated by the observation that speakers typically add more important information at the end of a description (where it is separated from the main part of the description and, thus, provides more emphasis). The rule in Figure 5 simply embodies the fact that relative clauses are found at the end of noun phrases, while adjectives are not and, thus, the features of a description that are provided adjectivally should be relaxed before those provided by a relative clause. However, a more general and more applicable rule is that information presented at the end of a description is usually more prominent (i.e., that information was placed more strongly *in focus* by the speaker).

Relax the features in the speaker's description in the order: adjectives, then prepositional phrases, and finally relative clauses and predicate complements.

E.g.,

```
Relax-Feature-Before(v1,v2)
←ObjectDescr(d),FeatureDescriptor(v1),
FeatureDescriptor(v2),
FeatureInDescription(v1,d),
FeatureInDescription(v2,d),
Equal(syntactic-form(v1,d),"ADJ"),
Equal(syntactic-form(v2,d),"REL-CLS")
```

Figure 5. A sample linguistic relaxation rule.

Figure 6 provides an example of a couple of the discourse knowledge relaxation rules. The rules note when misfocus is likely. They simulate how a listener can detect confusion on the part of the speaker during the search for a referent if the speaker interrupts his own utterance.¹⁶ An interruption can come about with a false start or a self-correction. A false start occurs when the speaker goofs on his initial description, stops, and then restarts the description (also see Polanyi (1978) on false starts). For example, exclamations like *oops*, *never mind*, *oh no*, and so on, are signals of false starts meant to inform the listener that there is a problem, though not stating precisely where the problem occurred. The problem could be due to the current utterance or a previous one. Speakers often (falsely) assume the listener "knows" just where the speaker means. Typically, a listener presumes the problem is with the current utterance. A listener should, however, note that a false start has occurred at this point in the dialogue and be prepared to back up to the same place later on. Self-corrections are less interruptive than false starts and more explicit about the source of the problem. They are redescriptions of a piece of the speaker's utterance that occur as it is spoken. Descriptions like *it—the tube* or *the large blue—uh violet tube* are typical ones that occur. As with false

starts, such places are conducive to confusion and should be noted by the listener.

Focus shift relaxation rules:

MarkForPossibleConfusion(u)

← Utterance(u), FalseStart(u)

MarkForPossibleConfusion(d)

← ObjectDescr(d), Self-Correction(d)

where

FalseStart[u]: This predicate determines whether or not a false start has occurred in some utterance, u. Such false starts have to be caught by the parser.

Self-Correction[d]: This predicate looks for self-corrections in a description, d. As with FalseStart, it would have to be implemented inside the parser.

Figure 6. Two discourse knowledge relaxation rules.

Each knowledge source produces its own partial ordering of features.¹⁷ Each partial ordering is topologically sorted to provide a consistent format for comparison. The partial orderings are then considered together. For example, perceptual knowledge may say to relax color. However, if the color value was asserted in a relative clause, linguistic knowledge would rank color lower, i.e., placing it later in the list of things to relax.

Since different knowledge sources generally produce different partial orderings of features, these differences can lead to a conflict over which features to relax. It is the job of the best candidate algorithm to resolve these disagreements among knowledge sources. Its goal is to order the referent candidates, C_1, C_2, \dots, C_n , so that relaxation is attempted on the best candidates first. Those candidates are the ones that conform best to the proposed feature orderings. To start, the algorithm examines candidates and the feature orderings from each knowledge source. For each candidate C_j , the algorithm scores the effect of relaxing the speaker's original description D to C_j , using the feature ordering from one knowledge source. The score reflects the goal of minimizing the number of features relaxed while trying to relax the features that are "earliest" in the feature ordering.¹⁸ Thus, these heuristics provide a simple way to reflect in the score how well a particular candidate fits a feature ordering. Notice that such scoring could very well favor a candidate C_1 that requires *more* features to be relaxed in D than another candidate C_2 if those features are *earlier* in the feature ordering than those required by C_2 . The algorithm repeats its scoring of C_j for each

knowledge source, and sums up its scores to form C_j 's total score. The C_j s are then ordered by that score (with the lower scores first).

Figure 7 provides a graphic illustration of what the best candidate algorithm does. The speaker's description is represented at the top of the figure. The set of specified features and their assigned feature value (e.g., the pair Color: Maroon) are also shown there. A set of objects in the real world are selected by the partial matcher as potential candidates for the referent. These candidates are shown near the top of the figure (C_1, C_2, \dots, C_n). Inside each box is a set of features and feature values that describe that object. A set of partial orderings are generated that suggest which features in the speaker's description should be relaxed first — one ordering for each knowledge source (shown as "Linguistic," "Perceptual," and "Hierarchical" in the figure). For example, linguistic knowledge recommends relaxing Color or Shape before Function, and relaxing Function before Size. Finally, the referent candidates are reordered using the information expressed in the speaker's description and in the partial orderings of features.

DETERMINE WHICH RELAXATION METHODS TO APPLY

Once a set of ordered, potential candidates is selected, the relaxation mechanism begins step 3 of relaxation; it tries to find proper relaxation methods to relax the features that have just been ordered (success in finding such methods *justifies* relaxing the speaker's description to a particular candidate). It stops at the first candidate in the list of candidates to which methods can be successfully applied. This step is the second place where the knowledge sources are useful.

Relaxation can take place with many aspects of a speaker's description: with complex relations specified in the description, with individual features of a referent specified by the description, and with the focus of attention in the real world where one attempts to find a match. Complex relations specified in a speaker's description include spatial relations (e.g., *the outlet near the top of the tube*), comparatives (e.g., *the larger tube*), and superlatives (e.g., *the longest tube*). These can be relaxed. The simpler features of an object (such as size or color) specified in the speaker's description are also open to relaxation.

Relaxation of a description has a few global strategies that can be followed for each part of the description:

1. drop the errorful feature value from the description altogether,
2. weaken or tighten the feature value in a principled way keeping its new value *close* to the specified one (e.g., movement within a subsumption hierarchy of feature values), or
3. try some other feature value based on some outside information (e.g., knowing that people often confuse opposite word pairs such as using *hole* for *peg* as illustrated in Excerpt 10).

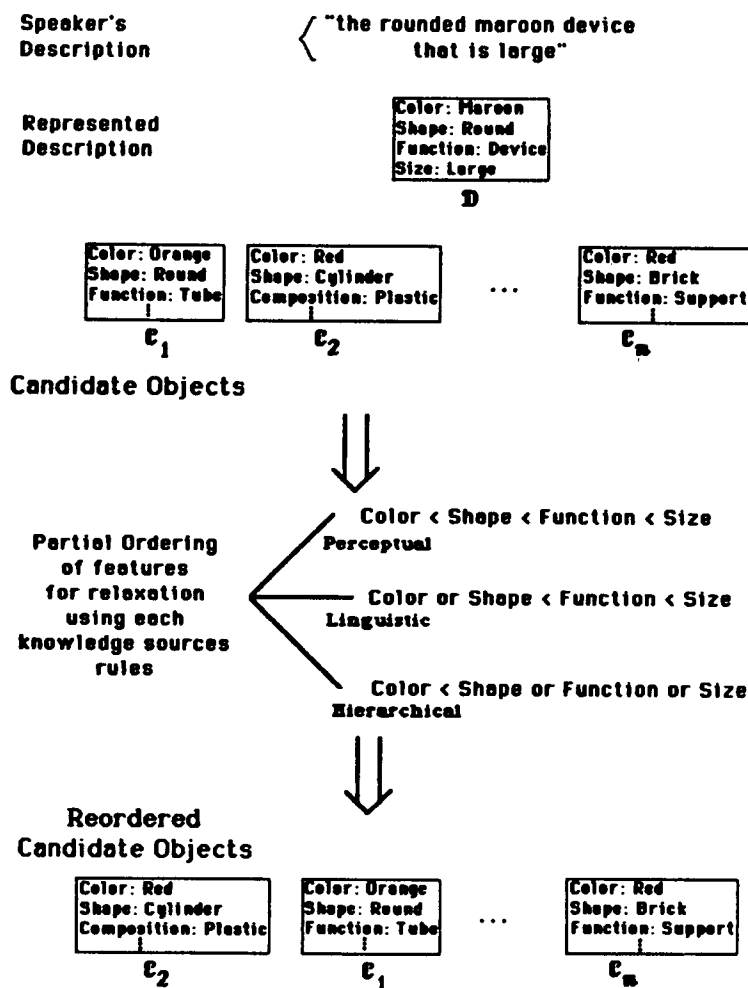


Figure 7. Reordering referent candidates.

When performing relaxation, one would attempt to use the least drastic measures first. (1) is the most drastic, while (2) is the least; (3) is in between.


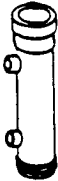


Often the objects in focus in the real world implicitly cause other objects to be in focus (Grosz 1977, Webber 1978). The subparts of an object in focus, for example, are reasonable candidates for the referent of a failing description and should be checked. At other times, the speaker might attribute features of a subpart of an object to the whole object (e.g., describing a plunger that is composed of a red handle, a metal rod, a blue cap, and a green cup as *the green plunger*). In these cases, the relaxation mechanism utilizes the part-whole relation in object descriptions to suggest a way to relax the speaker's description.

These strategies are realized through a set of procedures (or *relaxation methods*) that are organized hierarchically. Each procedure is an expert at relaxing its particular type of feature and draws on the knowledge sources for its expertise. For example, a Generate-Simi-

lar-Feature-Values procedure is composed of procedures like Generate-Similar-Shape-Values, Generate-Similar-Color-Values and Generate-Similar-Size-Values. Each of those procedures is a specialist that attempts to first relax the feature value to one "near" or somehow "related" to the current one (e.g., one would prefer to first relax the color *red* to *pink* before relaxing it to *blue*) and then, if that fails, to try relaxing it to any of the other possible values.¹⁹ The effect of the latter case is really the same as if the feature was simply ignored.

4.3 AN EXAMPLE OF MISREFERENCE RESOLUTION

This section describes how a referent identification system can recover from a misreference using the scheme outlined in the previous section. For the purposes of this example, assume that the water pump objects currently in focus include the CAP, the MAINTUBE, the AIRCHAMBER and the STAND. Assume also that the speaker tries to describe two of the objects – the MAINTUBE and the AIRCHAMBER.

 STAND	<p>DescrA: "... two devices that are clear plastic.</p> <p>DescrB: One of them has two openings on the outside with threads on the end, and it's about five inches long.</p> <p>DescrC: The other one is a rounded piece with a turquoise base on it.</p> <p>DescrD: Both are tubular.</p> <p>DescrE: The rounded piece fits loosely over ..."</p>	 MAINTUBE
 CAP		 AIRCHAMBER

The reference system can find a unique referent for the first object (described by **DescrA**, **DescrB** and **DescrD**) but not for the second (described by **DescrA**, **DescrC**, **DescrD**, and **DescrE**), since none of the focused objects are TURQUOISE. The relaxation algorithm is shown below to reduce the number of referent candidates for the second object to two. It, then, requires the system/listener to try out those candidates to determine if one, or both, fits loosely. The protocols exhibit a similar result when the listener uses "fits loosely" to get the correct referent (e.g., Excerpt 6 exemplifies where "fit" is used by the speaker to help confirm that the proper referent was found). Our system simulates this test by asking the user about the fit.

Figure 8 provides a simplified and linearized view of the actual KL-One representation of the speaker's descriptions after they have been parsed and semantically interpreted.

A representation of each of the water pump objects currently under consideration (i.e., in focus) is presented in Figure 9. Each provides a physical description of the object – in terms of its dimensions, the basic 3-D shapes composing it, and its physical features – and a basic functional description of the object. The first upper case entry in each representation in Figure 9 defines the basic kind of entity being described (e.g., TUBE means that the object being described is some kind of tube). The words in mixed case refer to the names of features, and the other upper case words refer to possible fillers of those features from things in the water pump world. The "Subpart" feature provides a place for an embedded description of an object that is a subpart of a parent object. Such subparts can be referred to on their own or as part of the parent object. The "Orientation" feature, used in the representations in Figure 9, provides a rotation and translation of the object from some standard orientation to the object's current orientation in 3-D space. The standard orientation provides a way to define relative positions such as *top*, *bottom*, or *side*. Figure 10

shows the KL-One taxonomy representing the same objects.

```

DescrA: (DEVICE (Transparency CLEAR)
          (Composition PLASTIC))
DescrB: (DEVICE (Transparency CLEAR)
          (Composition PLASTIC)
          (Subpart (OPENING))
          (Subpart (OPENING))
          (Subpart
            (THREADS (Rel-Position END)))
          (Dimensions (Length 5.0)))
DescrC: (DEVICE (Transparency CLEAR)
          (Composition PLASTIC)
          (Shape ROUND)
          (Subpart (BASE (Color TURQUOISE))))
DescrD: (DEVICE (Transparency CLEAR)
          (Composition PLASTIC)
          (Subpart (OPENING))
          (Subpart (OPENING))
          (Subpart
            (THREADS (Rel-Position END)))
          (Dimensions (LENGTH 5.0))
          (Analogical-Shape TUBULAR))
          (DEVICE (Transparency CLEAR)
          (Composition PLASTIC)
          (Shape ROUND)
          (Analogical-Shape TUBULAR)
          (Subpart (BASE (Color TURQUOISE))))
DescrE: (FIT-INTO
          (Outer (DEVICE (Transparency CLEAR)
                        (Composition PLASTIC)
                        (Shape ROUND)
                        (Analogical-Shape TUBULAR)
                        (Subpart
                          (BASE (Color TURQUOISE))))))
          (Inner . . .)
          (FitCondition LOOSE))
    
```

Figure 8. The speaker's descriptions.

The first step in the reference process is the actual search for a referent in the knowledge base. In people, the reference identification process is incremental in nature, i.e., the listener can begin the search process before he hears the complete description. This was observed throughout the videotape excerpts where an apprentice would commence his search after just a few words in a description. We try to simulate this incremental nature in our algorithm. It is readily apparent when considering the placement of the first description in **DescrD** into the KL-One taxonomy shown in Figure 10. **DescrD** is incrementally defined by first adding **DescrA** – as shown in Figure 11 – and then **DescrB** – as shown in Figure 13 – to the taxonomy. The KL-One Classifier compares the features specified in the speaker's descriptions with the features specified for each element in the KL-One taxonomy that corresponds to one of the current objects of interest in the real world. Notice that some features are directly comparable. For example, the "Transparency" feature of **DescrA** and the "Transparency" feature of MAINTUBE are both equal to "CLEAR." All the other features specified in **DescrA** fit the MAINTUBE so the MAINTUBE can be described by **DescrA**. This is illustrated in Figure 12, where MAINTUBE is shown as a subconcept of **DescrA**. STAND also is shown as a subconcept of **DescrA**. AIRCHAMBER is shown as a *possible* subconcept (with the dotted arrow) because **DescrA** mismatches with it on one of its subparts.²⁰ CAP#1 is not shown as a subconcept of **DescrA** since its "Transparency" feature is OPAQUE and not CLEAR. Other features require in-depth processing – which is outside the capability of the KL-One classifier – before they can be compared. The OPENING value of "Subpart" in **DescrB** provides a good example of this. Consider comparing it to the "Subpart" entries for MAINTUBE shown in Figure 9. An OPENING, as seen in Figure 14, is thought of primarily as a 2-D cross-section (such as a "hole"), while the three CYLINDER subparts of MAINTUBE (labelled as *Lip*, *Outlet1*, and *Outlet2* in Figure 9) are viewed as (3-D) cylinders that have the "Function" of being outlets, i.e., OUTLET-ATTACHMENT-POINTS. To compare OPENING and one of the cylinders, say CYLINDER#1 (for *Lip*), the inference must be made that both things can describe the same thing (similar inferences are developed in Mark (1982)). One way this inference can occur is by recursively examining the subparts of MAINTUBE (and their subparts, etc.) with the KL-One partial matcher until the cylinders are examined at the 2-D level. At that level, an end of the cylinder will be defined as an OPENING. With that examination, the MAINTUBE can be seen as described by **DescrB**. This inference process is illustrated in Figure 14. There the partial matcher examines the roles *Lip*, *Outlet1*, and *Outlet2* of MAINTUBE, which represents its subparts, and determines the following:

- A CYLINDER can have an *End* which is either a *2D-End* (e.g., a lid or hole) or a *3D-End* (e.g., a lip).
- A *2D-End* is either an OPEN-2D-END (e.g., a hole) or a CLOSED-2D-END (e.g., a lid on a can).
- An OPEN-2D-END is a kind of OPEN-2D-OBJECT.

These facts imply that OPENING can match any of the subparts *Lip*, *Outlet1*, or *Outlet2* on MAINTUBE since those subparts are defined as cylinders that function as outlets (i.e., Outlet-Attachment-Points).

DescrC poses different problems. **DescrC** refers to an object that is supposed to have a subpart that is TURQUOISE. The Classifier determines that **DescrC** could not describe either the CAP or STAND because both are BLUE. It also could not describe the MAINTUBE²¹ or AIRCHAMBER since each has subparts that are either VIOLET or BLUE. The Classifier places **DescrC** as best it can in the taxonomy, showing no connections between it and any of the objects currently in focus. **DescrD** provides no further help and is similarly placed. This is shown in Figure 15. At this point, a probable misreference is noted. The reference mechanism now tries to find potential referent candidates, using the taxonomy exploration routine described in Section 4.2.3, by examining the elements closest to **DescrD** in the taxonomy and using the partial matcher to score how close each element is to **DescrD**.²² This is illustrated in Figure 16. The matcher determines MAINTUBE, STAND, and AIRCHAMBER as reasonable candidates by aligning and comparing their features to **DescrD**.

Scoring **DescrD** to MAINTUBE:

- a TUBE is a kind of DEVICE; (>)
- the Transparency of each is CLEAR; (+)
- the Composition of each is PLASTIC; (+)
- a TUBE implies Analogical-Shape TUBULAR, which implies Shape CYLINDRICAL, which is a kind of Shape ROUND; (>)
- the recursive partial matching of subparts: BASE is viewed as a kind of BOTTOM. Therefore, BASE in **DescrD** could match to the subpart in MAINTUBE that has a Translation of (0.0 0.0 0.0) – i.e., *Threads* of MAINTUBE. However, they mismatch since color TURQUOISE in **DescrD** differs from color VIOLET of MAINTUBE. (–)

Scoring **DescrD** to STAND:

- a TUBE is a kind of DEVICE; (>)
- the Transparency of each is CLEAR; (+)
- the Composition of each is PLASTIC; (+)
- a TUBE implies Analogical-Shape TUBULAR, which implies Shape CYLINDRICAL, which is a kind of Shape ROUND; (>)
- the recursive partial matching of subparts: BASE in **DescrD** could match to the subpart in STAND that has a Translation of (0.0 0.0 0.0) – i.e., *Base* of STAND. However, they mismatch since color TURQUOISE in **DescrD** differs from color BLUE of STAND. (–)

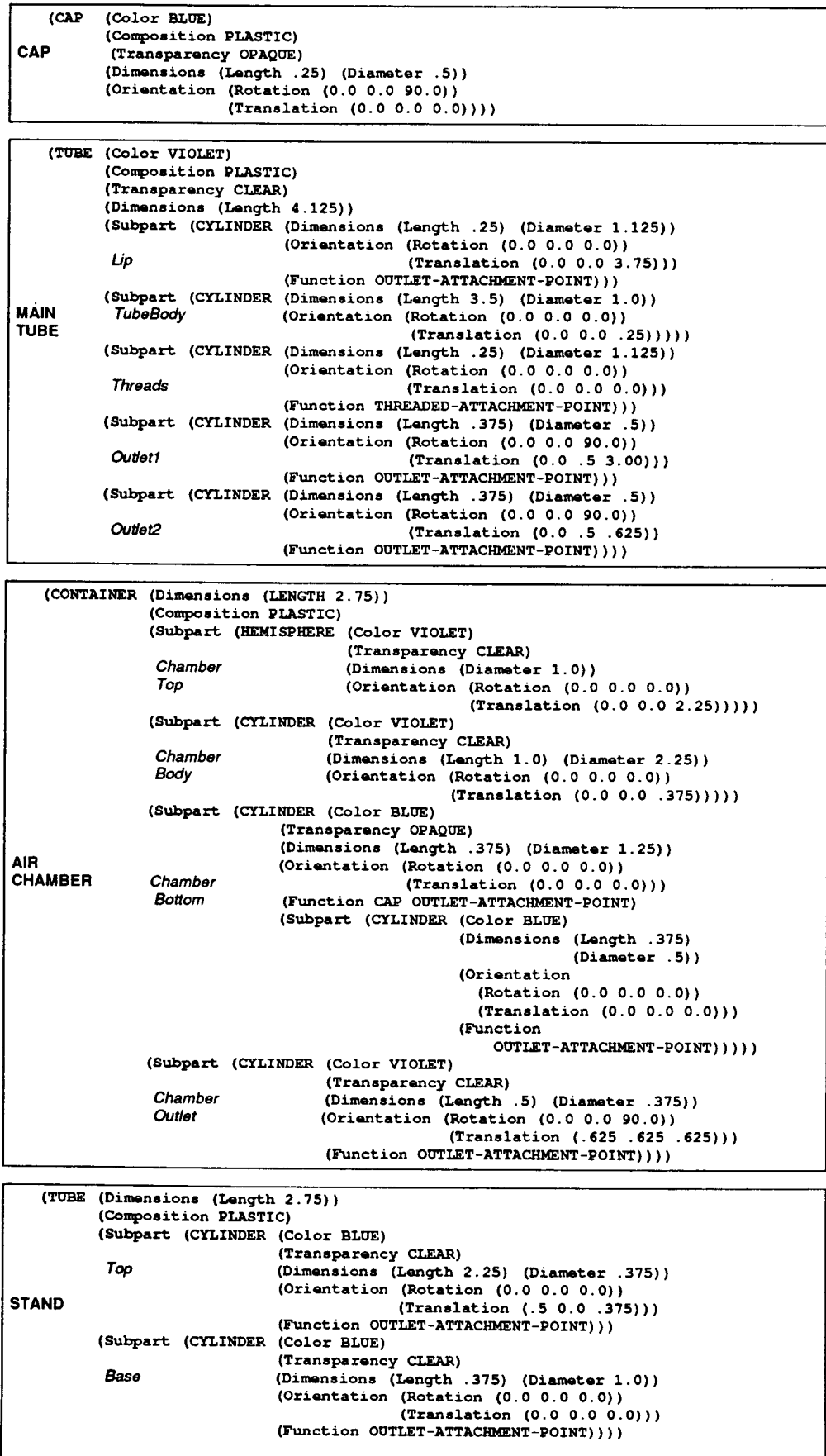


Figure 9. The objects in focus.

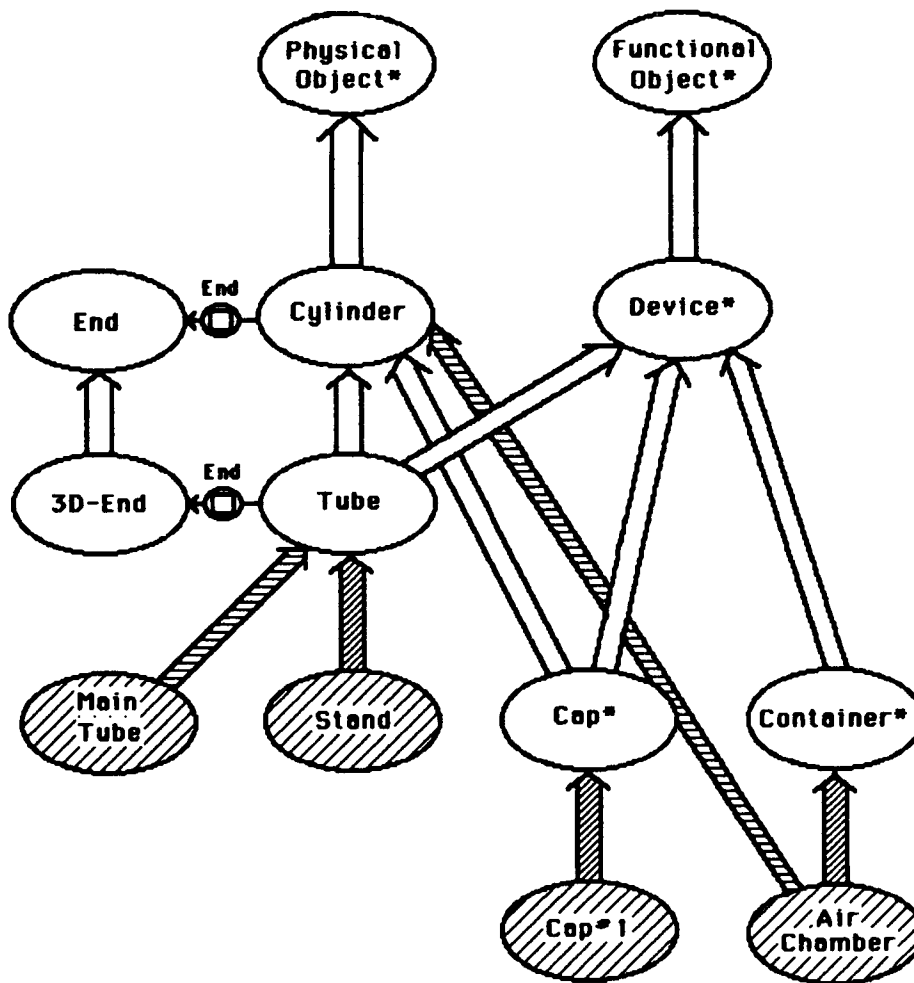


Figure 10. Taxonomy representing the objects in focus.

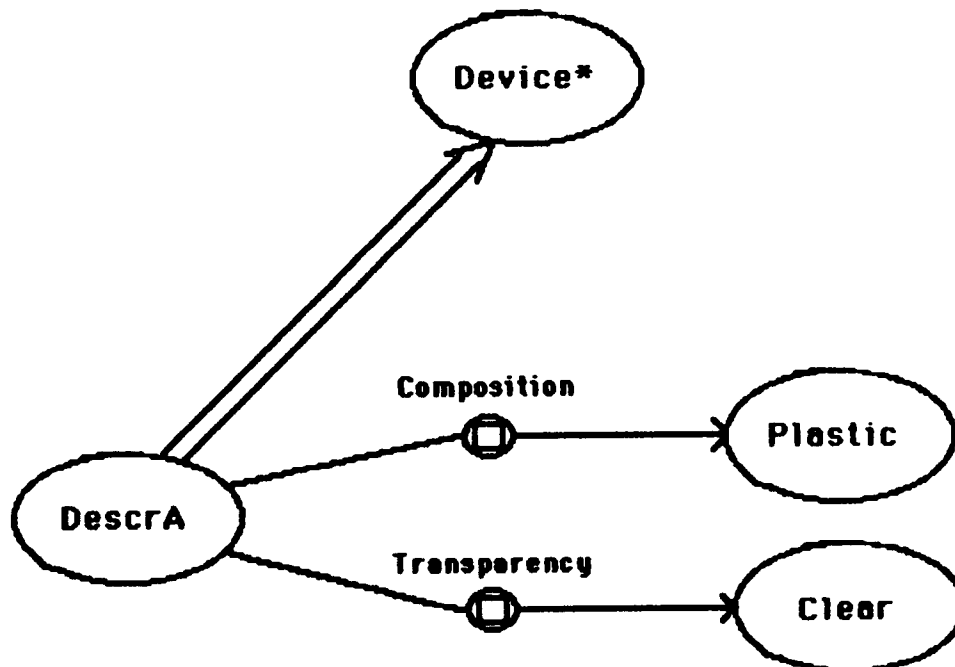


Figure 11. Adding DescrA to the taxonomy.

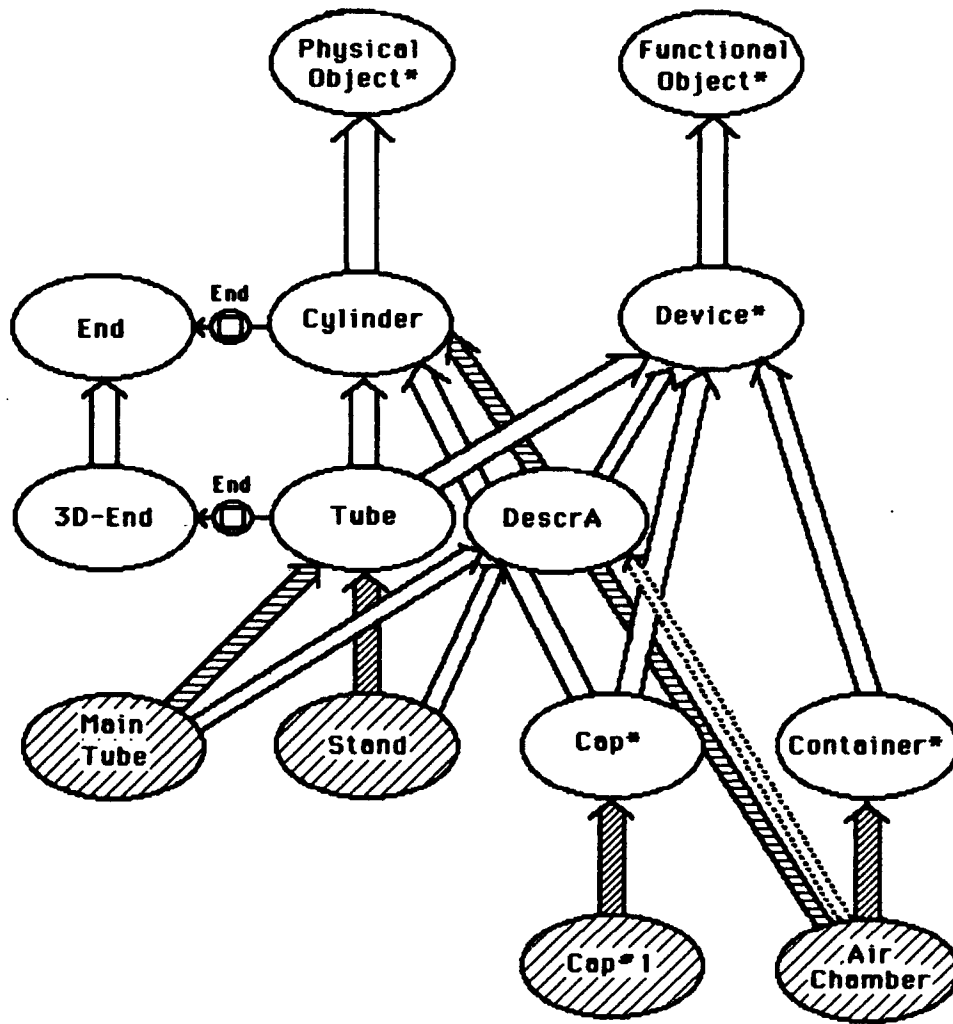


Figure 12. The classified DescrA.

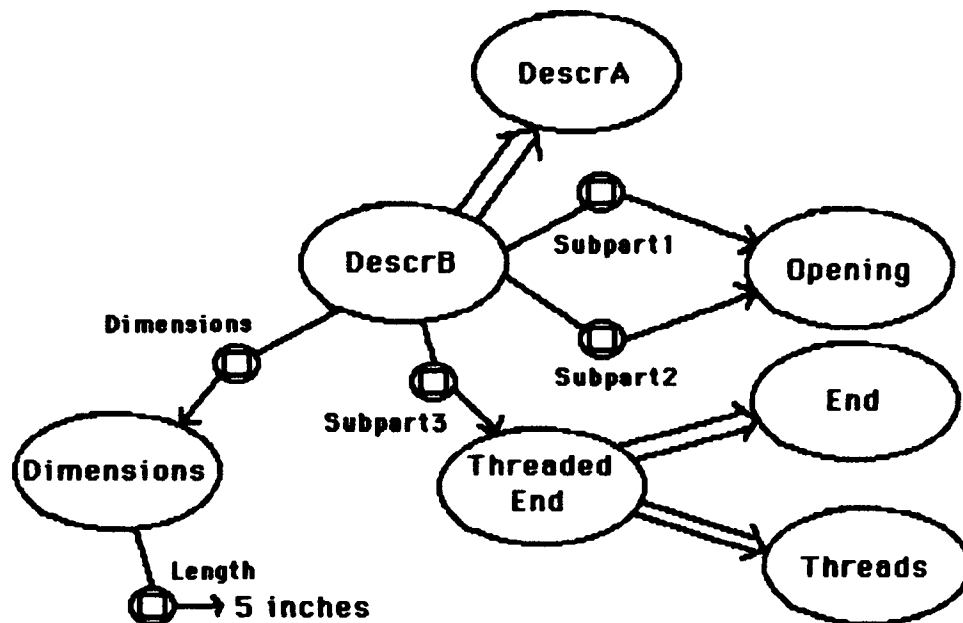


Figure 13. Adding DescrB to the taxonomy.

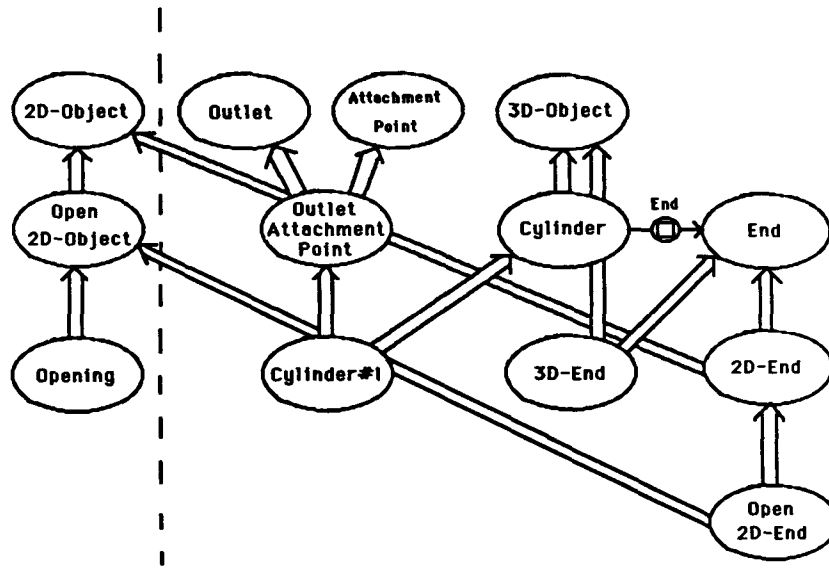


Figure 14. Attempt to match OPENING to CYLINDER#1.

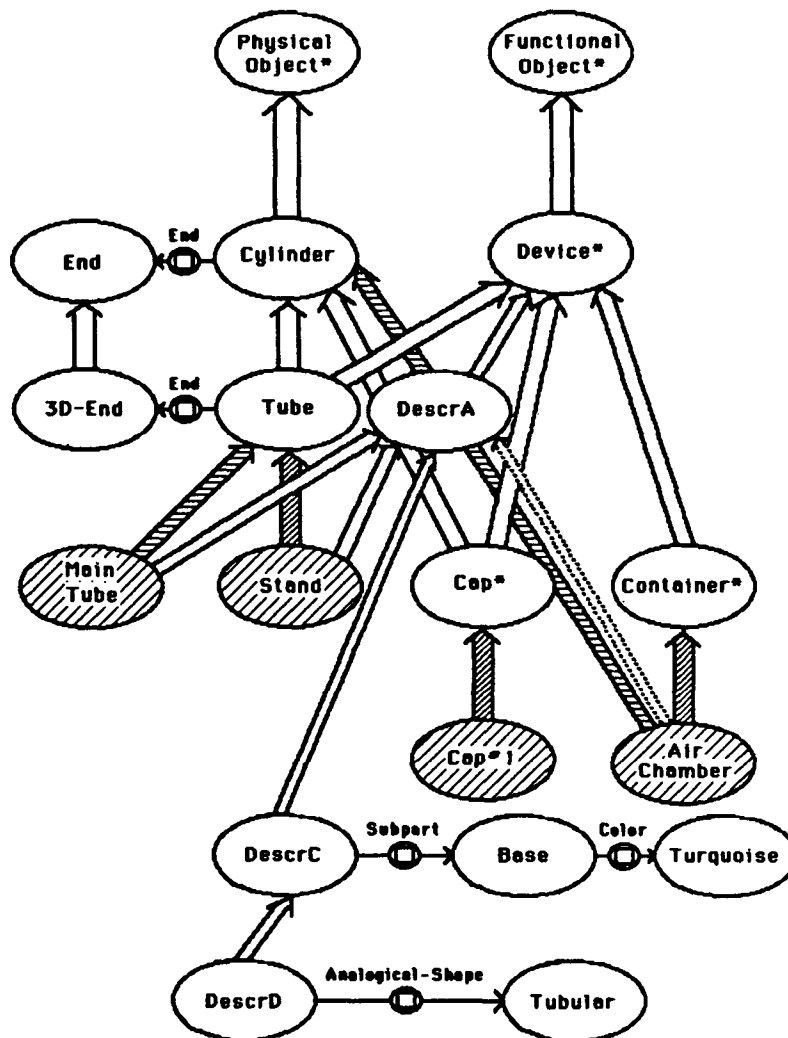


Figure 15. Adding DescrC and DescrD to the taxonomy.

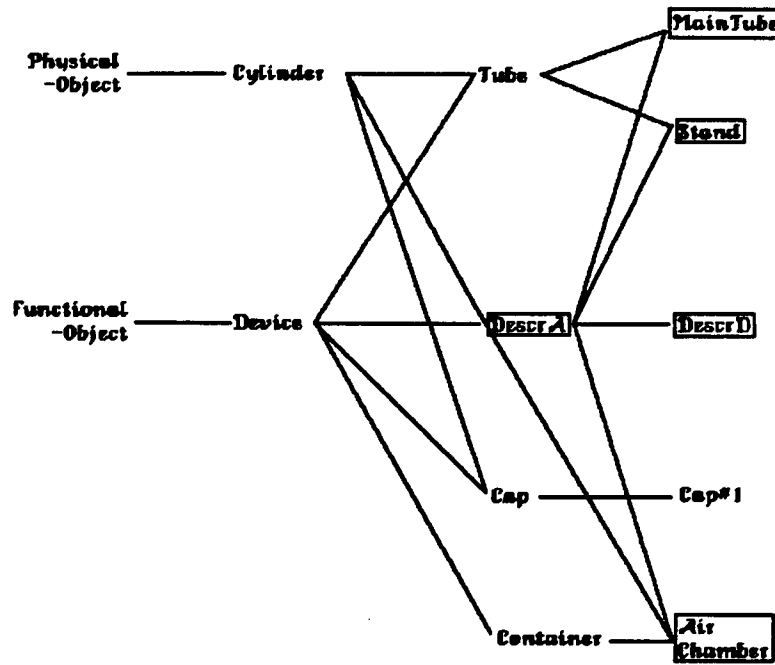


Figure 16. Exploring the taxonomy for referent candidates.

DescrD

	SuperC	Composition	Transparency	Shape	Subparts
Maintube	>	+	+	>	-
Stand	>	+	+	>	-
Air Chamber	>	+	?	>	-

Range of role scores:

Low Correlation - ? = < > + High Correlation

Figure 17. Scoring DescrD to the referent candidates.

Scoring DescrD to AIRCHAMBER:

- a CONTAINER is a kind of DEVICE; (>)
- the Transparency of DescrD, CLEAR, matches the Transparency of ChamberTop, ChamberOutlet, and ChamberBody of AIRCHAMBER but mismatches the Transparency of ChamberBottom of AIRCHAMBER. Therefore, the partial match is uncertain; (?)
- the Composition of each is PLASTIC; (+)
- the subparts of AIRCHAMBER have Shape HEMISPHERICAL and CYLINDRICAL which are each a kind of Shape ROUND; (>)

- the recursive partial matching of subparts: BASE in DescrD could match to the subpart in AIRCHAMBER that has a translation of (0.0 0.0 0.0) – i.e., ChamberBottom of AIRCHAMBER. However, they mismatch since color TURQUOISE in DescrD differs from color BLUE of AIRCHAMBER. (-)

Figure 17 summarizes the scoring. A weighted, overall numerical score is generated from the scores shown there.

The above analysis using the partial matcher provides no *clear* winner since the differences are so close, causing the scores generated for the candidates to be almost exactly the same (i.e., the only difference was in the score for Transparency). If there was a candidate that had a score significantly better than the others, then that candidate would be a clear winner. For example, a clear winner occurs if all but one of the candidates differ drastically in their feature values when compared to the feature values in the speaker's description. In that instance, it would be unnecessary to proceed further; we would assume the winner was our referent. For this example, however, all candidates will be retained.

At this point, the knowledge sources and their associated rules, mentioned earlier, apply. These rules attempt to order the feature values in the speaker's description for relaxation. First, we'll order the features in **DescrD** using linguistic knowledge. Linguistic analysis of **DescrD**, "... are clear plastic ... a rounded piece with a turquoise base ... Both are tubular ... fits loosely over ...," tells us that the features were specified using the following modifiers:

- Adjective: (Shape ROUND)
- Prepositional Phrase: (Subpart (BASE (Color TURQUOISE)))
- Predicate Complement: (Transparency CLEAR), (Composition PLASTIC), (Analogical-Shape TUBULAR), (Fit LOOSE)

Observations from the protocols (as described by the rules developed by Goodman (1984)) has shown that people tend to relax first those features specified as adjectives, then as prepositional phrases, and finally as relative clauses or predicate complements. Figure 5 shows this rule. The rule suggests relaxation of **DescrD** in the order:

```
{Shape} < {Color,Subpart}
      < {Transparency,Composition,
          Analogical-Shape,Fit}.
```

The set of features on the left side of a "<" symbol is relaxed before the set on the right side. The order that the features inside the braces, "{...}", are relaxed is left unspecified (i.e., any order of relaxation is all right). Perceptual information about the domain also provides suggestions. Whenever a feature has feature values that are close, then one should be prepared to relax any of them to any of the others (we call this the **clustered feature value rule**; it was motivated in Section 3.1.3). Figure 18 illustrates a set of assertions that compose a data base of similar color values in some domain. The Similar-Color predicate is defined to be reflexive and symmetric but not transitive. In this example, since a number of the color pairs are very close, color may be a reasonable thing to relax (see Figure 19). The clustered color rule defined in Figure 20 would suggest such a relaxation. It requires that at least three objects in the world have similar colors. It is meant as an exemplar for

a whole series of rules (e.g., ClusteredShapeValues, ClusteredTransparencyValues, and so on).

```
{Color} < {Shape,Subpart,Transparency,Composition,
            Analogical-Shape,Fit}.
```

```
Similar-Color ("BLUE","VIOLET")←
Similar-Color ("BLUE","TURQUOISE")←
Similar-Color ("GREEN","TURQUOISE")←
Similar-Color ("RED","PINK")←
Similar-Color ("RED","MAROON")←
Similar-Color ("RED","MAGENTA")←
...

```

Figure 18. Similar color values.

<i>Colors of Candidates & DescrD</i>	MainTube- violet Stand- blue Air Chamber- violet, blue DescrD- turquoise
--	---

Retrieve those Similar-Color assertions in the data base for the colors BLUE, VIOLET and TURQUOISE.

```
Similar-Color("BLUE","VIOLET")←
Similar-Color("BLUE","TURQUOISE")←
Similar-Color("GREEN","TURQUOISE")←
...

```

Figure 19. Objects with similar colors.

One can relax a feature whose feature values are clustered closely together before those of a non-clustered feature.

```
ClusteredFeatureValues(COLOR,w)
←Feature(COLOR),World(w),
ColorValue(c1),ColorValue(c2),ColorValue(c3),
WorldObj(o1,w),WorldObj(o2,w),WorldObj(o3,w),
Color(c1,o1),Color(c2,o2),Color(c3,o3),
Similar-Color(c1,c2),Similar-Color(c1,c3),
Similar-Color(c2,c3)

Relax-Feature-Before(v1,v2)
←ClusteredFeatureValues(feature(v1),w),
NOT(ClusteredFeatureValues(feature(v2),w))

```

Figure 20. The clustered color value rule.

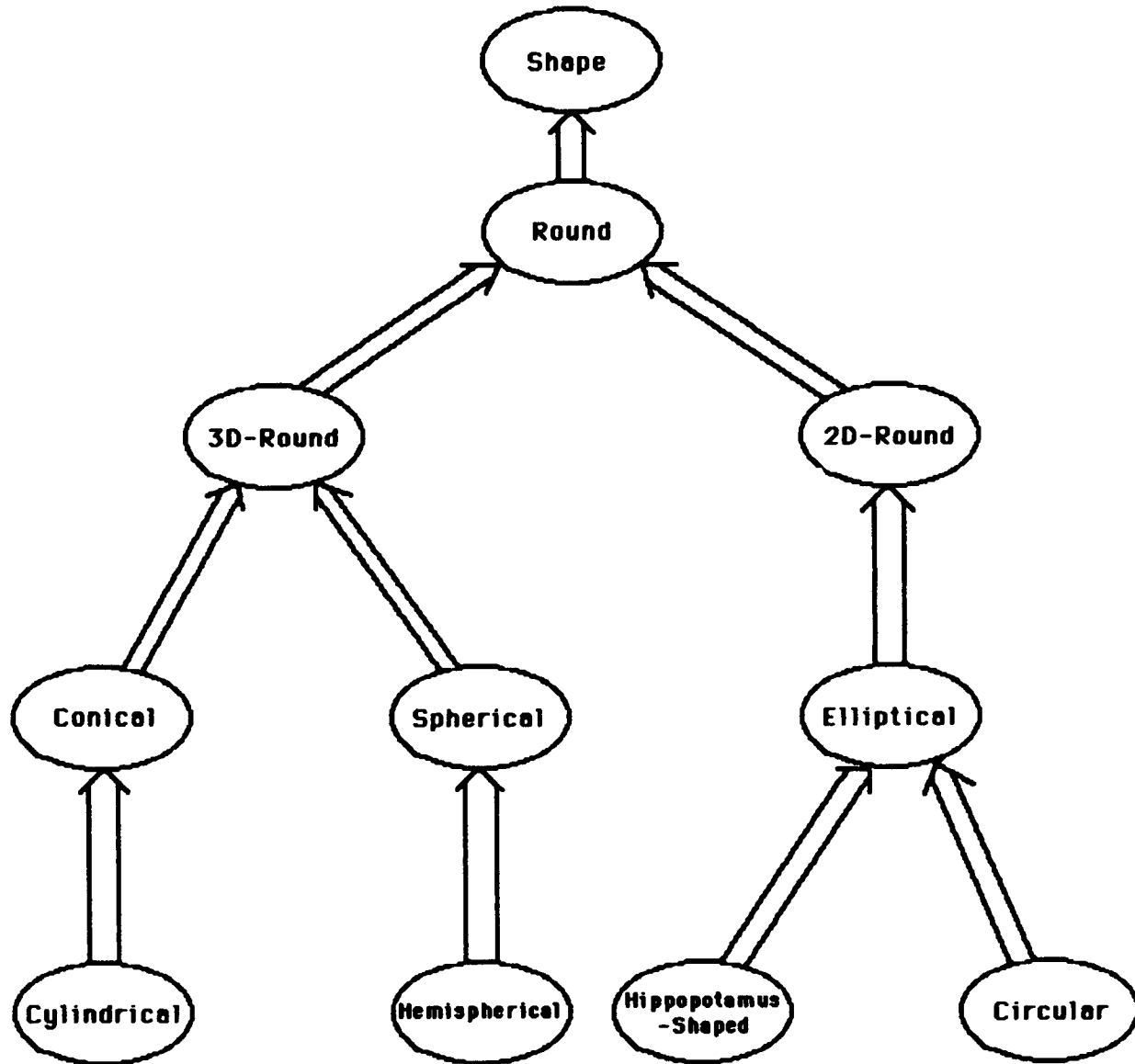


Figure 21. Hierarchical shape knowledge.

Hierarchical information about how closely related one feature value is to another can also be used to determine what to relax. The Shape values are a good example, as shown in Figure 21. A CYLINDRICAL shape is also a CONICAL shape, which is also a 3-D ROUND shape. Hence, it is very reasonable to match ROUNDED to CYLINDRICAL.

{Shape} < {Color,Subpart,Transparency,Composition, Analogical-Shape,Fit}.

The suggested orderings above can be merged. Since such orderings can have contradictory suggestions, we must describe the merging process. For example, in the above orderings, one says to relax Color first, while the

other says to relax Shape first. We combine both into one rule that says to relax *either* of them first: “{Shape,Color} < ...”. Another condition that occurs in the above orderings is when one rule says to relax a particular feature before others, while another rule does not care which of those features are relaxed first. In that case, one should use the more restrictive rule. Hence since one rule states that Subpart should be relaxed before the features Transparency, Composition, Analogical-Shape, and Fit, and the other rule does not care about their ordering, we split out Subpart and put it before the others:

{Subpart} < {Transparency,Composition, Analogical-Shape,Fit}.

Finally, all of these suggestions can be put together to form the order:

```
{Shape,Color} < {Subpart}
      < {Transparency,Composition,
          Analogical-Shape,Fit}.
```

The referent candidates MAINTUBE, STAND, and AIRCHAMBER can be examined and possibly ordered themselves using the above feature ordering. For this example, the relaxation of **DescrD** to any of the candidates requires relaxing their SHAPE and COLOR features. Since they each require relaxing the same features, the candidates can not be ordered with respect to each other (i.e., none of the possible feature orders is better for relaxing the candidates). Hence, no one candidate stands out as the most likely referent.

While no ordering of the candidates was possible, the order generated to relax the features in the speaker's description can still be used to guide the relaxation of each candidate. The relaxation methods mentioned at the end of the last section come into use here. Consider the shape values. The goal is to see if the ROUND shape specified in the speaker's description is similar to the shape values of each candidate. Generate-Similar-Shape-Values determines that it is reasonable to match ROUND to either the CYLINDRICAL or HEMISPHERICAL shapes of the AIRCHAMBER by examining the taxonomy shown in Figure 21 and noting that both shapes are below ROUND and 3D-ROUND. Notice that it is less reasonable to match CYLINDRICAL to HEMISPHERICAL since they are in different branches of the taxonomy. This holds equally true for the CYLINDRICAL shapes of the MAINTUBE and the STAND. Generate-Similar-Color-Values next tries relaxing the Color TURQUOISE. The assertions Similar-Color("BLUE","TURQUOISE")← and Similar-Color("GREEN","TURQUOISE")← are found as rules containing TURQUOISE. The colors BLUE and GREEN are, thus, the best alternates. Here only two clear winners exist – the AIRCHAMBER and the STAND – while the MAINTUBE is dropped as a candidate since it is reasonable to relax TURQUOISE to BLUE or to GREEN but not to VIOLET. Subpart, Transparency, Analogical-Shape, and Composition provide no further help (though the fact that the AIRCHAMBER has both CLEAR and OPAQUE subparts could be used to put it slightly lower than the STAND whose subparts are all CLEAR. This difference, however, is not significant.). This leaves trial and error attempts to try to complete the FIT action specified in **DescrE**. The one (if any) that fits – and fits loosely – is selected as the referent. The protocols showed that people often do just that – reducing their set of choices as best they can and then taking each of the remaining choices and trying out the requested action on them.

4.4 THE ACTUAL IMPLEMENTATION

The goal of our actual implementation of the reference and miscommunication mechanism was to provide a simulation of such a module in the context of a natural language system. We did not use an actual parser or semantic interpreter but assumed that we started with output expected from them. Such output was a representation in KL-One of the semantic interpretation of a description of an object in the water pump domain. We also built in KL-One a network of approximately 250 concepts to represent many of the water pump parts and their physical and functional features. A focus mechanism was simulated by a menu-driven routine that partitioned the network representation of the world to reflect focus spaces of referent candidates. We built a KL-One partial matcher and a network explorer to look for feasible referent candidates in the network. Finally, we wrote up a small batch of relaxation rules to test out our mechanism.

5 CONCLUSIONS

This paper had four objectives:

- to illustrate how complex reference really is;
- to show that previous referent identification paradigms don't suffice, given real world data;
- to isolate numerous kinds of knowledge people use for reference resolution; and
- to augment current reference algorithms to handle more real descriptions.

In this section, we provide a summary of our findings and present some reasonable future directions of this work.

5.1 SUMMARY

Our goal in this work is to build robust natural language understanding systems, allowing them to detect and avoid miscommunication. The goal is not to make a perfect listener but to make a more tolerant one that could avoid many mistakes, though it may still be wrong on occasion. In this paper, we introduced a taxonomy of miscommunication problems that occur in expert-apprentice dialogues. We showed that reference mistakes are one kind of obstacle to robust communication. To tackle reference errors, we described how to extend the succeed/fail paradigm followed by previous natural language researchers.

We represented real world objects hierarchically in a knowledge base using a representation language, KL-One, that follows in the tradition of semantic networks and frames. In such a representation framework, the reference identification task looks for a referent by comparing the representation of the speaker's input to elements in the knowledge base by using a matching procedure. Failure to find a referent in previous reference identification systems resulted in the unsuccessful termination of the reference task. We claim that people behave better than this and explicitly illustrated

such cases in an expert-apprentice domain about toy water pumps.

We developed a theory of relaxation for recovering from reference failures that provides a much better model for human performance. When people are asked to identify objects, they behave in a particular way: find candidates, adjust as necessary, re-try, and, if necessary, give up and ask for help. We claim that relaxation is an integral part of this process and that the particular parameters of relaxation differ from task to task and person to person. Our work models the relaxation process and provides a computational model for experimenting with the different parameters. The theory incorporates the same language and physical knowledge that people use in performing reference identification to guide the relaxation process. This knowledge is represented as a set of rules and as data in a hierarchical knowledge base. Rule-based relaxation provided a methodical way to use knowledge about language and the world to find a referent. The hierarchical representation made it possible to tackle issues of imprecision and over-specification in a speaker's description. It allows one to check the position of a description in the hierarchy and to use that position to judge imprecision and over-specification and to suggest possible repairs to the description.

Interestingly, one would expect that "closest" match would suffice to solve the problem of finding a referent. We showed, however, that it doesn't usually provide you with the correct referent. Closest match isn't sufficient because there are many features associated with an object and, thus, determining which of those features to keep and which to drop is a difficult problem due to the combinatorics and the effects of context. The relaxation method described circumvents the problem by using the knowledge that people have about language and the physical world to prune the search space.

We feel this research's implications for computational linguistics has to do primarily with the pragmatics of reference. In the past, when the logical form representing the speaker's description failed to denote a referent in the world, reference systems either failed or requested another possible logical form (i.e., they performed backtracking). We modify the reference architecture to perform non-equivalence transformations on the logical form to generate new ones using pragmatics.

5.2. FUTURE DIRECTIONS

This paper mentioned only a small aspect of what needs to be done with miscommunication. There are much broader problems that we also want to address. We alluded in the paper to problems due to metonymy – the use of the name of one thing for that of another – but never really tried in this work to handle more than a few special cases of it. Consider the three descriptions below.²³ Notice how the noun phrase *the window* refers to three different things in each utterance.

The window was broken.	(the glass)
The window was boarded up.	(the opening)
Open the window.	(the glass/frame inset)

Any reasonable reference mechanism must be able to distinguish such differences.

We neglected to discuss the effect of quantifiers in a description and how they affect the relaxation mechanism. A numerical quantifier could be specified that is incorrect; for example, *two pegs* when only *one* is found. Vague quantifiers such as *a bit*, *a few*, or *a piece* can be part of a description. Measurements like *a liter of beer* or *a pound of chicken* could also be used. A listener can estimate measurements when looking for the referent or be very precise and measure them. In all these examples, the relaxation of the quantifier could be required before a referent can be found. Our relaxation mechanism could be extended to handle many of these examples.

The FWIM reference identification system we developed models the reference process by the classification operation of KL-One. We need a more complicated model for reference. That model might need a complete identification plan that requires making inferences beyond those provided by classification. The model could also require the execution of a physical action by the listener before determining the proper referent. Cohen (1984:101) gives two excellent examples of such reference plans. The first, "the magnetic screwdriver, please," requires the listener to place various screwdrivers against metal to determine which is magnetic. The second, "the three two-inch long salted green noodles" requires the listener to count, examine, measure and taste to discover the proper referent.

The FWIM reference system uses relaxation rules to compile knowledge source information. These rules provide a convenient forum for evaluating a description with respect to language and physical knowledge about the world. However, reasoning mechanisms that "think" about these knowledge sources should really replace the rules and become part of the negotiation process.

There are also miscommunication problems that are outside of the reference area. We need to consider full utterances and the associated discourse in which they appear. Utterances can be imprecise or ill-formed with respect to the current discourse. The goals specified by a speaker through a particular utterance or discourse could be confused. For example, a speaker's requested goal could be outside the scope of the domain being discussed. We believe that our model will help solve the problem for this bigger picture. In particular, we feel the negotiation method will be important here, too. The negotiation process will become part of the plan recognition section of a natural language system. There a search of the plan space for the set of plans that might fit the utterance or sequence of utterances would be performed. A relaxation component related in style to the one outlined in this paper could be invoked to provide an orderly relaxation of the speaker's utterances to fit the plans and the

domain world. This process will require more interaction with the speaker through the use of clarification dialogues.

ACKNOWLEDGMENTS

I want especially to thank Candy Sidner for her insightful comments and suggestions during the course of this work. I'd also like to acknowledge the helpful comments of George Hadden, Diane Litman, Marie Macaaisa, Remko Scha, Marc Vilain, Dave Waltz, Bonnie Webber, and Bill Woods on this paper. Special thanks also to Phil Cohen, Scott Fertig, and Kathy Starr for providing me with their water pump dialogues and for their invaluable observations on them. I would also like to thank the reviewers for their valuable comments.

REFERENCES

- Agin, Gerald J. 1979 Hierarchical Representation of Three-Dimensional Objects Using Verbal Models. Technical Note 182, SRI International, Menlo Park, California.
- Allen, James F. 1979 A Plan-Based Approach to Speech Act Recognition. D.Phil. dissertation, University of Toronto, Toronto, Canada.
- Allen, James F.; Frisch, Alan M.; and Litman, Diane J. 1982 ARGOT: The Rochester Dialogue System. In *Proceedings of AAI-82*, Pittsburgh, Pennsylvania: 66-70.
- Appelt, Douglas E. 1981 Planning Natural Language Utterances to Satisfy Multiple Goals. D.Phil. dissertation, Stanford University, Palo Alto, California.
- Brachman, Ronald J. 1977 A Structural Paradigm for Representing Knowledge. D.Phil. dissertation, Harvard University, Cambridge, Massachusetts. Also, Technical Report No. 3605, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Brown, John Seely and VanLehn, Kurt. 1980 Repair Theory: A Generative Theory of Bugs in Procedural Skills. *Cognitive Science* 4(4): 379-426.
- Burling, R. 1970 *Man's Many Voices: Language in its Cultural Context*. Holt, Rinehart and Winston, New York, New York.
- Carberry, Mary Sandra. 1985 Pragmatic Modeling in Information System Interfaces. D.Phil. dissertation, University of Delaware, Newark, Delaware.
- Clark, Herbert H. and Clark, Eve V. 1977 *Psychology and Language*. Harcourt Brace Jovanovich, Inc., New York, New York.
- Cohen, Philip R. 1978 On Knowing What to Say: Planning Speech Acts. D.Phil. dissertation, University of Toronto, Toronto, Canada.
- Cohen, Philip R. 1981 The need for Referent Identification as a Planned Action. In: *Proceedings of IJCAI-81*, Vancouver, B.C., Canada: 31-35.
- Cohen, Philip R. 1984 The Pragmatics of Referring and the Modality of Communication. *Computational Linguistics* 10(2): 97-146.
- Cohen, Philip R.; Fertig, Scott; and Starr, Kathy. 1982 Dependencies of Discourse Structure on the Modality of Communication: Telephone vs. Teletype. In *Proceedings of ACL*, Toronto, Ont., Canada: 28-35.
- Cohen, P.; Perrault, C.; and Allen, J. 1981 Beyond Question Answering. In: Lehnart, W. and Ringle, M., Eds., *Strategies for Natural Language Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Gentner, Dedre. 1980 The Structure of Analogical Models in Science. Report 4451, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Goodman, Bradley A. 1981 The Representation of Three-Dimensional Objects. Unpublished manuscript, KRNL Group Working Paper, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Goodman, Bradley A. 1982 Miscommunication in Task-Oriented Dialogues. Unpublished manuscript, KRNL Group Working Paper, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Goodman, Bradley A. 1984 Communication and Miscommunication. D.Phil. dissertation, University of Illinois, Urbana, IL. To appear as a book in the Association of Computational Linguistics series of Cambridge University Press, London, England.
- Grice, H. P. 1975 Logic and Conversation. In Cole, P. and Morgan, J., Eds., *Syntax and Semantics*. Academic Press, New York, New York: 41-58.
- Grosz, Barbara J. 1977 The Representation and Use of Focus in Dialogue Understanding. D.Phil. dissertation, University of California, Berkeley, California. Also, Technical Note 151, Stanford Research Institute, Menlo Park, California.
- Grosz, Barbara J. 1978 Focusing in Dialog. In *Theoretical Issues in Natural Language Processing-2*. Urbana, Illinois: 96-103.
- Grosz, Barbara J. 1981 Focusing and descriptions in natural language dialogues. In Joshi, Webber, and Sags, Eds., *Elements of Discourse Understanding*. Cambridge University Press, Cambridge, England: 84-105.
- Hoepfner, W.; Christaller, T.; Marburger, H.; Morik, K.; Nebel, B.; O'Leary, M.; and Wahlster, W. 1983 Beyond Domain-Independence: Experience with the Development of a German Language Access System to Highly Diverse Background Systems. In *Proceedings of IJCAI-83*, Karlsruhe, West Germany: 588-594.
- Lipkis, Thomas. 1982 A KL-One Classifier. In *Proceedings of the 1981 KL-One Workshop*. Jackson, New Hampshire: 128-145. Report No. 4842, Bolt Beranek and Newman Inc., Cambridge, Massachusetts. Also Consul Note #5, USC/Information Sciences Institute, October 1981.
- Litman, Diane. 1983 Discourse and Problem Solving. Report No. 5338, Bolt Beranek and Newman Inc., Cambridge, Massachusetts. Also, TR130, University of Rochester, Department of Computer Science, Rochester, New York.
- Litman, Diane J. 1985 Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues. D.Phil. dissertation, University of Rochester, Rochester, New York. Also, TR170, University of Rochester, Dept. of Computer Science, Rochester, New York.
- Litman, Diane J. and Allen, James F. 1984 A Plan Recognition Model for Clarification Subdialogues. In *Proceedings of Coling84*. Stanford University, Stanford, California: 302-311.
- Mark, William. 1982 Realization. In *Proceedings of the 1981 KL-One Workshop*. Jackson, New Hampshire: 78-89. Report No. 4842, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- McCoy, Kathleen F. 1985a The Role of Perspective in Responding to Property Misconceptions. In *Proceedings of IJCAI-85*. Los Angeles: 791-793.
- McCoy, Kathleen F. 1985b Correcting Object-Related Misconceptions. D.Phil. dissertation, University of Pennsylvania, Philadelphia, Pennsylvania. Also, MS-CIS-85-57, University of Pennsylvania, Department of Computer and Information Science, Philadelphia, Pennsylvania.
- McDonald, David D. and Conklin, E. Jeffery. 1982 Salience as a Simplifying Metaphor for Natural Language Generation. In *Proceedings of AAI-82*. Pittsburgh, Pennsylvania: 75-78.
- McKeown, Kathleen R. 1983 Recursion in Text and Its Use in Language Generation. In *Proceedings of AAI-83*. Washington, D.C.: 270-273.
- Perrault, C. Raymond and Cohen, Philip R. 1981 It's for your Own Good: a Note on Inaccurate Reference. In Joshi, Webber, and Sags, Eds., *Elements of Discourse Understanding*. Cambridge University Press, Cambridge, England: 217-230.
- Polanyi, Livia 1978 False Starts Can Be True. In *Proceedings of Fourth Annual Meeting of Berkeley Linguistics Society*. University of California, Berkeley, California: 628-639.
- Polanyi, Livia and Scha, Remko. 1984 A Syntactic Approach to Discourse Semantics. In *Proceedings of Coling84*. Stanford University, Stanford, California: 413-419.
- Pollack, Martha E. 1986 Inferring Domain Plans in Question-Answering. D.Phil. dissertation, University of Pennsylvania, Philadelphia, Pennsylvania. Also, Report MS-CS-86-40 of the Department of Computer and Information Science, University of Pennsylvania.
- Reichman, Rachel. 1978 Conversational Coherency. *Cognitive Science* 2(4): 283-327.

- Reichman, Rachel. 1981 Plain Speaking: A Theory and Grammar of Spontaneous Discourse. D.Phil. dissertation, Harvard University, Cambridge, Massachusetts. Also, Technical Report No. 4861, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Ringle, Martin and Bruce, Bertram. 1981 Conversation Failure. In Lehnart, W. and Ringle, M., Eds., *Strategies for Natural Language Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Sidner, Candace Lee. 1979 Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. D.Phil. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts. Also, Report No. TR-537, MIT AI Lab, Cambridge, Massachusetts.
- Sidner, Candace L. 1985 Plan Parsing for Intended Response Recognition in Discourse. *Computational Intelligence* 1(1): 1-10.
- Sidner, C. L., and Israel, D.J. 1981 Recognizing Intended Meaning and Speaker's Plans. In *Proceedings of IJCAI-81*, Vancouver, B.C.: 203-208.
- Sidner, C. L.; Bates, M.; Bobrow, R. J.; Brachman, R. J.; Cohen, P. R.; Israel, D. J.; Schmolze, J.; Webber, B. L.; and Woods, W. A. 1981 Research in Knowledge Representation for Natural Language Understanding. Report No. 4785, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Sidner, C. L.; Bates, M.; Bobrow, R.; Goodman, B.; Haas, A.; Ingria, R.; Israel, D.; McAllester, D.; Moser, M.; Schmolze, J.; and Vilain, M. 1983 Research in Knowledge Representation for Natural Language Understanding - Annual Report, 1 September 1982 - 31 August 1983. Technical Report 5421, BBN Laboratories Inc., Cambridge, Massachusetts.
- Tversky, A. 1977 Features of Similarity. *Psychological Review* 84: 327-352.
- Webber, Bonnie Lynn. 1978 A Formal Approach to Discourse Anaphora. D.Phil. dissertation, Harvard University, Cambridge, Massachusetts. Also, Technical Report No. 3761, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Whorf, Benjamin Lee. 1956 *Language, Thought, and Reality*. The MIT Press, Cambridge, Massachusetts.
- Winograd, Terry. 1971 Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. D.Phil. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts. Also, Report No. TR-84, Project MAC, MIT, Cambridge, Massachusetts.
- Woods, W.A.; Kaplan, R.M.; and Nash-Webber, B.L. 1972 The Lunar Sciences Natural Language Information System: Final Report. BBN Report 2378, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

NOTES

1. This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by ONR under Contract No. N00014-77-C-0378 and N00014-85-C-0079. The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.
2. We will call these errors **misreference**.
3. An analysis of clarification subdialogues can be found in Litman and Allen 1984, Litman 1985.
4. Of course, there are some situations - such as teaching - where the hearer would be more willing to tolerate overspecific descriptions.
5. *Chamber* was interpreted here in a broader sense by the listener because it was used right at the beginning of the dialogue. This was before the speaker introduced other terms, such as *tube*, that would have helped distinguish the pieces better. The example demonstrates how discourse affects reference.
6. The whole word here is *plastic*. In these protocols, people often guess before hearing the whole utterance or even whole words.
7. Grosz (1977, 1981) would describe this as a difference in "task plans" while Reichman (1978, 1981) would say that the "communicative goals" differed.
8. Here we assume that either the speaker and the hearer have a shared perceptual context or the speaker has an extensive model of the hearer's perceptual context.
9. For example, Burling (1970) and Clark (1977) contrasted vocabulary in Garo, a language spoken in Burma, with English. Burling found that some words in English were accounted for in Garo by many words. The word *rice* was represented in Garo by different names for "husked", "unhusked", "cooked", "uncooked", and other forms of rice. Distinguishing the object types referred to by such specialized names would be more difficult for non-Burmese. Whorf (1956) found similar results in his studies.
10. Other descriptions such as *the second one from the left* are usable only when the speaker and listener are sharing the same perceptual view. Even when the same view is shared, the underlying task may also affect whether such a description is sufficient. For example, if the speaker is trying to teach an intractable robot how to perform a task, then a description such as *the second one from the left* may not be properly generalized by the robot for use in future perceptual views of the world.
11. We want to thank one of the reviewers for this example.
12. In more complex domains - such as ones requiring tools - the actions themselves may be helpful both in finding the referent and in confirming whether the choice was correct. For example, if a listener is told to use a screwdriver to screw one object onto another, the listener would expect to find threads on the object.
13. Note that the postcondition need not always be specified explicitly since some postconditions automatically come with an action. If the speaker said the utterance *fit the red gizmo into the bottom side outlet of the main tube*, the listener would expect that the red gizmo would fit snugly into the outlet. If, however, it fit loosely, than the listener may feel a mistake has occurred.
14. We have only written rules for focus so far.
15. Of course, once one particular candidate is selected, then deciding which features to relax is relatively trivial - one simply compares feature by feature between the candidate description (the target) and the speaker's description (the pattern) and notes any discrepancies.
16. These interruptions are more typical of spoken rather than written language.
17. Actually the orderings are not necessarily partial since it is possible within one knowledge source to have conflicting relaxation rules - i.e., one says relax feature f_1 before feature f_2 and the other says the opposite. We get around this problem by splitting the knowledge source ordering into two or more orderings that are partial.
18. The topological distance of a feature f in a feature ordering F_j is determined by counting how far f is from the nearest start node in F_j . A start node is a node that has no links coming into it but can have links coming out of it. Distance is then measured by counting the links between f and the start node.
19. The latter case is there primarily for the times when one can't easily define a similarity metric for a feature. McCoy (1985a) and Tversky (1977) provide additional discussions about similarity metrics.
20. We are stretching the definition of KL-One here with the dotted subsumption arrow. The point we want to make is that AIRCHAMBER is similar to DescrA because their descriptions are almost exactly the same.
21. Since DescrB refers to MAINTUBE, MAINTUBE could be dropped as a potential referent candidate for DescrC. We will, however, leave it as a potential candidate to make this example more complex. CAP could also be considered, but it is dropped since it was already eliminated by DescrA and it does not contain any subparts.
22. The partial matcher scores are numerical scores computed from a set of role scores that indicate how well each feature of the two descriptions match. Those feature scores are represented on a scale: {+}, {>} or {<}, {=}, {?}, {-}. + is the highest and - is the lowest score. > and < have the same score but the algorithm can distinguish between them.
23. These examples were suggested to me by Bonnie Webber.