DOCUMENT RESUME

ED 281 264 CS 505 609

AUTHOR	Goodman, Bradley A.
TITLE	Rule-Based Relaxation of Reference Identification Failures. Technical Report No. 396.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Dec 86
CONTRACT	400-81-0030
NOTE	73p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Artificial Intelligence; *Cognitive Processes; Cognitive Structures; *Communication Problems; Communication Research; Computational Linguistics; *Language Processing; Language Research; Listening Skills; *Man Machine Systems; Misconceptions; Programing Languages
IDENTIFIERS	Computer Models; Computer Users; Natural Language Communication; *Natural Language Processing; *Referents (Linguistics)

ABSTRACT

In a step toward creating a robust natural language understanding system which detects and avoids miscommunication, this artificial intelligence research report provides a taxonomy of miscommunication problems that arise in expert-apprentice dialogues (including misunderstandings, wrong communication, and bad analogies), and proposes a flexible extension of the succeed/fail paradigm to handle reference mistakes. Extended examples of these reference failures are provided. A theory of relaxation (similar to human referent identification processes) for recovering from reference failures is then developed using the representational language "KL-One," which represents real-world objects hierarchically (in the tradition of semantic networks and frames). Rule-based relaxation is described as an integral part of the process whereby people who are asked to identify objects behave in a particular way: by finding candidates, re-trying, and if necessary, giving up and asking for help. The study models the relaxation process and provides a computational model for experimenting with the different parameters of the relaxation process. Extensive examples show how the model handles problems with imprecision and over-specification in a speaker's description using rules in a hierarchical knowledge base. Finally, the relaxation model is shown to be preferable to "closest match" models, which fail because looking for the closest match cannot determine the most salient features of a possible referent. The report indicates that the relaxation model avoids this problem by determining salient features using the kinds of knowledge humans have about language and the physical world. Many computer-language examples and diagrams supplement the discussion. A 6-page reference list is appended. (SKC)

CENTER FOR THE STUDY OF READING

Technical Report No. 396

RULE-BASED RELAXATION OF REFERENCE IDENTIFICATION FAILURES

Bradley A. Goodman
BBN Laboratories
Cambridge, MA 02238

December 1986

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The work upon which this publication was based was performed pursuant to Contract No. 400-81-0030 of the National Institute of Education. It does not, however, necessarily reflect the views of the agency. Special thanks to Candy Sidner for her insightful comments and suggestions during the course of this work. George Hadden, Diane Litman, Remki Scha, Marc Vilain, Dave Waltz, Bonnie Webber, and Bill Woods made many helpful comments on this paper. Many thanks also to Phil Cohen, Scott Fertig and Kathy Starr for providing the water pump dialogues and for their invaluable observations on them.

Abstract

The goal of this work is the enrichment of human-machine interactions in a natural language environment.[1] We want to provide 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. A speaker and listener cannot be assumed to have the same beliefs, contexts, perceptions, backgrounds or goals at each point in a conversation. As a result, 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 down and possibly break down communication. Our goal is to recognize and isolate such miscommunications and circumvent them. This paper will highlight a particular class of miscommunication--reference problems--by describing a case study and techniques for avoiding failures of reference.

1. Introduction

Cohen, Perrault and Allen argued in their paper "Beyond Question Answering" (1981) 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. This paper summarizes the results of a dissertation (Goodman, 1984) that investigates the kinds of miscommunication that occur in human communication with a special emphasis on reference problems, i.e., problems a listener has determining whom or what a speaker is talking about. We have written computer programs and algorithms that demonstrate how one could solve such problems in a natural language understanding system. The study of miscommunication is a necessary task for 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.

Our current research (Sidner, Bates, Bobrow, Brachman, Cohen, Israel, Schmolze, Webber, & Woods, 1981; Sidner, Bates, Bobrow, Goodman, Haas, Ingria, Israel, McAllester, Moser, Schmolze, & Vilain, 1983) views most dialogues as being cooperative and goal-directed, i.e., a speaker and listener work together to achieve a common goal. The interpretation of an utterance involves identifying the underlying plan or goal that the utterance reflects (Cohen, 1978; Allen, 1979; Sidner & Israel, 1981; and Sidner, 1985). This plan, however, is rarely, if ever, obvious at the surface sentence level. A central issue is to transform sequences of complex, imprecise, or ill-devised utterances into well-specified plans that might be carried out by dialogue participants. Within this context, miscommunication can occur.

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 a listener knows about a topic, it is easy to make statements that can be misinterpreted or not understood by the listener 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 imprecise, confused, ambiguous or overly specific, or might be interpreted in the wrong context. As a result, the listener cannot determine what object is being described (we will call these errors "misreference"). The descriptions, which cause reference identification failure, are "ill-formed." The blame for ill-formedness may lie partly with the speaker and partly with the listener. The speaker may have been sloppy or not taken the hearer into consideration; the listener may be either remiss or unwilling to admit he can't understand the speaker and to ask the speaker for clarification, or may simply believe that he has understood when he, in fact, has not.

This 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 the previous paradigms proved inappropriate in the real world, given the data we've analyzed. We introduce a new process called "negotiation" that is used when reference fails, and we illustrate this by introducing a new computational model called FWIM, for "Find What I Mean." We develop a theory called extensional reference miscommunication that will help explain how people successfully use imperfect descriptions.

The last part of this section provides an introduction to the work and the methodology used. Section 2 of this paper 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 illustrating them with examples. Section 3 presents initial solutions to some of the problems of miscommunication.

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 that are 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 (1981, 1984; Cohen, Fertig, & 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 elements in it. The data provide examples of imprecision, confusion, and ambiguity, as well as attempts to correct these problems.

In the following exchange, A is instrucing J to assemble part of the water pump. Refer to Figure 1(a) for a picture of the pump. A and J 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 (1) listeners repair the speaker's description in order to find a referent, (2) they repair their initial reference choice once they are given more information, and (3) they can fail to choose a proper referent. In Line 7, A describes the two holes on the BASEVALVE as "the little hole." J, realizing that A doesn't really mean "one" hole but "two," must repair the description. J does, since he doesn't complain about A's description, and correctly attaches the BASEVALVE to the TUBEBASE. Figure 1(b) shows the pump after the TUBEBASE is attached to the MAINTUBE in Line 10. In Line 13, J interprets "a red plastic piece" to refer to the NOZZLE. When A adds the relative clause "that has four gizmos on it," J is forced to drop the NOZZLE as the referent and to select the SLIDEVALVE. In Lines 17 and 18, A's description "the other--the open part of the main tube, the lower valve" is ambiguous, and J selects the wrong site, namely the TUBEBASE, in which to insert the SLIDEVALVE. Since the SLIDEVALVE fits, J doesn't detect any trouble. Lines 20 and 21 keep J from thinking that something is wrong because the part fits loosely. In Lines 27 and 28, J indicates that A has not given him enough information to perform the requested action. In Line 30, J further compounds the error in Line 18 by putting the SPOUT on the TUBEBASE.

Excerpt 1 (Telephone)

A: 1. Now there's a blue cap
[J grabs the TUBEBASE]
2. that has two little teeth sticking
3. out of the bottom of it.

J: 4. Yeah.

A: 5. Okay. On that take the
6. bright shocking pink piece of plastic
[J takes BASEVALVE]

7. and stick the little hole over the teeth.
[J starts to install the BASEVALVE, backs off, looks at it again and then goes ahead and installs it]

J: 8. Okay.

A: 9. Now screw that blue cap onto
10. the bottom of the main tube.
[J Screws TUBEBASE onto MAINTUBE]

J: 11. Okay.

A: 12. Now, there's a--
13. a red plastic piece
[J starts for NOZZLE]

14. that has four gizmos on it.
[J switches to SLIDEVALVE]

J: 15. Yes.

A: 16. Okay. Put the ungizmoed end in the uh
17. the other--the open
18. part of the main tube, the lower valve.
[J puts SLIDEVALVE into hole in TUBEBASE, but A meant OUTLET2 of MAINTUBE]

J: 19. All right.

A: 20. It just fits loosely. It doesn't
21. have to fit right. Okay, then take
22. the clear plastic elbow joint.
[J takes SPOUT]

J: 23. All right.

A: 24. And put it over the bottom opening, too.
[J tries installing SPOUT on TUBEBASE]

J: 25. Okay.

A: 26. Okay. Now, take the--

J: 27. Which end am I supposed to put it over?
28. Do you know?

A: 29. Put the--put the--the big end--
30. the big end over it.
[J 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 but that they try to plow ahead anyway. The rest of this paper will formalize the kinds of problems that occur during reference and then extend the reference paradigm to get around many of them.

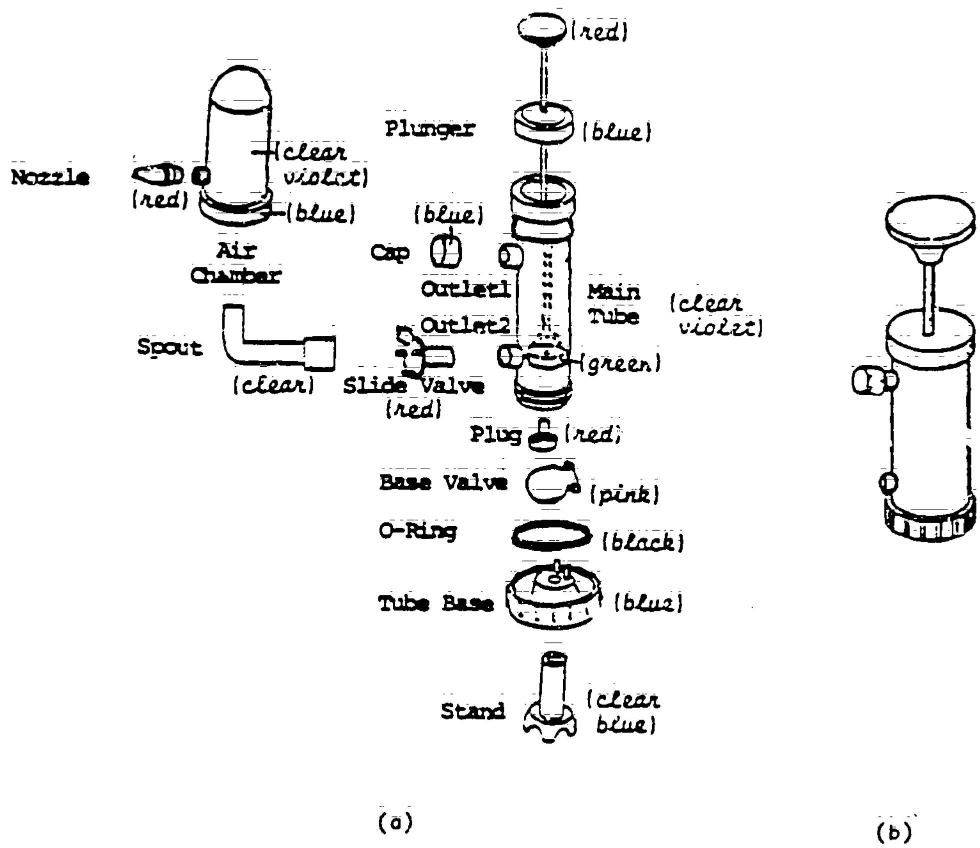


Figure 1: The Toy Water Pump

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.[2]

There are many aspects of an utterance that can confuse the listener and lead to miscommunication. The listener can become confused about what the speaker intends for the objects, the actions, and the goals described by the utterance. Confusions often appear to result from conflict between the current state of the conversation, the overall goal of the speaker, and the manner in which the speaker presented the information. However, when the listener steps back and is able to discover what kind of confusion is occurring, then that can be resolved.

2.1 Causes of Miscommunication

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 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, too.

The viewpoints of the expert and apprentice differ greatly in exchanges. The expert, understanding the functionality of the elements in the task, has more of a feel for how they work and go together, and how they can be used. The apprentice normally has no such knowledge and must base his decisions on his perceptions such as shape (Grosz, 1981).

The structure of the task affects the structure of the dialogue (Grosz, 1977), as the expert and apprentice accomplish each step of the task. The common center of attention of the dialogue participants is called the focus (Grosz, 1977; Reichman, 1978; and Sidner, 1979). Shifts in focus correspond to shifts between the tasks and subtasks. Focus and focus shifts are governed by many rules (Grosz, 1977; Reichman, 1978: and Sidner, 1979). Confusion may result when expected shifts do not take place. For example, if the expert changes focus to some object but does not talk about the object soon after its introduction (i.e., before it is used), without digressing in a well-structured way (see Reichman, 1978), or never discusses its subpieces (such as an obvious attachment surface), then the apprentice may become confused, leaving him ripe for miscommunication. 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 itself to the apprentice's world will also perplex him.

Focus also influences descriptions (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 formulating the description or may consider shifting focus away from the confusing objects.

2.1.2 Discrepancies in knowledge and miscommunication. Just as with discrepancies in focus, discrepancies in knowledge between the speaker and listener can cause miscommunication. These disagreements can occur because the listener does not bring sufficient knowledge and the speaker fails to convey enough information to give him the knowledge sufficient to perform the task (that knowledge becomes shared or mutually believed knowledge (Clark & Marshall, 1981; Perrault & Cohen, 1981; Joshi, 1982; Nadathur & Joshi, 1983). The speaker and listener could also have different beliefs. For example, they could differ on what each believes about the other, which can lead to false assumptions that each may use when interpreting the other's utterances. Knowledge differences, though, can sometimes provide a means to help detect miscommunication. For example, a listener's knowledge about the world in which the task is taking place can provide a way of checking whether or not a speaker's utterance is realistic.

Knowledge the listener brings to the task. In apprentice-expert dialogues such as those about the water pump, the knowledge brought to the task by a naive apprentice is limited to four principal areas: (1) language abilities, (2) perceptual abilities to identify objects, (3) past experience and knowledge in assembling objects, and (4) the ability to perform trial-and-error tests in the real world. The language abilities of the apprentice allow him to follow the flow of information provided by the expert in his utterances and descriptions. This knowledge about language is syntactic, semantic and pragmatic.

Perceptual abilities include recognizing physical features of an object such as its size, shape, color, location, composition and transparency. The fineness of each category's partitioning varies among individuals. For example, some people know more color values than others. An expert, if he wishes to prevent misreference, may choose to use only basic level descriptions in each category until the apprentice demonstrates a broader knowledge, or the expert can familiarize the apprentice with other values.

The past experience someone has with objects provides a method for the expert to tie a description down to a common point of view. If an object has a familiar name, the expert can refer to it by that name. The expert can also refer by making analogies to everyday objects through shapes or functions as a model for the apprentice in his selection of a referent. The same holds true for actions--past experience makes it easier for the expert to describe an action to the apprentice.

Finally, the apprentice brings to a task the ability to perform simple tests. He can experiment to determine whether two pieces can be attached. In the water pump domain, attachment is performed by pushing, twisting or screwing one object into or onto another. How good a fit is can be determined by noting the compatibility of the shapes of the attaching surfaces (and this can be used to align the surfaces) and by checking the snugness of the fit once the objects are attached.

The knowledge transferred in an utterance. At least two kinds of knowledge are conveyed in an utterance. For this paper we will focus on task knowledge and communicative knowledge. Task knowledge about the specific domain is used to fill the propositional content of an utterance. In the water pump domain it refers to: (1) the objects, the set of parts available to accomplish the task (i.e., the "real world" which is the physical environment around the conversational participants); (2) the actions, the set of physical actions available to the listener; and (3) instructions linking objects and actions together to achieve some goal.

Communicative knowledge consists of speech acts, communicative goals, and communicative actions. Speech acts are underlying forms that are performed by the speaker in expressing an utterance (e.g., REQUEST, INFORM) (Searle, 1969; Cohen, 1978; and Allen, 1979). They provide an illocutionary force that is applied to the proposition expressed. Communicative goals reflect the structure of the discourse (e.g., setting up a topic, clarifying, or adding more information (Allen, Frisch, & Litman, 1982)). They express how an utterance is to be understood with respect to the high-level communicative goals reflected in the structure of the dialogue and, hence, how the task the utterance examines is performed. A communicative act is a way of accomplishing the goal that one wants to (e.g., communicate the goal, communicate the object's description, communicate the action). Only some of the possible acts may be reasonable at any one time to reach the current communicative goal (Reichman, 1981; Allen, Frisch, & Litman, 1982; Litman, 1983).

Miscommunication can occur because of the way the information was transferred (e.g., communicative knowledge) or the content (e.g., task knowledge). Task knowledge-based miscommunication occurs when the speaker is unaware that the listener (1) has a different view of the task, (2) is considering a different subset of objects, or (3) is considering a different subset of actions, and so on. Difficulties with communicative knowledge can occur when the speaker uses the wrong speech act (e.g., utters something inadvertently that would be conventionally interpreted as an INFORM when meant as a REQUEST) or when the listener errs in interpreting the speaker's intention (e.g., the speaker may be INFORMing the listener that the blue cap fits around the end of the tube but the listener might interpret the utterance as a REQUEST to actually place the cap around the end of the tube). In both cases it is the effect of the speech act that causes the trouble since it influences what the listener will do (i.e., determine the intended responses). Finally, communicative knowledge can cause mistakes and confusion if the listener and speaker differ on the goal (e.g., the listener might think the speaker is clarifying previous information when, in fact, the speaker is adding new information). They will feel they are communicating at cross purposes--leading to frustration.

2.2 Instances of Miscommunication

In this section we will present evidence that people do miscommunicate and yet they often manage to repair reference failures. We will look at specific forms of miscommunication and describe ways to detect them and will demonstrate ways for resolving some miscommunication problems.

There are many ways hearers can get confused during a conversation. Figure 2 outlines some of them that were derived from analyzing the water pump protocols. We will only discuss referent confusion in this paper. The other forms of confusion--Action, Goal, and Cognitive Load--are described in Goodman, (1982, 1984). Another categorization of confusions that lead to conversation failure can be found in Ringle and Bruce (1981).

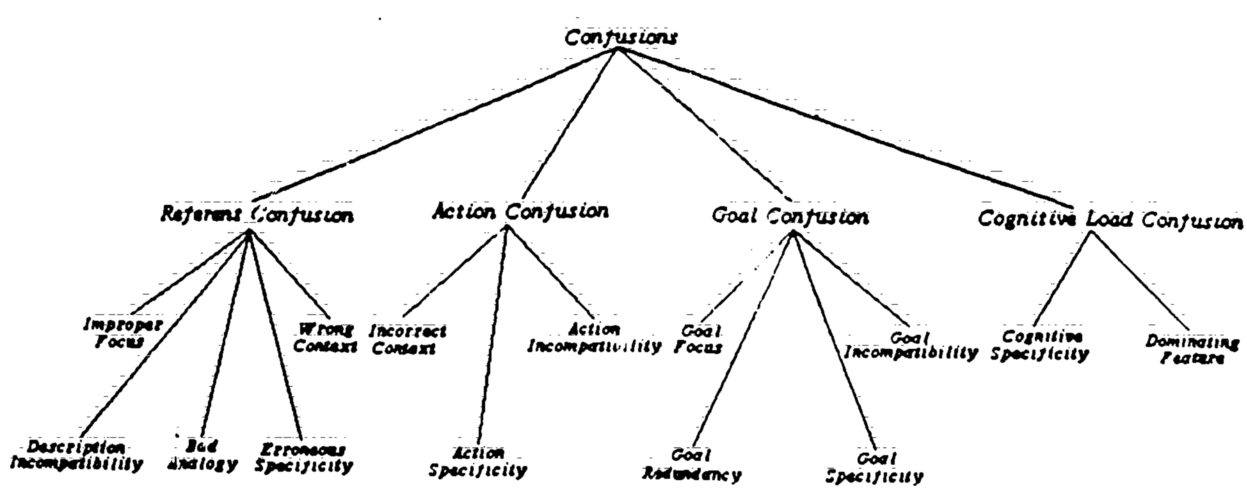


Figure 2: A taxonomy of confusions

Referent confusion occurs when the listener is unable to determine correctly what the speaker is referring to. It may occur when the descriptions in the utterance are ambiguous or imprecise, when there is confusion between the speaker and listener about what the current focus or context is, or when the descriptions are either incorrect or incompatible with the current or global context.

This section defines and illustrates many of the confusions through numerous excerpts. Each excerpt has marked in parentheses the 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.

<u>Erroneous specificity</u>. A speaker's overspecific or underspecific descriptions 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, he might think that he had done something incorrectly as the task seemed easier than the one apparently described by the speaker.[3] For example, in Excerpt 2, S'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)

S: 1. Okay?

2. Now you have two devices that
3. are clear plastic

[J picks up MAINTUBE and SPOUT]

J: 4. Okay.

S: 5. One of them has two openings
6. on the outside with threads on
7. the end, and its about five
8. inches long.

[J rotates MAINTUBE confirming S's description]

9. Do you see that?

J: 10. Yeah.

S: 11. Okay,

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

[J looks at AIRCHAMBER]

15. Do you see it?

16. About two inches long.

[J picks up STAND and drops MAINTUBE]

17. Both of these are tubular.

[J puts down SPOUT]

J: 18. Okay.

19. not the bent one.

[J puts down SPOUT]

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

Excerpt 3 (Face-to-Face)

S: 1. And now take the little red
2. peg,
[P takes PLUG]
3. Yes,

4. and place it in the hole at the
5. green end,
[P starts to put PLUG into OUTLET2 of MAINTUBE]

6. no

7. the--in the green thing
[P puts PLUG into green part of PLUNGER]

P: 8. Okay.

In Lines 4 and 5, S 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 <u>at the green end</u>," 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, S is able to correct the ambiguity in Lines 6 and 7.

An underspecified description can be imprecise in many possible ways. It 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 A provides a fuller description in Lines 5 to 8 that E is able to select the proper piece.

A description 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, Line 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 down to two contenders. Another use of imprecise feature values occurs when enough feature values are provided but at least one is too imprecise. In Excerpt 6, Line 3, the use of "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.[4] A more precise shape descriptor such as "bell-shaped" or "cylindrical" would have been 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.

[E 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)

A: 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

R: 7. ok

A: 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

R: 14. ok

Excerpt 6 (Teletype)

S: 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, without giving the listener any indication of the switch. 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 the listener.

Excerpt 7 below illustrates an instance of the first type of focus confusion. In the excerpt, the speaker (S) shifts focus without notifying the listener (P) of the switch. As the excerpt begins, P is holding the TUBEBASE. S provides in Lines 1 to 16 instructions for P to attach the CAP and the SPOUT to OUTLET1 and OUTLET2, respectively, on the MAINTUBE. When P successfully completes these attachments, S switches focus in Lines 17 to 20 to the TUBEBASE assembly and requests P to screw it on to the bottom of the MAINTUBE. While P completes the task, S realizes 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. S attempts to correct her mistake by requesting P to remove "the plas"[5] piece in Lines 22 and 23. Since S never indicated a shift in focus from the TUBEBASE back to the SPOUT, P interprets "the plas" to refer to the TUBEBASE.

Excerpt 7 (Face-to-Face)

S: 1. And place
2. the blue cap that's left
[P takes CAP]
3. on the side holes that are
4. on the cylinder,
[P lays down TUBEBASE]
5. the side hole that is farthest
6. from the green end.
[P puts CAP on OUTLET1 of MAINTUBE]
P: 7. Okay.

S: 8. And take the nozzle-looking
9. piece,
[P grabs NOZZLE]

10. no

11 I mean the clear plastic one,
[P takes SPOUT]

12. and place it on the other hole
[P identifies OUTLET2 of MAINTUBE]
13. that's left,

14. so that nozzle points away
15. from the
[P installs SPOUT on OUTLET2 of MAINTUBE]

16. right.

P: 17. Okay.

S: 18. Now

19. take the

20. cap base thing
[P takes TUBEBASE]
21. and screw it onto the bottom,
[P screws TUBEBASE on MAINTUBE]
22. ooops,
[S realizes she has forgotten to have P put SLIDEVALVE into OUTLET2 of MAINTUBE]
23. un-undo the plas
[P starts to take TUBEBASE off MAINTUBE]

24. no

25. the clear plastic thing that I
26. told you to put on
[P removes SPOUT]

27. sorry.

28. And place the little red thing
[P takes SLIDEVALVE]

29. in there first,

[J inserts SLIDEVALVE into OUTLET2 of MAINTUBE]

30. it fits loosely in there.

Excerpt 8 below demonstrates the focus confusion that occurs when the speaker (S) sets up one focus--the MAINTUBE, the correct focus in this case--but then proceeds in such a manner that the listener (J) 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 S and "the hole in the TUBEBASE" for J. J 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 J doesn't sense any problem) or if, later in the exchange, a conflict arises due to the mistake (e.g., a requested action cannot be performed). In Line 31, J inserts a piece into the wrong hole because of the misunderstanding in Line 15. Line 31 hints that J 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 S did not provide any clarification, he assumed he was correct.

Excerpt 8 (Telephone)

S: 1. Um now.
2. Now we're getting a little
3. more difficult.

J: 4. (laughs)

S: 5. Pick out the large air tube

[J picks up STAND]

6. that has the plunger in it.

[J puts down STAND, takes PLUNGER/MAINTUBE assembly]

J: 7. Okay.

S: 8. And set it on its base,

[J puts down MAINTUBE, standing vertically, on the TABLE]

9. which is blue now,
10. right?

[J has shifted focus to the TUBEBASE]

J: 11. Yeah.

S: 12. Base is blue.
13. Okay,
14. Now
15. You've got a bottom hole still
16. to be filled,
17. correct?

J: 18. Yeah.

[J 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]

S: 19. Okay.
20. You have one red piece
21. remaining?

[J 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]

J: 22. Yeah.

S: 23. Okay.
24. Take that red piece.

[J takes SLIDEVALVE]

25. It's got four little feet on
26. it?

J: 27. Yeah.

S: 28. And put the small end into
29. that hole on the air tube--

30. on the big tube.

J: 31. On the very bottom?
[J starts to put it into the bottom hole of TUBEBASE--though he indicates he is unsure of himself]

S: 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. Excerpt 8 above is an example of the first type because S failed to notice that an ambiguity existed since he never explicitly brought the TUBEBASE either into or out of focus. He just assumed that J had the same perspective as he had--a perspective in which there was no ambiguity.

Wrong context. Context differs from focus. The context of a portion of a conversation is concerned with the intention of the discussion and with the set of objects relevant to that discussion, though not attended to currently. Focus pertains to the elements which are currently being attended to in the context. For example, two people can share the same context but have different focus assignments within it--we are both talking about the water pump but you are describing the MAINTUBE and I am describing the AIRCHAMBER. Alternatively, we could just be using different contexts--I think you are talking about taking the pump apart but you are 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 different from one another.[6] The kinds of misunderstandings that can occur because of context inconsistencies are similar to those for focus problems: (1) the speaker might set up or use one context for a discussion and then proceed in another one without letting the listener know of the change, (2) the listener may feel that a change in context has taken place when in fact the speaker never intended one, or (3) the listener may fail to recognize that the speaker has indicated a switch in context. 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 may result.

Bad analogy. An analogy (see Gentner, 1980, for a discussion) 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, an additional way referent confusion can occur is to describe 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. Alternatively, the analogy may fail because it is too difficult to discover a mapping between the analogous object and something in the environment. In Excerpt 9, J at first has trouble correctly satisfying A'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 J was that A's functional analogy was not specific enough. It would have been better to use "cap" instead of "stopper."

Excerpt 9 (Telephone)

A: 1. Okay. Now,
2. take the big blue
3. stopper that's laying around
[J grabs AIRCHAMBER]

4. ... and take the black
5. ring--

J: 6. The big blue stopper?
[J is confused and tries to communicate it to A; he is holding the AIRCHAMBER here]

A: 7. Yeah,

8. the big blue stopper

9. and the black ring.
[J drops AIRCHAMBER and takes the O-RING and the TUBEBASE]

In other cases the analogy might be too specific and would confuse the listener because none of the available referents appear to fit it. In Line 8 of Excerpt 7, "nozzle-looking" is poor 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 think it is the intended referent. He assumes that the expert would just directly describe the object as a member of the class and not bother to form an analogy. 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" in Lines 2 and 3, and "champagne top" in Line 9, are too specific and the listener is unable to 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 M's descriptions, that he exclaimed "What!" when he heard them and was unable to proceed.

Excerpt 10 (Audiotape)

M: 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

[M 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

[M is trying to refer to TUBEBASE]

Description incompatibility. Descriptions incompatible with the scene can also lead to confusion. A description is incompatible when it does not agree with the current state of the world: (1) when one or more of the specified conditions, i.e., the feature values, do not satisfy _any_ of the pieces; (2) when one or more specified constraints do not hold (e.g., saying "the _loose_ one" when all objects are tightly attached); or (3) if no _one_ object satisfies _all_ of the features specified in the description. In Lines 7 and 8 of Excerpt 10 above, M'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. M actually meant "two pegs."

2.3 _Detecting Miscommunication_

Part of our research has been to examine how a listener discovers the need to repair an utterance or description during communication. The incompatibility of a description or action with the scene is one signal of possible trouble. The appearance of a goal incompatibility such as an obstacle or redundancy that blocks one from achieving a goal is another indication of a potential problem.

Description and action incompatibility. As we pointed out earlier, there are three kinds of possible i compatibility with the scene--description, action and goal. 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 (1) 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); (2) 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 (3) 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 that exhibits all the values).

An impossible reference could indicate an earlier action error (e.g., two parts were put together that never should have been). An action incompatibility problem is likely if (1) the listener cannot perform the action specified by the speaker because of some obstacle; (2) the listener performs the action but does not arrive at its intended effect (i.e., a specified or default constraint isn't satisfied); or (3) the current action affects a previous action in an adverse way, yet the speaker has given no sign that this side effect is important. Action incompatibility might indicate an earlier misreference (e.g., you chose the wrong part and used it in an earlier action).

Goal obstacle. A goal obstacle occurs when a goal (or subgoal) one is trying to achieve is blocked. This can result in confusion for the listener because in general listeners do not expect speakers to give them tasks that cannot be achieved. Often, though, it points out for the listener that some miscommunication, such as misreference, has occurred

Goal redundancy. Goal redundancy occurs when the requested goal (or subgoal) is already satisfied. This is a simple kind of goal obstacle where the goal to be fulfilled is blocked because it is already true and nothing has to be done to get around it. However, it can lead to confusion on the part of listeners because they may suspect that they misunderstood what the speaker has requested since they wouldn't expect a reasonable speaker to request them to perform an already completed action. It provides a hint that miscommunication has occurred.

3. Repairing Reference Failures

3.1 Introduction

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 identifications has evolved from work by other artificial intelligence researchers (see Grosz, 1977; Hoeppner, Christaller, Marburger, Morik, Nebel, O'Leary, & Wahlster, 1983). That traditional approach succeeds if a referent is found and fails if no referent is found (see Figure 3(a)). However, a reference identification component must be more versatile than those previously constructed. The excerpts above show that the traditional approach is inadequate because people's real behavior is much more complex. 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 that the turquoise block was the one the speaker meant since turquoise and blue are similar colors.

A key feature to reference identification is "negotiation" which, 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 that we are most concerned with in this research. The listener considers aspects of the speaker's description, the context of the communication, his 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 may overrule parts of the description because he cannot 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 3(b)).

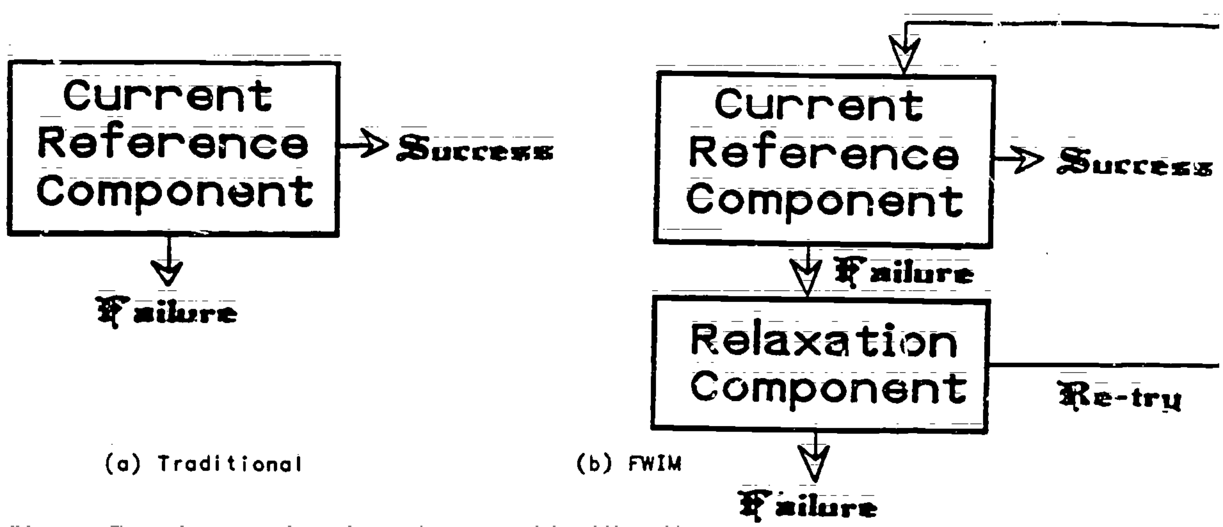


Figure 3: Approaches to reference identification

One means of making sense of a failed description is to delete or replace portions of it that cause it not to match objects in the hearer's world. In our program we are using "relaxation" techniques for this. 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 done blindly but is modelled on a person's behavior. We have developed a computational model that can relax aspects of a description using many of the sources of knowledge used by people. Relaxation then becomes a form of communication repair (in the style of the work on repair theory found in Brown & VanLehn, 1980).

3.2 The Referent Identifier and Relaxation Component

When a description fails to denote a referent in the real world properly, it is possible to repair it by a relaxation process that ignores or modifies parts of it. Since a description can specify many features of an object and relaxing in different orders could yield matches to different objects, the order in which parts of it are relaxed is crucial. There are several kinds of relaxation possible. One can ignore a constituent, replace it with a related value, or change focus (i.e., consider a different group of objects). This section describes the overall relaxation component of the referent identifier and how it draws on knowledge sources about descriptions and the real world as it tries to relax an errorful description and find one for which a referent can be identified.

3.2.1 *Find a referent using a reference mechanism.* Identifying the referent 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 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. 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 ("==>") is used between concepts to show set inclusion. It defines a property 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.

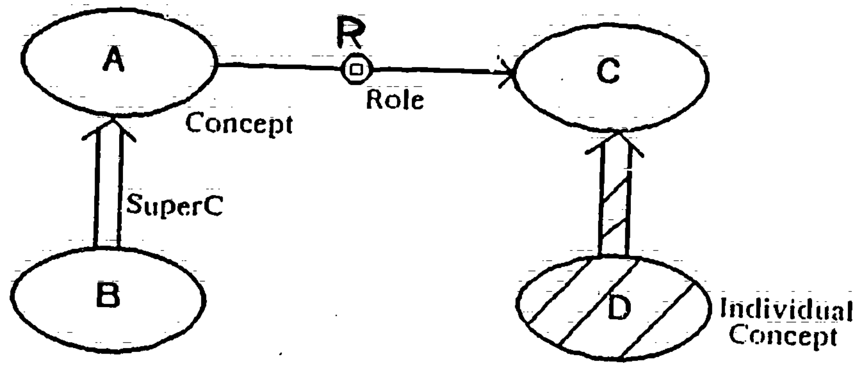


Figure 4: A KL-One Taxonomy

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 and 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 due to pragmatic issues. We will only consider the no reference case in the rest of the paper.

3.2.2 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 outside forces. 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, metonomy and synecdoche are considered to see if they affected the referent search. These rules will not be discussed here; we will assume that the problem lies in the speaker's description.

3.2.3 Perform the relaxation of the description. If relaxation is demanded, then the system must (1) find potential referent candidates, (2) determine which features in the speaker's description to relax and in what order, and use those to order the potential candidates with respect to the preferred ordering of features, and (3) determine the proper relaxation technique 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. 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,[7] and generates a master ordering of features for relaxation. Once the features are in order, the reference system uses that ordering to determine the order in which to try relaxing the candidates.

We draw primarily on sources of linguistic, pragmatic, discourse, domain, perceptual, and hierarchical knowledge, as well as trial and error during this repair process. A detailed treatment of all of them can be found in Goodman (1983-84) and Sidner, Goodman, Haas, Moser, Stallard, and Vilain (1984). 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. Speakers typically add more important information at the end of a description where it is separated from the main part 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.

Each knowledge source produces its own partial ordering of features which are then integrated together. For example, perceptual knowledge may say to relax color. However, if the color value was asserted in a relative clause, linguistic

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 relaxation rule

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, this 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 and to order the referent candidates, C_1, C_2, . . ., C_n, so that relaxation is attempted on the best candidates first, the ones that conform best to a proposed feature ordering. To start, the algorithm examines candidates in pairs and the feature orderings from each knowledge source. For each candidate C_i, the algorithm scores the effect of relaxing the speaker's original description to C_i, 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. It repeats its scoring of C_i for each knowledge source, and sums up its scores to form C_i's total score. The C_i's are then ordered by that score.

Figure 6 provides a graphic illustration of what the best candidate algorithm does. A set of objects in the real world are selected by the partial matcher as potential candidates for the referent These candidates are shown across the top of the figure. The lines on the right side of each box correspond to the set of features that describe that object. The speaker's description is represented in the center 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 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). These are put together to form a directed graph that represents the possible, reasonable ways to relax the features specified in the speaker's description. This graph isn't actually built by the best candidate algorithm, but helps to illustrate here the consideration of all the partial orderings by the algorithm. Finally, the referent candidates are reordered using the information expressed in the speaker's description and in the directed graph of features.

<u>Determine which relaxation methods to apply</u>. Once a set of ordered, potential candidates is selected, the relaxation mechanism begins step 3 of relaxation; it tries to find proper methods to relax the features that have just been ordered (success in finding such methods "justifies" relaxing the

speaker's description to the candidate). It stops at the first candidate in the list of candidates to which methods can be successfully applied.

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, or 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, as can simpler features of an object (such as size or color) that are specified in the speaker's description.

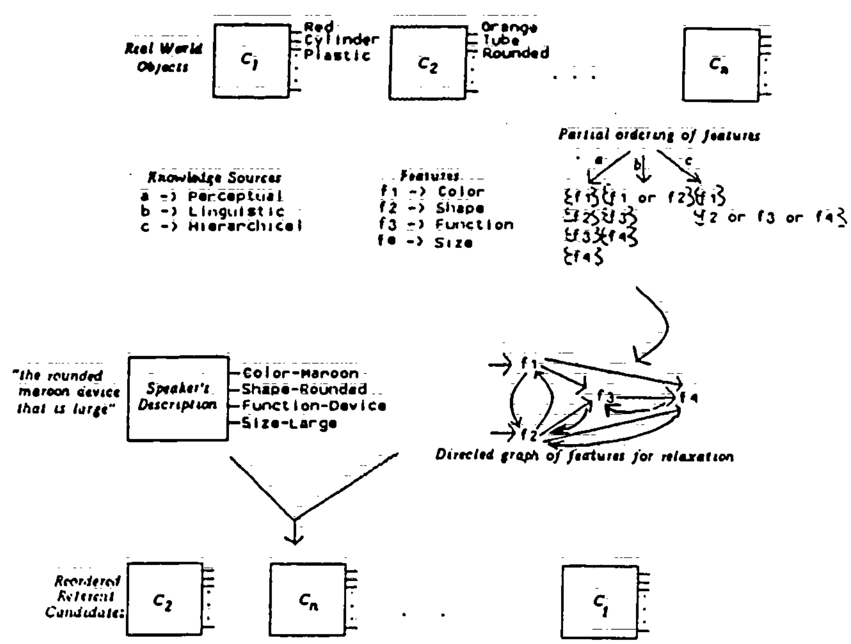


Figure 6: Reordering referent candidates

Relaxation has a few global strategies that people can follow for each part of the description. They can (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 features 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).

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, 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 relaxes its particular type of feature. For example, a Generate-Similar-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 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.[8] The effect of the latter case is really the same as if the feature was simply ignored.

3.3 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.

DescrA:
"...two devices that are clear plastic.

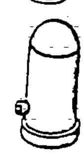
Main Tube

Stand 

DescrB:
One of them has two openings on the outside with threads on the end. and its about five inches long.

Cap 

DescrC:
The other one is a rounded piece with a turquoise base on it.



Air Chamber

DescrD:
Both are tubular.

DescrE:
The rounded piece fits loosely over..."

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**). The relaxation algorithm, shown below, reduces the set of referent candidates for the second one down 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 7 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 that are currently under consideration (i.e., in focus) is presented in Figure 8. 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 functional description of the object. The first entry in each representation in Figure 8 (shown in uppercase) defines the basic kind of entity being described. The words in mixed case refer to the names of features and the words in uppercase 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 which can be referred to either on its own or as part of the parent object. The "Orientation" feature, used in the representations in Figure 8, provides a rotation and translation of the object from some standard orientation, which provides a way to define relative positions such as "top," "bottom," or "side," to the object's current orientation in 3-D space. Figure 9 shows the KL-One taxonomy representing the same objects.

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, i.e., the listener can begin the search process before he hears the complete description, as was observed in the videotape excerpts. We try to simulate this incremental nature in our algorithm, as is apparent from the placement of the first description in **DescrD** into the KL-One taxonomy shown in Figure 9. **DescrD** is incrementally defined by first adding **DescrA**--as shown in Figure 10--and then **DescrB**--as shown in Figure 12--to the taxonomy. The KL-One Classifier compares the features specified in the speaker's descriptions with the features 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 11 where MAINTUBE is shown as a subconcept of **DescrA**.

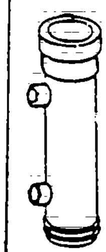
```
DescrA: (DEVICE (Transparency CLEAR)
                (Composition PLASTIC))
DescrB: (DEVICE (Transparency CLEAR)
                (Composition PLASTIC)
                (Subpart (OPENING))
                (Subpart (OPENING))
                (Subpart
                  (THREADS (Rel-Position END)))
                (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 7: The speaker's descriptions

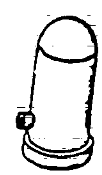
STAND also is shown as a subconcept of **DescrA**. AIR CHAMBER is shown as a <u>possible</u> subconcept (with the dotted arrow) because **DescrA** mismatches with it on one of its subparts.[9] Other features require in-depth processing--that is outside the capability of the KL-One classifier--before 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 8. An OPENING, as seen in Figure 13,

```
      (CAP  (Color BLUE)
            (Composition PLASTIC)
CAP         (Transparency OPAQUE)
            (Dimensions (Length .25) (Diameter .5))
            (Orientation (Rotation (0 0 0 0 90 0))
                         (Translation (0.0 0.0 0.0))))
```

```
      (TUBE (Color VIOLET)
            (Composition PLASTIC)
            (Transparency CLEAR)
            (Dimensions (Length 4.125))
            (Subpart (CYLINDER (Dimensions (Length .25) (Diameter 1.125))
                               (Orientation (Rotation (0.0 0.0 0.0))
             Lip                            (Translation (0.0 0.0 3.75)))
                               (Function OUTLET-ATTACHMENT-POINT)))
            (Subpart (CYLINDER (Dimensions (Length 3.3) (Diameter 1.0))
MAIN         TubeBody          (Orientation (Rotation (0.0.0.0 0.0))
TUBE                                        (Translation (0 0 0.0 .25)))))
            (Subpart (CYLINDER (Dimensions (Length .25) (Diameter 1.125))
                               (Orientation (Rotation (0.0.0.0.0.0))
             Threads                        (Translation (0.0.0.0 0.0)))
                               (Function THREADED-ATTACHMENT-POINT)))
            (Subpart (CYLINDER (Dimensions (Length .375) (Diameter .5))
                               (Orientation (Rotation (0.0.0.0 90.0))
             Outlet1                        (Translation (0.0 .5 3.00)))
                               (Function OUTLET-ATTACHMENT-POINT)))
            (Subpart (CYLINDER (Dimensions (Length .375) (Diameter .5))
                               (Orientation (Rotation (0.0 0.0 90.0))
             Outlet2                        (Translation (0.0 .5 .625))
                               (Function OUTLET-ATTACHMENT-POINT))))
```



```
         (CONTAINER (Dimensions (LENGTH 2.75))
                    (Composition PLASTIC)
                    (Subpart (HEMISPHERE (Color VIOLET)
                                         (Transparency CLEAR)
                     Chamber             (Dimensions (Diameter 1.0))
                     Top                 (Orientation (Rotation (0.0 0 0 0.0))
                                                      (Translation (0.0 0.0 2.25)))))
                    (Subpart (CYLINDER (Color VIOLET)
                                       (Transparency CLEAR)
                     Chamber           (Dimensions (Length 1.0) (Diameter 2.25))
                     Body              (Orientation (Rotation (0.0.0.0.0.0))
                                                    (Translation (0.0 0.0 .375)))))
                    (Subpart (CYLINDER (Color BLUE)
                              (TransParency OPAQUE)
                              (Dimensions (Length .375) (Diameter 1.25))
AIR                           (Orientation (Rotation (0.0.0.0 0.0))
CHAMBER              Chamber                 (Translation (0.0 0.0 0.0)))
                     Bottom   (Function CAP OUTLET-ATTACHMENT-POINT)
                              (Subpart (CYLINDER (Color BLUE)
                                                 (Dimensions (Length .375)
                                                             (Diameter .5))
                                                 (Orientation
                                                   (Rotation (0.0 0.0 0.0))
                                                   (Translation (0.0 0.0 0.0)))
                                                 (Function
                                                    OUTLET-ATTACHMENT-POINT)))))
                    (Subpart (CYLINDER (Color VIOLET)
                                       (Transparency CLEAR)
                     Chamber           (Dimensions (Length .5) (Diameter .375))
                     Outlet            (Orientation (Rotation (0.0 0.0 90.0))
                                                    (Translation (.625 .625 .625)))
                                       (Function OUTLET-ATTACHMENT-POINT))))
```



```
      (TUBE (Dimensions (Length 2.75))
            (Composition PLASTIC)
            (Subpart (CYLINDER (Color BLUE)
                               (Transparency CLEAR)
             Top               (Dimensions (Length 2.25) (Diameter .375))
                               (Orientation (Rotation (0 0 0 0 0.0))
STAND                                       (Translation (.5 0 0 .375)))
                               (Function OUTLET-ATTACHMENT-POINT)))
            (Subpart (CYLINDER (Color BLUE)
                               (Transparency CLEAR)
             Base              (Dimensions (Length .375) (Diameter 1.0))
                               (Orientation (Rotation (0.0 0.0 0.0))
                                            (Translation (0.0 0 0 0.0)))
                               (Function OUTLET-ATTACHMENT-POINT))))
```

Figure 8: The objects in focus

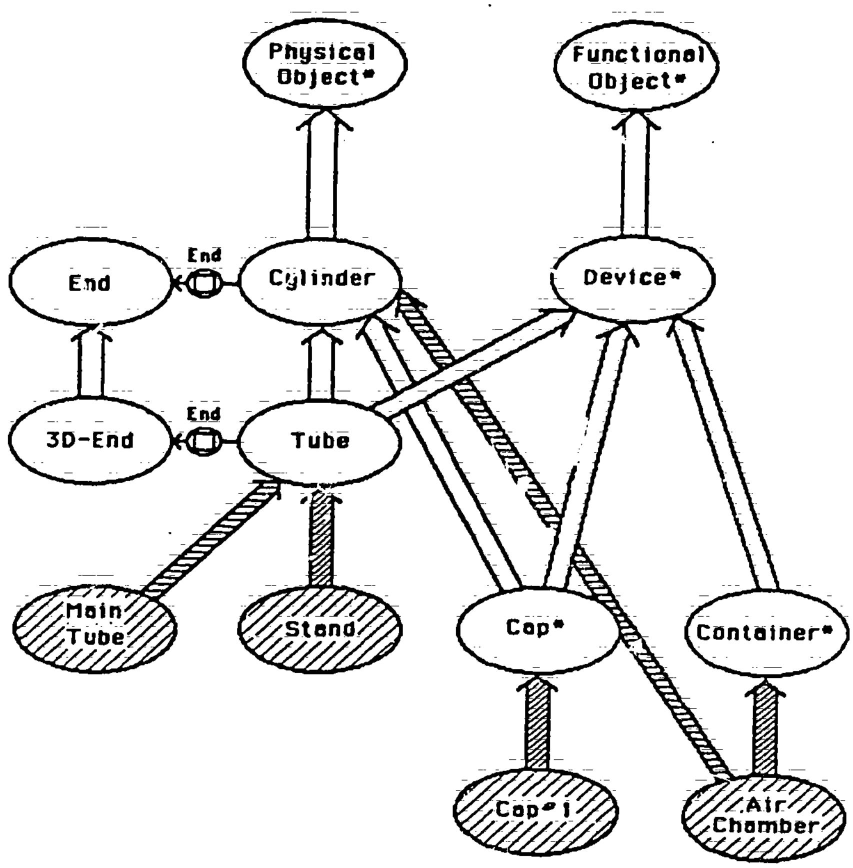


Figure 9: Taxonomy representing the objects in focus

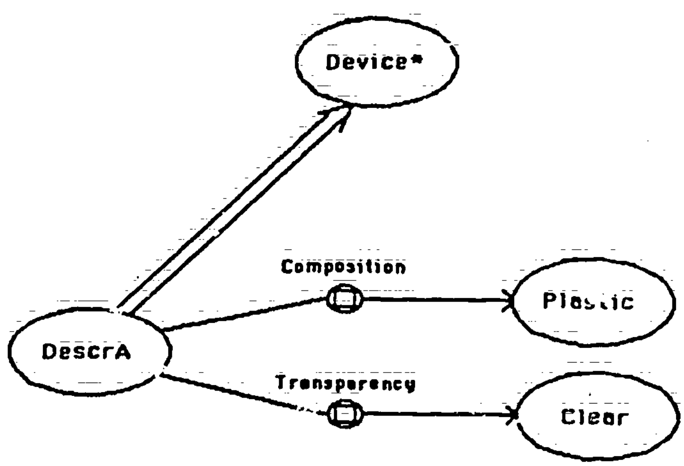


Figure 10: Adding DescrA to the taxonomy

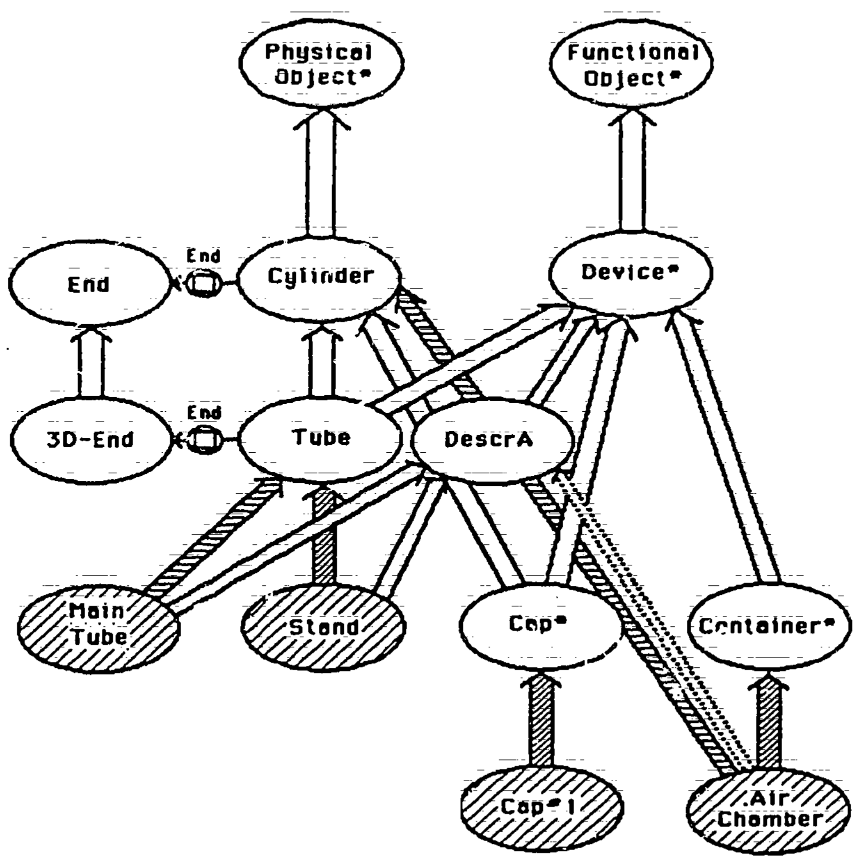


Figure 11: The classified DescrA

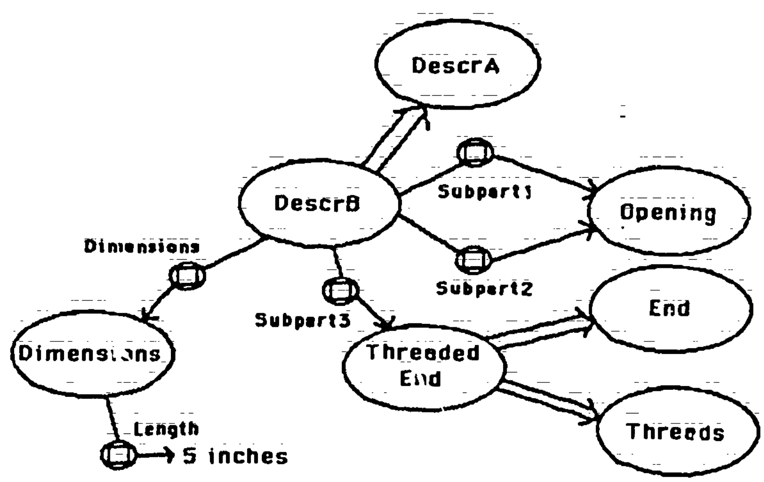


Figure 12: Adding DescrB to the taxonomy

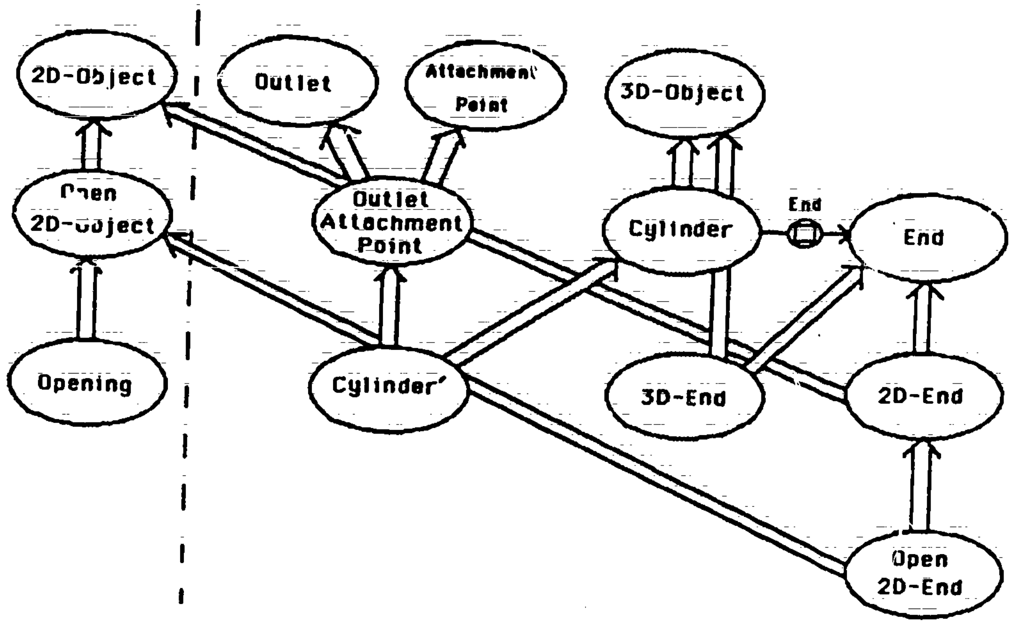


Figure 13: Attempt to match OPENING to CYLINDER'

is thought of primarily as a 2-D cross-section (such as a "hole"), while the two CYLINDER subparts of MAINTUBE 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, 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 13. 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 <u>End</u> which is either a <u>2D-End</u> (e.g., a lid or hole) or a <u>3D-End</u> (e.g., a lip).
>
> A <u>2D-End</u> 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., <u>Outlet-Attachment-Points</u>).

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 MAINTUBE[10] or AIR CHAMBER since each has subparts that are either VIOLET or BLUE. The Classifier places **DescrC** as best it can in the taxonomy, showing no connection between it and any of the objects currently in focus. **DescrD** provides no further help and is similarly placed. This is shown in Figure 14. 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 3.2.3, by examining the elements closest to **DescrD** in the taxonomy and using the partial matcher to score how close each element is to **DescrD**.[11] This is illustrated in Figure 15. The matcher determines MAINTUBE, STAND, and AIR CHAMBER 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: A 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 of BLUE of STAND. (-)

Scoring DescrD to AIR CHAMBER:

a CONTAINER is a kind of DEVICE; (>)

the Transparency of **DescrD**, CLEAR, matches the Transparency of ChamberTop. ChamberOutlet and ChamberBody of AIR CHAMBER, but mismatches the Transparency of ChamberBottom of AIR CHAMBER. Therefore, the partial match is uncertain; (?)

the Composition of each is PLASTIC; (+)

the subparts of AIR CHAMBER have Shape HEMISPHERICAL and CYLINDRICAL which are each a kind of Shape ROUND; (>)

Figure 14: Adding DescrC and DescrD to the taxonomy

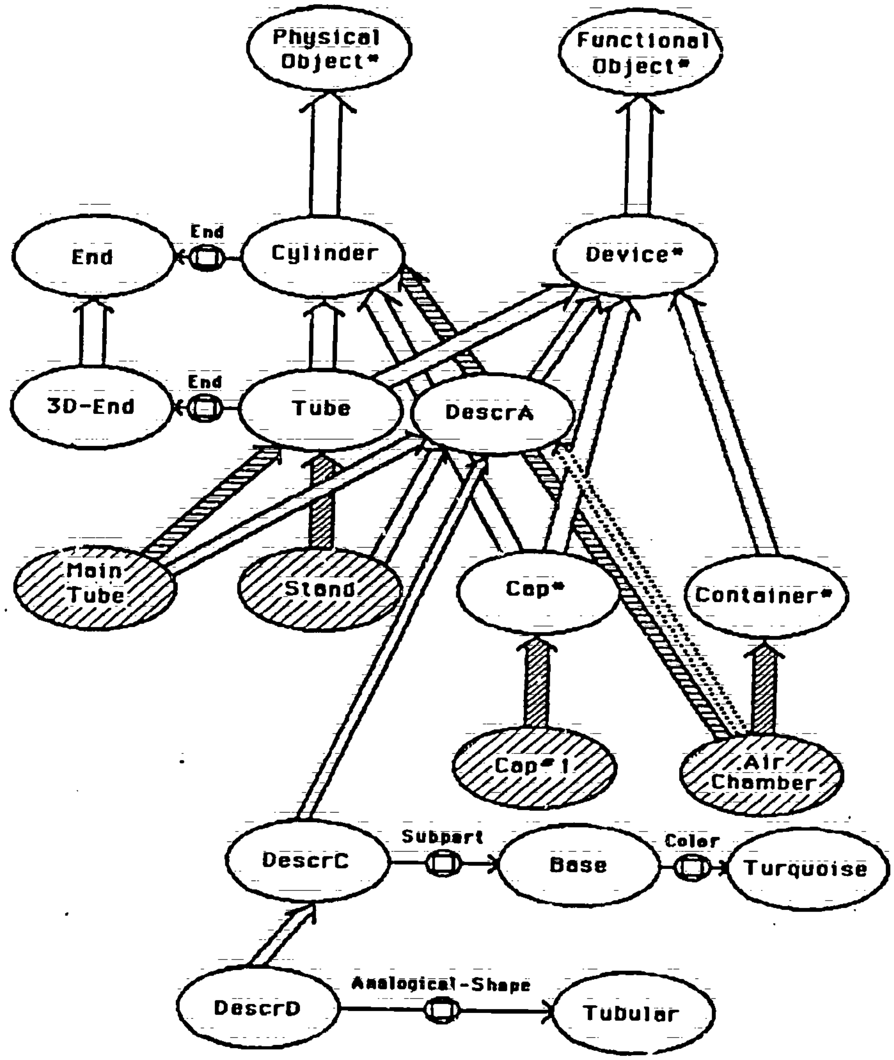
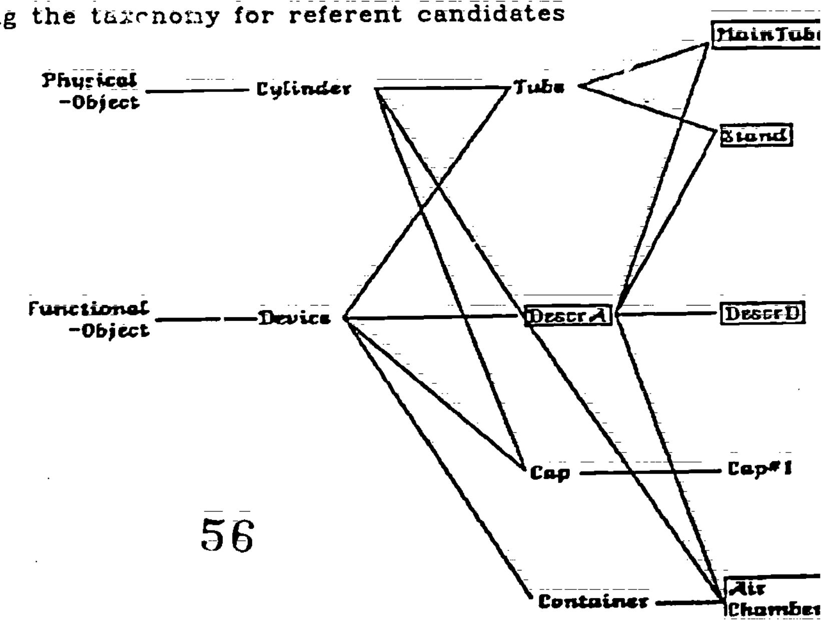


Figure 15: Exploring the taxonomy for referent candidates



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

Figure 16 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). All candidates, hence, will be retained for now.

At this point, the knowledge sources and their associated rules that were mentioned earlier apply. These rules attempt to order the feature values in the speaker's description for relaxation. First, we 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 above) 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:

DescrD

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

Range of role scores:

Low Correlation − ? = < > + High Correlation

Figure 16: Scoring DescrD to the referent candidates

{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 in which the features inside the braces, " {... }," are relaxed is not specified (i.e., any order of relaxation is alright). 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"). Figure 17 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 18). The clustered color rule defined in Figure 19 would suggest such a relaxation. It requires that there are at least three objects in the world that have similar colors. It is meant as an exemplar for a whole series of rules (e.g., Clustered Shape Values, Clustered Transparency Values, and so on). Hierarchical information about how closely relatec 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 20. 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. All of these suggestions can be put together to form the order:

```
(Shape, Color)  <  (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 17: Similar color values

The referent candidates MAINTUBE, STAND, and AIR CHAMBER 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 cannot be ordered with respect to each other. Hence, no one candidate stands out as the most likely referent.

Colors of Candidates & DescrD

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 18: 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 19: The clustered color value rule

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.

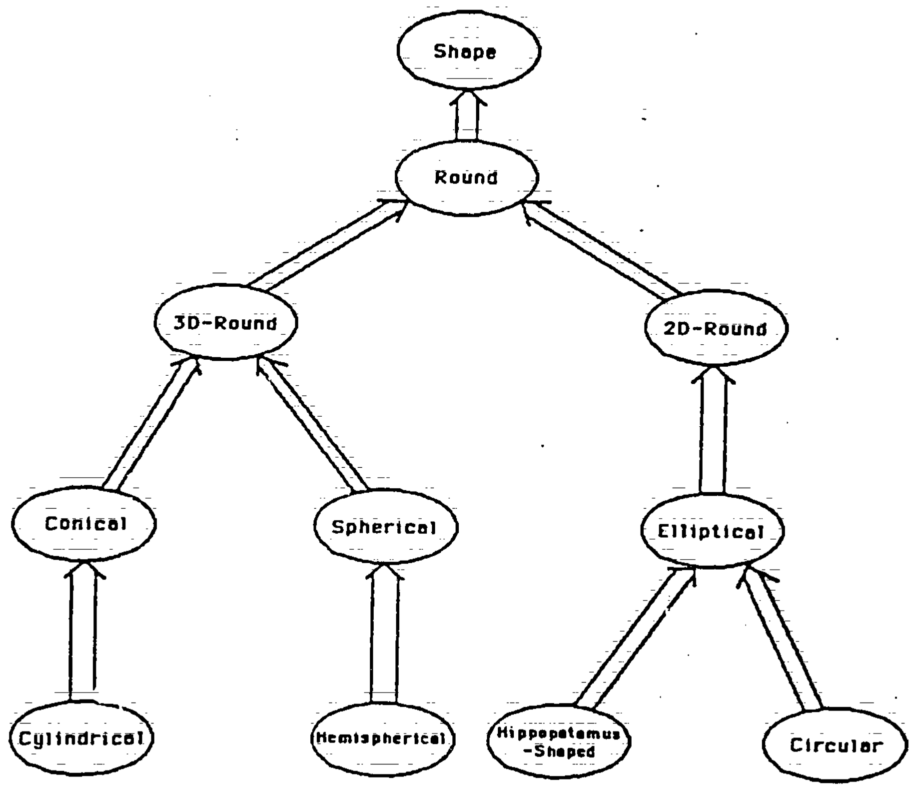


Figure 20: Hierarchical shape knowledge

Generate-Similar-Shape-Values determines that it is reasonable to match ROUND to either the CYLINDRICAL or HEMISPHERICAL shapes of the AIR CHAMBER by examining the taxonomy shown in Figure 20 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 AIR CHAMBER 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 AIR CHAMBER 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 down as best they can and then taking each of the remaining choices and trying out the requested action on them.

4 Conclusion

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 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 down the search space.

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 metonomy--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. 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.

References

Allen, J. F. (1979). A plan-based approach to speech act recognition. Ph.D. Thesis, University of Toronto.

Allen, J. F., Frisch, A. M., & Litman, D. J. (1982). ARGOT: The Rochester dialogue system. In Proceedings of AAI-82 (pp. 66-70). Pittsburgh, PA.

Appelt, D. E. (1981). Planning natural language utterances to satisfy multiple goals. Ph.D. Thesis, Stanford University.

Brachman, R. J. (1977). A structural paradigm for representing knowledge. Ph.D. Thesis, Harvard University. (Also, T. Rep. No. 3605, Bolt Beranek and Newman Inc.)

Brown, J. S. & Vanlehn, K. (1980). Repair theory: A general theory of bugs in procedural skills. Cognitive Science, 4(4), 379-426.

Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In Joshi, Webber, & Sags (Eds.), Elements of discourse understanding (pp. 10-64). Cambridge, MA: Cambridge University Press.

Cohen, P. R. (1978). On knowing what to say: Planning speech acts. Ph.D. Thesis, University of Toronto.

Cohen, P. R. (1981). The need for referent identification as a planned action. In Proceedings of IJCAI-81 (pp. 28-35). Vancouver, B.C., Canada.

Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. Computational Linguistics 10(2), 97-146.

Cohen, P. R., Fertig, S., & Starr, K. (1982). Dependencies of discourse structure on the modality of communication: Telephone vs. teletype. In *Proceedings of ACL* (pp. 28-35). Toronto, Ontario, Canada.

Cohen, P., Perrault, C., & Allen, J. (1981). Beyond question answering. In W. Lehnart & M. Ringle (Eds.), *Knowledge representation and natural language process*. Hillsdale, NJ: Erlbaum.

Gentner, D. (1980). *The structure of analogical models in science*. Cambridge, MA: Bolt Beranek and Newman Inc.

Goodman, B. A. (1982). *Miscommunication in task-oriented dialogues*. Cambridge, MA: KRNL Group Working Paper, Bolt Beranek and Newman Inc.

Goodman, B. A. (19[illegible]). Repairing miscommunication: Relaxation in reference. In *Proceedings of AAAI-83* (pp. 134-138). Washington, D.C.

Goodman, B. A. (1984). *Communication and miscommunication*. Ph.D. Thesis, University of Illinois.

Grosz, B. J. (1977). *The representation and use of focus in dialogue understanding*. Ph.D. Thesis, University of California, Berkeley. (Also, Tech. Note 151, Stanford Research Institute.)

Grosz, B. J. (1978). *Focusing in dialogue: Theoretical issues in natural language processing-2* (pp. 96-103). Urbana, IL.

Grosz, B. J. (1981). Focusing and descriptions in natural language dialogues. In Joshi, Webber, & Sags (Eds.), Elements of discourse understanding (pp. 84-105). Cambridge MA: Cambridge University Press.

Hoeppner, W., Christaller, T., Marburger, H., Morik, K., Nebel, B., O'Leary, M., & Wahlster, W. (1983). Beyond domain-independence: Experience with the development of German language access system to highly diverse background systems. In Proceedings of IJCAI-83 (pp. 588-594). Karlsruhe, West Germany.

Joshi, A. K. (1982). Mutual beliefs in question-answer systems. In N. Smith (Ed.), Mutual beliefs (pp. 181-197). New York: Academic Press.

Lipkis, T. (1982). A KL-ONE Classifier. In Proceedings of the 1981 KL-One Workshop (pp. 128-145). (Also, Report No. 4842, Bolt Beranek and Newman Inc. Also Consul Note #5, USC/Information Sciences Institute.)

Litman, D. (1983). Discourse and problem solving (Report No. 5338). Cambridge, MA: Bolt Beranek and Newman Inc. (Also, TR 130, University of Rochester, Department of Computer Science.)

Litman, D. J., & Allen, J. F. (1984). A plan recognition model for clarification subdialogues. In Proceedings of Coling84 (pp. 302-311). Stanford, CA: Stanford University.

Mark, W. (1982). Realization. In Proceedings of the 1981 KL-One Workshop (pp. 78-89). Cambridge, MA: Report No. 4842, Bolt Beranek and Newman Inc.

McCoy, K. F. (1985). The role of perspective in responding to property misconceptions. In Proceedings of IJCAI-85 (pp. 791-793). Los Angeles, CA.

McKeown, K. R. (1983). Recursion in text and its use in language generation. In Proceedings of AAAI-83 (pp. 270-273). Washington, D.C.

Nadathur, G., & Joshi, A. K. (1983). Mutual beliefs in conversational systems: Their role in referring expressions (pp. 603-605). Karlsruhe, West Germany.

Perrault, C., Cohen, R., & Cohen, P. R. (1981). It's for your own good: A note on inaccurate reference. In Joshi, Webber, & Sags (Eds), Elements of discourse understanding (pp. 217-230). Cambridge, MA: Cambridge University Press.

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

Reichman, R. (1981). Plain speaking: A theory and grammar of spontaneous discourse. Ph.D. Thesis, Harvard University. (Also, Tech. Rep. No. 4861, Bolt Beranek and Newman Inc.)

Ringle, M. & Bruce, B. (1981). Conversation failure. In W. Lehnart & M. Ringle (Eds.), Knowledge representation and natural language processing. Hillsdale, NJ: Erlbaum.

Searle, J. R. (1969). Speech acts. Cambridge, MA: Cambridge University Press.

Sidner, C. L. (1979). Towards a computational theory of definite anaphora comprehension in English discourse. Ph.D. Thesis, Massachusetts Institute of Technology. (Also, Rep. No. TR-537, MIT AI Lab.)

Sidner, C. L. (1985). Plan parsing for intended response recognition in discourse. Computational Intelligence 1, 1-10.

Sidner, C. L., Bates, M., Bobrow, R. J., Brachman, R. J., Cohen, P. R., Israel, D. J., Schmolze, J., Webber, B. L., & Woods, W. A. (1981). Research in knowledge representation for natural language understanding (Report No. 4785). Cambridge, MA: Bolt Beranek and Newman Inc.

Sidner, C. L., Bates, M., Bobrow, R., Goodman, B., Haas, A., Ingria, R., Israel, D., McAllester, D., Moser, M., Schmolze, J., & Vilain, M. (1933). Research in knowledge representation for natural language understanding--Annual Report, 1 September 1982--31 August 1983 (Tech. Rep. 5421). Cambridge, MA: BBN Laboratories.

Sidner, C., Goodman, B., Haas, A., Moser, M., Stallard, D., & Vilain, M. (1984). Research in knowledge representation for natural language understanding--Annual Report, 1 September 1983--31 August 1984 (Tech. Rep. 5694). Cambridge, MA: BBN Laboratories Inc., Cambridge, MA.

Sidner, C. L, & Israel, D. J. (1981). Recognizing intended meaning and speaker's plans. In Proceedings of the International Joint Conference in Artificial Intelligence, The International Joint Conferences on Artificial Intelligence (pp. 203-208). Vancouver, B.C.

Tversky, A. (1977). Features of similarity. _Psychological Review_, _84_, 327-352.

Webber, B. L. (1978). _A formal approach to discourse anaphora_. Ph.D. Thesis, Harvard University. (Also, Tech. Rep. No. 3761, Bolt Beranek and Newman Inc.)

Footnotes

[1]This research was supported in part by the Defense Advanced Research Project Agency under contract N00014-77-C-0378.

[2]An analysis of clarification subdialogues can be found in (Litman & Allen, 1984).

[3]Of course, there are some situations--such as teaching--where the hearer would be more willing to tolerate overspecific descriptions.

[4]"Chamber" was interpreted here in a broader sense by the listener because it was used right at the beginning of the dialogue before the speaker introduced other terms such as "tube" that would have better helped to distinguish the pieces. The example demonstrates how discourse affects reference.

[5]The whole word here is "plastic." In these protocols, people often guess before hearing the whole utterance or even whole words.

[6]Grosz (1977, 1981) would describe this as a difference in "task plans" while Reichman (1978, 1981) would say that the "communicative goals" differed.

[7]Of course, once one particular candidate is selected, then deciding which features to relax is relatively trivial--one simply compares features of the candidate description (the target) to the speaker's description (the pattern) and notes any discrepancies.

[8]The latter case is there primarily for the times when one can't easily define a similarity metric for a feature. McCoy, (1985) and Tversky (1977) provide additional discussions about similarity metrics.

[9]We are stretching the definition of KL-One here with the dotted subsumption arrow. The point we want to make is that the AIRCHAMBER is <u>similar</u> to **DescrA** because their descriptions are almost exactly the same.

[10]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.

[11]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.