REPAIRING REFERENCE IDENTIFICATION FAILURES BY RELAXATION

Bradley A. Goodman
BBN Laboratories
10 Moulton Street
Cambridge, Mass. 02238

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 assured to have the same beliefs, contexts, 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 constitute a kind of "ill-formed" input that 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, including techniques for avoiding failures of reference.

1 Introduction

Cohen, Perrault and Allen showed in their paper "Beyond Question Answering" [6] 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 problems in communication. This paper summarizes the results of a dissertation [13] 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 handle such problems in the context of a natural language understanding system. The study of miscommunication is a necessary task within such a context since any computer capable of communicating with humans in natural language must be tolerant of the imprecise, ill-devised or complex utterances that people often use.

Our current research [25, 26] 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 [5, 1, 23]. This plan, however, is rarely, if ever, obvious at the surface sentence level. A central issue in the interpretation of utterances is the transformation of sequences of imprecise, ill-devised or complex 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 [2, 19]. 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 description constructed by the speaker to refer to an actual object in the world. The description can be imprecise, confused, ambiguous or overly specific, it might be interpreted under the wrong context. This leads to difficulty for the listener when figuring out what object is being described, that is, reference identification errors. Such descriptions are "ill-formed" input. The blame for ill-formedness may lie partly with the speaker and partly with the listener. The speaker may have been sloppy or 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 feel that he has understood when he in fact has not.

This work is part of an on-going effort to develop a reference identification and plan recognition mechanism that can exhibit more "human-like" tolerance of such utterances. Our goal is to build a more robust system that can handle errorful utterances, and that can be incorporated in existing systems. As a start, we have concentrated on reference identification. In conversation people use imperfect descriptions to communicate about objects, sometimes their partners succeed in understanding and occasionally they fail. Any computer hoping to play the part of a listener must be capable of taking what the

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

speaker says and either deleting, adapting or clarifying it. We are developing a theory of the use of extensional descriptions that will help explain how people successfully use such imperfect descriptions. We call this the theory of reference miscommunication.

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 motivating many of them with illustrative protocols. Section 3 presents possible ways around some of the problems of miscommunication in reference. Motivated there is a partial implementation of a reference mechanism that attempts to overcome many reference problems.

We are following the task-oriented paradigm of Grosz [14] 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 (see [7, 8, 9]). 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.

The following exchange exemplifies one such situation. Here A is instructing 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 when listeners repair the speaker's description in order to find a referent, when they repair their initial reference choice once they are given more information, and when they fail to choose a proper referent. In Line 7, A describes the two holes on the *BASEVALVE* as "the little hole." J must repair the description, realizing that A doesn't really mean "one" hole but is referring to the "two" holes. J apparently does this since he doesn't complain about A's description and correctly attaches the *BASEVALVE* to the *TUBEBASE*. Figure 1(b) shows the configuration of 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 did not give 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]

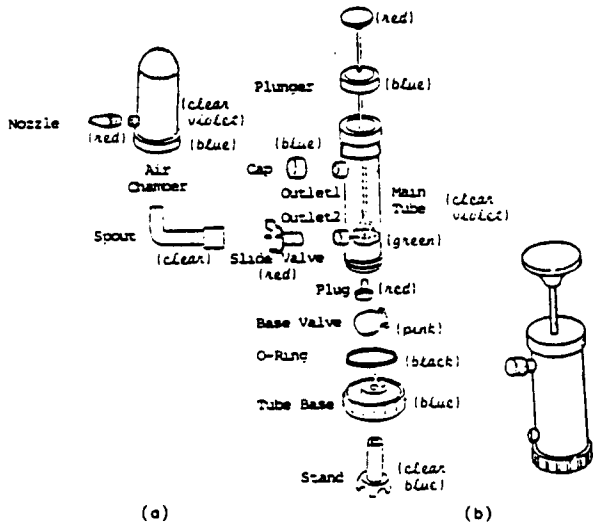

Figure 1: The Toy Water Pump

2 Miscommunication

People must and do manage to resolve lots of (potential) miscommunication in everyday conversation. Much of it is resolved subconsciously - with the listener unaware 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 the listener can become confused about and that can lead to miscommunication. The listener can become confused about what the speaker intends for the referents, 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, or 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 the confusion can quite possibly be resolved.

2.1 Causes of miscommunication

This section attempts to motivate a paradigm for the kinds of conversation that we studied and tries to point out places in the paradigm that leave room for miscommunication.

2.1.1 Effects of the structure of task-oriented dialogues

Task-oriented conversations have a specific goal to be achieved: the performance of a task (e.g., [14]). The participants in the dialogue can have the same skill level and they can simply work together to accomplish the task; or one of them, the expert, could know more and could direct the other, the apprentice, to perform the task. We have concentrated primarily on the latter case - due to the protocols that we examined - but many of our observations can be generalized to the former case, too. We will refer to this as the apprentice-expert domain.

The viewpoints of the expert and apprentice differ greatly in apprentice-expert exchanges. The expert, having an understanding of the functionality of the elements in the task, has more of a feel for how the elements work together, how they go together, and how the individual elements can be used. The apprentice normally has no such knowledge and must base his decisions on perceptual features such as shape [15].

The structure of the task affects the structure of the dialogue [14], particularly through the center of attention of the expert and apprentice. This is the phenomenon called focus [14, 20, 24], which, in task-oriented dialogues is a very real and operational thing (e.g., focus is used in resolving anaphoric references). Shifts in focus correspond directly to the task, its subtasks, the objects in a task and the subpieces of each object. Focus and focus shifts are governed by many rules [14, 20, 24]. Confusion may result when expected shifts do not take place. For example, if the expert changes focus to an object but never discusses its subpieces (such as an obvious attachment surface) or never bothers to talk about the object reasonably soon after its introduction (i.e., between the time of its introduction and its use, without digressing in a well-structured way in between (see [20])), 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 in the apprentice's world will also perplex the apprentice.

Focus also influences how descriptions are formed [15, 2]. The level of detail required in a description depends directly on the elements currently highlighted by the focus. If the object to be described is similar to other elements in focus, the expert must be more specific in the formulation of the description or may consider shifting focus away from the possibly ambiguous objects to one where the ambiguity won't occur.

2.2 Consequences of miscommunication

In this section we will make it clear that people do miscommunicate and yet they often manage to fix things. We will look at specific forms of miscommunication and describe ways to detect them. We will highlight relationships between different miscommunication problems but won't necessarily demonstrate ways to resolve each of them.

[2] An analysis of clarification subdialogues can be found in [17].

2.2.1 Instances of miscommunication

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. This section defines and illustrates many of them through numerous excerpts. Each excerpt is marked in parentheses to show what modality of communication was used (see [9] for a description about the collection of these excerpts). Each bracketed portion of the excerpt explains what was occurring at that point in the dialogue. The confusions themselves, coupled with the description at the end of this section on how to recognize when one of them is occurring, provides motivation for the use of the algorithm outlined in Section 3 as a means for repairing communication problems. We will only discuss referent confusion in this paper. The other forms of confusion – Action, Goal, and Cognitive Load – are described in [11, 13]. Another categorization of confusions that lead to conversation failure can be found in [22].

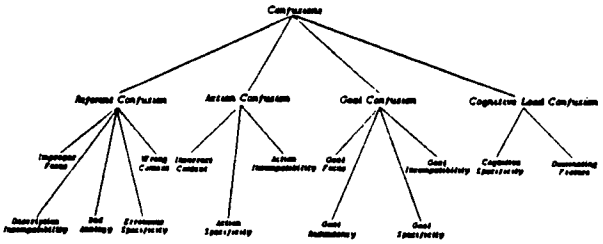

Figure 2: A taxonomy of confusions

Referent confusion occurs when the listener is unable to correctly determine what the speaker is referring to with a particular description. It occurs 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 in the utterance are either incorrect or incompatible with the current or global context.

Erroneous Specificity

Ambiguous (and, thus, imprecise) descriptions can cause confusion about the referent. Excerpt 2 below illustrates a case where the speaker's description is underspecified – it does not provide enough detail to prune the set of possible referents down to one.

Excerpt 2 (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 Line 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 at the green end", it defines a region where the peg might go instead of a specific location. In this particular case, there are three possible holes to choose from that are near the green end. The listener chooses one – the wrong one – and inserts the peg into it. Because this dialogue took place face to face, S is able to correct the ambiguity in Lines 6 and 7.

A speaker's description can be imprecise in several possible ways. (1) It may contain features that do not readily apply in the domain. In Line 3, Excerpt 3, the feature "funny" has no relevance to the listener. It is not until A provides a fuller description in Lines 5 to 8 that E is able to select the proper piece. (2) It may use a vague head noun coupled with few or no feature values (and context alone does not necessarily suffice to distinguish the object). In Excerpt 4, Line 9, "attachment" is vague 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. (3) Enough feature values are provided but at least one value is too vague leading to trouble. In Excerpt 5, Line 3, the use of the attribute value "rounded" to describe the shape does not sufficiently reduce the set of four possible referents (though, in this particular instance, A correctly identifies it) because the term is applicable to numerous parts in the domain. A more precise shape descriptor such as "bell-shaped" or "cylindrical" would have been more beneficial to the listener.

Excerpt 3 (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 4 (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 5 (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

Focus confusion can occur when the speaker sets up one focus and then proceeds with another one without letting the listener know of the switch (i.e., a focus shift occurs without any indication). An 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 6 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 outlets *OUTLET1* and *OUTLET2*, respectively, on the *MAINTUBE*. Upon P's successful completion of 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"[3] 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 6 (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,
[P inserts SLIDEVALVE into OUTLET2 of MAINTUBE]
30. it fits loosely in there.

Excerpt 7 below demonstrates the latter type of focus confusion that occurs when the speaker (S) sets up one focus - the *MAINTUBE*, which is 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 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

[3]The whole word here is "plastic." People in general tend to be good at proceeding before hearing the whole utterance or even the whole word.

due to the mistake (e.g., a requested action can not 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 but since the task was 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 7 (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 indicated by the speaker. Excerpt 7 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 him – a perspective in which no ambiguity occurred.

Wrong Context

Context differs from focus. The context of a portion of a conversation is concerned with the point of the discussion in that fragment and with the set of objects relevant to that discussion, though not attended to currently. Focus pertains to the elements 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're both talking about the water pump but you're describing the *MAINTUBE* and I'm describing the *AIRCHAMBER*. Alternatively, we could just be using different contexts – I think you're talking about taking the pump apart but you're talking about replacing the pump with new parts – in both cases we may be sharing the same focus – the pump – but our contexts are totally off from one another.[4] The kinds of misunderstandings that can occur because of context problems are similar to those for focus problems: (1) the speaker might set up or be in one context for a discussion and then proceed in another one without effectively letting the listener know of the change, (2) the listener may feel a change in context has taken place when in fact the speaker never intended one, or (3) the listener fails to recognize an indicated context switch by the speaker. Context affects reference because it helps define the set of available objects that are possible contenders for the referent of the speaker's descriptions. If the contexts of the speaker and listener differ, then misreference might result.

Bad Analogy

An analogy (see [10] for a discussion on analogies) is a useful way to help describe an object by attempting to be more precise by using shared past experience and knowledge – especially shape and functional information. If that past experience or knowledge doesn't contain the information the speaker assumes it does or isn't there, then trouble occurs. Thus, one more way referent confusion can occur is by describing an object using a poor analogy. An analogy used to describe an object might not be specific

[4] Grosz [14, 15] would describe this as a difference in "task plans" while Reichman [20, 21] would say that the "communicative goals" differed.

enough – confusing the listener because several pieces might conform to the analogy or, in fact, none at all appear to fit because discovering a mapping between the analogous object and some piece in the environment is too difficult. In Excerpt 8, 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".

Excerpt 8 (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 it might be too specific – confusing the listener because none of the available referents appear to fit it. In Line 8 of Excerpt 6, "nozzle-looking" forms a poor shape analogy because the object being referred to actually is an elbow-shaped spout. The "nozzle-looking" part of the description convinced the listener that what he was looking for was something specific like a nozzle (which is a small spout). Sometimes, when an object is a clear representative of a specified analogy class, the apprentice may become confused, wondering why the expert bothered to form an analogy instead of just directly describing the object as a member of the class. Hence, it would not be surprising if the apprentice ignored the best representative of the class for some less obvious exemplar. Thus, for example, it is better to say "nozzle" instead of "nozzle-looking." In Excerpt 9, the description "hippopotamus face shape" (a shape analogy) in Lines 2 and 3, and "champagne top" (a shape analogy) in Line 9, are too specific and the listener is unable to easily find something close enough to match either of them. He can't discover a mapping between the object in the analogy and one in the real world.

Excerpt 9 (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

Incompatible descriptions can lead to confusion also. A description is incompatible when (1) 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 9 above, M's use of "the two peg holes" leads to bewilderment for the listener because the described object has no holes in it. M actually meant "two pegs".

2.2.2 Detecting miscommunication

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

Incompatibility

Two kinds of incompatibility, action or referent, appear in the taxonomy of confusions. The strongest hint that there is a reference problem occurs when the listener finds no real world object to correspond to the speaker's description. 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 object exhibits all of the values). An action 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 of any importance to this side-effect.

Goal obstacle

A goal obstacle occurs when a goal (or subgoal) one is trying to achieve is blocked. This blockage can result in confusion for the listener because he did not expect the speaker to give him tasks that could not 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. In some sense, it is a special kind of goal obstacle where the goal to be fulfilled is blocked because it is already satisfied. It is a simple goal obstacle because nothing has to be done to get around it. However, it can lead to confusion on

the part of listeners because they may suspect they misunderstood what the speaker has requested since they wouldn't expect a reasonable speaker to request the performance of an already completed action. It provides a hint that miscommunication has occurred.

3 Repairing Reference Failures

3.1 Introduction

The previous section illustrated how task-oriented natural language interactions in the real world can induce contextually poor utterances. Given all the possibilities for confusion, when confusions do occur, they must be resolved if the task is to be performed. This section explores the problem of fixing reference failures.

Reference identification is a search process where a listener looks for something in the world that satisfies a speaker's uttered description. A computational scheme for performing reference has evolved from work by other artificial intelligence researchers (e.g., see [14]). That traditional approach succeeds if a referent is found, or fails if no referent is found (see Figure 3(a)). However, a reference identification component must be more versatile than those constructed in the traditional manner. The excerpts provided in the previous section show that the traditional approach is wrong because people's real behavior is much more elaborate. In particular, listeners often find the correct referent even when the speaker's description does not describe any object in the world. For example, a speaker could describe a blue block as the "turquoise block." Most listeners would go ahead and assume that the blue block was the one the speaker meant.

A key feature to reference identification is "negotiation." Negotiation in reference identification comes in two forms. First, it can occur between the listener and the speaker. The listener can step back, expand greatly on the speaker's description of a plausible referent, and ask for confirmation that he has indeed found the correct referent. For example, a listener could initiate negotiation with "I'm confused. Are you talking about the thing that is kind of flared at the top? Couple inches long. It's kind of blue." Second, negotiation can be with oneself. This type of negotiation, called 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, and the listener's own abilities. He then applies that deliberation to determine whether one referent candidate is better than another or, if no candidate is found, what are the most likely places for error or confusion. Such negotiation can result in the listener testing whether or not a particular referent works. For example, linguistic descriptions can influence a listener's perception of the world. The listener must ask himself whether he can perceive one of the objects in the world the way the speaker described it. In some cases, the listener's perception may overrule the description because the listener can't perceive it the way the speaker described it.

To repair the traditional approach we have developed an algorithm that captures for certain cases the listener's ability to negotiate with himself for a referent. It can look 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 I call "FWIM" for "Find What I Mean", is more faithful to the data than the traditional model (see Figure 3(b)).

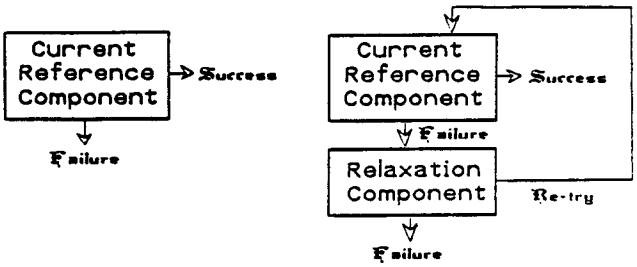

Figure 3: Approaches to reference identification

One means of making sense of an approximate description is to delete or replace portions of it that don't match objects in the hearer's world. In our program we are using "relaxation" techniques to capture this behavior. Our reference identification module treats descriptions as approximate. It relaxes a description in order to find a referent when the literal content of the description fails to provide the needed information. Relaxation, however, is not performed blindly on the description. We try to model a person's behavior by drawing on sources of knowledge used by people. We have developed a computational model that can relax aspects of a description using many of these sources of knowledge. Relaxation then becomes a form of communication repair [4] that hearers can use.

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

3.2.1 Find a referent using a reference mechanism

Identifying the referent of a description requires finding an element in the world that corresponds to the speaker's description (where every feature specified in the description is present in the element in the world but not necessarily vice versa). The initial task of our

reference mechanism is to determine whether or not a search of the (taxonomic) knowledge base that we use to model the world is necessary. For example, 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 do not clearly describe an object because they are composed of imprecise feature values). A number of aspects of discourse pragmatics can be used in that determination (e.g., 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) but we will not examine them here.

The knowledge base contains linguistic descriptions and a description of the listener's visual scene itself. In our implementation and algorithms, we assume it is represented in KL-One [3], a system for describing taxonomic knowledge. KL-One is composed of CONCEPTs, ROLEs on concepts, and links between them. A CONCEPT is like a set, representing those elements described by it. A SUPERC link ("==>") is used between concepts to show set inclusion. For example, consider Figure 3. 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 subset specified by a concept is unique. The Individual Concept D shown in the figure is defined to be a unique member of the subset specified by Concept C. ROLEs on concepts are like normal attributes and slot fillers in other knowledge representation languages. They define a functional relationship between the concept and other concepts.

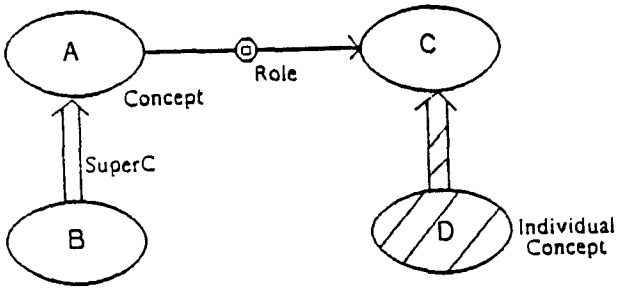

Figure 4: A KL-One Taxonomy

Assuming that a search of the knowledge base is considered necessary, then a reference search mechanism is invoked. The search mechanism uses the KL-One Classifier [16] to search the knowledge base taxonomy. This search is constrained by a focus mechanism based on the one developed by Grosz [14]. The Classifier's purpose is to discover all appropriate subsumption relationships between a newly formed description and all other descriptions in a given taxonomy. With respect to reference, this means that all possible (descriptions of) referents of the description will be subsumed by it after it has been classified into the knowledge base taxonomy. If more than one candidate referent is below (when a description 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 description 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 is invoked. We will only consider the last case in the rest of the paper.

3.2.2 Collect votes for or against relaxing the description

It is necessary to determine whether or not the lack of a referent for a description has to do with the description itself (i.e., reference failure) or outside forces that are causing reference confusion. For example, the problem 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. These rules will not be discussed here so 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 ordered features to order the potential candidates with respect to the preferred ordering of features, and (3) determine the proper relaxation techniques to use and apply them to the description.

Find potential referent candidates

Before relaxation can take place, potential candidates for referents (which denote elements in the listener's visual scene) must first be found. These candidates are discovered by performing a "walk" in the knowledge base taxonomy in the general vicinity of the speaker's classified description. 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 and does not take into account any information about the task domain. The ordering of features and candidates for relaxation described below takes into account the task domain. The set of best descriptions returned by the matcher (as determined by some cutoff score) are selected as referent candidates.

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,[5] and generates a master ordering of features for relaxation. Once the feature order is created, the reference system uses

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

that ordering to determine the order in which to try relaxing the candidates.

We draw primarily on sources of linguistic knowledge, pragmatic knowledge, discourse knowledge, domain knowledge, perceptual knowledge, hierarchical knowledge, and trial and error knowledge during this repair process. A detailed treatment of all of them can be found in [12, 27, 13]. 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. These rules are written in a PROLOG-like language. Figure 5 illustrates one such linguistic knowledge relaxation rule. This rule is motivated by the observation in the excerpts that speakers typically add more important information at the end of a description (where they are separated from the main part of the description and thus provided more emphasis). Since the syntactic constituents often at the end are relative clauses or predicate complements, we created this more specific relaxation rule. However, a more general and more applicable rule is that information presented at the end of a description is usually more prominent.

```
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

Each knowledge source produces its own partial ordering of features. The partial orderings are then integrated to form a directed graph. For example, perceptual knowledge may say to relax color. However, if the color value was asserted in a relative clause, linguistic knowledge would rank color lower, i.e., placing it later in the list of things to relax.

Since different knowledge sources generally have different partial orderings of features, these differences can lead to a conflict over which features to relax. It is the job of the best candidate algorithm to resolve the disagreements among knowledge sources. It's goal is to order the referent candidates, C_i, so that relaxation is attempted on the best candidates first. Those candidates are the ones that conform best to a proposed feature ordering. To start, the algorithm examines pairs of candidates 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 description of this process. 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. Finally, the referent candidates are reordered using the information expressed in the speaker's description and in the directed graph of features.

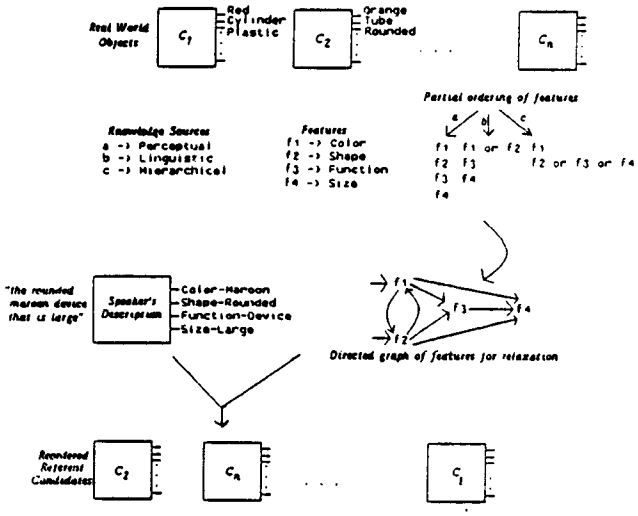

Figure 6: Reordering referent candidates

Once a set of ordered, potential candidates are selected, the relaxation mechanism begins step 3 of relaxation, it tries to find proper relaxation methods to relax the features that have just been ordered (success in finding such methods "justifies" relaxing the description). It stops at the first candidate which is reasonable.

Determine which relaxation methods to apply

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

Often the objects in focus in the real world implicitly cause other objects to be in focus [14, 28]. The subparts of an object in focus, for example, are reasonable candidates for the referent of a failing description and should be checked. At other times, the speaker might attribute features of a subpart of an

object to the whole object (e.g., describing a plunger that is composed of a red handle, a metal rod, a blue cap, and a green cup as "the green plunger"). In these cases, the relaxation mechanism utilizes the part-whole relation in object descriptions to suggest a way to relax the speaker's description.

Relaxation of a description has a few global strategies that can be followed for each part of the description: (1) drop the errorful feature value from the description altogether, (2) weaken or tighten the feature value but keep its new value close to the specified one, or (3) try some other feature value.

These strategies are realized through a set of procedures (or *relaxation methods*) that are organized hierarchically. Each procedure is an expert at relaxing its particular type of feature. 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 are specialists that attempt to first relax the feature value to one "near" 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. If those fail, the feature would simply be ignored.

3.3 An example on handling a misreference

This section describes how a referent identification system can handle 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* (see Figure 1(a) for a picture of these parts). Assume also that the speaker tries to describe two of the objects. "...two devices that are clear plastic. One of them has two openings on the outside with threads on the end, and its about five inches long. The other one is a rounded piece with a turquoise base on it. Both are tubular. The rounded piece fits loosely over...". The reference system can find a unique referent for the first object but not for the second. The relaxation algorithm will be shown below to reduce the set of referent candidates for the second description 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 the "fit" can confirm that the proper referent was found).

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 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 basic functional description of the object. The first entry in each representation in Figure 8 (that entry is shown in uppercase) defines the basic kind of entity being described (e.g., "TUBE" means that the object being described is some kind of tube). The words in mixed case refer to the names of features and the 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. Such subparts can be referred to on their own or as part of the parent object. The "Orientation" feature, used in the representations in Figure 8, provides a rotation and translation of the object from some standard orientation to the object's current orientation in 3-D space. The standard orientation provides a way to define relative positions such as "top," "bottom," or "side."

```
Descr1:
 (DEVICE (Transparency CLEAR)
         (Composition PLASTIC)
         (Subpart (OPENING))
         (Subpart (OPENING))
         (Subpart (THREADS (Rel-Position END)))
         (Dimensions (Length 5.0))
         (Analogical-Shape TUBULAR))


Descr2:
 (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

The first step in the reference process is the actual search for a referent in the knowledge base. The reference identification process is incremental in nature, i.e., the listener can begin the search process before he hears the complete description. This was observed throughout the videotape excerpts and the algorithm presented here is actually designed to be incremental. The KL-One Classifier compares the features specified in the speaker's descriptions (**Descr1** and the "Outer" feature of **Descr2** in Figure 7) with the features specified for each element in the KL-One taxonomy that corresponds to one of the current objects of interest in the real world. Notice that some features are directly comparable. For example, the "Transparency" feature of **Descr1** and the "Transparency" feature of *MAINTUBE* are both equal to "CLEAR." Other features require further processing before they can be compared. The OPENING value of "Subpart" in **Descr1** is thought of primarily as a 2-D cross-section (such as a "hole"), while 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 CYLINDER, the inference must be made that both things can describe the same thing (similar inferences are developed in [18]). One way this inference can occur is by recursively examining the subparts of *MAINTUBE* with the 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 **Descr1**.

Descr2 presents different problems. **Descr2** refers to an object that is supposed to have a subpart that is TURQUOISE. The Classifier determines that **Descr2** could not describe either the *CAP* or *STAND* because both are BLUE. It also could not describe the *MAINTUBE*[6] or *AIR CHAMBER* since each has subparts that are either VIOLET or BLUE. The Classifier places **Descr2** as best it can in the taxonomy, showing no connections between

[6] Since **Descr1** refers to *MAINTUBE*, *MAINTUBE* could be dropped as a potential referent candidate for **Descr2**. We will, however, leave it as a potential candidate to make this example more complex.

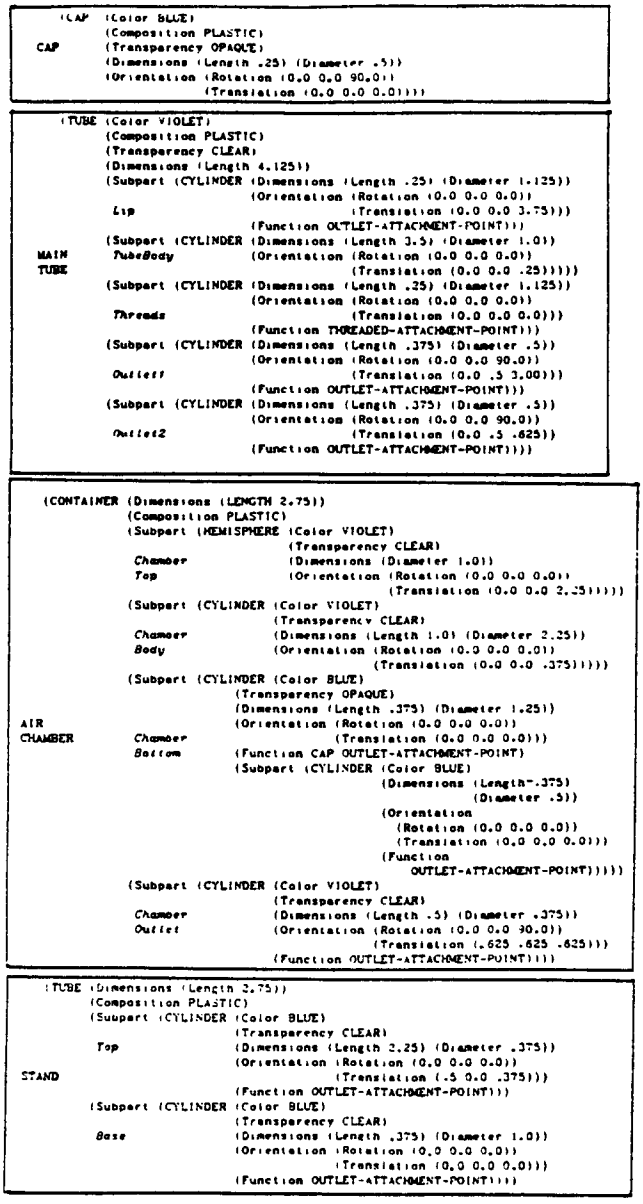

Figure 8: The objects in focus

it and any of the objects currently in focus. 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 **Descr2** in the taxonomy and using the partial matcher to score how close each element is to **Descr2**.[7] The matcher determines *MAINTUBE*, *STAND*, and *AIR CHAMBER* as reasonable candidates by aligning and comparing their features to **Descr2**.

Scoring **Descr2** 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 **Descr2** 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 **Descr2** differs from color VIOLET of *MAINTUBE*. (-)

Scoring **Descr2** 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 **Descr2** 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 Descr2 differs from color BLUE of *STAND*. (-)

Scoring **Descr2** to *AIR CHAMBER*:

- a CONTAINER is a kind of DEVICE. (>)
- the Transparency of **Descr2**, 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. (>)
- the recursive partial matching of subparts. BASE in **Descr2** 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 **Descr2** differs from color BLUE of *AIR CHAMBER*. (-)

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.

[7] 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 as a scale: HIGHEST {+}, {> <}, {=}, {?}, {-} LOWEST.

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'll order the features in **Descr2** using linguistic knowledge. Linguistic analysis of **Descr2**, "... are clear plastic ... a rounded piece with a turquoise base ... Both are tubular ... fits loosely over ...," tells us that the features were specified using the following modifiers.

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

Observations from the protocols (as described by the rules developed in [13]) has shown that people tend to relax first features specified as adjectives, then as prepositional phrases and finally as relative clauses or predicate complements. This suggests relaxation of **Descr2** in the order:

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

The set of features on the left side of a "<" symbol is relaxed before the set on the right side. The order that the features inside the braces, "}..{", are relaxed is left unspecified (i.e., any order of relaxation is 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"). In this example, since the colors are all very close - BLUE, TURQUOISE, and VIOLET - then Color may be a reasonable thing to relax. Hierarchical information about how closely related one feature value is to another can also be used to determine what to relax. The Shape values are a good example. 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}.
```

The referent candidates *MAINTUBE*, *STAND*, and *AIR CHAMBER* can be examined and possibly ordered for relaxation using the above feature ordering. For this example, the relaxation of **Descr2** to any of the candidates requires relaxing their SHAPE and COLOR features. Since they each require relaxing the same features, the candidates can not be ordered with respect to each other (i.e., none of the possible feature orders is better for relaxing the candidates). Hence, no one candidate stands out as the most likely referent.

While no ordering of the candidates was possible, the order generated to relax the features in the speaker's description can be used to guide the relaxation of each candidate. The relaxation methods mentioned at the end of the last section come into use here. Generate-Similar-Shape-Values can determine that HEMISPHERICAL and CYLINDRICAL shapes of the *AIR CHAMBER* are close to the 3D-ROUND shape. This holds equally true for the cylindrical shapes of the *MAINTUBE* and the *STAND*. Generate-Similar-Color-Values next tries relaxing the Color TURQUOISE. It determines the colors BLUE and GREEN as 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 might 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. 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 still wrong on occasion. In Section 2, 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 problems, 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 go about it in a certain 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.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Allen, James F. *A Plan-Based Approach to Speech Act Recognition*. Ph.D. Th., University of Toronto, 1979.

[2] Appelt, Douglas E. *Planning Natural Language Utterances to Satisfy Multiple Goals*. Ph.D. Th., Stanford University, 1981.

[3] Brachman, Ronald J. *A Structural Paradigm for Representing Knowledge*. Ph.D. Th., Harvard University, 1977. Also, Technical Report No. 3605, Bolt Beranek and Newman Inc.

[4] Brown, John Seely and Kurt VanLehn. "Repair Theory A Generative Theory of Bugs in Procedural Skills." *Cognitive Science 4*, 4 (1980), 379-426.

[5] Cohen, Philip R *On Knowing What to Say. Planning Speech Acts*. Ph.D. Th., University of Toronto, 1978.

[6] Cohen, P., C Perrault and J. Allen. Beyond Question Answering. In *Knowledge Representation and Natural Language Processing*, W. Lehnart and M. Ringle, Ed.,Lawrence Erlbaum Associates, 1981.

[7] Cohen, Philip R. The need for Referent Identification as a Planned Action. Proceedings of IJCAI-81, Vancouver, B.C., Canada, August, 1981, pp. 31-35.

[8] Cohen, Philip R., Scott Fertig and Kathy Starr. Dependencies of Discourse Structure on the Modality of Communication. Telephone vs. Teletype. Proceedings of ACL, Toronto, Ont., Canada, June, 1982, pp. 28-35.

[9] Cohen, Philip R. "The Pragmatics of Referring and the Modality of Communication." *Computational Linguistics 10*, 2 (April-June 1984), 97-146.

[10] Gentner, Dedre. The Structure of Analogical Models in Science. Bolt Beranek and Newman Inc., July, 1980.

[11] Goodman, Bradley A. Miscommunication in Task-Oriented Dialogues. KRNL Group Working Paper, Bolt Beranek and Newman Inc., April 1982.

[12] Goodman, Bradley A. Repairing Miscommunication: Relaxation in Reference. Proceedings of AAAI-83, Washington, D.C., August, 1983, pp. 134-138.

[13] Goodman, Bradley A. *Communication and Miscommunication*. Ph.D. Th., University of Illinois, Urbana, 1984.

[14] Grosz, Barbara J. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. Th., University of California, Berkeley, 1977. Also, Technical Note 151, Stanford Research Institute.

[15] Grosz, Barbara J. Focusing and descriptions in natural language dialogues. In *Elements of Discourse Understanding*, Joshi, Webber and Sags, Ed.,Cambridge University Press, 1981, pp. 84-105.

[16] Lipkis, Thomas. A KL-ONE Classifier. Proceedings of the 1981 KL-One Workshop, June, 1982, pp. 128-145. Report No. 4842, Bolt Beranek and Newman Inc. Also Consul Note # 5, USC/Information Sciences Institute, October 1981.

[17] Litman, Diane J. and James F. Allen. A Plan Recognition Model for Clarification Subdialogues. Proceedings of Coling84, Stanford University, Stanford, CA., July, 1984, pp. 302-311.

[18] Mark, William. Realization. Proceedings of the 1981 KL-One Workshop, June, 1982, pp. 78-89. Report No. 4842, Bolt Beranek and Newman Inc.

[19] McKeown, Kathleen R. Recursion in Text and Its Use in Language Generation. Proceedings of AAAI-83, Washington, D.C., August, 1983, pp. 270-273.

[20] Reichman, Rachel. "Conversational Coherency." *Cognitive Science 2*, 4 (1978), 283-327.

[21] Reichman, Rachel. *Plain Speaking: A Theory and Grammar of Spontaneous Discourse*. Ph.D. Th., Harvard University, 1981. Also, Technical Report No. 4861, Bolt Beranek and Newman Inc.

[22] Ringle, Martin and Bertram Bruce. Conversation Failure. In *Knowledge Representation and Natural Language Processing*, W. Lehnart and M. Ringle, Ed.,Lawrence Erlbaum Associates, 1981.

[23] Sidner, C. L., and Israel, D.J. Recognizing intended meaning and speaker's plans. Proceedings of the International Joint Conference in Artificial Intelligence, The International Joint Conferences on Artificai Intelligence, Vancouver, B.C., August, 1981, pp. 203-208.

[24] Sidner, Candace Lee. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse*. Ph.D. Th., Massachusetts Institute of Technology, 1979. Also, Report No. TR-537, MIT AI Lab.

[25] Sidner, C. L., M. Bates, R. J. Bobrow, R. J. Brachman, P. R. Cohen, D. J. Israel, J. Schmolze, B. L. Webber, W. A. Woods. Research in Knowledge Representation for Natural Language Understanding Report No. 4785, Bolt Beranek and Newman Inc., November, 1981.

[26] Sidner, C. L., Bates, M., Bobrow, R., Goodman, B., Haas, A., Ingria, R., Israel, D., McAllester, D., Moser, M., Schmolze, J., Vilain, M. Research in Knowledge Representation for Natural Language Understanding - Annual Report, 1 September 1982 - 31 August 1983. Technical Report 5421, BBN Laboratories, Cambridge, MA, 1983.

[27] Sidner, C., Goodman, B., Haas, A., Moser, M., Stallard, D., Vilain, M. Research in Knowledge Representation for Natural Language Understanding - Annual Report, 1 September 1983 - 31 August 1984. Technical Report 5694, BBN Laboratories Inc., Cambridge, MA, 1984.

[28] Webber, Bonnie Lynn. *A Formal Approach to Discourse Anaphora*. Ph.D. Th., Harvard University, 1978. Also, Technical Report No. 3761, Bolt Beranek and Newman Inc.