

REPAIRING MISCOMMUNICATION: RELAXATION IN REFERENCE*

Bradley A. Goodman

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, MA. 02238

ABSTRACT

In natural language interactions, a speaker and listener cannot be assured to have the same beliefs, contexts, backgrounds or goals. This leads to difficulties and mistakes when a listener tries to interpret a speaker's utterance. 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 paper explores the problem of resolving such reference failures in the context of the task of assembling a toy water pump. We are using actual protocols to drive the design of a program that plays the part of an apprentice who must interpret the instructions of an expert and carry them out. A primary means for the apprentice to repair such descriptions is by relaxing parts of the description.

I INTRODUCTION

Consider the dialogue below which exemplifies some kinds of complex descriptions used in utterances. Here A is instructing B to assemble part of a toy water pump [9, 7]. Refer to Figure 1 for a picture of the pump. A and B are communicating verbally but neither can see the other. (The bracketed text in the excerpt shows 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. In Line 1, B interprets "the long blue tube" to refer to the STAND. When A adds the relative clause "that has two outlets on the side," B is forced to drop the STAND as the referent, to relax the color "blue" to "violet," and to select the MAINTUBE. In Line 6, A's description "the nozzle-locking piece" is too specific and B selects the NOZZLE instead of the SPOUT. A's addition of "the clear plastic one" in line 7 rules out the NOZZLE - which is red and opaque - in favor of the SPOUT. Line 16 demonstrates a case where A previously focused B's attention on one object and intends to switch that focus to another one. In this case, B doesn't shift focus. This lack of agreement on what is in focus leads to confusion later on in the dialogue.

- A: 1. Take the long blue tube
[B reaches toward STAND]
2. that has two outlets on the side -
[B takes MAINTUBE]
3. that's the main tube.
4. Place the small blue cap
[B takes CAP]
5. over the hole on the side of that tube.
[B pushes CAP on OUTLET1]

6. Take the nozzle-locking piece,
[B grabs NOZZLE]
7. the clear plastic one,
[B takes SPOUT]
8. and place it on the other hole
[B identifies OUTLET2 of MAINTUBE]
9. that's left, so that the nozzle
10. points away.
[B installs SPOUT on OUTLET2 of MAINTUBE]
11. Okay?
B: 12. Okay.
A: 13. Now take the blue lid type thing
[B takes TUBEBASE]
14. and screw it onto the bottom
[B screws TUBEBASE on MAINTUBE]
15. ooops,
[A realizes he has forgotten to have B put
SLIDEVALVE into OUTLET2 of MAINTUBE]
16. undo the plastic thing
[B removes TUBEBASE but A meant the SPOUT]

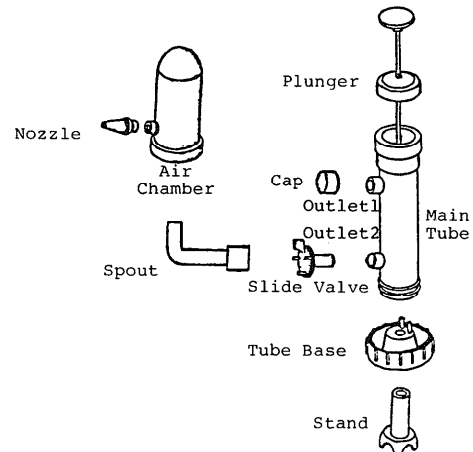


Figure 1: The Toy Water Pump

In conversation people use imperfect descriptions to communicate about objects; sometimes their partners succeed in understanding and occasionally they fail. I am working on a theory of the use of extensional descriptions that will explain how people successfully use such imperfect descriptions.

One means of making sense of an approximate description is to relax portions of it that don't match objects in the hearer's world. Relaxation then is a form of communication repair [5] that hearers can use. As part of my work I am developing a reference identification module for a natural language system that will treat descriptions as approximate. It can relax a description in order to find a referent when the literal content of the description fails to provide

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

the needed information. In this paper I will describe the relaxation component of the reference identification module and illustrate some of the sources of knowledge that guide it in relaxing a description.

II THE KINDS OF PROBLEMS

Part of my research has been an examination of how a listener discovers that a repair of a description is needed, and how the listener discovers the source of the problem in the communication.

- o How the problems are discovered:
 1. The listener finds no Real World object to correspond to the speaker's description;
 2. the listener finds other than the requested number of Real World objects (i.e., too many or too few);
 3. the listener cannot perform the action specified by the speaker because of some obstacle; or
 4. the listener performs the action but does not arrive at its intended effect.
- o Where the problems may reside:
 1. In the speaker's description of an object presented in the utterance;
 2. in the speaker's description of a physical action presented in the utterance;
 3. with the set of Real World objects that have been brought into attention (the speaker's set may differ from the listener's set);
 4. with the set of Real World actions that have been brought into attention (the speaker's set may differ from the listener's set);
 5. in the interpretation of the underlying force of the utterance (i.e., does the speaker want the listener to simply note the information in the utterance or to use it to do something); or
 6. with the hearer's concentration (e.g., the hearer may fail to pay attention, missing or mishearing a word or the like).*

These observations signal conditions in which a mistake might occur and where it might be found. We will now explore what a listener has available for resolving miscommunication.

III KNOWLEDGE FOR REPAIRING DESCRIPTIONS

When things go wrong during a conversation, people have lots of knowledge that they bring to bear to get around the problem (see [16]). Much of the time the repairs are so natural that we aren't conscious that they have taken place. At other times, we must make an effort to correct what we have heard, or determine that we need clarification from the speaker. This repair process involves the use of knowledge about conversation, its social conventions and the world around us.

In this work, I chose to consider the repair of descriptions rather than complete utterances. The most relevant knowledge for repair depends on the

*I am including this kind of problem because I have been talking about human dialogues. I will not, however, pursue it any further.

conversation itself and the Real World described therein. There are numerous sources of knowledge to consider that drive the reference repair process. We will look at two sources, linguistic and perceptual knowledge. Linguistic knowledge is the use of the structure and meaning of a description. Perceptual knowledge is a person's abilities to distinguish feature values, one's preferences in features by considering which seem more important (with respect to the person and the domain), and one's perception of an object. Other knowledge sources, such as discourse knowledge [9, 14, 18, 17, 15, 2, 11], pragmatic knowledge [6, 1, 13, 3], trial and error knowledge, hierarchical knowledge, and domain knowledge [9] will not be covered here. A more detailed treatment can be found in [8].

A. Linguistic Knowledge in Reference

Different linguistic structures can be utilized to describe objects in the extensional world. This section outlines some of these structures and their meanings and shows how they can be used to guide repairs in the description.

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

Relative clauses are well suited for expressing complicated information since they are separate from the main part of the noun phrase and can be arbitrarily complex themselves.

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

Prepositional phrases are better fitted for simpler pieces of information. They are often used to express predicative relationships.

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

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

Adjectives are used to express almost any perceptual feature - though complex relations can be awkward. Usually they modify the noun phrase directly, but sometimes they are expressed as a

predicate complement. In those situations, the complement describes the subject of the linking verb (e.g., "the tube is large"). As with some of the relative clauses above, predicate complements have an assertional nature to them.

B. Relaxing a Description Using Linguistic Knowledge

The relaxation process attempts to relax features in a description in the order: adjectives, then prepositional phrases and finally relative clauses and predicate complements. This order was chosen by examining the water pump protocols and by noting where the linguistic forms come into play during reference resolution. Adjectives and prepositional phrases play a more central role while relative clauses usually play a secondary role during referent identification. Relative clauses and predicate complements exhibit an assertional nature that reduces their usefulness for resolving the current reference (whereas the information they express can be useful in subsequent references). The head noun can also be relaxed. It normally is relaxed last but could be relaxed prior to a relative clause (especially in the instances where the relative clause expresses confirmational information).

For example, consider the description "the large violet cylinder that has two outlets." Here, the features size, color and shape are described in the adjectives and head noun of the description, and the two subparts' function in the relative clause. Following the above rules, the relaxation of size, color and shape should be attempted before either the number of subparts or the subparts' functions. The relaxation order is influenced by the other knowledge sources so the order proposed here is not hard and fast.

C. Perceptual Knowledge in Reference

A major factor involved here is how people perceive objects in the world and how this can be simulated in my system. Each object for my system is denoted by two forms: a spatial (3-D) representation and a cognitive/linguistic form that shows how the system could actually talk about the object. The spatial description is a physical description of the object in terms of its dimensions, the basic 3-D shapes composing it, and its physical features. The cognitive/linguistic form is a representation of the parts and features of the object in linguistic terms. It overlaps the visual form in many respects but it is more suggestive of the listener's perceptions. The cognitive/linguistic form describes aspects of an object such as its subparts by its position on the object ("top", "bottom") and its functionality ("outlets", "places for attachment"). More than one cognitive/linguistic form can refer to the same physical description. Some properties of an object differ in how they are expressed in the two forms. In the 3-D form, there are primarily properties such as numerical dimensions (e.g., "3 feet by 5 feet") and basic shapes (e.g., generalized cylinders), while, in the cognitive/linguistic form, there are relative dimensions (e.g., "large") and analogical shapes (e.g., "the L-shaped tube").

Perception, hence, may involve interpretation. This can lead to discrepancies between individuals. People usually agree on the spatial representation but not necessarily on the cognitive/linguistic description and this can lead to problems. For example, misjudgements by the speaker in calling an object "large" can cause the hearer to fail to find an object in the visual world that has dimensions that are perceptually "large" to the listener.

To prevent confusion of the listener, a speaker must distinguish the objects in the environment from each other. The perceptual features of an object provide people with a way to discriminate one object from another. A speaker must take care when selecting from these features since they can induce their own confusion. Perceptual features may be inherently confusing because a feature's values are difficult to judge (e.g., is the tube a cylinder or a slightly tapering cone?). They may

also be confusing because the speaker and listener may have differing sets of values for a feature (e.g., what may be blue for someone may be turquoise for another). These characteristics affect the salience of a feature (see [12]) which in turn determines the feature's usefulness in a description. A feature that is common in everyday usage (e.g., color, shape and size) is salient because the listener can readily distinguish the feature's possible values from each other. Of course, very unusual values of a feature can stand out, making it even easier to discriminate a unique object from all other objects [12].

The objects in the world may exhibit a feature whose possible values are difficult to distinguish. This occurs when a perceived feature does not have much variability in its range of values: all the values are clustered closely together making it hard to tell the difference between one value and the next.* This increases the likelihood of confusion because the usefulness of specifying the feature to a non-expert is diminished (especially if the speaker is more expert than the listener in distinguishing feature values). Hence, if one of these difficult feature values appears in the speaker's description, the listener, if he isn't an expert, will often relax the feature value to any of the members of the set of feature values.

D. Relaxing a Description Using Perceptual Knowledge

When examining the features presented in a speaker's description, one can consider perceptual aspects to determine which features are most likely in error. Such an inspection can generate a partial ordering of features for use during the repair process to determine which feature in a description to relax. As shown below, the relaxation ordering suggested by the inspection of features interacts with ordering proposals from other knowledge sources.

Active features are ones that require a listener to do more than simply recognize that a particular feature value belongs to a set of possible values - the listener must perform some kind of evaluation. When considering the water pump domain, it seems that one should first relax those features that require less active consideration such as color (though it is easier to relax red to orange than red to blue), composition, transparency, shape and function. Only after this should one relax those features that require active consideration of the object under discussion and its surroundings (such as superlatives, comparatives, and relative values like size, length, height, thickness, position, distance and weight). People tend to be casual with less active features while the active ones require their full attention. Hence, in a reference failure the source of the problem is likely to be the less active ones.

IV THE RELAXATION COMPONENT

I have discussed some of the numerous kinds of knowledge available to a listener to interpret a speaker's description. I pointed out places where that knowledge affects the listener's ability to interpret a description and ways in which it is helpful to the listener for overcoming poor descriptions. When a description fails to denote properly a referent in the Real World, it is possible to repair it by a relaxation process that drops 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. There are several kinds of relaxation possible. One can ignore a constituent, replace it with something close, replace it with a related

*For example, certain Eskimo languages have names for many different grades of snow that may be difficult for most non-Eskimos to distinguish [19].

value, and change focus (i.e., consider a different group of objects.). In this section, I will describe the overall relaxation component that draws on the knowledge sources as it tries to relax the errorful description to one that suffices.

A. Find a Referent Using a Reference Mechanism

Identifying the referent of a description entails finding an element in the world that corresponds to the speaker's description (where "described by" means 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* is necessary. A number of aspects of discourse pragmatics can be used in that determination but I will not examine them here.

If a search of the knowledge base is considered necessary, then the reference search mechanism is invoked. The search mechanism uses the KL-One Classifier [10] to search** the knowledge base taxonomy. The Classifier uses the subsumption relationships inherent in the taxonomy to place the description in the correct spot [10]. What this means with respect to reference is that the possible referents of the description will be found below the description after it has been classified into the knowledge base taxonomy. If more than one referent is below the classified description, then, unless a quantifier in the description specified more than one element, the speaker's description is ambiguous. If 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, then the relaxation component is invoked.

B. 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.*** Pragmatic rules are invoked to decide whether or not the description should be relaxed. These rules will not be discussed here.

C. Perform the Relaxation of the Description

If relaxation is voted for, then the system must (1) find potential referent candidates, (2) determine which features to relax and in what order, and use that 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.

*The knowledge base contains linguistic descriptions and a description of the listener's visual scene itself. Here it is represented in KL-One [4], a system for describing inheritance taxonomies.

**This search is constrained by a focus mechanism [9, 14, 17].

***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.

1. 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 scoring KL-One partial matcher is used to determine how close candidate descriptions found during the walk are to the speaker's description. The partial matcher generates a score to represent how well the descriptions match (it also generates scores at the feature level to help determine how the features are to be aligned and how well they match). The best of the descriptions returned by the matcher are selected as referent candidates.

2. Order the features and candidates for relaxation

At this point the reference system inspects the speaker's description and the candidates and decides which features to relax and in what order.* Once the feature order is created, it determines the order in which to try relaxing the candidates.

Various knowledge sources are consulted to determine the relaxation ordering. These include the perceptual and linguistic knowledge sources that were described above, as well as others not discussed here. The suggestions from the knowledge sources are then integrated. This integration requires evaluating the partial orderings imposed by each knowledge source. For example, perceptual knowledge says to relax color. However, if the color value was asserted in a relative clause, linguistic knowledge would rank color lower. This leads to a conflict. Thus, the relaxation of some other feature may win out over color should it cause less conflict.

Thus, the feature ordering can be used to order candidates: choose first those candidates that best follow the feature order when determining changes that must be made to the speaker's description. The control structure to enforce this rule examines each candidate and assigns a higher priority to those candidates that exhibit a feature ranked higher in the order of features. Hence, the candidates with the least important features slip to the back of the queue.

Once a potential candidate is selected by the controller, 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).

3. Determine which relaxation methods to apply

Relaxation can take place with many aspects of a speaker's description: with the focus of attention in the Real World where one attempts to find a match, with complex relations specified in the description, and with individual features of a referent specified by the description.

Often the objects in focus in the Real World implicitly cause other objects to be in focus [9, 18]. The subparts of an object in focus, for example, are reasonable candidates for the referent of a bad 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 follows the part-whole

*Of course, once a 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).

relation.

Complex relations specified in a speaker's description can also be relaxed. These relations 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").

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

Relaxation of a description has a few global strategies that can be followed: (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.

The realization of these strategies is 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, the Generate-Similar-Feature-Values procedure is composed of procedures like Generate-Similar-Shape-Values and Generate-Similar-Size-Values. Each of those procedures are further divided into specialists that first attempt to 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 could be dropped out of consideration.

CONCLUSIONS

Natural language interactions in the Real World invite contextually poor descriptions. This paper sketches the ideas behind an on-going effort to develop a reference identification mechanism that can exhibit more "human" tolerance of such descriptions. My goal is to build a more robust system that can handle errorful descriptions when looking for a referent, and that is adaptable to existing systems. My work tackles the use of descriptions referring to the Real World and the repair of problems in those descriptions.

The work attempts to provide a computational scheme for handling noun phrases (following the work on noun phrases by [9, 18, 14, 17]) that is robust enough to provide human-like performance. When people are asked to identify, they go about it in a certain way: find candidates, adjust as necessary, re-try, and, if necessary, give up and ask for help. I 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. My work provides a forum for trying out the different parameters.

ACKNOWLEDGEMENTS

I want to especially thank Candy Sidner for her insightful comments and suggestions during the course of this work. I'd also like to acknowledge the helpful comments of Chip Bruce, Jeff Gibbons, Diane Litman, Jim Schmolze, Marc Vilain and Dave Waltz on portions of 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] Allen, James F., Alan M. Frisch, and Diane J. Litman. ARGOT: The Rochester Dialogue System. Proceedings of AAAI-82, Pittsburgh, Pa., August, 1982, pp. 66-70.

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

[4] Brachman, Ronald J. A Structural Paradigm for Representing Knowledge. Ph.D. Th., Harvard University, 1977.

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

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

[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] Goodman, Bradley A. Miscommunication and Reference. KNRL Group Working Paper, Bolt Beranek and Newman Inc., January 1983.

[9] Grosz, Barbara J. The Representation and Use of Focus in Dialogue Understanding. Ph.D. Th., University of California, Berkeley, 1977.

[10] Lipkis, Thomas. A KL-ONE Classifier. Proceedings of the 1981 KL-One Workshop, June, 1982, pp. 128-145.

[11] Litman, Diane. Discourse and Problem Solving. KNRL Group Working Paper, Bolt Beranek and Newman Inc., August 1982.

[12] McDonald, David D. and E. Jeffery Conklin. Saliency as a Simplifying Metaphor for Natural Language Generation. Proceedings of AAAI-82, Pittsburgh, Pa., August, 1982, pp. 75-78.

[13] Perrault, C. Raymond and Philip R. Cohen. It's for your own good: a note on inaccurate reference. In Elements of Discourse Understanding, Joshi, Webber and Sags, Ed., Cambridge University Press, 1981, pp. 217-230.

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

[15] Reichman, Rachel. Plain Speaking: A Theory and Grammar of Spontaneous Discourse. Ph.D. Th., Harvard University, 1981.

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

[17] Sidner, Candace Lee. Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. Ph.D. Th., Massachusetts Institute of Technology, 1979.

[18] Webber, Bonnie Lynn. A Formal Approach to Discourse Anaphora. Ph.D. Th., Harvard University, 1978.

[19] Benjamin Lee Whorf. Language, Thought, and Reality. The M.I.T. Press, 1956.