Reference and Reference Failures

Bradley A. Goodman

BBN Laboratories Inc.
10 Moulton Street
Cambridge, Ma. 02238

Introduction

Reference in the real world differs greatly from the reference processes modelled in current natural language systems. A speaker in the real world is a rational agent who must make a decision about his description in a limited time, with limited resources, knowledge, and abilities. In particular, the speaker's perceptual and communicative skills are imperfect or his model of the listener is erroneous or incomplete. Additionally, a speaker can also be sloppy in his description. Since the speaker's goal in the reference process is to construct a description that "works" for the listener, the listener, from his viewpoint, must take these imperfections into account when trying to interpret the speaker's utterances. Yet, listeners, too, have imperfect perceptual or communicative skills and can be sloppy. Hence, they must be prepared to deal with their own imperfections when performing reference identification. In real reference, listener's often recover from initial misunderstandings with or without help from the speaker. Natural language understanding systems must do this, too. Therefore, in performing the reference process, a system should assume and expect problems.

The focus of my work in [3, 4, 5] was to study how one could build robust natural language processing systems that can detect and recover from miscommunication. I investigated how people communicate and how they recover from problems in communication. That investigation centered on reference problems, problems a listener has determining whom or what a speaker is talking about. A collection of protocols of a speaker explaining to a listener how to assemble a toy water pump were studied and the common errors in speakers' descriptions were categorized. The study led to the development of techniques for avoiding failures of reference that were employed in the reference identification component of a natural language understanding program.

The traditional approaches to reference identification in natural language systems were found to be less flexible than people's real behavior. In particular, listeners often find the correct referent even when the speaker's description does not describe any object in the world. To model a listener's behavior, a new component was added to the traditional reference identification mechanism to resolve difficulties in a speaker's description. This new component uses knowledge about linguistic and physical context in a negotiation process that determines the most likely places for error in the speaker's utterance. The actual repair of the speaker's description is achieved by using the knowledge sources to guide relaxation techniques that delete or replace portions of the description. The algorithm developed more closely approximates people's behavior than reference algorithms designed in the past. The next section describes in more detail my work on reference.

Reference

Communication involves a series of utterances from a speaker to a hearer. The hearer uses these utterances to access his own knowledge and the world around him. Some of these utterances are noun phrases that refer to objects, places, ideas and people that exist in the real world or in some imaginary world. They cannot be considered in isolation. For example, consider the utterance "Give me that thing." It can be uttered in many different situations and can result in different referents of "that thing." Understanding such referring expressions requires the hearer to take into account the speaker's intention, the speaker's overall goal, the beliefs of the speaker and hearer, the linguistic context, the physical context, and the syntax and semantics of the current utterance. The hearer could misinterpret the speaker's information in any one of these parts of communication. Such misunderstandings constitute miscommunication. In my research I focused primarily on effects of the linguistic context and the physical context.

To explore such reference problems, the following method was devised and followed. First, protocols

of subjects communicating about a task were analyzed. Knowledge that people used to recover from reference miscommunications - knowledge about the world and about language - was then isolated. Algorithms were designed to apply a person's knowledge about linguistic and physical context to determine the most likely places for error in the speaker's utterance. Then, computer programs were written: (1) to represent a spatially complex physical world, (2) to manipulate the structure of that representation to reflect the changes caused by the listener's interpretation of the speaker's utterances and by physical actions to the world, (3) to perform referent identification on noun phrases, and, when referent identification failed, (4) to search the physical world for reasonable candidates for the referent. These programs form one component of a natural language system.

One goal in this summary of my research is to illustrate how my views on reference identification departed from views held by other researchers in artificial intelligence. Another goal is to show where my research fits in the scheme of natural language understanding by computers. My last goal is to summarize the approach of my research.

A new reference paradigm from a computational viewpoint

Reference identification is a search process where a listener looks for something in the world that satisfies a speaker's uttered description. A computational scheme for performing such reference identifications has evolved from work by other artificial intelligence researchers (e.g., see [6]). That traditional approach succeeds if a referent is found, or fails if no referent is found (see Figure 1(a)). However, a reference identification component must be more versatile than those previously constructed. The excerpts provided in [3] show that the traditional approach is inadequate because people's real behavior is much more elaborate. In particular, listeners often find the correct referent even when the speaker's description does not describe any object in the world. For example, a speaker could describe a turquoise block as the "blue block." Most listeners would go ahead and assume that the turquoise block was the one the speaker meant since turquoise and blue are similar colors.

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

To repair the traditional approach I developed an algorithm that captures for certain cases the listener's ability to negotiate with himself for a referent. It can search for a referent and, if it doesn't find one, it can try to find possible referent candidates that might work, and then loosen the speaker's description using knowledge about the speaker, the conversation, and the listener himself. Thus, the reference process becomes multi-step and resumable. This computational model, which I call "FWIM" for "Find What I Mean", is more faithful to the data than the traditional model (see Figure 1(b)).

One means of making sense of a failed description is to delete or replace the portions that cause it not to match objects in the hearer's world. In my program I am using "relaxation" techniques to capture this behavior. My 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. I try to model a person's behavior by drawing on sources of knowledge used by people. I have developed a computational model that can relax aspects of a description using many of these sources of knowledge. Relaxation then becomes a form of communication repair (in the style of the work on repair theory found in [1]). A goal in
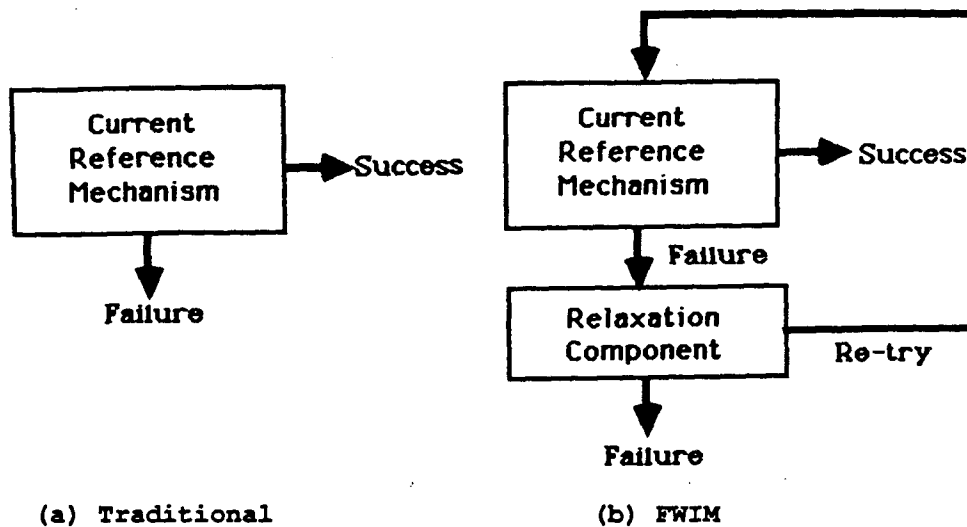
172

(a) Traditional	**(b) FWIM**

Figure 1: Approaches to reference identification

my model is to use the knowledge sources to reduce the number of referent candidates that must be considered while making sure that a particular relaxation makes sense. A brief description of it follows.

The component works by first selecting with a partial matcher a set of reasonable referent candidates for the speaker's description (see also [7]). The candidates are selected by searching the knowledge base, scoring partial matches of each candidate to the speaker's description, and selecting those with higher scores. The component then generates, using information from the knowledge sources, a relaxation ordering graph that describes the order to relax features in the speaker's description. Finally, it combines the candidates with the ordering to yield the most likely referent. An ordered relaxation of parts of the speaker's description can be provided by consulting knowledge known about linguistics (the actual form of the speaker's utterance), perception (physical aspects of the world and the listener's ability to distinguish different feature values in that world), specificity (hierarchical knowledge to judge how vague or specific a particular feature value is), and others. In other words, the algorithm attempts to show how a listener might judge the importance of the features specified in a speaker's description using knowledge about linguistic and physical context. Figure 2 illustrates this process. The speaker's description is represented at the top of the figure. The set of specified features and their assigned feature value (e.g., the pair Color: Maroon) are also shown there. A set of objects in the real world are selected by the partial matcher as potential candidates for the referent. These candidates are shown near the top of the figure $(C_1, C_2, ..., C_n)$. Inside each box is a set of features and feature values that describe that object. A set of partial orderings are generated that suggest which features in the speaker's description should be relaxed first - one ordering for each knowledge source (shown as "Linguistic," "Perceptual," and "Hierarchical" in the figure). For example, linguistic knowledge recommends relaxing Color or Shape before Function, and relaxing Function before Size. A control structure was designed that takes the speaker's description, puts all the (partial) orders together, and then attempts to satisfy them as best it can. This is illustrated at the bottom of the diagram by the reordered referent candidates.

Summary

My goal in this work is to build robust natural language understanding systems, allowing them to detect and avoid miscommunication. The goal is not to make a perfect listener but a more tolerant one that could avoid many mistakes, though it may still be wrong on occasion. In this summary of my research, I indicated that problems can occur during communication. I showed that reference mistakes are one kind of obstacle to robust communication. To tackle reference errors, I described how to extend the succeed/fail paradigm followed by previous natural language researchers.

I represented real world objects hierarchically in a knowledge base using a representation language, NIKL, 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. I claim that people behave better than this and explicitly illustrated such cases in an expert-apprentice domain about toy water pumps [3].
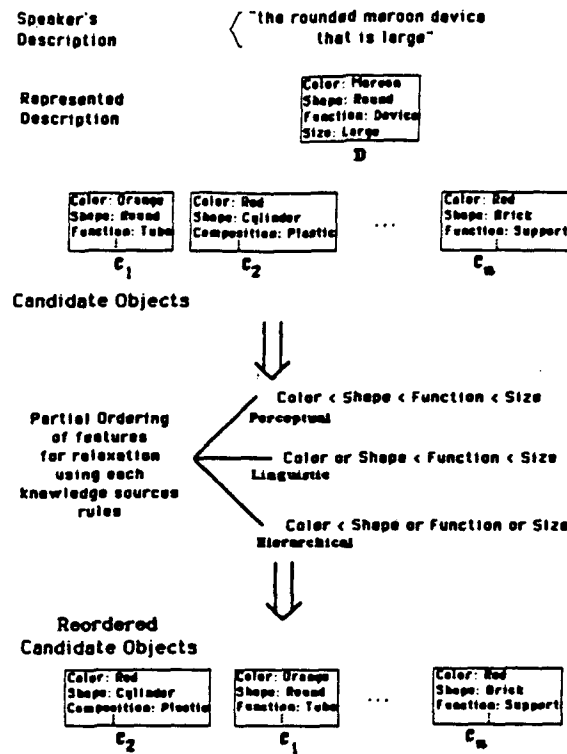
173

Figure 2: Reordering referent candidates

I 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 appear to behave in a particular 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 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. I 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.

Future directions

The FWIM reference identification system I developed models the reference process by the classification operation of NIKL. I need a more complicated model for reference. That model might need a complete identification plan that requires making inferences beyond those provided by classification. The model could also require the execution of a physical action by the listener before determining the proper referent. Cohen gives two excellent examples of such reference plans (pg. 101, [2]). The first, "the magnetic screwdriver, please," requires the listener to place various screwdrivers against metal to determine which is magnetic. The second, "the three two-inch long salted green noodles" requires the listener to count, examine, measure and taste to discover the proper referent.

ACKNOWLEDGEMENTS

References

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

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

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

[4] Goodman, Bradley A. Repairing Reference Identification Failures by Relaxation. Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics, Chicago, Illinois, July, 1985, pp. 204-217.

[5] Goodman, Bradley A. Reference Identification and Reference Identification Failures. Accepted for publication in *Computational Linguistics*, 1986.

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

[7] Joshi, Aravind K. A Note on Partial Match of Descriptions: Can One Simultaneously Question (Retrieve) and Inform (Update)? Theoretical Issues in Natural Language Processing-2, Urbana, Ill., July, 1978, pp. 184-186.