Collaborating on Referring Expressions

Peter A. Heeman
Department of Computer Science
University of Toronto
Toronto, Canada, M5S 1A4
heeman@ai.toronto.edu

Abstract

This paper presents a computational model of how conversational participants collaborate in making referring expressions. The model is based on the planning paradigm. It employs plans for constructing and recognizing referring expressions and meta-plans for constructing and recognizing clarifications. This allows the model to account for the generation and understanding both of referring expressions and of their clarifications in a uniform framework using a single knowledge base.

I. Introduction

In the dialogue below[1], person A wants to refer to some object and have person B identify it. Person A does this by uttering a referring expression; however, A's expression fails to allow B to uniquely identify the object. Person B then tries to clarify A's referring expression by expanding it. A rejects B's clarification and replaces it, which allows B to identify the referent of the refashioned referring expression.

A: [1] See the weird creature
B: [2] In the corner?
A: [3] No, on the television
B: [4] Okay.

This paper presents a computation model of Clark and Wilkes-Gibbs's work on how conversational participants collaborate in forming referring expressions [2]. Our model takes the role of one of the participants, either the participant who initiates the referring expression, the *initiator*, or the one who is trying to identify the referent, the *responder*. It accounts for how the initiator constructs the initial referring expressions and how she and the responder then collaborate in clarifying the referring expression until it is acceptable. Each step of the collaboration consists of a clarification of the referring expression and a subsequent understanding of the clarification.

This work is based on the planning paradigm. The knowledge that is needed to choose the content of a referring expression is encoded in plans. This allows an agent to use the same knowledge base for both constructing and recognizing initial referring expressions. Furthermore, the knowledge needed to clarify a referring expression is encoded as plans. These are meta-plans that take an instantiated plan corresponding to a referring expression as a parameter. The meta-plans reason about the failed constraints or effects of the instantiated plan in order to clarify it. These repairs can subsequently be understood by performing plan recognition. This approach allows the entire collaborative process to be expressed in a uniform framework with a single knowledge base.

[1] This example is a simplified version of [6] S.2.4a (1-8).

II. Referring as Action

Plans encode a relationship between goals and the primitive actions that will accomplish these goals. Hence, a set of primitive actions is needed that is relevant in the domain of referring expressions [1]. We use the primitive actions **s-refer** and **s-attr**. **S-refer** is performed by the initiator to signal to the responder that she is referring to an object, and that she intends him to identify the object. **S-attr** ascribes some attribute to an object, for instance its category, color, or shape.

III. Initial Referring Expression

Constructing: When an initiator wants to refer to an object, she can do so by constructing a **refer** plan. This plan consists of two steps, the action **s-refer**, mentioned above, and the subplan **describe**. **Describe**, through its subplans **headnoun** and **modifiers**, constructs a description of the object that is intended to allow the responder to identify the object. **Headnoun** decomposes into an **s-attr** action that ascribes to the object the head noun chosen by the constraints of the plan. The **modifiers** plan is more complicated. Through its constraints, it ensures that the referring expression is believed to allow the responder to uniquely identify the object. The **modifiers** plan achieves this by decomposing into the **modifier** plan a variable number of times (through recursion). Each instance of the **modifier** plan constructs an individual component of the description, such as the object's color, shape, or location (through an **s-attr** action).

Recognizing: The responder, after hearing the initial referring expression, tries to recognize the intention behind the initiator's utterance. Starting with the set of primitive actions that he observed, the responder employs plan

recognition to determine a plan that accounts for them. This process will lead him to ascribe the `refer` plan to the initiator, including the intention for the responder to identify the referent of the description. Plan recognition, by analyzing the constraints and effects of the inferred plan, lets the responder attempt to identify the referent of the description.

There are two reasons why the responder might be unable to identify the referent. Either the responder is unable to find any objects that satisfy the referring expression or he is able to find more than one that satisfies it. This situation might arise if the initiator and responder have different states of knowledge or belief about the world. For instance, in the dialogue above the responder might think that several objects are "weird". The constraint or effect that was violated in the inferred plan is noted by the plan recognizer, and this knowledge is used to repair the plan. This approach is motivated by Pollack's treatment of ill-formed domain plans [5].

IV. Clarifications

Constructing: If the responder was unsuccessful at inferring the referent of the referring expression, he will plan to inform the initiator that her referring expression was not successful. As Clark and Wilkes-Gibbs [2] point out, the responder will try to refashion the referring expression in order to minimize the collaborative effort, and hence he will prefer to replace or expand the referring expression rather than just rejecting it or postponing the decision.

The responder has several different clarification plans [4] at his disposal and they take as a parameter the inferred plan corresponding to the referring expression. These plans correspond to Clark and Wilkes-Gibbs's analysis of the repair process. One of these plans is `reject-replace`. This plan rejects the step of the inferred referring expression plan that has a constraint violation and replaces it by a similar step but with the violated constraint relaxed (relaxing a description is due to [3]). A second plan is `postpone-expand`, which is used to further qualify a referring expression that a participant found to match several objects. This plan is used by the responder in (2) in the dialogue above.

Recognizing: If the responder clarifies the referring expression, the initiator will have to infer that the responder is unable to identify the referent of the expression. Furthermore, the initiator must determine how the clarification will affect the underlying referring expression. The responder might have rejected or postponed his decision, as well as proposed a correction to the underlying referring expression by replacing or expanding it. Following Litman's work on understanding clarification subdialogues [4], this process is achieved through plan recognition.

Continuing On: Clarification subdialogues might extend beyond the responder's clarification of the initial referring expression. For instance, in the above dialogue, after the initiator inferred the responder's clarification, she found the resulting referring expression plan ill-formed. Hence, she constructed a subsequent clarification—"No, on the television". Then, the responder had to infer this clarification. In general, this process will continue until both participants accept the referring expression. The analysis involved with these subsequent turns of the dialogue is similar to the analysis given in the preceding two subsections. There may be differences between how the initiator and responder clarify a referring expression, since the initiator knows the identity of the referent. Also, there may be differences between a clarification following the initial referring expression and one following another clarification, since, in the latter case, the referring expression may have already been partially accepted.

V. Belief Revision

As was mentioned earlier, the initiator and responder might have different states of knowledge or belief about the world, and these differences will be a cause of clarification subdialogues. In the process of collaborating to make referring expressions, these differences in belief will arise in the replacements and expansions that the two participants propose. Hence, they will need a way of resolving their differences in beliefs about the world if they are to both accept the referring expression. Hence the model proposed in this paper will need to incorporate belief revision.

VI. Conclusion

This paper has presented a computational model of how conversational participants collaborate in making referring expressions. However, it is hoped that the ideas presented in this paper are of relevance to a much larger range of collaborative processes.

The work outlined in this paper is in progress. At present, a computer system has been implemented in Prolog that can construct and recognize initial referring expressions, and that can construct clarifications. In terms of the dialogue above, the system can model both the initiator and responder for the first line and can model the responder for the second.

References

[1] D. E. Appelt. Planning English referring expressions. *Artificial Intelligence*, 26(1):1–33, April 1985.

[2] H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[3] B. A. Goodman. Repairing reference identification failures by relaxation. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pages 204–217, 1985.

[4] D. J. Litman and J. F. Allen. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11(2):163–200, April–June 1987.

[5] M. E. Pollack. Inferring domain plans in question-answering. Technical Note 403, SRI International, 1986.

[6] J. Svartvik and R. Quirk. *A Corpus of English Conversation*. Lund Studies in English. 56. C.W.K. Gleerup, Lund, 1980.