



IMPROVING DISCOURSE MANAGEMENT IN TRIPS-98

Donna K. Byron

University of Rochester Department of Computer Science
P.O. Box 270226, Rochester NY 14627, U.S.A.
dbyron@cs.rochester.edu

Abstract

The discourse manager is the component of the TRIPS-98 system that maintains a list of candidate antecedents to be used for resolving anaphoric referring expressions. The model currently implemented in the system is a monologue-based model of context that has a variety of limitations. This paper discusses changes to the underlying model to make it handle a much broader range of referring expressions. Some of these changes have already been implemented in the system and some are the subject of a long-term project.

INTRODUCTION

The TRIPS98 system, developed at the University of Rochester, is a collaborative planning assistance that accepts spoken input from the user and generates multi-modal responses [9]. As in any language understanding system, assigning the correct referent to anaphoric referring expressions is a vital task of the system. An important component of this process is a data structure called the Discourse Context (DC), which contains a representation of the common ground of the discourse participants [7]. Items that have been mentioned in the past and are available for anaphoric reference are listed in the DC in order of salience.

The DC is intended to contain data elements for all entities that can be the subject of anaphoric reference. But the implementation has been found to have a variety of problems that result in the system being unable to resolve specific types of anaphoric reference. This paper describes an ongoing project to improve the range of referring expressions handled by the system by improving the discourse context management component. The set of issues that present themselves to a dialog understanding system are

This material is based on work supported by USAF/Rome Labs contract F30602-95-1-0025, ONR grant N00014-95-1-1088, and Columbia University grant OPG:1307. Any opinions, findings, conclusions and recommendations expressed in this material are those of the author and do not necessarily reflect the views of the granting institutions.

somewhat different than those encountered in text understanding systems. In this paper we present several types of referring behavior that are problematic for our system and that do not arise in text understanding systems. We hope this discussion proves to be of interest to other system builders who face similar problems in their own dialog understanding systems.

EXAMPLES

This section contains sample interactions between a user (the author) and the TRIPS-98 system that demonstrate some of the referential behavior that the current model cannot handle. The user's utterances are preceded by (U) and the system's by (S). The system's verbal response is shown first, then GUI modifications. The user's graphical display initially consists of a map with cities and roads. In the sample dialogs, Abyss, Barnacle, and Exodus are cities on the map. Other displays are added to the user's screen as needed to support the collaboration. For example, when a plan is under consideration, a time-line plan viewer pops up to show the time required for plan segments and for the overall plan.

Reference to Display Properties

Example 1 shows a sample dialog in which the user of the system attempted to refer to a route based on its color. Because the system color-matches actions on the plan viewer with the map objects used in that plan, it is natural for the user to notice the information content behind the color choice and use it as an efficient way to refer to the plan or plan objects.

Example 1: Reference to display property

-
- 1) U: Where are the trucks?
 - 2) S: There are two trucks at Delta.
(system displays yellow truck icons on the map at Delta.)
 - 3) U: Send a truck to Abyss.
 - 4) S: Done.
(the route used to go to Abyss is shown in orange on the map, and an orange truck icon is shown at Abyss. The plan viewer displays an orange time-line representing the planned journey to Abyss).
 - 5) U: How long does the orange route take?
 - 6) S: **I don't understand your reference to the orange route.**
-

Answering Questions

The old discourse context model only included entities mentioned by the user. That model might be appropriate for a system to understand text or monologue (where only one speaker's prior speech needs to be modeled). In fact, it was adequate in the previous application domain (TRAINS-97). But in the TRIPS-98 domain, there are many more entities in the world, and reference has become slightly more complicated. The system now needs to know about items mentioned in its own generated turns in order to support anaphoric reference by the user. Especially problematic are entities mentioned as answers to the user's questions. Entities in answers are highly focused, and the user is likely to refer to them with an anaphoric referring expression in the next utterance. Example 2 is typical.

Example 2: Reference after a question

- 1) U: Are there any people
- 2) S: People 4, people 3 and people 1
- 3) U: Where are they?
- 4) S: I don't understand what you are referring to**

Example 3 is a more egregious error. The system's use of *it* in utterance 4 gives a false impression that it knows of an entity that can be referred to with a pronoun, namely the action of traveling to Exodus. The pronominal reference in subject position explicitly thrusts the action into focus, and the user naturally refers to it with a pronoun in his following utterance. The system's rejection of that utterance is bizarre and confusing.

Example 3: Reference after a question

- 1) U: Send a truck to Exodus
- 2) S: Done
- 3) U: How long does that take
- 4) S: It will take 5 hours and 12 minutes
- 5) U: How long would it take to use the chopper instead
- 6) S: I don't understand your reference to 'it'**

The status of entities mentioned within the question itself (i.e. whether they can be the subject of anaphoric reference) is an open research problem. It seems that different categories of questions must be treated differently, and the existential presupposition carried by the matrix verb and phrasing of the sentence is important [3]. However, after the question is answered, entities in the answer must definitely be included in the DC.

Referring to Abstract Entities

Example 3 hints at another problem with the current model: the system has a problem resolving reference to the plan that the user is constructing. The reference resolution portion does contain a hack that allows use of the pronoun *that* to refer to an action mentioned in the previous utterance, but slight changes in the phrasing or reference with *it* instead of *that* cannot be resolved. This is just one example of an entire class of entities that is missing from the DC. Entities that were not mentioned as noun phrases, but that are evoked into the common ground by virtue of being mentioned as events, routes, facts, etc., are

not currently captured as candidates for anaphoric reference. These *abstract entities* must be added to the DC in a separate process.

Building Sets for Plural Pronouns

When the user refers to a plural entity, the system looks in the DC for entities it can coalesce into a set. Example 4 shows a confusing failure of this process. The referring expression *them* in utterance 5 caused the system to create a SET token containing all the people and the truck, but the user clearly had only the people in mind as the referent. When the system tries to process utterance 7, the semantics of the verb *unload* require its direct object to be something that can be carried around in a vehicle. Because the set token contains the truck as well as the people, it cannot be used to resolve this pronoun.

Example 4: Plural Pronouns

- 1) U: Send a truck to Barnacle
- 2) S: Alright
- 3) U: Load the people into the truck
- 4) S: Ok
- 5) U: Take them to Abyss
- 6) S: Alright
- 7) U: Unload them
- 8) S: I don't understand what you're referring to**

Linguistic studies, as well as a corpus analysis of the TRAINS93 dialogs, have shown that there is a pragmatic contrast between the use of definite (eg. *it/them*) and demonstrative (eg. *these/that*) pronouns as they are used to refer to sets, and also that demonstrative pronouns are much more common in spoken dialog than in text [6, 4]. A demonstrative pronoun is preferred for sets composed of heterogeneous entities - entities of mixed semantic type (eg. a vehicle and its cargo). The user's choice of a definite pronoun in utterance 5 should cause the system to build a set with homogeneous semantic properties, in this case just the people.

PREVIOUS MODEL

This section describes the current model of discourse context implemented in the system. The discourse context data structure contains several different categories for previously mentioned entities. These categories are used to support resolution of different referring expression surface forms. The categories of interest in this paper are:

- **focus** - the most salient object
- **location** - focus location, used to resolve *there*
- **time** - focus time, used to resolve *now*
- **relevant-objects** - initially, all objects in the (plan) world
- **mentioned-objects** - salient objects, actually objects mentioned in user's last utterance
- **last-action** - last plan action
- **objects-in-last-action** - objects in the last action
- **plan-objects** - all objects in the entire plan

After the user's utterance is parsed, entities he mentioned are added to the focus and mentioned-objects. Items that were previously in mentioned-objects are faded back to relevant-objects. If the utterance causes the addition of a new action to the plan, it replaces the previous contents of last-action. These categories are used to resolve different sorts of referring expressions. The following parts of the context are searched for each referring expression type:

- **Definite Descriptions** - 1) mentioned-objects 2) objects in last action 3) plan objects 4) relevant objects
- **Definite Pronouns** - 1) Focus 2) treat like definite description
- **Demonstrative Pronouns** - *That* is always bound to last-action
- **Names** - 1) relevant objects
- **Temporal anaphors** - focused time
- **Locative anaphors** - focused location

The system contains a set of discourse manager rules that encode how to increment the discourse context based on the user's speech act. These updates only include entities mentioned by the user, not entities mentioned or displayed by the system. Some examples are:

| User Speech Act | Context Effect |
|-----------------------------------|-----------------------|
| Accept a proposed option | update context |
| Ask a WH- or Yes/No question | none |
| Command to undo plan step or plan | back to prior context |
| Reject a proposed option | none |
| Answer a clarification request | none |
| Command to start over | Re-initialize |

The system maintains internal symbol names for all entities, so the process of reference resolution requires matching the surface form used by the speaker to a set of internal symbols.

ISSUES

Because this is a spoken dialog system, phenomena specific to dialog, such as turn-taking behavior, acknowledgments, overlapping speech, etc., can be incorporated into the model. Previous authors have observed that acknowledgments in spoken dialog impact the referring process [8]. In Eckert and Strube's model, entities are not added to the DC until the addressee acknowledges the utterance, thereby explicitly entering the entity into common ground. In our examination of human/human dialogs, this model is preferred over one in which entities are added to the DC as soon as they are mentioned [11]. Unfortunately, users of the TRIPS-98 system do not use as many acknowledgments when speaking to the system compared to the frequent use of acknowledgments when speaking to another human. So we were unable to factor acknowledgments into the model. In the future, when the language handled by the system becomes more natural, this point should be re-visited.

Another question appropriate to dialog systems is whether both participants might be maintaining separate

context structures. Is the user's anaphoric reference to entities mentioned by the system different than his anaphoric reference to entities previously mentioned by himself? A previous study of social dialog found this not to be the case [2]; therefore, the system does not model which speaker's utterance evoked an entity into the DC.

Our model also makes no assumptions about the amount of intervening talk that can occur between one mention of an entity and another reference to the same entity. All previously mentioned objects fade backwards in the salience ranking and remain in the relevant objects list for the duration of the dialog.

NEW MODEL

The new model includes many additional types of entities in the DC. New interface modules have been created to give the discourse manager access to additional classes of information from the surrounding system. The DC categories, such as focus, mentioned-objects, etc., were found to be sufficient and have not been changed. The model has been only partially implemented at this point.

Referring to Display Properties

A new component has been added to the realization manager of TRIPS-98. Previously, it was responsible for deciding what objects to add to the display but not their display properties. A separate, encapsulated component within the Java graphics builder determined display properties such as color. The realization component was modified so that it now chooses display properties. The command sent to the JAVA front end now includes what object to build and also its display properties. An interface module was added to the realization component to handle queries about display properties, for example "Which trucks are orange?". This interface is called by the reference resolution modules whenever the user mentions a display property in a definite description.

Answers to Questions

Questions typically contain a definite description portion and an aspectual portion. For example, to interpret the question "Are there any trains at Abyss" the system must resolve the referring expression "trains at Abyss" and then calculate that the aspect under question is their existence. So the definite description portion of the user's utterance can be used to evoke entities into the DC. To reprise Example 2, the system resolves the user's reference to "any people" to people1, people3, and people4. Tokens for those entities should be added to the DC. To support this, many of the question speech-act rules for context effects will be changed (this portion of the implementation is ongoing). Some examples are shown in Table 1.

As a long-term strategy, entities mentioned in all content-bearing utterances by the system must be added to the DC. At this point, the system has very little initiative of its own, so it never refers to objects unless the user asks a question about them. In the future, as the system's ability to initiate plan actions is extended, an interface will

| User Speech Act | Context Effect |
|-------------------------------------|-------------------------------------|
| Ask a WH- or Yes/No question | update |
| Yes/No Query affirmative answer | update |
| Yes/No Query negative answer | none |
| Reject a proposed option | Remove action and action objects |
| Answer a clarification request | update |
| Command to use a new option instead | replace |

Table 1: New Speech Act Rules

be needed between the realization component and the discourse manager. Objects that are added to or highlighted on the display, or that are mentioned in an utterance, must be added to the DC. This is a point of future work.

Referring to Abstract Entities

This is the most ambitious portion of our remodeling efforts. The DC currently contains entities that were mentioned by the user as noun phrases (eg. trains, people, cities). Typical models of anaphoric reference in the computational linguistics literature include only these entities [1, 10]. However, propositions, properties of entities, events, etc. can also be the subject of anaphoric reference even though their corresponding surface constituents are not noun phrases. An extensive analysis of TRAINS93 dialogs reveals around fifteen categories of such entities that must be added to the DC [5]. To accomplish this will require modification to many components of the parser, including grammar rules, reference resolution, and the interface between syntactic analysis in the parser and the discourse context manager. These modifications are the subject of a long term project.

Building Sets for Plural Pronouns

The system must treat definite and demonstrative pronouns differently. In the new model, definite pronouns can only refer to semantically homogeneous sets, and demonstrative pronouns can only refer to semantically heterogeneous sets. An additional change is that sets are added to the DC during interpretation of each sentence rather than upon encountering the pronoun. Sets are built whenever multiple objects of the same semantic type are present in a sentence, for example, a complex utterance such as “pick up the people at Abyss and the people at Exodus with truck-1 and truck-2, respectively” triggers the creation of several sets. Not only sets for {people at abyss, people at exodus} and {truck1, truck2} which were mentioned in conjunctions, but also {abyss, exodus}, which were not. Heterogeneous objects that are syntactically conjoined also make sets, such as “send the helicopter and the other truck to Abyss”. Plan actions which result in physical connection of semantically heterogeneous objects, such as loading people into a truck, also create set entities. These heterogeneous sets can be the referent of demonstrative pronouns.

FUTURE WORK

The work reported in this paper represents the start of a longer-term project to make the reference resolution component of TRIPS more robust. Besides the categories of

reference failure mentioned here, several additional issues need to be addressed to make the system handle referring expressions in a more natural way. They include:

- The model must be extended to add demonstrative reference to GUI objects (eg. “that route” after the system highlights a route).
- The model must be extended to contain tokens for referents at all communication levels, including information about the surface form. This information is needed to support surface anaphora such as “the former...the latter” and also to interpret some descriptive reference. For example, users commonly refer to the surface form of a previous turn when it seems that the system has lost track of the task. For example, “Remember awhile ago when I asked about sending the helicopter to Abyss?” If tokenized representations of the surface form of utterances were kept in the DC, reference to surface form properties could be handled.

REFERENCES

- [1] Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (ACL '87)*, pages 155–162, 1987.
- [2] D. Byron and A. Stent. A preliminary model of centering in dialog. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL '98)*, 1998.
- [3] D. Byron and A. Stent. Referential properties of questions. In *ACL '99 Workshop: The Relationship Between Discourse/Dialog Structure and Reference*, 1999.
- [4] Donna K. Byron. Analysis of pronominal reference in two spoken language collections: Trains93 spontaneous task-oriented dialog and Boston University radio news prepared monologue. Technical Report 703, Department of Computer Science, University of Rochester, March 1999.
- [5] Donna K. Byron. Resolving pronominal reference to abstract entities: Thesis proposal. Technical Report 714, Department of Computer Science, University of Rochester, 1999.
- [6] Robert Channon. Anaphoric *that*: A friend in need. In Jody Kreiman and Almerindo E. Ojeda, editors, *Papers from the Parasession on Pronouns and Anaphora*, pages 98–109. Chicago Linguistic Society, The University of Chicago Classics, 1980.
- [7] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie Lynn Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 10–62. Cambridge University Press, 1981.
- [8] Miriam Eckert and Michael Strube. Dialogue acts, synchronising units and anaphora resolution. In *Amstelveen '99*, 1999.
- [9] George Ferguson and James F. Allen. Trips: An intelligent integrated problem-solving assistant. In *Proceedings of AAAI '98*, 1998.
- [10] Ruslan Mitkov. Robust pronoun resolution with limited knowledge. In *Proceedings of ACL '98*, pages 869–875, 1998.
- [11] Michael Strube. Personal communication, December, 1998.