

DISCOURSE STRUCTURE IN THE TRAINS PROJECT

James F. Allen

Department of Computer Science
University of Rochester
Rochester, NY 14627

ABSTRACT

In a natural dialog, a considerable proportion of the utterances actually relate to the maintenance of the dialog itself rather than to furthering the task or goals motivating the conversation. For example, many utterances serve to acknowledge, clarify, correct a previous utterance rather than pursue some goal in the domain. In addition, natural dialog is full of false starts, ungrammatical sentences and other complexities not found in written language. This paper describes our recent efforts to define and construct a model of discourse interaction that handle dialogs that are rich in these natural dialog-related phenomena.

INTRODUCTION

The TRAINS project involves building an intelligent planning assistant that is conversationally-proficient in natural language. The name of the project comes from the domain used to test and demonstrate the ideas: the system acts as an intelligent assistant to a person attempting to solve transportation problems involving freight trains and factories in a simulated world. The system assists in formulating plans, and monitors these plans as they are executed by the simulated agents in a simulated TRAINS world, providing updates and support to the human in replanning as necessary. The human should be able to communicate using unconstrained natural, spoken language.

We have started to collect dialogs using a wizard scenario in the TRAINS domain and have built an initial prototype system. The current system uses keyboard input derived from the transcripts of actual spoken dialogs. The dialogs exhibit complex behavior as both the human and system take initiative at times in the dialog, and there are a large number of clarifications, corrections and acknowledgements.

This short paper describes both our empirical work in analyzing the transcripts, and our theoretical work in defining a computational discourse model.

THE DATA

We decided to collect our own data rather than using an existing source, such as the ATIS corpus, for several reasons. First, the dialogs in ATIS are structured to emphasize question-answering rather than interactive problem solving. More importantly, the mixed modality interaction of the ATIS scenario inhibits most natural spoken dialog phenomena. In particular, the long pauses before responses and the table-based system output

prevent natural follow-up, such as acknowledgements, clarifications and confirmations that are common in spoken dialog. Almost 50% of the speech collected in our more natural setting was of these types.

The TRAINS domain was carefully designed so that a significant part of it is within reach of current (or near future) capabilities of plan reasoning systems. Because of this, we should be able to fully specify and implement the reasoning underlying the "system" in the dialogs. If ATIS were extended to be a travel-planner rather than a database, the domains would be comparable.

We have collected an initial corpus of natural spoken conversations between two people engaged in complex problem solving in the TRAINS world. One person (the "system") has most of the information and detail about the domain, but the other has the problem to solve. The two are in different rooms and so have no visual contact, but they both have the same map from which to work. A fragment from one of the dialogs shown in Figure 1. Each utterance is roughly classified as to its function: whether it is primarily concerned with making progress on solving the problem (plain text), or whether it is primarily concerned with maintaining the conversation itself (in bold). The agents are labelled <H> (for human) and <S> (for system), even though the system here was simulated by a person. Comments on the possible discourse function of the utterances concerned with maintaining the conversation are presented in italics.

As can be seen, approximately half of the utterances are concerned with maintaining the communication process. There are utterances that identify the goals of the next stretch of discourse, and a large number of utterances that pertain to acknowledging the other agents utterances and in maintaining a smooth flow of control (i.e. identifying whose turn it is to speak). It has been our claim for some time that this level of discourse interaction must be explicitly modelled if we are to build systems that can converse in natural language, and in previous papers we have described a plan-based model that accounted for clarification subdialogs among other things (Litman & Allen, 1990, Litman & Allen, 1987). We are now attempting to develop an extended model that can account for all the discourse-level interactions found in the corpus.

The project is pursuing two main thrusts. First, we are developing a database for studying discourse phenomena. To do this, we are developing a taxonomy of discourse-level acts with which different people can independently classify each utterance reliably. Using this classification,

<H> ok, now uhh, let me, let me check on the uhh
 <H> where the.. where the engines are and the.. the boxcars are uhh
setting the immediate conversation goals for the following dialog fragment

<H> I'm assuming,
indicating that <H> is asking for confirmation

<H> let's see, that uhh
<H> is holding the turn while he examines the map

<H> I have two engines to work with, engine E2 which is at city D
 <H> and engine E3 which is at city A
 <S> aah, yes.
the "aah" probably indicates that <S> is thinking about the answer (and acknowledging that the question was understood)

<H> and uh, I've got two tankers, tanker t1 is at city A,
 <H> and tanker t2 is at city B
 <S> that's right. hnn, hnn.
 <H> ok
<H> indicates that he has accepted <S>'s reply

<S> there're.. there're other tankers as well.
 <H> ok
<H> acknowledges <S>'s introduction of new information

<S> there're actually four tankers at city E
"actually" indicates that <S> believes <H> doesn't know about these tankers

<H> four tankers at city E, ok
<H> acknowledges hearing the new information, and then accepts it

<H> uhh so, tankers t3, t4, t5, and t6 are all at city E.
<H> confirms his understanding of <S>'s assertions

<S> that's right
<S> confirms <H>'s confirmation

<H> ok. and just uh
<H> acknowledges the previous exchange and signals a move to a new topic

<H> I have four boxcars, b6 at city H, b5 at city F,
 <H> b7 at city B, and b8 at city I.

we are building a database of dialogs with each utterance annotated by its discourse function. In addition, we are analyzing the tapes and extracting prosodic information (primarily pitch contours, speech rate) and adding this information to the database as well. We have started some preliminary studies on prosodic cues to the discourse acts in our taxonomy, but need to analyze additional data before we have significant results. Second, we are developing a system that implements the discourse model together with full natural language processing and plan reasoning in the domain. In this paper, I will mainly describe the problems we are facing and the initial taxonomy developed so far. At the end, I will briefly describe the discourse model in the current implementation.

THE TAXONOMY

Rather than analyze the dialogs in terms of abstract discourse relations, our taxonomy is based entirely on the intentions of the speaker. This allows us to integrate well with previously developed computational speech act models, and provides a slightly different view from the other approaches. It is important to remember that just because a speaker intended an utterance in a certain way, it doesn't mean that the hearer understands it that way. Establishing agreement between the speaker and hearer as to what was intended is the primary reason for acknowledgements, clarifications and corrections. In addition, even if an utterance is understood correctly, this

doesn't commit the hearer to accepting the intended consequences of the act (e.g. believing the speaker's assertion, or performing the requested act). Acceptance involves yet additional mechanisms to acknowledgment.

As we define the set of speech act types, it is important to realize that nearly every speech act can be used at different levels of the conversation: they can involve the plan in the TRAINS world (the **domain level**), or the problem solving process that the two agents are engaged in (the **problem solving level**), or the understanding and managing of the conversation itself (the **discourse level**). We will try to give examples of the acts at each level as they are defined. Because of the focus on the discourse-level acts in this paper, we will often distinguish these as separately named acts.

The speech acts themselves break into three major classes: the **understanding** acts, which include acknowledgements and confirmations, the **information** acts, which involve imparting information and include informs, elaborations, clarifications, corrections and summarizations, and the **co-ordination** acts, which involve co-ordinating the activities of the two agents and include requests, suggestions, acceptances and so on. Throughout we will refer to the agent performing the speech act as the *speaker* and the other agent as the *other agent*.

There is not the space to precisely define each act, but I would like to present the entire taxonomy. To do this, some of the acts will simply be presented by an example.

THE UNDERSTANDING ACTS

The understanding acts specifically relate to indicating the successful hearing of the other agent's utterances.

Acknowledgment (Ack)

An acknowledgment indicates that the speaker has understood the other agent's previous utterance, but does not necessarily commit the speaker to agreeing with the other agent. An acknowledgement that is not an acceptance of the other agent's request is shown in italics:

<H> unload b8 processing orange juice and load
t2
<S> *OK*, ah but tanker t2 is currently full of beer

Confirmation (Conf)

A confirmation act is a special form of acknowledgment that involves restating or paraphrasing information established previously in the conversation. If there is any doubt implied in the utterance, say by using a question intonation, then the utterance is a clarification request rather than a confirmation.

<H> Can you have city I fill B8 with oranges
please?
<S> *OK. We're gonna fill b8 with oranges at city
I*

Completion (Compl)

A completion occurs when the speaker completes the other agent's utterance rather than waiting for the agent to finish.

<S> which should leave us plenty of time to uhhd
<H> *get to city H*
<S> city H after that

KeepTurn

This is a wide-ranging class and includes any utterances whose main purpose is to maintain the speakers turn, although they may also serve as an acknowledgement.

<S> where it will then pick up the orange juice
<S> *and uhhd ...*
<S> and then take that to city G

THE INFORMATION ACTS

Information acts involve making claims about the state of the world. The prototypical speech act in this class in the speech act literature in the inform act. We will break down informs at the discourse level into clarifications, corrections, elaborations and summarizations.

Inform (Inf)

An inform act in the TRAINS domain is generally either in response to a question, or is a situation setting action that describes background information necessary to understand the problem. Inform is the default assignment for acts in this class if none of the following acts seem appropriate.

<H> Where's e3?
<S> *e3 is just coming in to city A.*

Clarifications (Clr)

A clarification is an utterance that provides additional information to help the interpretation of the previous utterance. Utterances that provide information which is not necessary to understand the previous utterance are not clarifications, but rather elaborations. Examples are

<H> great, have them unload b6
<H> *have D unload b6*
and

<H> Let's do that
<H> *Let's move E2 to city E*

Intent Clarification (Tag)

An intent clarification utterance clarifies the intention of a previous utterance. The name Tag is assigned as this is the role that is played by tags in sentences such as *John is coming to the party, isn't he?*. The tag indicates that the utterance is a question rather than an assertion. A tag can be deleted without affecting the dialog (if the previous utterance is treated appropriately as indicated by the tag)

<H> It is 2PM
<H> *Is that right?*

Corrections (Cor)

A correction is a special form of clarification that replaces some earlier information with the new information. Corrections often follow utterances that signal some problem, such as *No*, or *opps*, and so on. Corrections also can appear mid-way through an utterance when the speaker needs to make a correction of something uttered earlier in the sentence.

<S> e3 is on its way to t1
<S> *opps with tanker t1*
<S> full of orange juice

Elaboration (Elab)

An elaboration is an inform that further develops a previous topic. The information is *not* needed in order to understand the previous sentence (in which case it would be a clarification).

<S> The quickest route would be to go through
city C
<H> OK
<S> *uhh that should take six hours*

Summary (Sum)

A summary act is an inform that restates what has been asserted or decided upon in the previous utterances, or draws conclusions from what was previously asserted.

<S> uh we can actually have the orange juice
made by uhh twelve pm tonight

<S> *so there should be plenty.. plenty of time*

THE CO-ORDINATION ACTS

These acts involve the two agents co-ordinating their activities by making requests and suggestions and reaching agreement after negotiation. As mentioned above, this co-ordination can occur at the three different levels of conversation. As before, the acts at the discourse level will be given special treatment as subclasses of the general cases.

Request (Req)

A request involves one agent attempting to get the other agent to do something by direct means. If a request is not taken up, it must be explicitly denied by the hearer either by stating that he won't comply or by suggesting a modification to the requested action. The requested action may be either a domain act, as in:

<H> *Can you have city I fill B6 with oranges,
please*

or a problem solving act as in

<H> *Let me know when E3 has B6 loaded.*

Particular subclasses of requests involving questions are treated individually as they have their own specific syntactic markers in language.

Wh Question (WHQ)

Wh-questions are true question where the speaker is actually asking for information about a specific entity from the hearer. An example at the domain level is:

<H> *How much does it cost to dunk it on the
ground?*

and at the problem solving level is

<H> *What should we do?*

Yes-No Question (YNQ)

These are true yes/no questions, where the speaker would be content with a simple yes or no answer. If additional information does seem to be required, then the original question was probably an indirect request or WHQ. An example at the domain level is:

<H> *Is e3 at city I?*

and at the problem solving level is

<H> *are you uhh trying to compute the time to
take E2 with T3 and T4*

Requests and questions at the discourse level are typically clarification requests, which are marked in their own category below.

Clarification Request (reqClr)

A clarification request is a request for information to help interpret some previous utterance(s), i.e. a request for a clarification. In the following example, the clarification request is in bold italics, and the ensuring clarification in italics:

<S> *To city I?*

<H> *yes*

Here's a clarification request (bold italics) that was answered with a correction (italics):

<S> I just found city E2

<H> *city E2?*

<S> uhh .. *engine E2*

Suggest (Sug)

A Suggestion in this domain also involves getting the other agent to do something, but is weaker than a request. Suggestions explicitly leave open an option of negotiation between the agents, often by using the first person plural to include both agents in the suggested action. An example at the domain level is:

<<S> *Why don't we begin loading oranges in
boxcar B6*

and at the problem solving level:

<S> *Shall we look at the other engine?*

and at the discourse level:

<H> *Well, let's talk about orange juice*

Correction Suggestion (sugCor)

Other suggestions at the discourse level may be correction suggestions. In the example, the correction suggestion is in bold italics, and the acceptance (i.e. a correction) is in italics.

<S> second engine E3 is going to uhh city H to
pick up the bananas

<S> back to A, dro.....

<H> *H to pick up the oranges*

<S> sorry, *pick up the oranges*

<S> back to A to drop the oranges off

Accept (Acc)

An accept indicates that the hearer has accepted the act in the previous utterance, be it a request, suggest, inform of whatever. After an agent has done an accept, they are committed to whatever the speech act that was accepted requires. Accepts can also be implicit if the agent

continues on without explicit denial. Examples often overlap with acknowledgments. Here is a suggestion at the domain level that is accepted:

<H> and in the mean time, it would be nice if city
H could be filling B6 with oranges
<S> OK, *it looks like we can do that*

Denial or Rejection (Den)

A Denial is the opposite of an acceptance. As with accept acts, one can deny requests, suggestions or many other acts. There are not many denials in the current dialogs as the conversants are quite co-operative!. But they do occur occasionally. Here's an acknowledgement of a request followed by a denial.

<H> have city B prepare for its arrival. it should
unload b8 processing orange juice and load
t2.
<S> OK, *ah but tanker t2 is currently full of beer.*

Evaluative Statement (Eval)

An evaluative statement describes the reaction of the speaker to the current situation. Such statements serve a confirmation or denial role, or express more subtle shades in between.

Typical Phrases: *great!, terrific!, yuk!, how nice!*

<S> looks like we can do that
<H> *Terrific*

THE CURRENT SYSTEM

Eventually, we intend to develop a model that defines each of the above discourse acts in terms of the changes that the act makes to the shared and individual beliefs and goals of the two participants in the dialog. The current system, however, is quite simple and was constructed mainly to define the overall architecture of the system. The current discourse model has the following basic capabilities

- maintaining knowledge of the turn taking (i.e. whose responsibility is it to speak next);
- tracking the status of each fragment of the plan as it is suggested and discussed;
- tracking and responding to simple discourse obligations (e.g. answering questions)..

The discourse module uses the domain plan reasoner, which uses planning and plan recognition techniques, to maintain the domain plans. It calls the domain reasoner to verify hypotheses about the discourse function of the utterances, and to update the state of the plan as needed. Plan fragments in the knowledge base are characterized by six modalities that are used to indicate the status of parts of the plans being discussed. These are organized hierarchically with inheritance so that we can examine the

full plan from either human's of the system's perspective as shown in Figure 2.

The modalities include:

- the plan fragment suggested by the human but not yet acknowledged by the system (Human-Proposed-Plan-Private);
- the plan fragment suggested by the system and not yet acknowledged by the human (System-Proposed-Plan-Private);
- the plan fragment suggested by the human and acknowledged but not yet accepted by the system (Human-Proposed-Plan);
- the plan fragment suggested by the system and acknowledged but not yet accepted by the human (System-Proposed-Plan);
- the plan fragment that is shared between the two (i.e. accepted by both) (Shared-Plan); and
- the plan fragment constructed by the system but not yet suggested (System-Private-Plan).

Each context is associated with a particular form of plan reasoning as indicated in the figure. In particular, the plan in the System-Private-Plan context is extended by plan construction (essentially classical planning), where the plans in all the other contexts are extended by plan recognition relative to the appropriate set of beliefs.

Figure 2 also shows how plan fragments may move between the various contexts. A suggestion from the human enters a new plan fragment into the Human-Proposed-Plan-Private context and initiates plan recognition with respect to what the system believes about the human's private beliefs. Once acknowledged, this suggestion becomes "public" (i.e. it is in Human-Proposed-Plan). An acceptance from the system would then move that plan fragment into the Shared-Plan context, again invoking plan recognition.

Planning by the system results in new actions in the System-Private-Plan context. To make these actions part of the Shared-Plan context, the system must suggest the actions and then depend on the human to acknowledge and accept them. This model, while still crude by philosophical standards, is rich enough to model a wide range of the discourse acts involving clarification, acknowledgment and the suggest/accept speech act cycle ever-present in dialogs in this setting.

Because of the inheritance through the spaces, when the system is planning in the System-Private-Plan context, it sees a plan consisting of all the shared goals and actions, what it has already suggested, and all the new actions it has introduced into the plan privately but not yet suggested.

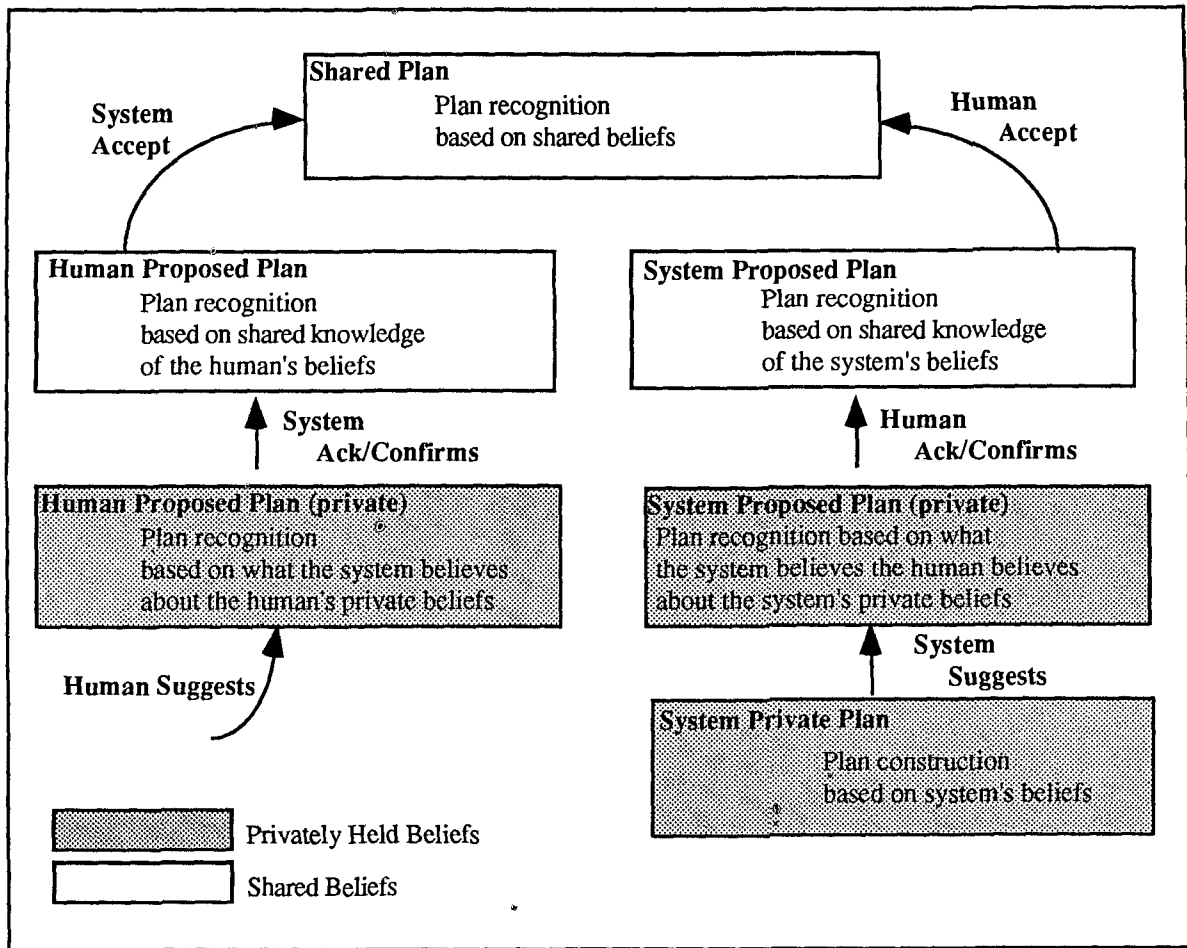


Figure 2: The different plan modalities from the system's perspective

Consider an example. Assume that the Shared-Plan context contains a plan to move some oranges to a factory at B, but there is no specification of the engine to be used. The system might plan to use engine E3. At this stage, the plan from the System-Private-Plan context involves E3. The plan in the System-Proposed context, however, is still the same as the plan in the Shared-Plan context, which still does not identify which engine to use. When the system makes the suggestion, the plan fragment involving E3 is added to the system proposed plan (private). An acknowledgment from the human results in this plan fragment being added to the system-proposed plan known to both agents. If the human then accepts this, it then becomes part of the shared plan. If the human rejects the suggestion, then E3 does not become part of the shared plan (at least, not without further discussion).

The prototype system can handle simple examples along these lines where the two agents are free to accept or reject suggestions as they are made in the dialog. The system under development will extend the current one to support some forms of negotiation between the agents in order to arrive at a mutually agreeable plan.

Acknowledgements

This work has been done in conjunction with Shin'ya Nakajima and David Traum. It was supported in part by ONR/DARPA contract number N00014-82-K-0193.

References

- Allen, J.F. & Perrault, C.R. Analyzing intention in utterances, *Artificial Intelligence* 15, 1980
- Litman, D.J. & Allen, J.F. A plan recognition model for subdialogues in conversation, *Cognitive Science* 11, 1987
- Litman, D.J. & Allen, J.F. Discourse processing and commonsense plans, in *Intentions in Communication*, P. Cohen, J. Morgan and M. Pollack (eds), MIT Press, 1990.