Speech Actions and Mental States in Task-Oriented Dialogues*

Peter Heeman
Department of Computer Science
University of Rochester
Rochester, New York, 14627
heeman@cs.rochester.edu

Abstract

In this paper, we propose a set of speech actions to account for how conversants collaborate in building a domain plan. The actions that we propose are indexical on the *current plan*, which is part of the mental states of the conversants. The current plan not only includes the actions that have been proposed to solve the problem, but also beliefs about the plans validity and how each action contributes to the overall goal of the plan. The intended effects of the actions are that the current plan is updated.

Introduction

In a task oriented dialogue, conversants are jointly building a domain plan to accomplish the task at hand. They do this through their speech actions. Some of the actions deal with exchanging information about the state of the world or about the plan that is being built. But others are intended to change the plan. For instance, a speaker can suggest that a certain action be added to the plan, can express a judgment about the validity of the plan, and can propose a refashioning of the plan. To understand these dialogues, we need to formalize such speech actions and their intended effects.

Our earlier work (Heeman, 1991; Heeman and Hirst, 1992) focused on how agents collaborate in making referring expressions. We are now extending that work to deal with how agents collaborate in building a domain plan. In this paper, we focus on giving definitions for the speech actions that are used. These actions presuppose that the mental state of the conversants includes a *current plan*, which is composed of the actions to be performed to achieve the goal, how each action contributes to the goal (as in Pollack's EPLAN (Pollack, 1990)), and beliefs about the plan's validity. The speech actions that we propose are indexical on the current plan, and their intended effects are to update the current plan. This view differs from other approaches, such as Sidner (1992), in which judgments and refashionings are in terms of the previous utterance, rather than of the current plan or a part of it.

The research described in this paper is based on the speech act interpretation work being done as part of the TRAINS project at the University of Rochester (Allen and Schubert, 1991). The speech act interpretation comes after the syntactic, semantic, and definite description interpretation (Poesio, 1993), and it uses surface features of the utterance to determine possible meanings of the utterance. The resulting acts are then passed to the speech act pruner, which decides which ones make sense given the context. The output of the pruner is then fed to the discourse reasoner (Traum, 1993), which interacts with the plan reasoner in order to take the appropriate actions.

In the rest of the paper, we first look at a sample dialogue to illustrate the speech actions that conversants use when collaborating. Second, we give an overview of our plan-based model of collaboration. Third, we give definitions for the speech actions that we use. Fourth, we return to the sample dialogue and give the speech act interpretations for each utterance and sketch how these interpretations are used by the discourse reasoner.

An Example

Consider the following constructed dialogue (Allen and Schubert, 1991) about the Trains-World scenerio given in Figure 1.

(1) **M:**[1] *We have to make OJ.*
M:[2] *There are oranges at I and an OJ factory at B.*
M:[3] *Engine E3 is scheduled to arrive at I at 3pm.*
M:[4] *Shall we ship the oranges?*
S: [5] *Ok.*
S: [6] *Shall we load the oranges in the empty car at I?*
M:[7] *No, use the car attached to E3.*

This example shows two conversants, the manager (**M**) and the system (**S**), collaborating on a plan to achieve the goal of making orange juice. Since utterances (2)

*Presented at the AAAI 1993 Spring Symposium on Reasoning about Mental States: Formal Theories & Applications.

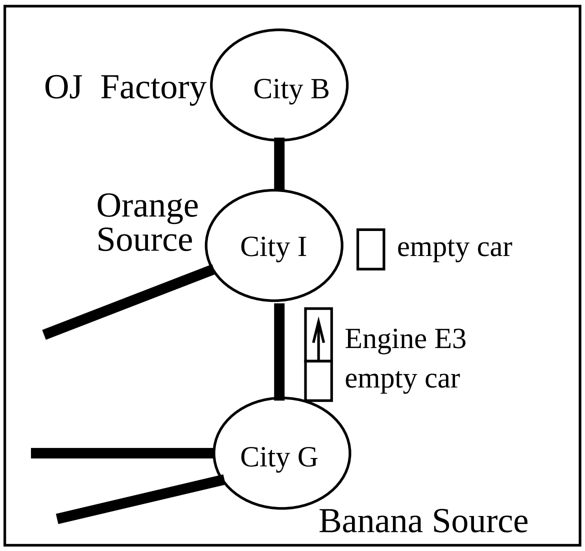

Figure 1: Trains World Scenario

and (3) are about facts and events that are already mutually believed, the manager must be suggesting that these facts and events are relevant to achieving the goal of making orange juice. So, they are best viewed as suggestions to incorporate the facts or events into the plan.

Utterance (4) is also such a suggestion, but is uttered as a yes-no question. What is it questioning? One gloss of this utterance is that the yes-no question is questioning whether the suggestion is acceptable. But this interpretation ignores how the suggestion is embedded in the current plan: that the oranges will be shipped with engine E3 from city I to city B and that E3 will be at I by 3 pm. It is this context, all of the facts and events that are related to the suggestion by way of the current plan, that is being question, not just whether the oranges should be shipped. So, this utterance should also be interpreted as a request for judgment of the part of the plan that deals with shipping the oranges.

In utterance (5), the system utters 'ok,' thereby accepting the shipping subplan. Then in (6), the system suggests that they load the oranges into the empty car at I. This suggestion is uttered as a yes-no question, and as above, we view it as also a request for a judgement of the part of the plan that involves loading the oranges into the empty car. In utterance (7), the manager rejects the proposed subplan, and modifies it by suggesting an alternate car to load the oranges into, the car attached to E3.

With our model, the current plan is being updated after each utterance. Suggestions add additional actions, facts, or subgoals to the plan; judgments express the speaker's beliefs about the validity of the plan, and requests for judgments add discourse obligations for the hearer to express his views about the plans validity.

A Plan-Based Model of Collaboration

Clark and Wilkes-Gibbs (1986) investigated how participants in a conversation collaborate in making a referring action successful. They found that after the initial referring expression was presented, the other participant would pass judgment on it, either *accepting* it, *rejecting* it, or *postponing* his decision. If it was rejected or the decision postponed, then one participant or the other would *refashion* the referring expression. This would take the form of either *expanding* it by adding further qualifications, or *replacing* the original expression with a new expression. The referring expression that results from this is then judged, and the process continues until the referring expression is acceptable enough to the participants for current purposes.

From the above description, we can see that each step in the acceptance process has a referring expression associated with it. The judgment moves serve to judge the expression, while the refashioning moves serve to refashion it, resulting in the referring expression that is associated with the next turn. This view led us (Heeman, 1991; Heeman and Hirst, 1992) to propose that the *current* referring expression is part of the state of the collaborative process, along with beliefs about its validity. Besides having an intention to achieve the goal of the collaborative activity, the participants, in order to co-ordinate their activity, have intentions to keep the state in their common ground. So, the judgment and refashioning moves serve to fulfill these intentions.

Our earlier work further lends itself to task-oriented dialogues since we viewed referring expressions as plans that conversants collaborate upon. A judgment move is uttered with the intent to make it mutually believed whether the current plan is valid or not, and a refashioning move serves to update the current plan. Since the two agents are collaborating, neither one is explicitly controlling the conversation. So, we proposed that as long as a judgment or refashioning is understood, the hearer will tend to accept it, and update the common ground accordingly. Disagreements are handled by subsequent judgments and refashioning of the plan, rather than by judgments and refashionings of the previous utterance. So, the current plan, which is in the common ground of the participants, need not be valid, only that the system be able to find it coherent (Pollack, 1990). The plan representation proposed by Ferguson (1992) allows such plans to be represented, as well as partial plans.

The Speech Actions

We now need to adapt the model of Heeman and Hirst for use in collaborative task-oriented dialogues in general, rather than just referring expressions. Our approach is to give definitions for the speech actions, and to state their intended effect on the mental state of the participants. This of course will not be a full classification of the speech actions that can be realized in such dialogues, but only the more central ones that are needed to account for the conversation given above.

We need to account for four different operations that speakers can perform on the current plan. They can suggest (or request) additions to the plan, request judgment of the plan (or subplan), make judgments, and make replacements. Furthermore, they can restrict their attention to just a portion of the plan, usually the part of the plan that is relevent to the last addition to the plan. We will refer to this as the *subplan in focus*. So, the discourse reasoner, with the help of the plan reasoner, will need to keep track of this. Also, to account for the coherency across multiple speech actions, discourse expectations and obligations will need to be inferred from the speech actions. For instance, a request for judgment of the subplan in focus gives rise to a discourse expectation in the speaker that the hearer will judge the subplan. Likewise, the hearer will have a discourse obligation. This will also need to be kept track of by the discourse reasoner.

Our definition of the speechacts uses a typed Prolog that allows functional roles (variables are indicated with a leading '?'). Each speech act type has a role for the speaker, hearer, time, and its content. Below, we will give the content of the speech actions. The content is expressed in a logical form language (Allen, 1992), which will be explained as needed. Definitions of the speech acts have been simplified by removing discourse markers and references to time.

Additions to the Plan

When a speaker mentions a fact, event, or goal, one interpretation for this utterance is that it is a suggestion to *use* that fact, event, or goal in the plan.[1] We represent this by using the `suggest` speech act. The content of this act is a relationship between the content of the utterance and the current plan. For instance, a suggestion of a fact is expressed as the relationship that the current plan uses some object (which is derived using focus heuristics from the utterance that the fact was uttered in) in the context expressed by the fact. Given below is the speech act interpretation of a suggestion to use some object, `?obj` in the plan, where `?fact` corresponds to the context that the object was mentioned in.

```
(:content ?e*t-suggest
  (:the ?p*t-plan
    (:current-plan ?p)
    (:uses ?obj*t-object ?fact ?p)))
```

Note that the `:uses` predicate is embedded inside of the quantifier `:the`. This quantifier takes three parameters, the first is the variable, the second, the restrictions, and the third is the expression that is being quantified. This is used in the example above in order to refer to the current plan.

For expressing that an event should be added to the plan, we use the relation `:event-in`, which takes an event type as its first parameter, and a plan as its second.

Requests for Judgment

The second plan operation is a request for a judgment of a subplan of the current plan. These are usually uttered in combination with a suggestion, by using the modal verb 'shall' followed by some action. In this case, the request for judgment is of the subplan dominated by the suggestion. Such speech actions must add a discourse expectation/obligation that the next utterance will be a judgment of the subplan for which judgment was requested. Below is the content of a request for an acceptance.

```
(:content ?e*t-request
 (:occurs
  (:lambda ?acc*t-accept
    (:and
     (:focus-subplan ?subplan)
     (:content ?acc ?subplan)))))
```

The content uses the `:occurs` logical form, which takes as its argument a speech act type. In this case it is an accept speech act whose content is the subplan that is in focus. So, this speech action is requesting the occurrence of an acceptance of the subplan in focus.

Judgments

The intended effect of a judgment is to make it mutually believed that the current plan, or part of it, is or is not acceptable. Judgments typically follow a request for a judgment, and so the part of the plan that is being judged is given by the discourse expectations or obligations. Otherwise, the utterance itself will include some reference that indicates the scope of the judgment, such as "that sound's good". If the judgment is a rejection, then a discourse expectation/obligation will be added that the next utterance will be a replacement to the subplan. Given below is an example of the content of a rejection (acceptances are practically identical).

```
(:content ?e*t-reject
  (:lambda ?p*t-plan
    (:focus-subplan ?p)))
```

The content of a judgment uses the `:lambda` operator in order to specify what part of the plan is being judged.

Replacements

Replacements are used to fix a subplan that was rejected. At the surface level, they are identical to the suggestions mentioned earlier, and so we rely on discourse expectations in order to disambiguate them. Our

[1] Speakers can also be more direct and request that a fact, event, or goal should be used in the plan.

approach to replacements is to treat them as the conjunction of an addition to the plan and an argumentative speech action (Traum and Hinkelman, 1992). The argumentative action will be a `replace` speechact, which takes two roles, one for referring to the speech action that is being used to make the addition, and the second for referring to the judgment speech act. The judgment speech act is specified since it might indicate which part of the subplan that the replacement should happen in. Note that it is left to the plan reasoner to decide how to best incorporate the replacement into the subplan. Given below is the role information for a sample replacement.

```
(:role ?e :r-replaced ?er*t-reject)
(:role ?e :r-replacement ?es*t-suggest)
```

The Example

We now work through the example, giving the speech acts for each utterance. In addition, we take the viewpoint of the system, and show how the speech act interpretations can be used by the discourse reasoner, or generated by it.

In the first utterance the manager introduces the goal to be achieved, and so begins the collaborative activity.

The next utterance expresses knowledge about the world. This utterance could be either an inform, or a suggestion that the oranges and OJ factory are relevant to the plan. Since these facts are mutually believed by both participants, the speech act pruner will remove the former interpretation. So, we give only the speech acts for the latter.

M:[2] *There are oranges at I and an OJ factory at B.*

```
(:content [ST-SUGGEST-0006]
 (:the ?p*t-plan
  (:current-plan ?p)
  (:uses [X1.2] (:at [X1.2] [STN-I]) ?p)))
```

```
(:content [ST-SUGGEST-0009]
 (:the ?p*t-plan
  (:current-plan ?p)
  (:uses [X1.3] (:at [X1.3] [STN-B])  ?p)))
```

So, the speech act interpretation is that there are two suggestions, one that the current plan uses the oranges, whose discourse marker is `[X1.2]`, and that these oranges are at I, `[STN-I]`; and the second is that the current plan uses the OJ factory, `[X1.3]`, which is at B, `[STN-B]`.

The third utterance is similar to the second. Below is the speech act interpretation that it is a suggestion to use engine E3 in the plan.

M:[3] *Engine E3 is scheduled to arrive at I at 3pm.*

```
(:content [ST-SUGGEST-0012]
 (:the ?p*t-plan
  (:current-plan ?p)
  (:uses [ENG3]
   (:occurs
    (:lambda ?e*t-arrive
     (:and
      (:role ?e :r-object [ENG3])
      (:role ?e :r-location [STN-I])
      (:role ?e :r-time [3PM]))))
   ?p)))
```

The fact that the engine is schedule to arrive at I at 3pm is represented by the occurrence of an arrival event type. The operator `:occurs` takes an event type. Event types are built using the `:lambda` operator, which takes an event variable as its first parameter, and restrictions as its second. This allows an arrival event at `[STN-I]` at `[3PM]` of `[ENG3]` to be specified.

These speech act interpretations are given to the discourse reasoner, which will use the plan reasoner to incorporate them into the current plan. This will result in the plan having the fact that there are oranges at I, an orange juice factory at B, and that engine E3 is scheduled to arrive at I at 3pm. The plan reasoner, in order to make the plan coherent, will postulate that the oranges at I are to be used to make orange juice and that engine E3 will be used to move them from I to B (Ferguson, 1992).

In the fourth utterance, the manager asks if they should ship the oranges. Two of the speech act interpretations for this utterance are that it is a suggestion to ship the oranges and a request for a judgment of the subplan dominated by the action of shipping the oranges.

M:[4] *Shall we ship the oranges?*

```
(:content [ST-SUGGEST-0014]
 (:the ?p*t-plan
  (:current-plan ?p)
  (:event-in
   (:lambda ?e*t-bring-about
    (:and
     (:role ?e :r-agent [SYSHUM])
     (:role ?e :r-event
      (:lambda ?em*t-move-commodity
       (:role ?em :r-commodity [X1.2]))))))
   ?p)))
```

```
(:content [ST-REQUEST-0015]
 (:occurs
  (:lambda ?e*t-accept
    (:and
     (:focus-subplan ?subplan)
     (:content ?e ?subplan)
```

```
     (:role ?e :r-speaker [SYS])))))
```

These speech act interpretations will be passed to the discourse reasoner, which will use the plan reasoner to incorporate the suggestion into the plan. Assuming that this is successful and creates no difficulties, the discourse reasoner can accept the resulting subplan.

Since the system found the suggestion to be acceptable, it responds with "OK.", and so accepts the subplan. The content of the accept is the subplan in focus.

S: [5] *Ok.*

```
(:content [ST-ACCEPT-0017]
 (:lambda ?p*t-plan
  (:focus-subplan ?p)))
```

The system then reasons about the plan and suggests loading the oranges into a specific boxcar, and requests a judgment about this subplan.

S: [6] *Shall we load the oranges in the empty car at I?*

```
(:content [ST-SUGGEST-0019]
 (:the ?p*t-plan
  (:current-plan ?p)
  (:event-in
   (:lambda ?e*t-load
    (:and
     (:role ?e :r-agent [SYS])
     (:role ?e :r-commodity [X1.2])
     (:role ?e :r-car [C1])
   ?p)))))

(:content [ST-REQUEST-0020]
 (:occurs
  (:lambda ?e*t-accept
    (:and
     (:focus-subplan ?subplan)
     (:content ?e ?subplan)
     (:role ?e :r-speaker [SYS])))))
```

The manager however rejects the subplan and suggests, as a replacement to the subplan, to use an alternate boxcar.

M:[7] *No, use the car attached to E3.*

This is interpretated as a rejection of the subplan in focus, a request that the current plan uses the car attached to E3, [C2], and that this request should be a replacement to the subplan in focus in order to correct the plan.

```
(:content [ST-REJECT-0025]
 (:lambda ?p*t-plan
  (:focus-subplan ?p)))

(:content [ST-REQUEST-0029]
 (:the ?p*t-plan
  (:current-plan ?p)
  (:uses [C2] ?x*t-anything ?p)))

(:role [ST-REPLACE-0030]
  :r-replaced [ST-REJECT-0025])
(:role [ST-REPLACE-0030]
  :r-replacement [ST-REQUEST-0030)
```

The system invokes the plan reasoner in order to find an appropriate part of the subplan and to replace it by the specified boxcar.

Discussion

In this paper, we presented a set of speech actions for representing how conversants collaborate in task oriented dialogues. These actions are indexical on the current plan. Suggestions of facts and events are interpreted as suggestions that these facts and events be incorporated into the current plan. Judgments and requests for judgments are about the current plan, or at least that part which is relevant to the current conversation. Likewise refashionings are about replacing one part of the plan with something else. So, the current plan captures the state of the collaborative activity, and is assumed to be in the common ground of the conversants.

Our work differs from other approaches to modeling collaboration in dialogues. In particular, Sidner (1992) emphasizes the role of utterances. Her speech acts, which are part of an artificial language designed to model collaborative dialogues, are concerned with making individual beliefs mutually believed, without reference to a current plan. For instance, she has the action `ProposeForAccept`, which is used to propose that some belief be mutually believed; `Reject`, which is used to make it believed by the hearer that some belief that has been proposed is not believed by the speaker; and `Counter`, which is used to replace a belief that the hearer has proposed by one that the speaker is proposing. So, any notions of how actions relate to other actions that have already been proposed are not captured, and so she must rely on a general belief revision system rather than taking advantage of more directed plan reasoning mechanisms.

Our work also differs from Traum's work (1991) on reaching mutual understanding in dialogues. In representing the current state of a dialogue, Traum proposes a number of different plan spaces, corresponding to whether a plan (or action) is just privately held, or has been proposed, acknowledged, or accepted. Our work assumes a much simpler model. Once a proposal has been uttered, it is incorporated into the current plan, so long as this can be done coherently. So, judgments about the proposal and refashionings to it are carried out with respect to the current plan, which corresponds to Traum's *accepted* plan space. So, unless a proposal cannot be understood, it would move directly from *privately held* to the *accepted* plan space.

Future Work

Much work remains to be done. Foremost, we need to re-examine judgments and requests for judgments. We have proposed that these actions are about the subplan in focus. However, there is the other speech act interpretation: that the judgment or request for judgment is about the specific utterance, regardness of how it fits into the current plan. We need to account for how these two speech act interpretations affect each other. Second, we need to further explore the notion of a *subplan in focus*, and how this focus shifts during a conversation. Third, we need to give a formal account of the speech actions that we presented, including their effect on the mental state of the agent and the discourse expectations and obligations that can arise from them. Fourth, we need to verify our results against a corpus of task oriented dialogs, a task that has already been started.

Acknowledgments

I wish to thank my supervisor, James Allen, for enlightening discussions. Also, thanks to David Traum, Hannah Blau, George Ferguson, and Massimo Poesio. Funding gratefully received from the Natural Sciences and Engineering Research Council of Canada, from NSF under Grant IRI-90-13160, and from ONR/DARPA under Grant N00014-92-J-1512.

References

Allen, J. F. (1992). Users guide for using RHET logical forms. Unpublished Manuscript.

Allen, J. F. and Schubert, L. K. (1991). The TRAINS project. Technical Report 382, Department of Computer Science, University of Rochester.

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

Ferguson, G. M. (1992). Explicit representation of events, actions, and plans for assumption-based plan reasoning. Technical Report 428, Department of Computer Science, University of Rochester.

Heeman, P. A. (1991). Collaborating on referring expressions. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, Student Session*, pages 345–346.

Heeman, P. A. and Hirst, G. (1992). Collaborating on referring expressions. Technical Report 435, Department of Computer Science, University of Rochester.

Poesio, M. (1993). Definite descriptions and the dynamics of mental states. In *Working Notes AAAI Spring Symposium on Reasoning about Mental States: Formal Theories & Applications*, Stanford.

Pollack, M. E. (1990). Plans as complex mental attitudes. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, SDF Benchmark Series, pages 77–103. MIT Press.

Sidner, C. L. (1992). Using discourse to negotiate in collaborative activity: An artificial language. In *Proceedings of the Workshop on Cooperation among Heterogeneous Intelligent Agents*. AAAI-'92.

Traum, D. R. (1991). Towards a computational theory of grounding in natural language conversation. Technical Report 401, Department of Computer Science, University of Rochester.

Traum, D. R. (1993). Mental state in the TRAINS-92 dialogue manager. In *Working Notes AAAI Spring Symposium on Reasoning about Mental States: Formal Theories & Applications*, Stanford.

Traum, D. R. and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3).