

TRACKING USER GOALS
IN AN INFORMATION-SEEKING ENVIRONMENT

Sandra Carberry

Department of Computer Science
University of Delaware
Newark, Delaware 19711

ABSTRACT

This paper presents a model for hypothesizing and tracking the changing task-level goals of a speaker during the course of an information-seeking dialogue. It allows a complex set of domain-dependent plans, forming a hierarchical structure of component goals and actions. Our model builds the user's plan as the dialogue progresses, maintains both a local and a global plan context, and differentiates between past goals and goals currently pursued by the user. This research is part of a project to develop a robust natural language interface. If an utterance cannot be interpreted normally or a response cannot be generated due to pragmatic overshoot, the strong expectations about the utterance provided by our context model can be used as an aid in processing the input and producing useful responses.

I. INTRODUCTION

Determining the goals and plans of the speaker is essential in understanding natural language dialogue. A cooperative participant uses the information exchanged during the dialogue and his knowledge of the domain to hypothesize the speaker's goals and plans for achieving these goals. The speaker formulates his utterances under the assumption that they will be interpreted in this manner.

This context of goals and plans provides clues for interpreting utterances and formulating cooperative responses. In the following, the second utterance can only be interpreted within the context of the speaker's goal, as communicated in the first utterance.

"I want to cash this check. Small bills only, please."

Similarly, a useful response to the query

"Is Prof. Smith teaching Expert Systems next semester?"

might be (1) if the speaker wants to take Expert Systems with Dr. Smith, (2) if the speaker's primary interest is the Expert Systems course, or (3) if the speaker's primary interest is Dr. Smith :

1. "No, but Prof. Smith is scheduled to teach it next year."

2. "No, Prof. Jones is teaching Expert Systems next semester."

3. "No, Prof. Smith is teaching Natural Language Processing next semester."

This paper presents a model for hypothesizing and tracking the changing goals of a speaker during the course of an information-seeking dialogue. Our research differs from previous work in an information-seeking environment in three ways:

- [1] The knowledge base allows for a complex domain of goals and plans
- [2] The context mechanism builds the speaker's plan as the dialogue progresses and differentiates between local and global plan context.
- [3] The history mechanism incorporates previous plans into the overall plan context.

II. OVERVIEW

A plan in our system (called TRACK) is a hierarchical structure of component goals and actions, each of which has an associated plan or is a primitive in the domain. Plans are represented using a STRIPS formalism [Fikes and Nilsson, 1971]; each plan contains preconditions, a set of partially ordered actions, and effects. Such a plan can be expanded to any level of detail; its full expansion will contain goals and actions that are also components of other fully expanded plans. The existence of a goal/action as an entity within the hierarchy indicates that it has a well-defined plan which an agent may follow and captures the generality of this entity within several higher level plans.

In most cases, a complete plan for the speaker cannot be built during the first part of a dialogue. Our approach is to infer a lower-level goal, relate it to potential higher-level plans, and build the complete plan context as the dialogue progresses. The local context is the goal and associated plan upon which the speaker is currently focused; the global context includes higher level goals and plans which led to the current local context. The context mechanism distinguishes local and global contexts and uses these to predict new speaker goals from the current utterance.

TRACK was implemented for an information-seeking environment, as part of an ongoing project to develop a robust natural language interface. The domain is the courses, requirements, and policies for students at a university. It is assumed that the system and user share the belief that the user wants to obtain information relevant to a program of study and that the system is a capable and cooperative provider of such information. The context of speaker goals and plans constructed by TRACK can be used to interpret ill-formed input, handle pragmatic overshoot [Sondheimer&Weischedel, 1980], and produce helpful responses.

To transfer to another area, such as seeking information about real estate, only the corpus of domain-dependent plans and goals must be reconstructed; the decision-making heuristics need not be altered.

III. GOAL PROCESSING

At least three types of goal structures appear necessary: the immediate goal, derived goals/actions, and focused plans. The immediate goal is extracted directly from the semantic representation of the literal interpretation of the speaker's utterance. A derived goal or action is inferred from the immediate goal; it relates requests for information and indirect speech acts to task-dependent goals or actions. The focused goal is the goal that the speaker is currently pursuing; it has an associated focused plan. This focused plan produces the strongest expectations for understanding ellipsis and detecting unsignalled goal changes.

A. Derived Goals/Actions and Focused Plans

The inference rules to infer a derived goal or action from an immediate goal are based upon shared knowledge concerning the roles and capabilities of the speaker and the system. They represent compilations of some plan-recognition rules, including those responsible for indirect speech act interpretation [Allen,1980], [Sidner&Israel,1981]. The following are a few of the inference rules for producing derived goals/actions:

- [I1] If the speaker wants to know the $x:P(x)$ that comprise the possible choices for the parameter in a subaction specified by the speaker, then the speaker may want to perform an action whose plan contains that subaction.

Example: "What Science course must I take?"

- [I2] If the speaker wants to know the value of x , and x is a term in a precondition or subaction of a plan, then the speaker may want to perform the action represented by that plan.

Example: "What are the prerequisites of History 304?"

- [I3] If the speaker wants to know how to achieve an effect, then the speaker's goal may be to

achieve that effect.

Example: "How do I become a Computer Science major?"

Focused plans are constructed by relating the derived goals/actions to the domain-dependent set of plans. Candidate focused plans are produced by the following heuristics, in which DERIVED is a derived goal or derived action.

- [F1] If DERIVED is an action and there is no plan for that action, then candidate focused plans are any plans which include DERIVED.
- [F2] If DERIVED is an action and there is a plan for that action, then the candidate focused plan is that plan.
- [F3] If DERIVED is a true predicate which is a precondition in a plan or the effect of an action in a plan, then that plan is a candidate focused plan.
- [F4] If DERIVED is an unsatisfied predicate which is the effect of a plan, then that plan is a candidate focused plan.

B. Examples

Consider the query

"Do I have credit for French 112?"

The immediate goal is

Knowif(Agent,Earned-Credit(Agent,French112)).

One possible derived goal of the agent is that the agent have credit for French 112. However, as Allen points out [Allen,1980], one may ask if x is true when one wants x to be false, as in the query "Am I on probation?". Thus from the immediate goal of knowing if x is true, the inference rules produce the two derived goals:

- D1. Earned-Credit(Agent,French112)
- D2. Not-Earned-Credit(Agent,French112)

If the agent has credit for French 112, then rule F3 applies to derived goal D1 and no rule applies to derived goal D2. Derived goal D1 is the effect of the action Earn-Credit(Agent,French112) in the plan for Satisfy-Language(Agent); therefore the plan for Satisfy-Language(Agent) becomes a candidate focused plan.

If the agent does not have credit for French 112, then rule F4 applies to derived goal D1 and rule F3 applies to derived goal D2. The plan for Earn-Credit(Agent,French112) has D1 as its effect; therefore rule F4 produces this plan as a candidate focused plan. The predicate Not-Earned-Credit(Agent,French112) is a precondition in the plan for Earn-Credit(Agent,French112). Therefore rule F3 applied to derived goal D2 again produces Earn-Credit(Agent,French112) as a candidate focused plan.

IV. CONTEXT MECHANISM

Two different forms of context processing are necessary. The first constructs a context model at the start of a dialogue or when the speaker terminates the current dialogue and pursues an entirely new task. The second type of context processing updates the context model as the dialogue continues.

A. Hypothesizing Initial Context

The immediate goal is extracted directly from the first utterance, the derived goals are obtained from the inference rules of the previous section, and the focused plan is computed from the heuristics of the previous section.

The focused plan represents the local context. If only one focused plan exists and it is a plan for an action that appears in only one higher-level plan, then this higher-level plan forms part of a global context assumed obvious between speaker and hearer. This global context is built until a choice of higher-level plans must be made. The initial context tree contains this global context, if any, the focused plan, and the derived goal/action, all of which are marked as active constituents.

If only one focused plan and context tree exists, then we have definite expectations for future utterances. If more than one focused plan exists, we are uncertain of the speaker's goals. We would request clarification if this ambiguity prevented a response to the utterance or if we believed an obstacle existed to the speaker pursuing one of the focused goals. This relates to a general maxim of human behavior that one notifies another of obstacles to anticipated courses of action but does not request information about another's plans unless necessary for a response (i.e., don't be nosy).

B. Updating the Context Model

As each new utterance occurs, the context model must be expanded to reflect an updated hypothesis about the speaker's plans and goals. TRACK analyzes each utterance and constructs derived goals/actions and candidate focused plans. It selects a focused plan apropos to the current plan context and expands the context tree to include it. This expansion process spawns a new subtree. The history mechanism marks all plans along the path from the root of the context tree to the focused plan as active constituents. Previously active components are marked as inactive but retained in the context model for future reference.

Sidner[1981] and McKeown[1982] have investigated focusing constraints in anaphora resolution and natural language generation respectively. McKeown proposes an ordering on the speaker's choice of focus options. She claims that when faced with a choice of topic, the speaker will choose to move to a recently introduced topic if he has something further to say about it; other-

wise the speaker must reintroduce this topic at a later time. In addition, the speaker will choose to continue with the current topic before switching back to a previous one.

TRACK's heuristics for context processing are based on similar principles:

- [1] a user will generally obtain all desired information about the currently focused task and the most recently considered subaction before considering other tasks
- [2] the path of currently active plans forms a stack of potential focused tasks to which the user may return.

TRACK uses the first applicable heuristic in the following set to select an appropriate focused plan and adjust the context tree.

- [H1] If a breadth-first expansion of the most recently considered subaction within the current focused plan in a context tree includes a candidate focused plan, select the candidate focused plan that occurs earliest and expand the context tree to include it.
- [H2] If a breadth-first expansion of the other actions in the current focused plan in a context tree includes a candidate focused plan, select the candidate focused plan that occurs earliest and expand the context tree to include it.
- [H3] If an expansion of a plan in the active portion of a context tree includes a candidate focused plan, select that candidate focused plan. If more than one candidate focused plan meets this criteria, select the candidate focused plan that is a descendant of the plan deepest in the active portion of the context tree. Expand the active portion of the context tree to include the selected focused plan.
- [H4] If an expansion of a candidate focused plan includes the plan for the root of a context tree, select that candidate focused plan, form a new context tree for it, and expand that tree to include the old context tree.
- [H5] If an expansion of another plan includes both a current context tree and a candidate focused plan, select that candidate focused plan, form a context tree for it, and expand these two context trees upward until they meet as subtrees of the same higher-level plan.
- [H6] If none of the above apply, the speaker either has incorrect beliefs regarding how to achieve his goals or has begun planning for an entirely new and unrelated goal.

V. EXAMPLES

Consider the following sequence of three requests. In each illustration of a context tree,

the current focused plan is preceded by an asterisk and active constituents by the letter A. Derived and immediate goals are not shown.

- [1] "Can I obtain a Computer Science major?"

This utterance produces two potential context trees,

A* Satisfy-Dept(Agent,CS,BA)

and

A* Satisfy-Dept(Agent,CS,BS)

- [2] "What are my choices for the foreign language course I must take?"

The plan for the action Satisfy-Language(Agent) contains the disjunction of the two actions 1) Earn-Credit in an intermediate level foreign language course and 2) Pass-Skills-Test in a foreign language. Inference Rule I1 and heuristic F2 produce Satisfy-Language(Agent) as a candidate focused plan. Heuristic H5 applies to this candidate focused plan and the first context tree in Example 1. (The BS degree does not have a foreign language requirement.) The system should inform the user that it assumes he is pursuing the BA degree. The new context tree is shown in Figure 1.

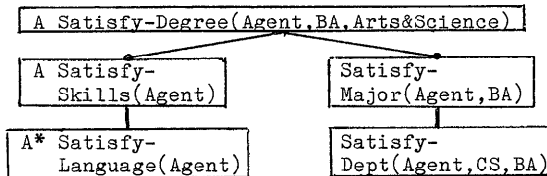


Figure 1. The context tree in Example 2

- [3] "What are the prerequisites of French 112?"

The term prerequisites-of(French112) appears in a precondition of three plans; each becomes a candidate focused plan. The action represented by one of these candidates, EARN-CREDIT(Agent,French112), is a descendant of the current focused plan and heuristic H1 applies. The new context tree is shown in Figure 2.

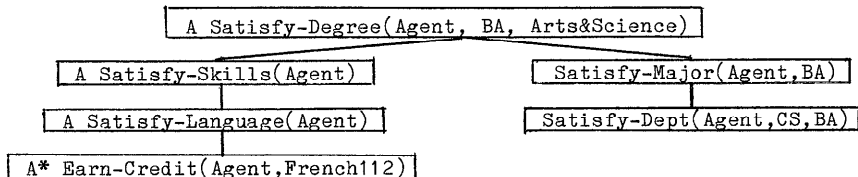


Figure 2. The context tree produced in Example 3

Each of the above is a working example from the TRACK system.

VI. RELATED WORK

Allen[1980] inferred the speaker's goal and plan in order to produce helpful responses to indirect speech acts. The goals were either MEET TRAIN or BOARD TRAIN and the plan for each consisted of only a few primitive actions. His plan recognition mechanism inferred a complete plan from the speaker's utterance. In more complex domains, the speaker's complete plan consists of a hierarchy of subplans and subgoals. Such a complete plan is not immediately evident; furthermore, the speaker's current goal within such a plan changes during the course of a dialogue.

The TDUS system acts as an expert guiding an apprentice in the assembly of an air compressor[Robinson et al.,1980]. Grosz[1977] developed the concept of a focus space hierarchy to represent those objects upon which the attention of the dialogue participants was centered. Her system tracked the shifting focus of the apprentice and expert and was used to determine the referents of definite noun phrases in a dialogue. Robinson[1981] constructed a model of the actions and goals of the apprentice as inferred from the dialogue. The model contained a goal-action tree which represented execution of the task and differentiated between background goals and the goal/action currently focused upon by the apprentice.

In contrast with such task-execution domains, the user in our environment is seeking information in order to formulate a plan for subsequent execution. The information domain is relevant to many diverse plans and the user's overall task-related goal is not obvious at the start of a dialogue. Since the plan is not actually executed during the dialogue with the system, the user's utterances are not as tightly constrained by the structure inherent in the plan as are the utterances in the apprentice-expert task dialogue. The user may investigate several low-level subgoals which could be part of many higher-level plans and only later relate them as components of a specific plan. Natural language understanding requires that enough of the plan structure be built to represent the speaker's communicated plans and goals and that the system track the speaker's focus of attention within this plan structure.

Reichman[1981] investigated social conversations and represented a participant's model of the

discourse as a hierarchical structure of "context spaces" with associated focusing information.

Mann, Moore, and Levin[1977] designed a model of human language interaction. Their system analyzed and structured dialogues according to linguistic goals, such as "Seek-Permission" or "Describe-Problem", not task goals.

VII. LIMITATIONS AND FUTURE WORK

The TRACK system has been implemented for a domain consisting of a subset of the courses, requirements, and policies for students at a University. The system is presented with a logical representation of the literal interpretation of a user's query and returns an updated context model. There are five areas for future work:

- [1] The system must be extended to handle recursively defined plans.
- [2] The system currently presumes that the user will seek information in a relatively coherent, organized manner. This restriction must be removed and provision made for stacking and later connecting sequences of utterances that at first appear unrelated.
- [3] Certain utterances, such as

"Is CS105 offered at night?"

express user preferences. The system should be extended to infer and represent such preferences in a user model. This model could then be used to produce helpful responses that address the particular user's desires.
- [4] Since the speaker may reconsider or refer back to an old deactivated goal, there must be heuristics for detecting this and merging new and old plans.
- [5] The context model can be used in robust understanding of natural language. If an utterance cannot be interpreted normally or a response cannot be generated due to pragmatic overshoot [Sondheimer&Weischedel 1980], the strong expectations about the utterance provided by TRACK's context tree can be used as an aid in processing the input and producing useful responses.

VIII. CONCLUSIONS

The TRACK system hypothesizes and tracks the changing task-level goals of a speaker during the course of a dialogue. It allows a complex set of domain-dependent plans, forming a hierarchical structure of component goals and actions. The system captures the generality of an inferred lower-level goal as a distinct entity within higher-level plans and builds the user's plan as the dialogue progresses. This eliminates the need for working with many separate complete plans at once. TRACK maintains both a local and a global plan context and differentiates between past goals

and goals currently pursued by the user. TRACK is part of a project to develop a robust natural language interface [Sondheimer&Weischedel, 1980].

ACKNOWLEDGMENTS

I would like to thank Ralph Weischedel for his encouragement and direction in this research and for his suggestions on the style and content of this paper.

REFERENCES

- Allen, J.F. and C.R. Perrault, "Analyzing Intention in Utterances", *Artificial Intelligence* 15,3,1980
- Birnbaum, L., "Argument Molecules: A Functional Representation of Argument Structure", *Proc. AAAI*, August 1982
- Cohen, R., "Investigation of Processing Strategies for the Structural Analysis of Arguments", *Proc. 19th Annual Meeting of ACL*, June 1981
- Fikes, R.E. and N.J. Nilsson, "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving", *Artificial Intelligence* 2, 1971
- Grosz, B.J., "The Representation and Use of Focus in a System for Understanding Dialogs", *Proc. of the IJCAI*, Pittsburgh, Pennsylvania, 1977
- Mann, W., J. Moore, and J. Levin, "A Comprehension Model for Human Dialogue", *Proc. of the IJCAI*, Cambridge, Massachusetts, Aug., 1977
- McKeown, K.R., "The Text System for Natural Language Generation: An Overview", *Proc. of the 20th Annual Meeting of the ACL*, Toronto, Ontario, Canada, June, 1982
- Perrault, C.R. and J.F. Allen, "A Plan-Based Analysis of Indirect Speech Acts", *American Journal of Computational Linguistics*, July 1980
- Reichman, R., "Conversational Coherency", *Cognitive Science* vol.2, 1978
- Robinson, A.E., "Determining Verb Phrase Referents in Dialogs", *American Journal of Computational Linguistics*, Jan. 1981
- Robinson, A.E., Appelt, D.E., Grosz, B.J., Hendrix, G.G., and Robinson, J.J., "Interpreting Natural-Language Utterances in Dialogs about Tasks", *Technical Note No.210*, SRI International, Menlo Park, California
- Sidner, C.L. "Focussing for Interpretation of Pronouns", *American Journal of Computational Linguistics*, October 1981
- Sidner, C. and D. Israel, "Recognizing Intended Meaning and Speakers' Plans", *Proc. IJCAI*, Aug. 1981
- Sondheimer, N. and R.M. Weischedel, "A Rule-Based approach to Ill-Formed Input", *Proc. 8th International Conf. on Computational Linguistics*, 1980