Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Inferring Appropriate Responses in Discourse

Permalink

https://escholarship.org/uc/item/2pf522fr

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 9(0)

Author

Chase, Melissa P.

Publication Date

1987

Inferring Appropriate Responses in Discourse*

Melissa P. Chase
The MITRE Corporation
Burlington Road, Bedford, MA 01730

Abstract

This paper discusses how *Scenes*, declarative representations of the intentional and attentional structure of discourse, facilitate the inference of appropriate responses.

1 Introduction

When people engage in conversation, one of the most striking features is the ability of the particpants to infer the meaning of utterances and respond appropriately. Many researchers in discourse processing explain this phenomenon through plan recognition: conversational participants generate and recognize plans to make and understand utterances designed to acheive certain goals. The most successful work along these lines has involved individual utterances [Allen & Perrault 80,Cohen & Perrault 79]. Recently, Grosz and Sidner[Grosz & Sidner 86] have suggested that we require a better understanding of discourse structure in order to extend this work to sequences of utterances in a large discourse.

Grosz and Sidner have proposed a model of discourse structure with three distinct, but interacting components:

---

*This work was supported by MITRE Sponsored Research Project 90780.

1. the structure of the actual sequence of utterances of the discourse
2. a structure of intentions
3. an attentional state

The linguistic structure of the discourse is composed of ***discourse segments (DSs)***, which are aggregates of the actual utterances of the discourse. The intentional structure of the discourse consists of ***discourse segment purposes (DSPs)***, which specify how the DS's contribute to the overall purpose of the discourse (the DP). Grosz and Sidner have identified two intentional relations that play a crucial role in the structure of a discourse: ***dominance*** and ***satisfaction-precedence***. If an action (physical or linguistic) which satisfies one intention, say DSP1, is intended to partially satisfy another intention, say DSP2, then we say that DSP2 ***dominates*** DSP1. If DSP1 must be satisfied before DSP2, then we say that DSP1 ***satifaction-precedes*** DSP2. The attentional structure of the discourse is represented by a collection of ***focus spaces***. The attentional state captures the salient objects, properties, and relations at each point of the discourse.

At the MITRE Corporation, we have devised a representation for the intentional and attentional structure of discourse based upon this model, called a ***scene***, which serves as the basis for the discourse component of the KING-KONG system (a transportable natural language interface for expert systems). By keeping track of the current intentional and attentional state discourse, KING-KONG is able to reason about the underlying goals and intentions of the user in order provide appropriate responses. In this paper, I would like to briefly describe scenes (for a fuller description, see[Zweben & Chase 87]) and then show how discourse structure contributes to the recognition of the speaker's plans and facilitates appropriate and intelligent response.

2 Scenes

A ***scene*** is a schema representation, similar in spirit to frames[Minsky 75], scripts[Schank & Abelson 77], and related formalisms[Bobrow & Collins 75].

THE INTENTIONAL STRUCTURE

Field	Description
Name	The type of scene
Roles	The prominent object classes
Inferiors	The scenes dominated by this one
Superiors	The scene dominating this one
Enables	The post-requisite scenes
Enabled-by	The pre-requisite scenes
Actions	The expert system operations appropriate to this scene

THE ATTENTIONAL STRUCTURE

Field	Description
Role-Fillers	The objects filling the roles
Predecessors	The scenes preceding this one in the actual discourse
Successors	The scenes following this one in the actual discourse
Focus Cache	The objects available for anaphoric references
Domain Goal	The current expert system task

Figure 1: The Slots of an Instantiated Scene

A scene defines the potential intentional structure of the discourse of an interaction with the expert system by defining the user's intended actions and their relationships. An *instantiated* scene captures the attentional structure of the discourse by recording which intentions have been satisfied, and what objects and expert system operations were involved.

In a typical expert system, the user wishes to carry out some task which is often decomposed into subtasks. The tasks involve a limited number of operations, which can be performed on a limited number of object classes. Thus, each scene contains information about the possible object classes involved in a task, called *roles*; the potential expert system operations on these roles, called *actions*; and the relations between different tasks, the *inferior*, *superior*, *enables*, and *enabled-by* scenes. Together, these pieces of information permit the computation of the intentional structure of the actual discourse.

When a scene is recognized as the current intentional state, it is instantiated in order to represent the attentional state. Its roles and focus cache are filled with the referents of the objects in the current utterance,

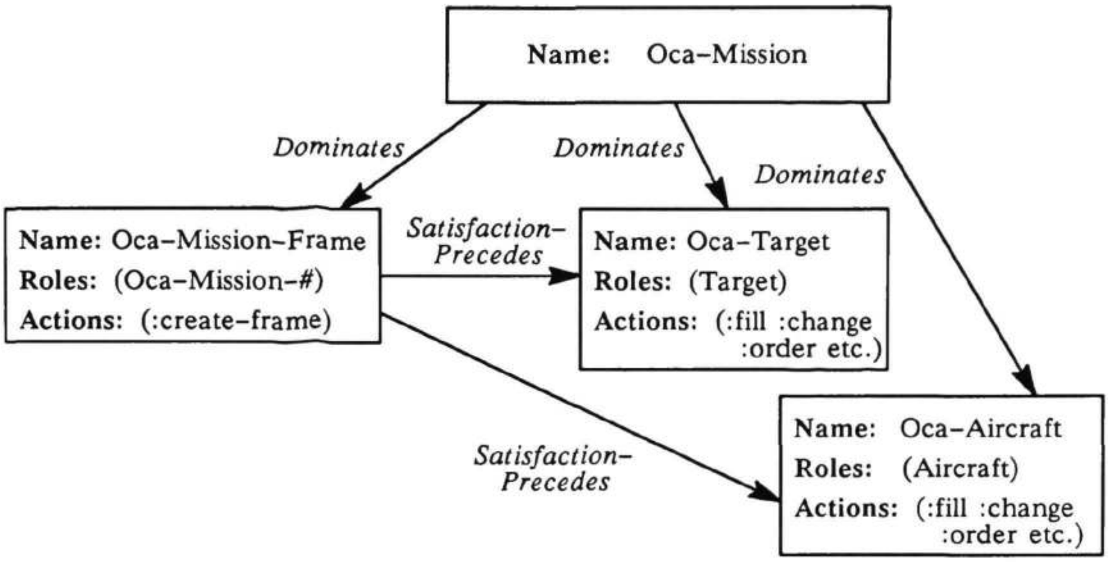
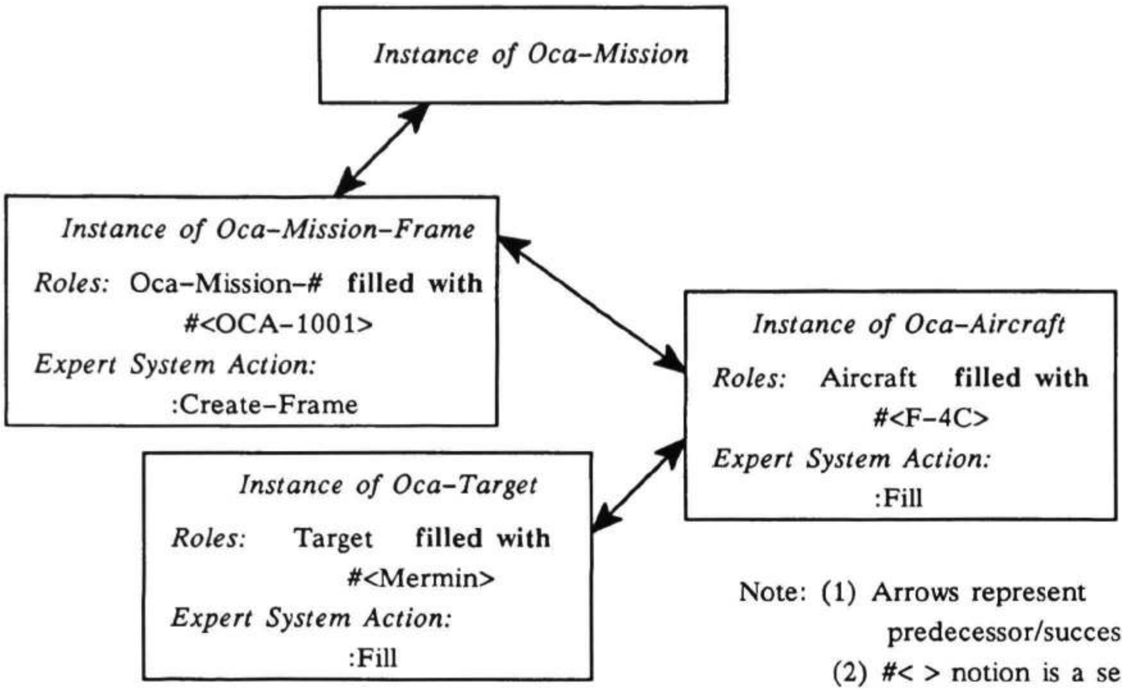


Figure 2: The Intentional Structure

predecessor/successor links are created which model the actual flow of the interaction, and the goal (which expert system task is involved) is recorded. See Figure 1 for a description of the most important information maintained in a scene.

To make this description a little more concrete, consider the following example drawn from the Knowledge-Based Replanning System (KRS) mission planning application. The primary goal of this application is to plan an Offensive Counter Air (OCA) mission. In order to achieve this goal, several choices must be made, such as the target and the type of aircraft. The system is a mixed-initiative expert system, which is capable of fully planning missions, or can guide a user planning a mission by verifying that the user has made appropriate choices and suggesting suitable choices on request. The root of the scene hierarchy corresponds to the DP of the overall discourse–planning an OCA mission. The inferior scenes model the intentions of fulfilling the subtasks of the mission planning task. Each scene has a single role and a number of expert system actions (for making a choice, changing a choice, requesting suitability information, etc.). Figure 2 depicts the intentional structure and Figure 3 the attentional structure generated



Note: (1) Arrows represent predecessor/successor links
(2) #< > notion is a semantic representation

Figure 3: The Attentional State

by the following discourse fragment:

> **User:** Create a mission.
>
> **Computer:** Displays a new mission template.
>
> **User:** Send F-4Cs.
>
> **Computer:** Fills in the aircraft slot of the template with F-4C.
>
> **User:** Attack Mermin.
>
> **Computer:** Fills in the target slot of the template with Mermin.

3 Inferring Appropriate Responses

In order to respond appropriately to an utterance, the system must recognize the user's intention in making this utterance within the context of the discourse. This involves recognizing, through use of the intentional and attentional structures of the discourse, both the linguistic act and the domain act intended by the user.

In an expert system interface, the linguistic acts performed by the user are typically requests to inform or requests to act; these requests are made to help satisfy the DP of the interaction, namely, to carry out some expert system task. To respond to these requests, the system must understand the intended meaning of the request, reason about the role of this intention in the overall intentional structure of the discourse, and select an appropriate response based upon the attentional structure of the discourse (the objects and DSPs in focus).

3.1 Recognizing the User's Intention

In many natural language systems, the process of recognizing the intentions of utterances involves describing the user's speech acts as plans and recognizing them[Allen & Perrault 80], [Cohen & Perrault 79], [Sidner & Israel 81], [Sidner 83b], [Litman 86]. In the KING-KONG interface, these acts are described as schemata of patterns to be matched, and the Response Handler (the portion of the system that determines the appropriate response to an utterance) uses the intentional and attentional information in the current scene and linguistic and semantic information from the utterance to recognize the user's plan and respond accordingly. Since each scene encapsulates a collection of objects (the prominent roles of the scene) and the domain actions that manipulate these objects, the amount of inferencing involved in recognizing the user's plan is constrained.

The patterns in the schemata essentially describe the pre-requisites and post-requisites of the plan corresponding to the user's intentions. Each action schema contains the following patterns to be matched:

1. *Scenes* – a list of scenes appropriate for the action, or a list of scenes inappropriate for the action
2. *Speech Act* – the linguistic action of the utterance

3. *Semantic Representation* – the semantic representation of the utterance, either a backend operation and its arguments, or a database relation and its arguments
4. *Scene Roles* – the roles from intentionally related scenes that must already be filled
5. *Effects* – the effects upon the intentional and attentional structure of the discourse, and the backend, of executing this action

When the Scene Controller, the portion of the system that recognizes the intentional and attentional states of the discourse, proposes a scene to the Response Handler, the Response Handler tries to match each action schema against information from the utterance and the current scene. First, the Response Handler checks to see if the scene is appropriate to the action. Then, linguistic features from the utterance are used to deduce the speech act. Next, the system infers the domain action of the request by examining the semantic interpretation of the utterance (determined, in part, by the scene). From the intentional structure encoded in the scene, and the linguistic act and the semantic interpretation of the utterance, the Response Handler infers the intention of the utterance; this includes determining whether the intention is part of the current discourse segment or another one. Finally, the system considers the attentional state contained in the scene (the roles) in order to determine how much of the user's plan has been satisfied. By default, all satisfaction-precedence intentions should have been satisfied, but the schema description permits one to relax this condition. If the match is successful, the Response Handler transmits the effects to the Scene Controller, so the intentional and attentional states can be updated, and issues the appropriate commands to the expert system to execute the specified action.

A couple of examples will clarify how this mechanism allows KING-KONG to provide intelligent responses. The first example illustrates how enabling actions are inferred and executed. The second example demonstrates how an underlying request for information is inferred.

3.2 Example: Implied Actions

In the KRS domain, the choice of target is so central to planning an OCA mission that specifying a target signals the intention of creating a new mission. So, if the user begins a planning session by saying "Make Mermin the target," the system should first create a new mission and then fill in the target slot with Mermin. KING-KONG does indeed carry out this interaction.

The intentional structure of this interaction, captured in the intentional relations encoded in the scene hierarchy is:

DSP(oca-mission) *dominates* DSP(oca-target)

DSP(oca-mission-frame) *satisfaction-precedes* DSP(oca-target)

These intentional relations describe this formal plan:

INTEND(user, INTEND(computer, DO(plan-mission))) $\wedge$
INTEND(user, INTEND(computer, DO(mission-frame))) $\wedge$
INTEND(user, INTEND(computer, DO(fill(target, Mermin))) $\wedge$
BELIEVES(user, GENERATES(plan-mission,
mission-frame $\wedge$ fill(target, Mermin))) $\wedge$
BELIEVES(user, ENABLES(mission-frame, fill(target, Mermin)))

When the Scene Controller proposes the *oca-target* scene, the following action schemata both match:

Name: fill-action
Scenes: (oca-target)
Speech Act: :act
Semantic Representation: (:fill (current-scene prominent-roles))
Scene Roles: (enabling-scene prominent-roles :optional)
Effects: Discourse: DSP(enabling-scene) satisfied $\wedge$
DSP(oca-target) satisfied
Backend: fill target in oca-mission frame

and

Name: create-mission-frame-action
Scenes: (oca-target)
Speech Act: :act
Semantic Representation: (:fill (current-scene prominent-roles))
Scene Roles: (enabling-scene prominent-roles :absent)
Effects: Discourse: DSP(enabling-scene) satisfied
Backend: create oca-mission frame

At this point in the discourse the intention of the utterance "Make Mermin the target" is INTEND(user, INTEND(computer, DO(fill(target, Mermin)))). This intention is recognized because features of the utterance and the state of the discourse represented in the scene match the specifications of both schemata. The domain action is recognized by matching the semantic representation specified in the schemata against the interpretation of the utterance (i.e., to fill the target slot with Mermin). The force of the utterance (a request to act) is deduced by matching the linguistic action of the utterance (an imperative) against the speech act specified in the schemata and determining how far the plan has progressed from the attentional information in the scene hierarchy (what scene is proposed and which intentionally related scenes have had their discourse purposes satisfied). Further, the *fill-action* schema indicates that the satisfaction-precedence intentions of the scene need not be filled (since the prominent-roles of the enabling scene–the scene that *satisfaction-precedes* the current scene–need not have been filled). Since the mission frame has not been created, the schema matches. Similarly, the *create-mission-frame-action* schema specifies that the satisfaction-precedence intention *has not* been satisfied, so this schema also matches. The effects specified in the two schemata ensure that the *create-mission-frame-action* is executed before the *fill-action*.

3.3 Example: Underlying Intentions for Information

Requests for information typically are made to support the satisfaction of the intentional structure of the planning session. Some of these questions directly involve roles and their attributes, and the intention of the request to obtain the information needed to fill a mission frame slot with an appropriate choice.

Other questions are more oblique and do not directly request information about a role. For example, consider the following discourse:

> **User:** Send 4 F-4Cs from Halfort to Mermin.
>
> **Computer:** Displays a new mission template and fills in the slots.
>
> **User:** Leave at 0330 hours.
>
> **Computer:** Fills in the time of departure slot.
>
> **User:** What is an F-4C?
>
> **Computer:** An F-4C is an oca-aircraft.
>
> **User:** Is the mission at night?
>
> **Computer:** Yes. A mission at 0330 hours should not be flown by any of Halfort's aircraft.

The intention underlying the final question reflects a concern about the night-flying abilities of the aircraft.

The intentional structure of this interaction is simply:

DSP(oca-mission) *dominates* DSP(oca-aircraft)

which represents the plan:

INTEND(user, INTEND(computer, DO(plan-mission))) $\wedge$
INTEND(user, INTEND(computer, DO(fill(aircraft, X)))) $\wedge$
BELIEVES(user, GENERATES(plan-mission, fill(aircraft, X)))

The first scene proposed by the Scene Controller is the *oca-aircraft* scene, since that was the previous intentional and attentional state of the discourse. In this context, the following schema matches:

Name: aircraft-inference
Scenes: (oca-aircraft)
Speech Act: :inform
Semantic Representation: (verify(possession(mission-time, night)))
Scene Roles: (:or (current-scene prominent-roles :present)
(constraining-scene prominent-roles :present))
Effects: Discourse: DSP(current-scene) satisfied
Backend: respond to inform request $\wedge$
verify role choices consistent with response

The ***aircraft-inference*** schema specifies that in the context of the *oca-aircraft* scene, requests for information about the mission time are relevant to the discourse purpose of the scene. Thus, the intentional structure of the interaction matches the intention of this request. From the attentional state of the discourse, either the prominent-role of the scene (the aircraft in this case) or the prominent-roles of constraining scenes (scenes that do not ***dominate*** or ***satisfaction-precede*** the current scene, because the order in which the scenes are traversed is flexible, but scenes whose intentions involve domain constraints; for example, the choice of airbase and aircraft constrain one another) must have been filled; that is the intentions of the current scene or constraining scenes must have been satisfied. This is the default specification for a request having an underlying intention, and the response it triggers is to answer the surface request for information, and to verify that the role choices specified by the Scene Roles is consistent with the result of the inform act. Hence, the system responds as shown in the above fragment.

This particular schema is extremely specific in its plan; currently we are generalizing this inferential ability. The schema will encode various relationships and attributes that constraint the task represented by a scene (described by the combination of roles and expert system operations), and the effect will be to verify that a particular choice satisfies these constraints, to generate choices and filter them by these constraints, and so on.

4 Conclusion

We have implemented a system based upon the discourse structure model proposed by Grosz and Sidner. In a restricted domain, such as an expert system interface, we have discovered that it is possible to use a representation, called *scenes* to capture the intentional and attentional state of a discourse, and use this information to reason about a user's intentions.

We do not suggest that this approach can solve the general discourse processing problem, but it provides a mechanism for tracking discourse and understanding intentions in a restricted context. To make this system more flexible, we intend to generalize the reponse component, investigate how to acquire new scenes[Mooney & DeJong 85] and reason about misconceptions[Pollack 86] within the scene framework.

References

[Allen & Perrault 80] Allen, J.F. & Perrault, C.R. (1980). Analyzing Intention in Utterances. *Artificial Intelligence*, 15:143–178.

[Bobrow & Collins 75] Bobrow, D. & Collins, A.M. (1975), editor. *Representation and Understanding.* Academic Press, New York.

[Cohen & Perrault 79] Cohen, P. & Perrault, C.R. (1979). Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science*, 3:177–212.

[Grosz & Sidner 86] Grosz, B.J. & C. L. Sidner, C.L. (1986). Attentions, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12:175–204.

[Litman 86] Litman, D.J. (1986). Linguistic Coherence: A Plan-Based Alternative. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 215–223, New York.

[Minsky 75] Minsky, M. (1975). A Framework for Representing Knowledge. In P. H. Winston, editor, *The Psychology of Computer Vision*, McGraw-Hill, New York.

[Mooney & DeJong 85] Mooney, R. & DeJong, G. (1985). Learning Schemata for Natural Language Processing. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 681–687, Los Angeles.

[Pollack 86] Pollack, M.E. (1986). A Model of Plan Inference that Distinguishes Between the Beliefs of Actors and Observers. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 207–214, New York.

[Schank & Abelson 77] Schank, R.C. & Abelson, R.P. (1977). *Scripts, Plans, Goals and Understanding.* Lawrence Erlbaum Associates, Hillsdale, N. J.

[Sidner & Israel 81] Sidner, C.L. & Israel, D.J. (1981). Recognizing Intended Meaning and Speakers' Plans. In *Proceedings of Seventh International Joint Conference on Artificial Intelligence*, pages 203–208, Vancouver, B. C.

[Sidner 83b] Sidner, C.L. (1983b). What the Speaker Means: The Recognition of Speakers' Plans in Discourse. *International Journal, of Computers and Mathematics*, 9:71–92.

[Zweben & Chase 87] Zweben, M. & Chase, M.P. (1987). Scenes: A Representation of Intentional and Attentional Structure. Submitted to AAAI-87, Seattle, WA.