Translation from Transactions of IPSJ

Natural Language Dialogue Understanding on a Four-layer Plan Recognition Model

HITOSHI IIDA* and HIDEKAZU ARITA*†

More often than written text, spoken Japanese dialogue often contains fragmental utterances in which predicate substitutional expressions, case element ellipsis, and contracted expressions are used. In the understanding and translation of such dialogues, it is necessary to interpret each utterance in line with the developmental flow of the dialogue. This paper proposes a method for understanding a task-oriented dialogue that employs a four-layer plan recognition model. The four layers of plans are (1) *interaction plans*, which contain knowledge that can be represented by ordered utterances between speaker and hearer for the exchange of certain information, (2) *communication plans*, which contain knowledge that allows actions for the purpose of information transfer through dialogue to be realized as a series of communicative acts, (3) *domain plans*, which contain knowledge that allows a given action to be achieved by carrying out an ordered set of acts, and (4) *dialogue plans*, which contain knowledge that allows a dialogue to be realized by means of pragmatic knowledge for dialogue development. Use of these plans makes it possible to relate each utterance to the topic of the discourse in line with the development of the dialogue, and build up structures that extend over the entire dialogue. Predicate substitutions and case element ellipsis can be resolved by interpreting the fragmental utterances within the structure of the dialogue.

1. Introduction

Spoken Japanese dialogue contains more fragmental utterances than written dialogue because there is more ellipsis of established information and parts of predicates, and more use of predicate substitutional expressions. Further, speech acts using various linguistic expressions are added. When attempting to understand and translate dialogue with such characteristics, it is necessary to interpret each fragmental utterance in terms of the developmental flow of the dialogue. Fragmental utterances can be interpreted by invoking domainspecific knowledge shared by the dialogue participants, pragmatic knowledge for the purpose of conducting the dialogue, and so on. Such knowledge makes possible the resolution of ellipses.

This type of problem has been addressed by several researchers. Allen et al. [1, 2] have demonstrated a model that analyzes speech acts within the framework of plans and forms a cooperative response. Their model allows responses that provide even more information than what was asked for and responses that are cooperative with fragmental and indirect utterances. Litman et al. [3] established the domain plan, which is constructed from the hierarchical and procedural nature of the dependency relationships of domain-dependent actions; they also introduced the discourse plan, which shows the relationship between the domain plan and each utterance. The discourse plan contains plans in which a plan is introduced to the dialogue by a speech act that requests a certain action, and plans that identify the values of the variables of the actions of a certain plan according to speech acts that provide a certain value. There is also the work of Grosz et al. [4], which involves a model of the dialogue structure constructed by looking at the purpose of each segment of the dialogue.

This paper proposes a mechanism for dialogue understanding. The object of this understanding model is task-oriented dialogues. The dialogue corpus consists of inter-keyboard dialogues between computer terminals for "queries and explanations regarding registration for an international conference." This model introduces interaction plans and communication plans for representing knowledge related to the execution of turn-taking and information transfer to clarify the correspondence of each utterance to the previously described domain plan. Corresponding to the earlier described discourse plan, a dialogue plan is introduced to represent knowledge related to the overall development of the dialogue Next, we present a method of applying extra-linguistic knowledge such as domain-specific knowledge. Through the intermediary pragmatic-level

This is a translation of the paper that appeared originally in Japanese in Transactions of IPSJ, Vol. 31, No. 6 (1990), pp. 810–821.

*ATR Interpreting Telephony Research Laboratories.
†Now with Mitsubishi Electric Corp.

process in the dialogue, this allows the process of plan recognition that arrives at the purpose of the utterance to be expressed. It thus becomes possible for the assignment of the relationships of each utterance to proceed flexibly along with the development of the dialogue, and structures extending across the entire dialogue can be formed. At the same time, it is possible to resolve the elements of ellipsis within each utterance.

The problems related to understanding fragmental utterances are described in Section 2 of this paper. In Section 3, a layered plan recognition model is described. In Section 4, an experimental system is outline, and experimental results are presented. Topics for future study are suggested in Section 5.

2. Fragmental Utterances

To achieve dialogue understanding, many problems must be solved, such as understanding various types of ellipsis, substitutional expressions, rephrasing, euphemistic expressions, and colloquial expressions, the means for which depend on prosodic information [5]. Of these, we address here the problem of understanding fragmental utterances that contain predicate substitutional expressions, ellipsis of case elements, and so on. In particular, we consider Japanese dialogue, which is especially context-dependent. The types of expressions mentioned above are efficient as communicative acts, because the knowledge behind the linguistic expressions is already known to the dialogue participants. However, this feature creates problems for machine understanding of such expressions. Here, the assumed degree of understanding is that in a limited discourse domain, contents that are not context-dependent are determined to have the same meaning when translated from Japanese into English. Next, we give examples of important types of expressions that pose problems in machine understanding.

In examples 1 and 2, "*onegaishimasu*," which means "please," is a request form in the dialogue, and serves as a substitute for repeating the predicate that appears directly or indirectly in the immediately preceding utterance. Accordingly, the interpretation of this expression varies with the context in which it appears. The sentence enclosed in parentheses has an identical interpretation.

Example 1: (substitutional expression)
"*Ginkoufurikomi-de onegaishimasu*"
[Bank transfer, please.]
("*Ginkoufurikomi-de (ryoukin-wo) shiharattekudasai*")
([Please pay (the fee) by bank transfer.])

Example 2: (substitutional expression)
"*Okurisaki-wo onegaishimasu*"
[Address, please]
("*Okurisaki-wo oshietekudasai*")
([Please tell me the address.])

Examples 3 and 4 also featur repetition in the dialogue. In these examples, the items of knowledge that are known to both speaker and hearer from the context of the dialogue up to the time of utterance are omitted from the utterance to avoid repetition. In example 3, the focussing object within the dialogue is omitted, as indicated by the parentheses. The same is true of example 4.

Example 3: (ellipsis of case element)
"( ) 12 *gatsu* 25 *nichidesu*"
[December 25]
("*Shimekiri-wa 12 gatsu 25 nichidesu*")
([The deadline is December 25.])

Example 4: (ellipsis of case element)
"*Shimekiri-ga sematteimasu node,* ( ) *oisogikudasai*"
[The deadline is coming up, so please hurry.]
("*Shimekiri-ga sematteimasu node, moushikomi-wo oisogikudasai*")
([The deadline is coming up, so please apply in a hurry.])

Rather than there being a fixed method of referring to objects that are newly introduced the dialogue, contracted expressions for established concepts are often used. In examples 5 and 6, the referent cannot be determined uniquely from the contents of the utterance only [6].

Example 5: (contracted expression)
"*Youshi-wo okuttekudasai*"
[Please send me the forms.]
("*Genkou-youshi-wo okuttekudasai*")
([Please send me the manuscript papers.])

Example 6: (contracted expression)
"*Youshi-wo ookurishimasu*"
[I'll send the form.]
("*Furikomi-youshi-wo ookurishimasu*")
([I'll send the bank transfer form.])

3. Layered Plan Recognition Model

3.1 Overview

In dialogue, there is more than one means of utterance turn-taking for the purpose of conveying intention or exchanging information. For example, if the office wants to convey an intention of sending registration forms to the inquirer, the commissive [7, 8] utterance (7-1) in example 7 is possible, but the directive [7, 8] utterance (8-1) in example 8 is also possible.

Example 7:
Office: "*Touroku-youshi-wo ookurishimasu*" (7-1)
[I'll send you the registration foms.]
Inquirer: "*Onegaishimasu*" (7-2)
[Thank you.]

Example 8:
Inquirer: "*Touroku-youshi-wo okuttekudasai*" (8-1)
[Please send me the registration forms.]
Office: "*Hai, wakarimashita*" (8-2)
[I will.]

If, for machine understanding of this dialogue, the

turn-taking conditions are stored one-by-one in memory, the number of combinations of conditions quickly becomes very large.

Thus, in this model, utterance interpretation proceeds while relationships are formed between domain-specific knowledge and social common-sense commitments concerning dialogue turn-taking for each utterance in a task-oriented dialogue, thus achieving flexible processing of the utterances. For this purpose, we introduce four types of plan: (1) *interaction plans* for expressing commitments related to dialogue turn-taking as order-of-appearance relationships between communicative act types, (2) *communication plans* for expressing actions for the exchange of information through dialogue as a series of communicative act types, (3) as plans for achieving actions within the dialogue domain, *domain plans* for representing subordinate actions and their order-of-appearance relationships, as well as their conditions of occurrence, and (4) *dialogue plans*, which concern the construction of the overall dialogue structure. These plans represent knowledge for forming cooperative responses and conducting cooperative dialogue, domain-specific knowledge, and knowledge for the general development of a dialogue.

Levinson [9] constructs dialogues from three structural elements: 'opening-section,' 'themes,' and 'closing-section.' The relationships among those structural elements is described as a dialogue plan. In 'themes,' the dialogue proceeds towards the dialogue purpose. The plan for achieving acts within the domain is described as the domain plan. Further, a plan for determining what kind of dialogue should be used to accomplish the acts of a given domain is called a communication plan. In addition, it is pointed out that each utterance consists of a 'turn-taking' pair [9]. These turn-taking pairs are described by the relationships between utterances governed by the interaction plan. When viewed in this way, a hierarchical relationship can be seen among these four plans.

Making use of this hierarchical feature, it is possible to represent utterances and their purposes more clearly. That is to say, the relationship between the pragmatic-level processes for performing the utterance and the actions realized through the dialogue become clear. Thus, the purpose of this dialogue model is to monitor a dialogue between two persons and interpret the utterances that occur. On the other hand, Wilensky [10] defines a metaplan, introducing common sense related to a planner's goal, and considers an action plan for the computer side of a manmachine information exchange. This metaplan is a description of a plan for each situation, and is not sufficient for preparing a plan that assigns a relationship between each utterance and the overall dialogue.

In this model, plans are described by means of plan schemata. The understanding of an utterance begins with a semantic description of the utterance. Plan schemata that constitute subordinate actions for the type of action indicated in that discription are then obtained. By further determining links between the plan schemata, the plan schema for the end of the chain is obtained. The action described there is interpreted as the goal of that utterance. This method of reasoning is basically the same as that described by Allen et al. [1, 2].

3.2 Communicative Acts and Turn-taking Pairs

For task-oriented dialogue, where communication and exchange of information are the main goals, this model limits the information processed to the propositional contents of the utterance and the communicative acts that affect the beliefs and intentions of the hearer. However, so that utterances related to the opening and closing of the dialogue can be recognized, communicative acts for the purpose of conducting a smooth and harmonious dialogue are also considered. Such communicative acts form utterance pairs in the patterns of questions, responses, and confirmations.

3.2.1 Types of Communicative Act

Communicative acts are described by schemata. Schemata consists of the following slots and values:

- •Prerequisites: Conditions that must exist before the communicative act can be performed.
- •Effects: The results of performing the communicative act.
- •Constraints: Restrictions on the performance of the communicative act. If these conditions do not exist, the communicative act cannot be performed.

For example, for proposition P (e.g., "*Touroku-youshi-wo okuru*" ["send registration forms"]), the directive communicative act "P *shitekudasai*" ["Please P."] has the prerequisite that the accomplishment of the speaker's plan P is included at the time of that utterance (P, captured as an action, is described as 'WANT (speaker, action)' in the same way as in Allen [11]). The intended effect of the utterance on the hearer is that the hearer should believe WANT (speaker, action), that is, 'BELIEVE (hearer, WANT (speaker, action))'. At this time, the requirement that the performer of action P is the hearer serves as a condition. This communicative act is an action that requests an action of the hearer.

Cohen [12] identified eight communicative act types in an analysis of dialogues containing "instructions for assembly of a pump located in front of the hearer." The acts included requests for acts of assembly, requests for identification of objects, and so on. In this paper, communicative acts collected from dialogues in the domain of inter-keyboard dialogue between computer terminals for "queries and explanations regarding registration for an international conference" are classified in terms of objects, actions, and dialogue control. Twenty-six dialogues were examined, including a total of 616 exchanges between the participants. The utterances included responses such as "*Hai*" ["Yes"] and

Table 1 Communicative Act Types (partial).

Category		Act Type	Explanation
Related to objects	Related to object attribute	Ask-Value	Requests an attribute value.
		Confirm-value	Confirms an attribute value.
		Inform-Value	Provides an attribute value.
		Affirmative-Value	Affirms an attribute value.
	Related to existence of object attribute	Ask-Exist-Value	Asks whether the object has the attribute.
		Affirmative-Exist-Value	Affirms that the object has the attribute.
Related to actions	Related to action contents	Ask-Action	Asks for the procedure for performing the action.
		Confirm-Action	Confirms the procedure for performing the action.
	Related to performance of action	Request-Action	Requests performance of the action.
		Offer-Action	Offers to perform the action.
		Ask-Acceptability1	Asks whether the speaker can perform the action.
		Ask-Acceptability2	Asks whether the hearer can perform the action.
		Accept-Action	Accepts performance of the action.
		Reject-Action	Rejects performance of the action.
		Confirm-Acceptance	Confirms acceptance of the action.
	Related to action prerequisites and effects	Ask-Truth	Asks whether the proposition is true.
		Inform-Truth	Informs that the proposition is true.
		Affirmative-Truth	Affirms that the proposition is true.
		Negative-Truth	Informs that the proposition is false.
Related to dialogue control	Related to object introduction and notification	Introduce-Object	Introduces an object to the dialogue.
		Notice-Value	Notifies that a value follows.
	Related to turn-taking	Require-Question	Elicits a question from the other party.
		Notice-Question	Notifies that a question follows.
	Related to beginning or ending of a dialogue	Greeting-Open	Begins a dialogue.
		Greeting-Close	Ends a dialogue.
	Others	Require-Question2	Etiquette

HEADER:	Request-Action(speaker, hearer, action)
PREREQUISITES:	WANT(speaker, action)
EFFECTS:	BELIEVE(hearer, WANT(speaker, action))
CONSTRAINTS:	hearer=AGENT(action)
HEADER:	Offer-Action(speaker, hearer, action)
PREREQUISITES:	BELIEVE(speaker, WANT(hearer, action))
EFFECTS:	BELIEVE(hearer, BELIEVE(speaker,WANT(hearer, action)))
CONSTRAINTS:	speaker=AGENT(action)

Fig. 1 Examples of communicative act definitions.

"*Iie*" ["No"], and connectives. Fine division of the classification items resulted in the classification of communicative acts shown in Table 1. 'Offer-Action' is a commissive type that corresponds to the directive type 'Request-Action.' An example of their use as schema headers in the description of communicative acts is shown in Fig. 1.

3.2.2 Turn-taking Pairs

The turn-taking pair in Example 8 consists of a request for action and a response indicating compliance. The corresponding communicative act types are 'Request-Action' and 'Accept-Action.' The inquirer who produces utterance (8-1) believes that the execution of that communicative act will generate 'BELIEVE (hearer, WANT (speaker, action))' as a mental state of the hearer, namely, the office, as shown in Section 3.2.1, and that the office will then cooperatively consent to perform the referent action 'action' as a variable in the predicate WANT. Corresponding to that belief, the office must, in a cooperative response, perform a communicative act that bears the meaning of either acceptance or rejection. If an utterance indicating acceptance is produced, that communicative act type is taken to be 'Accept-Action.' At this time, the mental state of the office is 'BELIEVE (hearer, WANT (hearer, action)),' and the constraint that the performer of this action should be the office is satisfied. Then, if a confirmation utterance is produced by the inquirer, that communicative act type is described as 'Confirm-Acceptance.' In

Table 2 Utterance Pairs Based on Communicative Acts (partial).

	"Request"	"Response"
p1	Ask-Value (e.g., "A *wa* WH *desuka*") [e.g., "What is A?"]	Inform-Value (e.g., "B *desu*") [e.g., "B"]
p2	Confirm-Value (e.g., "A *wa* B *desune*", "A *wa* B *desuka*") [e.g., "A is B, isn't it?", "Is A B?"]	Affirmative (e.g., "*Hai, soudesu.*") [e.g., "Yes, it is."] Negative (e.g., "*Iie, chigaimasu*") [e.g., "No, it isn't."]
p3	Ask-Exist-Value (e.g., "~ *wa arimasuka*", "*~wa irunodeska*") [e.g., "Do you have ~?", "Is ~ needed?"]	Affirmative-Exist-Value (e.g., "*Hai, arimasu.*", "*Hai, irimasu.*") [e.g., "Yes, we have it.", "Yes, it is needed."] Negative-Exist-Value (e.g., "*Iie, arimasen.*", "*Iie, irimasen.*") [e.g., "No, we don't have it.", "No, it is not needed."]
p4	Request-Action (e.g., "~ *shitekudasai.*") [e.g., "Please ~."]	Accept-Action (e.g., "*Hai, wakarimashita.*") [e.g., "We will ~."] Reject-Action (e.g., "*Moushiwake arimasenga, ~ dekimasen.*") [e.g., "We're very sorry, but we can't do that."]
p5	Offer-Action (e.g., "~ *shimashouka.*") [e.g., "Shall we ~."]	Accept-Offer (e.g., "*Hai, onegaishimasu.*") [e.g., "Yes, please ~."] Reject-Offer (e.g., "*Iie, kekkoudesu*") [e.g., "No, that's all right."]

this way, a series of turn-taking pairs becomes a chain of various communicative act types, and it is possible to invoke the plan initiated by the inquirer as a plan common to the inquirer and the office. This is assured by assuming a cooperative dialogue.

Examples of the turn-taking pairs specified by this model are given in Table 2. In turn-taking pair **p1**, the communicative act type 'Ask-Value,' in which speaker 1 asks for a value known to speaker 2, is paired with the communicative act type 'Inform-Value,' in which speaker 2 informs speaker 1 of the requested value. Many contiguous turn-taking pairs consisting of a request and a response occur, but at times they are nested. In this model, the interaction plan (described later) and the plan stack make it possible to process dialogues that have this sort of nested structure.

In Examles 7 and 8, the pairing of the commissive utterance (7-1) and its corresponding acceptance utterance (7-2), and the pairing of the directive utterance (8-1) and its corresponding acceptance utterance (8-2) can be described as interaction plans. Further, a communication plan for executing reference to actions in the domain of "the office sends registration forms to the inquirer" can be described by using those utterance pairs. Moreover, if those actions are described in a domain plan, these plans can be used to understand that the dialogues in Examples 7 and 8 are turntaking pairs that refer to actions in the same domain.

3.3 Utterance Representation

Utterances are represented in the form of `(communicative act type, speaker, hearer, topic, propositional contents)`. An NP marked with the special Japanese particle "*wa*" or by a clause that is equivalent to "*wa*," such as "*~nikanshite*" (as for, regarding, and so on) is taken to be a topic [13]. When such NPs are absent, the topic is taken to be the first NP in the utterance. The propositional contents are taken to be a description of the relationships centered around the main predicate in an utterance other than a communicative act type and the case elements of that predicate.

There is no one-to-one correspondence between the surface expressions of utterances and communicative act types. In the present model, the communicative act type is subject to constraints from the preceding context. If the communicative act type is not determined uniquely, it is left for the moment as a type with multiple possible interpretations, and later specified from subsequent dialogue.

The utterance expression "*Tourokuyoushi-wo okutekudasai*" ["Please send registration forms"] is

Table 3 Slots for Plan Schema.

Slot	Description
HEADER	This is the name of the plan. The plan is connected to other plans by means of parameters contained in the header.
PREREQUISITES	These are the conditions that must exist before the plan can be applied.
DECOMPOSITION	The plan is divided into sub-plans. At the level of HEADER description, a plan can be regarded as a single action that can be described by PREDICATE&CASES, but there are times when it can be further decomposed into a chain of several actions.
PREDICATE&CASES	These are the predicate that expresses the plan, and the case elements of that predicate. They are a description of the propositional contents.
EFFECTS	These are the results achieved if the plan is successful.
CONSTRAINTS	These are restrictions on the execution of the plan. If the CONSTRAINTS conditions are not present, the plan cannot be applied.
BELONG-TO	This indicates the set to which the plan belongs.

HEADER:	(Get-Value-Unit ?sp, ?hr, ?topic, (pred&cases (predicate DEARU) (object ?obj) (identifier ?id)))
DECOMPOSITION:	(Ask-Value ?sp, ?hr, ?topic, (pred&cases (predicate DEARU) (object ?obj) (identifier ?id))) (Inform-Value ?hr, ?sp, nil, (pred&cases (predicate DEARU) (object ?obj) (identifier ?id)))
EFFECTS:	(KNOW ?sp ?obj ?id)
HEADER:	(Request-Action-Unit ?sp, ?hr, ?topic, ?action)
DECOMPOSITION:	(Request-Action ?sp, ?hr, ?topic, ?action) (Accept-Action ?hr, ?sp, nil, (pred&cases (predicate ?pred) (object ?action)))
HEADER:	(Offer-Action-Unit ?sp, ?hr, ?topic, ?action)
DECOMPOSITION:	(Offer-Action ?sp, ?hr, ?topic, ?action) (Accept-Offer ?hr, ?sp, nil, (pred&cases (predicate ?pred) (object ?action)))

Fig. 2 Examples of interaction plans.

HEADER:	(Execute-Domain-Plan ?action)
DECOMPOSITION1:	(Request-Action-Unit ?hr, ?sp, ?topic, ?action)
DECOMPOSITION2:	(Offer-Action-Unit ?sp, ?hr, ?topic, ?action)
CONSTRAINTS:	?sp = AGENT(?action)
HEADER:	(Introduce-Object-Plan ?action)
DECOMPOSITION1:	(Get-Value-Unit ?sp, ?hr, ?topic, ?action)
DECOMPOSITION2:	(Confirm-Value-Unit ?sp, ?hr, ?topic, ?action)

Fig. 3 Examples of communication plans.

represented as

```
(Request-Action, Speaker, Hearer, tourokuyoshi,
    (pred&cases (predicate okuru) (agent Hearer)
                (recipient Speaker)
                (object tourokuyoushi)))
```

From the expression at the end of the example sentence, the communicative act type is Request-Action, requesting the other party to perform an action. No topic indicated by the special Japanese particle "*wa*" or an equivalent clause is present, so the noun phrase "*tourokuyoushi*," which appears at the beginning of the utterance, is taken to be the topic. The predicate of the proposition is "*okuru*" ["send"], the agent is "Hearer," the recipient is "Speaker," and the object is "*tourokuyoushi*" [registration forms].

3.4 Plans

3.4.1 Plan Descriptions

Plans represent knowledge that relates actions and situations by means of cause and effect relationships, time sequence relationships, and so on. They specify the search space for actions and situations. Plans are described by slots, such as HEADER, PREREQUISITES, DECOMPOSITION, PREDICATE&CASES, EFFECTS, and CONSTRAINTS, as shown in Table 3.

3.4.2 Interaction Plans

An interaction plan represents knowledge that makes possible the exchange of certain information between speaker and hearer by means of an ordered set of utterances. Examples of interaction plans are shown in Fig. 2. The plan Get-Value-Unit, which represents the utterance turn-taking concerning the attribute values of a given object, consists of a sequence of the communicative act types Ask-Value, which asks for the attribute value of the object, and Inform-Value, which provides the value. Here, replacing the variables **?sp** and **?hr**, which represent the speaker and the hearer, indicates a shift in the speaker of a question or response.

3.4.3 Communication Plans

A communication plan represents knowledge that allows acts for conveying information through a dialogue to be realized as a series of communicative act types. Examples of communication plans are shown in Fig. 3. The communication plan Execute-Domain-Plan for executing actions within the domain can be decomposed into the turn-taking pair represented by Request-Action-Unit (Example 8) or the turn-taking pair represented by Offer-Action-Unit (Example 7). In addition, there is a constraint condition that the agent (AGENT) of action **?action** should be **?sp**.

3.4.4 Dialogue Plans

A dialogue plan represents knowledge for establishing a dialogue construction. An example of a dialogue plan is shown in Fig. 4. The actions for carrying out a dialogue include opening the dialogue (Open-Dialogue-Unit), in which confirmation of the identity of the other party and other such tasks are accomplished, the body of the dialogue (Contents), in which communication and information exchange are accomplished, and the closing of the dialogue (Close-Dialogue-Unit), in which polite formalities and so on are exchanged.

HEADER:	(Dialogue ?sp, ?hr)
DECOMPOSITION:	(Open-Dialogue-Unit ?sp, ?hr) (Contents ?sp, ?hr) (Close-Dialogue-Unit ?sp, ?hr)

Fig. 4 Example of a dialogue plan.

HEADER:	(Make-Registration ?sp1, ?sp2, ?form)
DECOMPOSITION:	(Get-Form ?sp1, ?sp2, ?form) (Fill-Out-Form ?sp1, ?sp2, ?form) (Return-Form ?sp1, ?sp2, ?form)
PREDICATE&CASES:	(pred&cases (predicate MOUSHIKOMU) (agent ?sp1) (object ?conference))
HEADER:	(Get-Form ?sp1, ?sp2, ?form)
PREREQUISITES:	(KNOW ?sp2 name & address(?sp1) ?N&A)
DECOMPOSITION:	(Execute-Domain-Plan (pred&cases (predicate OKURU) (agent ?sp2) (recipient ?sp1) (object ?form)))
PREDICATE&CASES:	(pred&cases (predicate OKURU) (agent ?sp2) (recipient ?sp1) (object ?form))

Fig. 5 Examples of domain plans.

3.4.5 Domain Plans

A domain plan is knowledge that makes it possible to perform a given action as an ordered sequence of other actions. Examples of domain plans are shown in Fig. 5. The domain plan for registering for a conference, Make-Registration, can be decomposed into three subplans: obtaining the registration form (Get-Form), filling out the form (Fill-Out-Form), and sending the form in (Return-Form).

3.5 Mechanism for Understanding Dialogues by Means of Plan Recognition

The model proposed in this paper is for cooperative task-oriented dialogues that do not diverge from the focus [14] of the participants.

3.5.1 Plan Recognition

Plan recognition is accomplished by obtaining a chain of plan schema instances. The chain types are the same as those in Litman et al:

- Decomposition chain:

If the decomposition slot elements of plan **a** are unifiable with the header of plan **b**, then the two plans are linked by means of a decomposition chain. This chain indicates that the execution of a given plan is part of the execution of a plan above it.

- Prerequisite chain:

If condition **a** and the elements of a prerequisite slot of plan **b** are unifiable, then **a** and **b** are linked by a prerequisite chain. This chain indicates that when the conditions that are prerequisites for the plan exist, the plan can be executed.

- Effect chain:

If condition **b** and the elements of an effect slot of plan **a** are unifiable, then **a** and **b** are linked by an effect chain. This chain indicates that if the plan is executed, then the conditions that are effects will be brought about.

The chains of schemata are connected in the following order for utterance plan recognition: interaction plans for knowledge obtained from utterance expressions that is related to dialogue turn-taking, communication plans for knowledge that allows information to be exchanged through dialogue, domain plans for knowledge of the dialogue domain, and dialogue plans for knowledge related to dialogue development. The plan recognition process applies plans in the above order to obtain a chain of plan schema instances.

3.5.2 Management of Dialogue Conditions

Kato et al. [15] manage topics in a natural language interface by means of a topic stack and stack operations determined from input intentions such as continuation, advancing, branching, and backing-up. Elsewhere, Carberry has introduced the discourse stack [16]. In this paper, dialogue conditions are managed according to knowledge shared by the dialogue participants and a plan stack. The shared knowledge consists of the facts understood by the dialogue participants obtained through the dialogue, and is updated by description of the interaction plan effects. The plan stack holds schema instances that have a possibility of becoming the next dialogue topic.

The plan recognition of a given utterance is performed by taking the plan schema instance at the top of the plan stack as the goal, and obtaining a chain of instances. When a chain is not obtained, that goal is popped, and the next instance in the plan stack is taken as the goal in the same way.

The plan schema that is obtained as the goal is pushed onto the top of the plan stack. In this way, the dialogue topics can be searched in order, starting with the most recent topic contents. Further, the plans predicted from the prerequisite slot of that instance, denoted by $P_1, P_2, \ldots, P_n$, and the plans predicted from the decomposition slot, denoted by $D_1, D_2, \ldots, D_m$, are pushed onto the plan stack in the order $D_m, \ldots, D_1, P_n, \ldots, P_1$. This corresponds to priority searching of utterance turn-taking that satisfies the execution prerequisites of the given plan.

There may be more than one chain of plan schema instances because (1) there are times when multiple utterance representations correspond to the input surface expression, and (2) there are times when a given chain is possible for more than one plan. In such cases, it is considered that there are a number of possible understanding contents, and plan stacks corresponding to the respective understanding contents are generated. The understanding of ellipsis and so on is accomplished under each understanding content. On the other hand, understanding contents that do not produce a chain of schema instances for a given utterance are rejected.

Regarding plan control for introducing new topics, when utterances that contain clue phrases such as "tokorode" ["by the way"] or "*hanashi-wa chigaimasuga*" ["this is a different matter, but"] appear, a plan for introducing a topic is invoked with priority. These clue

phrases are surface expressions that suggest the discourse structure, and were used by Litman et al. for the control of plan searching.

3.5.3 Dialogue Understanding by Means of Plan Recognition

The correspondence among the plan schema instances in the chain of schema instances, which is the result of plan recognition, is achieved by unification of their variables.

For utterances in which there is ellipsis of case elements of the propositional contents, the parts that are omitted are represented by variables with no values in the utterance representation. However, after plan recognition these variables are assigned fixed values that are determined from preceding utterances and the plan schema. In this way, case element ellipsis is resolved.

For utterances that include substitutional expressions represented as 'noun+case particle+"*onegaishimasu*," plan recognition is performed as for an utterance in which the predicate has been omitted. Interpretation is done by resolution of the predicate in the same way as for case element ellipsis.

Ellipsis resolution by means of this model is not dependent on the recency [11] of the history list [17, 18] including sentence patterns or lexical similarity. Rather, by building up a dialogue structure, the meaning of the overall dialogue is captured, and ellipsis is analyzed from the interpretation of utterances obtained by indicating the position of each utterance within the dialogue. Thus, it is possible to resolve ellipsis in the utterances of dialogues that have a more complex structure and in utterances that contain substitutional expressions. Moreover, by means of a domain plan that describes the flow of actions related to a given topic, and a communication plan that determines how to execute or achieve a topic, ellipsis related to topics that do not appear earlier in the dialogue can also be resolved.

4. Implementation

4.1 Overview of Experimental System

An experimental system based on the plan recognition model described in Section 3 was implemented in Symbolic Common Lisp. An outline of the system is shown in Fig. 6.

The utterances that served as input in this experiment are assumed to be representations whose intentions, corresponding to communicative acts, are extracted by an

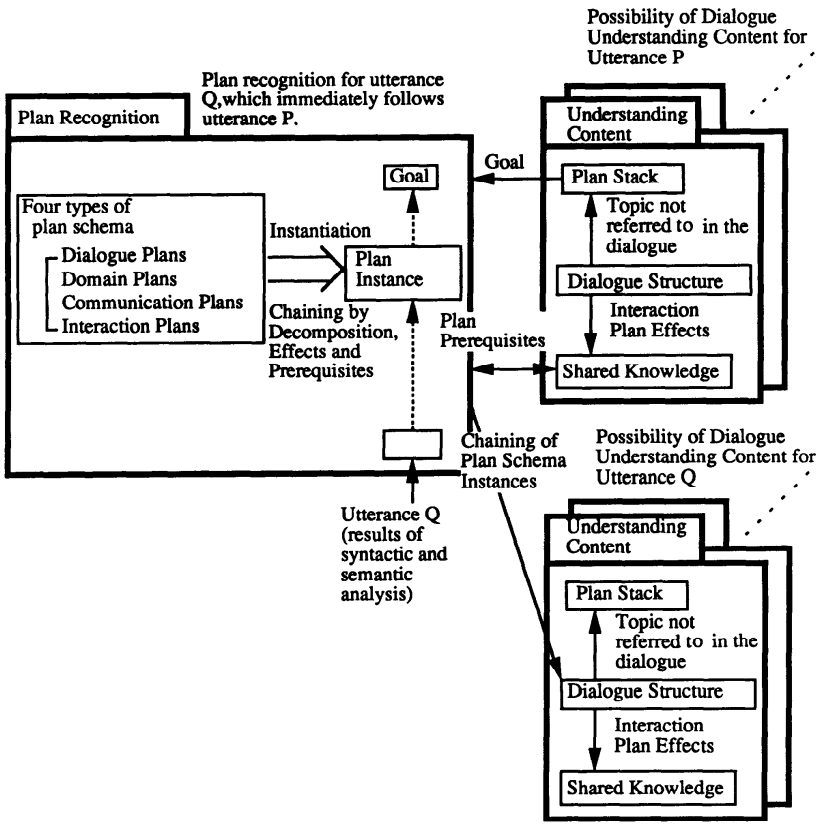


Fig. 6 Diagram of the experimental system.

analyzer of surface illocutionary speech act types [20][1] from the results output by an active chart parser [19] based on a unification-based grammar.[2] The plan recognition part of the system receives, as input, utterance representations from the analysis part. Taking a plan from the plan stack as the goal, the system obtains a chain of plan schema instances while referring to the items of shared knowledge. While the chain is being produced, the search space of possible chains is restricted by means of the hierarchical relationship described in Section 3, as follows:

Interaction Plans ≪ Communication Plans ≪ Domain Plans ≪ Dialogue Plans

Here, A ≪ B indicates that for any set of the plan A, the plan B, which ranks higher than A and includes higher sets than B, is searched, as well as the plan schemata that belong to A. In this way, excess searching of plan schemata is avoided, recognition proceeds on the basis of the defined hierarchical relationship, and the utterance plan recognition process is clarified.

The domain plans among the plan schemata used in the experiment contain knowledge that depends on the domain of the target dialogue. Thus five domain plans are described on the basis of the experiment's target dialogue: one for "conference registration," and one each for the four subordinate actions: "sending registration forms," "filling our registration forms," "sending back registration forms," and "paying participation fees." Other plans, which are not peculiar to the discourse domain of this experiment, are monitoring plans generally related to cooperative dialogue from the concepts of communicative acts and turn-taking pairs. Thus, 23 interaction plans, 17 communication plans, and one dialogue plan are specified, following the examples in Section 3.2 and Section 3.4.

The dialogue understanding contents consist of plan stacks which hold certain topics that may be referred to in the dialogue, discourse structures made up of chains of plan schemata, and items of knowledge shared by the dialogue participants.

Plan recognition is performed utterance by utterance, rather than after all utterances have been input. Thus, if plan recognition is performed by using the knowledge included the input utterance and in the utterances produced earlier, it is possible that multiple plan schema chains may be found. In this case, all possible interpretations are kept as understanding contents. If the next input utterance can be interpreted by any of the multiple candidate understanding contents, those understanding contents are kept and the others are discarded.

[1]By supplying this mechanism with rewriting rules, it is possible to extract intention from the syntactic and semantic information in the results of the analysis.

[2]This parser currently has a lexicon of about 300 words. All of the utterances in the dialogues treated in the experiment can be syntactically and semantically analyzed by the parser.

4.2 Experiment Results

The experiment was performed for four dialogues each consisting of 15 to 17 utterances, which are part of a dialogue corpus.[1] These four dialogues share the topic of "conference registration," but differ in the manner of dialogue development. They are cooperative dialogues, with no competition of goals between the participants. It is possible for either of the participants to take the initiative in these dialogues.

An example of the output of the experimental system is shown in Fig. 7. As output, the analysis results for the input utterance (interpretation window) and the discourse structure obtained from plan recognition (discourse structure window) are displayed.

The resolution of ellipsis achieved as results by this model is described for the following example of an experimental dialogue:
(Example 9)

⋮

Inquirer: *Tourokuyoushi-wo okuttekudasai.* [Please send the registration forms.]
Office: *Hai, wakarimashita.* [I will.]
Okurisaki-wo onegaishimasu. [Your name and address, please?]
Inquirer: *Oosaka higashiku* OOO, *Suzuki Mayumi desu* [Mayumi Suzuki, OOO Higashi-ku, Osaka.]
Shimekiri-wa itsu desuka. [When is the deadline?]
Office: 12 *gatsu* 25 *nichi desu node, oisogikudasai.* [December 25th, so please hurry.]

⋮

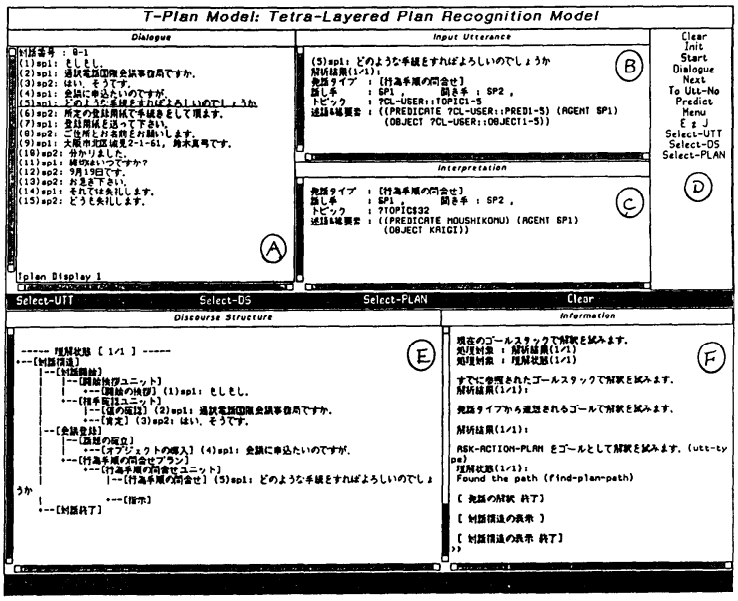
The discourse structure obtained in the experiment by plan recognition of the above dialogue is shown in Fig. 8. Each utterance is related to an interaction plan and a communication plan through a domain plan. That there is nothing below Fill-Out-Form in the diagram indicates that no utterance related to Fill-Out-Form, a sub-plan of Make-Registration, appeared in the dialogue.

By using the correspondences among the variables of the plan schema instances that make up the dialogue structure, the system was able to interpret each utterance as having restored to it the parts indicated by parentheses in the dialogue below:

⋮

Inquirer: *Tourokuyoushi-wo okuttekudasai.* [Please send the registration forms.]
Office: *Hai, (jimukyoku-ga shitsumonsha-ni tourokuyoushi-wo okurukoto-ga) wakari-*

[1]Regarding the collection and analysis of the corpus, see Arita et al.: "Media-dependent Conversation Modes—A Comparison of Telephone and Keyboard Conversations," IPS Japan SIG Reports NL61-5 (1987).



(A) Dialogue display window; The entire target dialogue is displayed.
(B) Input display window: The internal representation of the utterance to be processed is displayed.
(C) Interpretation display window: The utterance interpretation results are displayed. Case elements not included in the utterance are restored.
(D) Command Menu: A menu of the experimental system commands.
(E) Discourse structure display window: The discourse structures are displayed.
(F) Processing information window: The system processing status is displayed.

Fig. 7 Example of the output of the experimental system.

mashita. [Yes, I will (send the registration forms to the inquirer).]
Okurisaki-wo onegaishimasu. [Your name and address, please?]

Inquirer: (*Okurisaki-wa*) *Oosaka Higashiku* OOO, *Suzuki Mayumi desu* [(My name is) Mayumi Suzuki, (my address is) OOO Higashi-ku, Osaka.]
Shimekiri-wa itsudesuka. [When is the deadline?]

Office: (*Shimekiri-wa*) 12 *gatsu* 25 *nichi desu node*, (*shitsumonsha-ga jimukyoku-ni touroku-youshi-wo okurukoto-wo*) *oisogikudasai*. [(The deadline is) December 25th, so please hurry (to return the registration forms to us).]

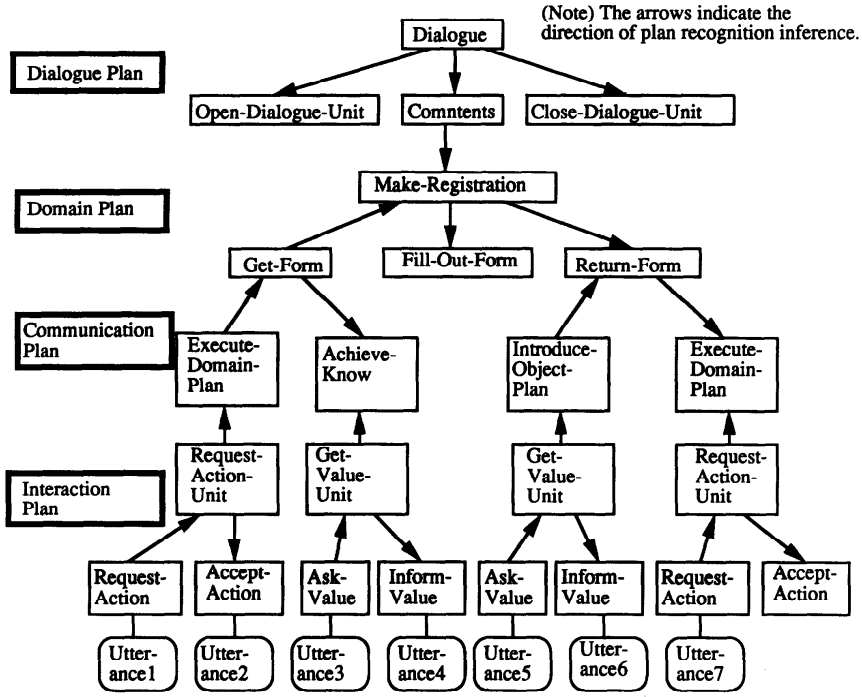
:
:

On the other hand, the confirmation utterance "*wakarimashita*" ["we understand/all right"], for example, permits multiple understanding contents. Thus, no actual interpretation is made and the processing efficiency decreases. That is, if in the dialogue preceding the appearance of "*wakarimashita*," exchanges concerning a number of topics have occurred, and there is no utterance of confirmation related to any of them, then rather than forming an interpretation in which all of the topics are confirmed as a whole, multiple understanding contents are generated, in which each of the exchanges is separately confirmed. In subsequent processing, these multiple understanding contents must be referred to, which decreases the efficiency.

5. Conclusion

This paper has described a dialogue understanding method that uses a fourlayered plan recognition model. Interaction plans, which represent knowledge about dialogue turn-taking, and communication plans, which represent knowledge related to execution of communicative acts, are used to clarify the correspondence between each utterance and domain plans, which hold knowledge about the dialogue domain. This makes it possible to understand fragmental utterances that use substitutional expressions, case element ellipsis, and so on. The effectiveness of this model in ellipsis resolution was demonstrated by a dialogue understanding experiment.

The interaction plan, communication plan, and dialogue plan desscribed in this paper do not depend on the dialogue domain; they represent pragmatic



utterance1: Inquirer: touroku-youshi wo okuttekudasai [Please send me the registration forms.]
utterance2: Office: hai, wakarimashita [All right.]
utterance3: okurisaki wo onegaishimasu [Will you give me your name and address ?]
utterance4: Inquirer: Oosakashi ....., Suzuki Mayumi desu
[My name is Suzuki Mayumi and address is .....]
utterance5: shimekiri wa itsu desuka [When is the deadline ?]
utterance6: Office: 12gatsu 25nichi desu [December 25.]
oisogi kudasai [Please return it as soon as possible.]

Fig. 8 Example of a discourse structure.

knowledge for achieving cooperative dialogue. Thus they can also be applied to other cooperative dialogues, such as consultation on travel or events, and instructions for object assembly and the like.

To extend the application range of this model, we suggest the following topics for future research. The domain plan reflects knowledge of the real world, so a mechanism that allows expansion while maintaining domain dependency is required. As regards plan search controls, it is necessary to allow controls that are limited to the subplans of a given plan to be applied to other plans. This is a matter of creating a flexible understanding mechanism that allows topic introduction with an interaction process. Further, there is a need for study of plan recognition methods that can handle discourse entities [21] in dialogue and Pollack's processing of dialogues containing errors in question premises [22]. In addition, research should be done on reasoning mechanisms that use analogical inference to understand topics not included in the domain plan.

Acknowledgement

We thank Dr. Akira Kurematsu, President of the ATR Interpreting Telephony Research Laboratories, and Dr. Teruaki Aizawa, the former Director of our Natural Language Understanding Department, for the opportunity to do this research and for their pertinent advice. We also thank our colleague Mr. Takayuki Yamaoka, for his unsparing help in working out the details of the model and in construction of the experimental system, and all of the other people in the department for their enthusiastic discussions.

References

1. ALLEN, J. F. and PERRAULT, C. R. Analyzing Intention in Utterances, *Artif. Intell.*, **15** (1980), 143–178.
2. PERRAULT, C. R. and ALLEN, J. F. A Plan-Based Analysis of Indirect Speech Acts, *Am. J. Comput. Linguist.*, **6** (1980), 167–182.
3. LITMAN, D. J. and ALLEN, J. F. A Plan Recognition Model for Subdialogues in Conversations, *Cognitive Science*, **11** (1987), 163–200.
4. GROSZ, B. and SIDNER, C. L. Attention, Intentions, and Structure of Discourse, *Computational Linguistics*, **12**, 3 (1986), 175–204.

5. IIDA, H. Pragmatic Characteristics of Natural Spoken Dialogues and Dialogue Processing Issues, *J. JSAI*, **3**, 4 (1988) (In Japanese), 445-452.
6. NOGAITO, I. and IIDA, H. Noun Phrase Identification in Dialogue and Its Application, *2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages* (1988).
7. LEECH, G. N. *Principles of Pragmatics*, p. 241, Longman (1983).
8. YAMANASHI, M. *Hatsuwa Koui* (Speech Acts), p. 245, Taishukan Shoten (1986) (In Japanese).
9. LEVINSON, S. C. *Pragmatics*, p. 420, Cambridge University Press (1983).
10. WILENSKY, R. *Planning and Understanding*, p. 168, Addison-Wesley (1983).
11. ALLEN, J. F. *Natural Language Understanding*, p. 574, Benjamin/Cummings Publishing Company (1987).
12. COHEN, P. R. The Pragmatics of Referring and the Modality of Communication, *Computational Linguistics*, **10**, 2 (1984), 97-146.
13. INOUE, K. Bun-Bunpou kara Danwa-Bunpou e (From Sentence Grammar to Dialogue Grammar), *Gengo*, **12** (1983) (In Japanese), 38-46.
14. GROSZ, B. J. Focusing and Description in Natural Language Dialogues, *Elements of Discourse Understanding*, Webber, B., Joshi, A. and Sag, I. (eds.), Cambridge University Press (1981), 84-105.
15. KATO, T. and NAKAGAWA, Y. Method for Intention Understanding and Topic Management in Natural Language Interface System, *Trans. IPS Japan*, **29**, 9 (1988) (In Japanese), 815-823.
16. CARBERRY, S. A Pragmatics-Based Approach to Ellipsis Resolution, *Computational Linguistics*, **15**, 2 (1989), 75-96.
17. WALTZ, D. L. An English Language Question Answering System for a Large Relational Database, *Comm. ACM*, **21**, 7 (1978), 526-539.
18. HENDRIX, G. G. Human Engineering for Applied Natural Language Processing, *Proceedings of 5th IJCAI* (1977), 183-191.
19. KOGURE, K. and IIDA, H. A Method of Analyzing Japanese Speech Act Types, *2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages* (1988).
20. KUME, M. et al. A Descriptive Framework for Translating Speaker's Meaning—Towards a Dialogue Translation System between Japanese and English, *European Chapter of ACL 89* (1989).
21. WEBBER, B. L. So What Can We talk About Now?, *Computational Models of Discourse*, Chapter 6, The MIT Press (1983), 331-371.
22. POLLACK, M. E. A Model of Plan Inference That Distinguishes between the Beliefs of Actors and Observers, *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics* (1986), 207-214.