Dialogue Management with the Semiotic Base: A Systemic Functional Linguistic Approach

Toru Sugimoto, Noriko Ito, Hiroko Fujishiro, and Michio Sugeno

RIKEN Brain Science Institute

2-1 Hirosawa, Wako, Saitama 351-0198, Japan

email:{sugimoto, itoh, msgn}@brain.riken.go.jp

***Abstract*—We propose a new dialogue management model that incorporates contextual and linguistic features of utterances. Our model is based on the use of a systemic functional linguistic resource, the Semiotic Base, which enables the unified systematic treatment of contextual, semantic and lexicogrammatical features of language. The Semiotic Base is used to identify the situation type of a user's input text, to set up resources in accordance with the situation type, and to analyze and generate the text. The result of the analysis is represented with a set of linguistic features, which is used by the plan module to identify the stage in a dialogue, and to select the following moves of the system. In this paper, we explain how our dialogue management model with the Semiotic Base is applied to an intelligent secretary agent system that helps a user manipulating application software, and discuss the expected advantages of our approach.**

***Index Terms*—Dialogue Management, Natural Language Processing, Planning, Systemic Functional Linguistic Theory.**

I. Introduction

Flexible natural language dialogue systems, in particular, multi-purpose dialogue systems such as intelligent secretary agents and dialogue systems that deal with complex tasks involving planning and negotiation, should be able to identify the current dialogue context and behave appropriately according to it. Some works have been done on multi-purpose dialogue systems [1][2] and portable dialogue systems with generic domain models [3]. However, they do not provide a well-motivated classification of contextual features and do not specify the systematic relationship between context and language use in it.

Systemic functional linguistic theory (hereafter, SFLT) deals with context, meaning and wording in a unified way by modeling them in system networks [4]. It describes language use with reference to the context or situation of a dialogue. It also distinguishes three metafunctions of language, called ideational, interpersonal and textual. Thus, the range of SFLT is comprehensive compared to other linguistic theories. Although SFLT has been used as the basis for many natural language generation systems [5][6], little work has been done for systems designed for natural language understanding.

Using SFLT, we are currently developing the *Semiotic Base* (hereafter, SB), which is a database of linguistic knowledge that can be used to understand and generate natural language texts [7]. The SB consists of the Context Base, the Meaning Base, the Wording Base, and the Expression Base, corresponding to the stratal organization of a language in context. This is a fundamental part of our research called Everyday Language Computing project [8], which aims at realizing a new computational framework where various types of computations can be done through everyday language.

In this paper, we introduce a model of dialogue management for an intelligent secretary agent that helps a user to manipulate application software through natural language. We give a brief description of the SB, and present a plan-based dialogue management model that uses the result of the linguistic analysis with the SB. Dealing with a short example dialogue about manipulation of a word processor, we explain how contextual and lexicogrammatical features of utterances are used to manage a dialogue between a user and a secretary.

II. Overview of the Model

The architecture of the secretary agent proposed in this paper is depicted in Fig. 1. Here, we give a rough explanation of the processing flow for a user's input text.

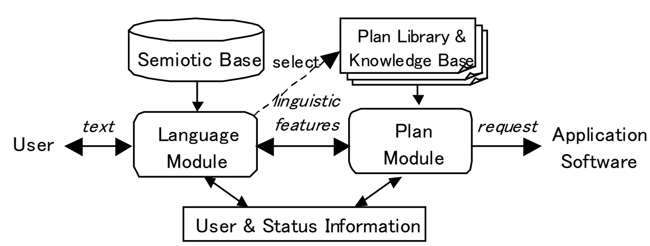

Fig. 1. Architecture of the secretary agent

A user enters a natural language text to demand assistance in doing his task with application software. He may inform the system of his goals, request the system to manipulate specific software on behalf of him, and/or ask a question about how to proceed or manipulate software. The input text is analyzed by the language module referring to the SB. At the same time, the current *situation type* is identified or updated on the linguistic basis, and it is used to select an appropriate part of the linguistic resource, the plan library and the knowledge base. After the analysis of the input text, a set of systemic functional linguistic features of the input text is sent to the plan module, where it is used to identify the current stage in the plan structure. Referring to the plan library, user information, and the current status information, the plan module selects the following moves. Each move may indicate a request to application software or a set of systemic functional linguistic features for a reply by the secretary to the user. Finally, the language module uses the SB again to realize the surface linguistic form from the features supplied from the plan module, and it outputs the resulting texts.

III. Semiotic Base

A. *Structure of the Semiotic Base*

As mentioned in Section I, the SB has the following four components: (i) the *Context Base* (hereafter, CB), which stores the features characterizing a given situation of dialogue and selection constraints on semantics specifying which semantic features are relevant to a given situation type: (ii) the *Meaning Base* (hereafter, MB), which stores features depicting the meanings associated with a situation type and constraints on lexicogrammar specifying which lexicogrammatical features are available in realizing a particular meaning in a situation type: (iii) the *Wording Base* (hereafter, WB), which stores the features to describe dialogue in terms of Japanese lexicogrammar and constraints specifying which graphological features are available for realizing a particular lexicogrammatical features in a situation type: and (iv) the *Expression Base* (hereafter, EB), which is currently designed to deal with written texts and stores graphological features associated with rules to lay out the word list using a conventional publication language, e.g., HTML.

In addition to these main bases, the SB accommodates a machine-readable dictionary (hereafter, MRD) and corpus. In the MRD, both ordinary dictionary information on lexical items and their systemic features are stored together. A corpus stores texts that have been already analyzed by means of the resources in the CB, the MB, the WB and the EB. As the corpus grows, it is more likely that the corpus has texts with the same or similar features to the texts to be processed. Thus, the dialogue system can simplify the process of language understanding and generation by case-based inference. Fig. 2 summarizes the structure of the SB.

Class		Instance
	Context Base	Annotated corpus
MRD	Meaning Base	
	Wording Base	
	Expression Base	

Fig. 2. The Semiotic Base (linguistic resource database)

Among the components, the CB is the most important and relevant part to the discussions in the following sections. The CB provides resource for characterizations of situations in which people communicate with natural language in daily life. Any instance of situations belongs to one of a set of similar situations, constituting *situation type*. We characterize situation type in terms of “what is happening (*Field*)”, "who are taking part (*Tenor*)”, and “what part the language is playing (*Mode*)” [9]. For instance, we can define the situation type of “command-compliance dialogue between a user and the secretary agent about creating a document with a word processor” as the following set of features of Field, Tenor, and Mode:[1]

1. Field: MATERIAL ACTION [non-present: deferred] & VERBAL ACTION [constitutive: practical: plan] & SPHERE OF ACTION [specialized]
2. Tenor: [unequal power] & [frequent contact] & [low affective involvement]
3. Mode: [one-way visual contact] & [no aural-oral] & [typed] & [immediate feedback]

These features of Field, Tenor and Mode are represented in the form of *system networks*. Fig. 3 represents a part of the system network for Field, where the features selected for the situation type mentioned above are in rectangle.[2]

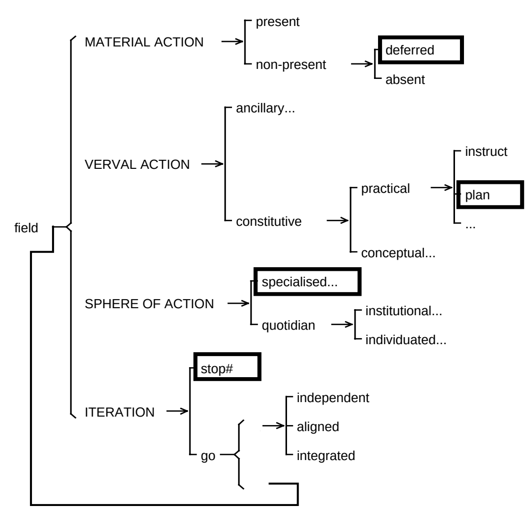

Fig. 3. System network of Field

B. *Characteristics of the Semiotic Base*

As explained above, the features in each base link to features in other bases. All the semantic, lexicogrammatical and graphological features are associated with situation types based on the results of corpus analysis. We call a set of linguistic features relevant to a situation type *register*. This is one of the most important links among the features, and referred to as situation type-linguistic register association. The same kind of association is established between the situation type and non-linguistic resource for the secretary agent such as concept-frame and plan. This assures that all the relevant

[1] For the definitions of the features concerning Field, see [10]. For Tenor and Mode, we adopt the features provided in [11].
[2] This is a simplified version of Hasan’s [10] figure of “field of discourse: some systemic choices in a language based conception of social activity.”

resource for a text processing is provided in concordance with the situation. On the other hand, it helps the language module to infer a situation type from the linguistic features of utterance. One of the ways to do this is to refer to the semantic features associated with the lexical items that we obtained through the morphological analysis in text understanding. Each semantic feature is also linked with a particular situation type. We regard the situation type associated with the most links as the current situation type.

Following SFLT, we take a distinctive perspective in design of the MB and the WB. While many linguistic theories provide different ways of representation of grammar and lexis, we describe grammar and lexis in the same manner. That is, all of the grammatical and lexical features are arranged in delicacy and represented in system network [12]. Lexicogrammatical system network in the WB starts with decision tree of grammatical features and ends with lexical features associated with lexical items (word). Accordingly, semantics is divided into two parts, grammatical-semantics and lexical-semantics. The system network in the MB starts with selection of grammatical-semantic features and ends with lexical-semantic features associated with concept realized through lexical items. This implies that concept-level ontology is linked with lexical-level ontology via grammar.

Our way of linking the two types of ontology seems plausible because we assume that mapping between lexical features and lexical-semantic ones varies according to situation types and this is the main factor that causes to ambiguity and misunderstanding. There is doubt that the mapping at the lexical and concept levels without taking into consideration the context of language use would be of use. We also assume that mapping between grammatical features and grammatical-semantic ones is rather stable and this is a basis of communication and mutual understanding among the same language speakers. This policy in modeling linguistic system is different from those taken in constructing conventional large machine readable linguistic database such as EDR [13] and Ikehara et al. [14], where their concept taxonomy is not sufficiently motivated by situation or grammar.

Moreover, we link lexical-semantic features to concept-frame in the knowledge base and plan in the plan library so that the secretary agent can utilize linguistic features of the SB in planning efficiently.

IV. DIALOGUE MANAGEMENT

In this section, we explain our plan-based model of dialogue management that uses the result of the linguistic analysis of a user's input text with the SB.

A. *Plan Structure*

The plan module maintains a plan structure, which is a model of the dialogue structure constructed by referring to the plan library prepared for the current situation type. Following the research in the area of task-oriented dialogue management where discourse-level and task-level plans are distinguished [15][16], we use a two-layered representation of plan structures: the top layer consists of *interaction plans*, which represent a discourse strategy for the secretary agent and a user to do a task cooperatively through a dialogue, and the bottom layer consists of *domain plans*, which represent possible processing flows of a task appearing in the context. Both layers of plan structures are constructed from unit templates called *plan units* stored in the plan library. A plan unit consists of a header, a body, and other features such as effects and constraints. The body of a plan unit is represented by a transition network, each of whose arc is called a *stage*. A stage is associated with a label, which is a specification of either a speech act, a domain-level action, or another plan unit. A stage is called a *move* if its label is not a plan unit. Each stage is associated also with a numeric value called a *preference value*, which represents the desirability (or the plausibility in the case of a user's speech act) of that stage. Preference values are used to control inference, and they are obtained from corpus analysis or learned through the dialogue. Fig. 4 and 5 illustrate the main parts (without detailed specifications of associated actions and preference values) of several interaction and domain plan units stored in the plan library for the situation type of command-compliance about creating a document with a word processor.

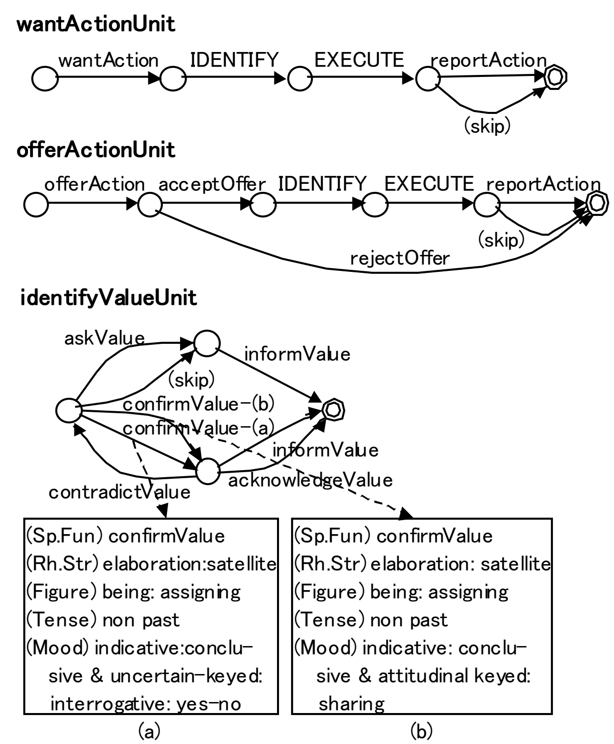

Fig. 4. Examples of interaction plans

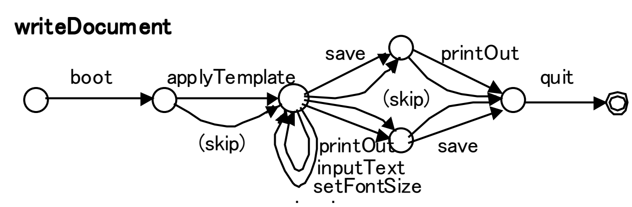

Fig. 5. Example of a domain plan

Our modeling of discourse strategies by interaction plans resembles conversational game theory [17], but it has three distinct features. First, the plan module uses only an appropriate set of plan units by selecting the plan library prepared for each situation type. In general, the content of the plan library is determined mainly by the VERBAL ACTION features in the Field and the Tenor of the current situation type, reflecting the roles of and the relationship between the secretary and the user in that context. Second, our model systematically incorporates linguistic features, which are not restricted to those traditionally used such as speech act types and propositional contents, by adding sets of systemic functional linguistic features in the SB to action specifications with which moves in interaction plans are associated. Thus, interaction plans correspond to *generic structure potential* [9], which is a kind of general linguistic structure of discourse in the context. Although we assume that linguistic features are added to each move occurring in Fig. 4, due to space limitations, we show only two sets of linguistic features, which are added to confirmValue speech acts so that more appropriate linguistic form can be selected and realized. Use of these features is discussed in Sect V.B. Third, execution of domain-level actions can be interleaved with execution of speech acts in the way determined by the context. In command-compliance situations, we use macro notations EXECUTE and IDENTIFY to connect the two layers of plan structures. For a domain plan *action*, EXECUTE(*action*) denotes the sequence of the disjunction of interaction plan units such as wantActionUnit(*a*) and offerActionUnit(*a*) for each arc *a* in the body of *action*. IDENTIFY(*action*) denotes the disjunction of all sequences composed from identifyValueUnit(*slot*) for all still unspecified *slots* of the *action*. Note that we may need other kinds of macro notations in other situations, for example, INSTRUCT(*action*) to connect a domain plan *action* with a series of speech acts for instruction prior to other's performing *action*.

The structure of a dialogue is represented by a *plan structure*, which is a tree with its nodes being stages from interaction plan units, with a pointer for the current position in the dialogue. As the dialogue proceeds, the plan structure is extended by adding nodes that are realized or planned by the secretary agent. To be precise, preparing for a user's arbitrary or unexpected switching of the topic, the plan module should maintain a *stack* ($<s_1, p_1>$, $<s_2, p_2>$, ..., $<s_n, p_n>$), where s_n is an identified situation type and p_n is a plan structure. It is also desirable for the plan module to maintain several candidates for the current dialogue context in order to enable backtracking when it recognizes an inconsistency in the dialogue.

B. Identifying and Selecting Moves

The plan module undertakes two tasks: identifying the current stage in the plan structure from a user's input text and selecting the following moves of the secretary agent itself. In doing both tasks, the plan module uses a common procedure, that is, it generates a set of moves that may follow the current pointer position of the plan structure. This procedure can be easily implemented in a similar way to top-down prediction and generation procedures for natural language processing.

After the language module analyzes a user's input text with the SB, it sends to the plan module the result of the analysis, which is represented with a set of systemic functional linguistic features including the situation type. The plan module first checks whether the situation type has been updated and then tries to match the received linguistic features with one of the predicted moves (associated with linguistic features as mentioned in the previous subsection) that may follow in the plan structure. If the situation type remains the same and the matching succeeds, then the identifying task completes. Otherwise, the plan module creates a new tree from the received linguistic features, and it tries to extend it upwards in a similar way to standard plan recognition procedures [18] using interaction plan units in the plan library, and then it pushes the identified situation type and the new tree into the stack of plan structures.

After identifying the current stage, the plan module selects the following moves from the possible candidates by considering their preference values given in the plan library and other kinds of criteria such as user information, the current status information, and linguistic features of a user's input text. If a selected move is a domain-level atomic action, then it is realized by sending a request to application software. If it is a secretary agent's speech act, then linguistic features are sent to the language module for text generation. This process is repeated until the following move is not the secretary agent's one.

V. EXAMPLE DIALOGUE

In this section, we present an example dialogue about manipulation of a word processor, which can be explained in our dialogue management model, to illustrate how contextual and linguistic features of utterances are used to manage a dialogue. Consider a dialogue between a user (U) and a secretary agent (S) shown in Fig. 6.

U1: カナダ出張の報告書を書きたい。(I want to write a report on the business trip to Canada.)
S2: ワープロを起動します。(I shall boot a word processor.)
S3: [The secretary agent boots a word processor.]
S4: 用意された書式を使いますか？ (Would you like to use a template?)
U5: はい。タイトルは「ISFC出張報告書」で。
(Yes. Please set the title "ISFC business trip report.")
S6: 日付は今日でよろしいですか？ (Is it all right if I set the date today?)
U7: いや、明日にして。(No, it should be tomorrow.)
S8: [The secretary agent applies a template.]
S9: 報告書の書式を適用しました。(O.K. I have applied a template for a report.)

Fig. 6. Example dialogue

The plan structure realized for the overall dialogue and part of the result of linguistic analysis of a user's first input text U1 with the SB are shown in Fig. 7 and Table I, respectively.

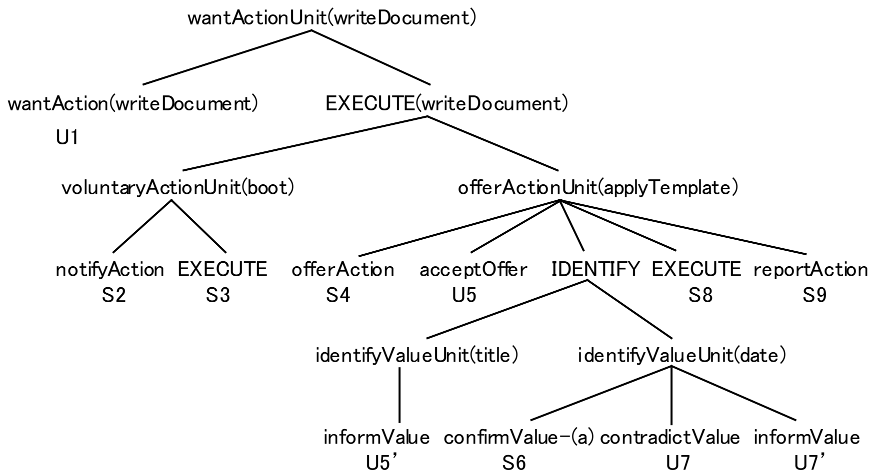

Fig. 7. Plan structure for the example dialogue

TABLE I
RESULT OF TEXT UNDERSTANDING OF U1

Contextual features	Situation Type		dialogue between a user and the secretary agent about creating a document with a word processor										
Contextual features	Field		non-present: deferred & practical & specialised (computing: with editor & creating document)										
Contextual features	Tenor		unequal power, frequent contact, low affective involvement (Speaker = user, Hearer = secretary agent)										
Contextual features	Mode		one-way visual contact, no aural-oral, typed, immediate feedback										
Semantic features	Ideational		figure: non-projected & doing: doing to sth with means>> writeDocument										
Semantic features	Interpersonal		Initiating & Giving & Information: action in the user's territory >> wantAction (speaker= user, hearer= secretary, action= wd1)										
Semantic features	Textual	Rhetorical Structure	hypotactic relation & external & expanding: enhancing: purposive → solutionhood: satelite										
Semantic features	Textual	Lexical-Conceptual Chain	N4: user= agent (name= "Riken Hanako") & trip1.participant & doc1.author & wd1.client	–	N1: canada= country (name= "Canada")	N2: trip1= businessTrip (participant= user <N4>, place= canada <N1>, fromDate= 07/22/2001 toDate= 07/29/2001, conference= isfc)	–	N3: doc1= tripReport (author= user <N4>, fileName=, wp= wp1 <N5>, trip= trip1 <N2>)	–	N5: wp1= jwp= wordProcessor (name= "javaWordProcessor", started=false)	–	V1: wd1= writeDocument (client= user <N4>, doc= doc1 <N3>, wp= wp1 <N5>)	–
LG features	Ideational		Actor/Agent		Goal/Medium					Circumstance: Manner: Means		Process: material	
LG features	Interpersonal		Subject		Complement					Adjunct		Predicator	
LG features	Interpersonal		Tense: non past, Phase: unspecified, Mood: indicative: conclusive, Modal Operator: modulation: inclination: optative: action of self										
LG features	Textual		implicit-stage Theme		Rheme								
Wording			watasi	wa	kanada	syuttyo	no	hokokusyo	o	wapuro	de	kaki	tai

A. Using Contextual Features

As dialogue unfolds, the situation becomes clear or switches to another type. As the situation is specified (traversing the network from left to right), a subset of the current resource will be foregrounded. Consider the beginning of our example dialogue. Before the user's first input U1, the situation type is considered as a default one, that is, "*dialogue between a user and the secretary agent*" with the following Field features: MATERIAL ACTION [non-present] & VERBAL ACTION [constitutive] & SPHERE OF ACTION [unspecified].[3] After the linguistic analysis of U1 with the SB (Table I), MATERIAL ACTION and SPHERE OF ACTION are specified. In particular, the specification of SPHERE OF ACTION narrows the range of domain knowledge to that of creating a document with a word processor. Finally, through identification of the stage in the plan module, the feature of VERBAL ACTION is recognized as [constitutive: practical: plan]. Consequently, the situation type is updated to "*command-compliance* dialogue between a user and the secretary agent about creating a document with a word processor", and the interaction plans associated with this situation type are activated.

When the situation switches to another one (as it moves into a different system in the network), a different set of resources will be provided. If the user inputs "Multexについて調べたい。(I want to search for Multex.)" in our document creating dialogue, the language module detects a context switching that reflects the SPHERE OF ACTION change to *specialised (computing: with browser & collecting information)*. The situation type is now identified as "command-compliance between a user and the secretary agent about *collecting information with browser*". A different set of domain knowledge concerning collecting

[3] Here, we do not mention Tenor or Mode because their features do not change in the sample dialogue.

information with a browser becomes available for further procedure. The plan module pushes a new tree into the stack of plan structures, and it manages the dialogue appropriately according to the new situation type.

B. Using Linguistic Features

The SB provides linguistic features for various purposes, for instance, identifying the current situation type and selecting the exact words to convey a given meaning. Some of the features are also useful for the plan module to identify and select moves. One of the features that influence domain plans is PHASE (interpersonal lexicogrammatical features). Assume that the speech act is recognized as wantAction or demandAction and the feature in PHASE is either *completion* (realized by "-teshimau") or *preparation in advance* (realized by "-teoku"). The secretary agent then selects moves and stages so that the user may save as much time as possible, for instance, by decreasing the preference values for optional actions. If the feature in PHASE is *conative* (realized by "-temiru"), the agent attempts to intervene in the user's decision making, by increasing the preference values for optional actions.

Other linguistic features may manifest differences in preference values. In order to convey the speaker's degree of certainty to the value with confirmValue, for example, we can select and combine such features as ATTITUDINAL-KEY and UNCERTAIN-KEY within the MOOD network (interpersonal lexicogrammatical features). The sentence S6 ends with "-desuka?", which is a lexical item that realizes *non-attitudinal-keyed & uncertain-keyed: interrogative: yes-no* (see also Fig. 4(a)). This indicates that the speaker is not confident with the value that he represents in the sentence. Instead of S6, the secretary agent could say "-desune?" that realizes *attitudinal-keyed: sharing & non-uncertain-keyed* (Fig. 4(b)). This selection implies that the agent has confidence with the value though not sufficient to skip comfirmValue action.

VI. CONCLUSION

We have proposed a new dialogue management model that exploits contextual and linguistic features of utterances. By means of the SB, the situation type of a user's input text is identified and its linguistic features are extracted. Then they serve to identify the stage in a dialogue and to select moves that the system is going to carry out.

Advantage of our model, which owe mainly to SFLT, is not only the richness of the contextual and linguistic features that we can deal with but also a mechanism for incremental specification of such features by means of system networks as exemplified in Section V.A. Moreover, the completeness of linguistic analysis, which may result in an extended version of Table I, is not required since our dialogue management model is fundamentally based on a simple matching between sets of linguistic features. These should become computational advantages of our approach when we try to develop a robust dialogue system. Furthermore, we expect the organization of the SB with such strong connections among contents inside and outside of the SB to make knowledge management and acquisition through texts readily.

We are now engaged in implementation of the text understanding algorithm with which the language module extracts contextual, semantic, and lexicogrammatical features of the input texts using the system networks and the links stored in the SB. As far as the current state of art in the area of natural language processing is concerned, complete automatic analysis of texts seems difficult due to the comprehensive nature of our description of language. We are investigating methods based on the analysis of our original corpus and should provide tools for this problem.

REFERENCES

[1] S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid, and V. Zue, "Galaxy-II: a reference architecture for conversational system development," in *Proc. of ICSLP 98*, Nov. 1998.

[2] M. Mori, H. Dohi, and M. Ishizuka, "A multi-purpose dialogue management system employing visual anthropomorphous agent," in *Proc. of RO-MAN '95*, pp.187-192, Tokyo, 1995.

[3] G. Ferguson and J. Allen, "TRIPS: an intelligent integrated problem-solving assistant," in *Proc. of 15th National Conference on Artificial Intelligence (AAAI-98)*, pp.567-573, 1998.

[4] M. A. K. Halliday, *An Introduction to Functional Grammar*. 2nd ed., Edward Arnold, London, 1994.

[5] M. A. K. Halliday and C. M. I. M. Matthiessen, *Construing Experience through Meaning: a Language-based Approach to Cognition*. Cassell, London, 1999.

[6] C. M. I. M. Matthiessen, L. Zeng, M. Cross, I. Kobayashi, K. Teruya, and C. Wu, "The Multex generator environment: application and development," *The 9th International Workshop on Natural Language Generation*, pp.228-237, Association for Computational Linguistics, 1998.

[7] N. Ito, I. Kobayashi, and M. Sugeno, "The Semiotic Base as a resource in text processing systems," *JASFL Occasional Papers*, vol.2, no.1, pp.63-71, 2001.

[8] I. Kobayashi, M. Iwazume, M. Sugeno, and N. Ito, "Toward a computational environment for everyday language communication," in *Proc. of Joint 9th IFSA World Congress and 20th NAFIPS*, International Conference, pp.663-668, 2001.

[9] M. A. K. Halliday and R. Hasan, *Language, Context, and Text: Aspects of Language in a Socio-semiotic Perspective*. Oxford University Press, London, 1989.

[10] R. Hasan, "Speaking with reference to context," in M. Ghadessy, ed., *Text and Context in Functional Linguistics*, pp.219-329, John Benjamins, Amsterdam Philadelphia, 1999.

[11] S. Eggins, *An Introduction to Systemic Functional Linguistics*. Pinter, London, 1994.

[12] R. Hasan, "The grammarian's dream: lexis as most delicate grammar," in *Ways of Saying, Ways of Meaning: Selected Papers of Ruqaiya Hasan*, C. Cloran, D. Butt, and G. Williams, Eds., London, Cassel, 1996, pp.73-103.

[13] EDR, *Electronic dictionary technical guide*. Japan Electronic Dictionary Research Institute Ltd., Tokyo TR-042, 1993.

[14] S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Oyama, and Y. Hayasi, *Japanese Lexicon*. Tokyo, Iwanami, 1997.

[15] D. J. Litman and J. F. Allen, "A plan recognition model for subdialogues in conversation," *Cognitive Science*, vol.11, pp.163-200, 1987.

[16] T. Yamaoka and H. Iida, "Dialogue interpretation model and its application to next utterance prediction for spoken language processing," in *Eurospeech '91, Proc. of 2nd European Conference on Speech Communication and Technology*, pp.849-852, Genova, Italy, 1991.

[17] J. C. Kowtko, S. D. Isard, and G. M. Doherty, "Conversational games within dialogue," Technical report, Human Communication Research Centre, University of Edinburgh, 2, Nov. 1993.

[18] S. Carberry, *Plan Recognition in Natural Language Dialogue*. MIT Press, 1990.