Dialog management in the Waxholm system

Rolf Carlson and Sheri Hunnicutt
Department of Speech Communication and Music Acoustics,
KTH, Box 70044, S 100 44 Stockholm, Sweden

ABSTRACT

Recently we have begun to build the basic tools for a generic speech-dialog system. A preliminary version of the system has been tested, using simplified versions of the modules. The dialog component of the system is described by a dialog grammar with the help of semantic features. Probabilities are also used in this process. We will give a general overview of the system and describe the dialog component in more detail.

INTRODUCTION

Our research group at KTH[1] is currently building a generic system in which speech synthesis and speech recognition can be studied in a man-machine dialog framework. In addition, the system should facilitate the collection of speech and text data that are required for development. The system was first presented at FONETIK 93 (Blomberg et. al. 1993a) and at the Eurospeech '93 conference (Blomberg et. al. 1993b). The dialog management component has recently been reformulated in a more general framework and is the focus of our presentation.

The demonstrator application, which we call WAXHOLM, gives information on boat traffic in the Stockholm archipelago. It references time tables for a fleet of some twenty boats from the Waxholm company which connects about two hundred ports. In addition to boat time-tables the database also contains information about port locations, hotels, camping places, and restaurants in the Stockholm archipelago.

Besides the speech recognition and synthesis components, the system contains modules that handle graphic information such as pictures, maps, charts, and time-tables. This information can be presented to the user at his/her request. The application has great similarities to the ATIS domain within the ARPA community and other similar tasks in Europe, for example SUNDIAL. The possibility of expanding the task in many directions is an advantage for our future research on interactive dialog systems.

NATURAL LANGUAGE COMPONENT

Our initial work on a natural language component is focused on a sublanguage grammar, a grammar limited to a particular subject domain: that of requesting information from a transportation database.

The fundamental concepts are inspired by TINA, a parser developed at MIT (Seneff 1989). Our parser, STINA, i.e., Swedish TINA, is knowledge-based and is designed as a

[1] The Waxholm group consists of staff and students at the Department of Speech Communication and Music Acoustics, KTH. Most of the efforts are done part time. The members of the group in alphabetic order are: Mats Blomberg, Rolf Carlson, Kjell Elenius, Björn Granström, Joakim Gustafson, Sheri Hunnicutt, Jesper Högberg, Roger Lindell, Lennart Neovius, Lennart Nord, Antonio de Serpa-Leitao and Nikko Ström.

probabilistic language model (Carlson and Hunnicutt 1992). It contains a context-free grammar which is compiled into an augmented transition network (ATN). Probabilities are assigned to each arc after training. Features of STINA are a stack-decoding search strategy and a feature-passing mechanism to implement unification.

In the implementation of the parser and the dialog management, we have stressed an interactive development environment. This makes it easier to have control over the system's progress as more components are added. It is possible to study the parsing and the dialog flow step by step when a tree is built. It is even possible to use the collected log files as scripts to repeat a collected dialog including all graphic displays and acoustic outputs.

Lexicon and Features

The lexicon entries are generated by processing each word in the Two-Level Morphology (TWOL) lexical analyzer (Koskenniemi 1983, Karlsson 1990). Each entry is then corrected by removing all unknown homographs. New grammatical and semantic features, which are used by our algorithm and special application, are then added.

The basic grammatical features can be positive, negative or unspecified. Unspecified features match both positive and negative features.

Semantic features can be divided into two different classes. The basic features like BOAT and PORT give a simple description of the semantic property of a word. These features are hierarchically structured. During the unification process in STINA, all features which belong to the same branch are considered. Thus, a unification of the feature PLACE engages all subordinate semantic features (in our case, REGION and ISLAND).

Another type of semantic feature controls which nodes can be used in the syntactic analysis. For example, the node DEPARTURE TIME cannot be used in connection with verbs that imply an arrival time. This is also a powerful method to control the analysis of responses to questions from the dialog module. The question "Where do you want to go?" conditions the parser to accept a simple port name as a possible response from the user.

Dialog Management

Dialog management based on grammar rules and lexical semantic features has recently been implemented in STINA. The notation to describe the syntactic rules has been expanded to cover some of our special needs to model the dialog. The STINA parser is running with two different time scales during data collection corresponding both to the words in each utterance and to the turns in the dialog. Syntactic nodes and dialog states are processed according to transition networks with probabilities on each arc.

Each dialog topic is explored according to the rules. These rules define which constraints have to be fulfilled and what action should be taken depending on the dialog history. Each dialog node is specified according to Figure 1.

The constraint evaluation is described in terms of features and the content in the semantic frame. If the frame needs to be expanded with additional information, a system question is synthesized. During recognition of a response to such a question the grammar is controlled with semantic features in order to allow incomplete sentences. If the response from the subject does not clarify the question, the robust parsing is temporarily disconnected so that specific information can be given to the user about syntactic or unknown word problems. At the same time a complete sentence is requested giving the dialog manager the possibility of evaluating whether the chosen topic is a bad choice.

Node types:	Node functions:
branching or preterminal	record utterance
	synthesize message
Constraint evaluation on:	test constraints
dialog flow features	data base search using SQL
semantic frame slots and features	graphic display table
....	graphic display picture
If more information needed:	
synthesize question to user	
accept incomplete sentence	

Figure 1. Dialog node specification.

A positive response from the constraint evaluation clears the way for the selected action to take place. The node function list in the figure gives examples of such actions.

Topic Selection

In Figure 2 some of the major topics are listed. The decision about which path to follow in the dialog is based on several factors such as the dialog history and the content of the specific utterance. The utterance is coded in the form of a "semantic frame" with slots corresponding to both the grammatical analysis and the specific application. The structure of the semantic frame is automatically created based on the rule system.

TIME TABLE: Goal: to get a time-table presented with departure and arrival times specified between two specific locations.
Example: När går båten? (When does the boat leave?)
SHOW MAP: Goal: to get a chart or a map displayed with the place of interest shown.
Example: Var ligger Vaxholm? (Where is Vaxholm?)
FACILITY: Goal: to display the availability of lodging and dining possibilities.
Example: Var finns det vandrarhem? (Where are there hostels?)
OUT OF DOMAIN: Goal: Inform the user that the subject is out of the system's domain.
Example: Kan jag boka rum. (Can I book a room?)

Figure 2. Topic examples.

Each semantic feature found in the syntactic and semantic analysis is considered in the form of a conditional probability to decide on the topic. (See Figure 3.) The probability for each topic is expressed as: p(topic|F), where F is a feature vector including all semantic features used in the utterance. Thus, the TIME feature can be a strong indication for the TIME-TABLE topic but this can be contradicted by a HOTEL feature.

Introduction of a New Topic

The rule-based and to some extent probabilistic approach we are exploring makes the addition of new topics relatively easy. Suppose we want to create a topic called "weather information." First a topic node is introduced in the rule system. Some words will need to be included in the lexicon and labelled with a semantic feature showing that the system does not know how to deal with the subjects these words relate to. Then a synthesis node might be added with a text informing the user about the situation. Example sentences must

FEATURES ↓ / TOPICS →	TIME TABLE	SHOW MAP	FACILITY	NO UNDER-STANDING	OUT OF DOMAIN	END
OBJECT	.062	.312	.073	.091	.067	.091
QUEST-WHEN	.188	.031	.024	.091	.067	.091
QUEST-WHERE	.062	.688	.390	.091	.067	.091
FROM-PLACE	.250	.031	.024	.091	.067	.091
AT-PLACE	.062	.219	.293	.091	.067	.091
TIME	.312	.031	.024	.091	.067	.091
PLACE	.091	.200	.500	.091	.067	.091
OOD	.062	.031	.122	.091	.933	.091
END	.062	.031	.024	.091	.067	.909
HOTEL	.062	.031	.488	.091	.067	.091
HOSTEL	.062	.031	.122	.091	.067	.091
ISLAND	.333	.556	.062	.091	.067	.091
PORT	.125	.750	.244	.091	.067	.091
MOVE	.875	.031	.098	.091	.067	.091

Figure 3. Topic probability matrix.

be created that illustrate the problem. The dialog parser must be trained with these sentences labelled with the "weather information" topic.

FINAL REMARKS

No module in the Waxholm system is yet considered complete. The dialog management module still needs to be tested in a more hostile environment.

In addition we are currently testing a simple application-independent grammar on unlimited text. This system will also be used as part of our general text-to-speech system, which is outside the scope of this presentation.

ACKNOWLEDGEMENT

This work has been supported by The Swedish National Language Technology Program.

REFERENCES

Blomberg, M., Carlson, R., Elenius, K., Granström, B., Gustafson, J., Hunnicutt, S., Lindell, R., and Neovius, L. (1993b): "An experimental dialog system: WAXHOLM," Proceedings of Eurospeech '93. pp 1867-1870.

Blomberg, M., Carlson, R., Elenius, K., Granström, B., Hunnicutt, S., Lindell, R., and Neovius, L. (1993a): "An experimental dialog system: WAXHOLM," The Seventh Swedish Phonetics Conference, RUUL 23, Ling. Dept., Uppsala University, pp. 49-52.

Carlson, R., & Hunnicutt, S. (1992): "STINA: A probabilistic parser for speech recognition," Sixth Swedish Phonetics Conference, Technical Report No. 10, Dept. of Information Theory, Chalmers University of Technology, Göteborg. pp 23-26.

Karlsson, F. (1990): "A Comprehensive Morphological Analyzer for Swedish," manuscript, University of Helsinki, Department of General Linguistics.

Koskenniemi, K. (1983): "Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production," University of Helsinki, Department of General Linguistics, Publications No. 11.

Seneff, S. (1989): "TINA: A Probabilistic Syntactic Parser for Speech Understanding Systems," Proceedings ICASSP-89, pp. 711-714.