THE DIALOG COMPONENT IN THE WAXHOLM SYSTEM

Rolf Carlson
Department of Speech, Music and Hearing,
Box 70014 KTH,
S-10044 Stockholm, Sweden
rolf@speech.kth.se, www.speech.kth.se

ABSTRACT

In this paper we give an overview of the NLP and dialog component in the Waxholm spoken dialog system. We will discuss how the dialog and the natural language component are modeled from a generic and a domain-specific point of view. Dialog management based on grammar rules and lexical semantic features is implemented in our parser. The notation to describe the syntactic rules has been expanded to cover some of our special needs to model the dialog. The parser is running with two different time scales corresponding to the words in each utterance and to the turns in the dialog. Topic selection is accomplished based on probabilities calculated from user initiatives. Results from parser performance and topic prediction are included in the presentation.

INTRODUCTION

Our research group at KTH has, for some years, been building a generic system in which speech synthesis and speech recognition can be studied in a man-machine dialog framework. The demonstrator application, 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. The system has been presented on several occasions, for example, the Eurospeech '93 conference (Blomberg et al., 1993), the ARPA meeting '94 (Carlson, 1994) and the ETRW on Spoken Dialog Systems (Bertenstam et al., 1995a).

Besides the speech recognition and synthesis components, the system contains modules that handle graphic information such as pictures, maps, charts, and time-tables (Figure 2). This information can be presented to the user at his/her request. The possibility of expanding the task in many directions is an advantage for our future research on interactive dialog systems. In addition to boat time-tables, the database also contains information about port locations, hotels, camping places, and restaurants in the Stockholm archipelago. This information is accessed by SQL, the standardized query language. An initial version of the system has been running since September 1992.

The application has similarities to the ATIS domain within the ARPA community, the Voyager system from MIT (Glass et al., 1995) and similar tasks in Europe, for example SUNDIAL (Peckham, 1993), the systems for train timetables information developed by Philips (Aust et al., 1994; Oerder and Aust, 1994) and CSELT (Clementino, and Fissore, 1993; Gerbino and Danieli, 1993; and flight information in the Danish Dialog Project, Dalsgaard and Baekgaard, 1994).

Spoken dialog management has attracted considerable interest during the last years. Special workshops and symposia, for example the special workshop at Waseda University, Japan 1993 (Shirai and Furui, 1995), the Twente Workshops on Language Technology in Enschede, The Netherlands, and the 1995 ESCA workshop on Spoken Dialog Systems in Vigsø, Denmark, have all been arranged to forward research in this field. We will not attempt to review this growing field in this paper. We will, however, describe in some detail the current effort to model the dialog in the Waxholm system.

Our objective is to develop a dialog management module which can handle the type of interaction that can occur in our chosen domain. The Waxholm system should allow user initiatives, without any specific instructions to the user, complemented by system questions to achieve the user’s goal. We will discuss how the dialog and the natural language component are modeled from a generic and a domain-specific point of view.

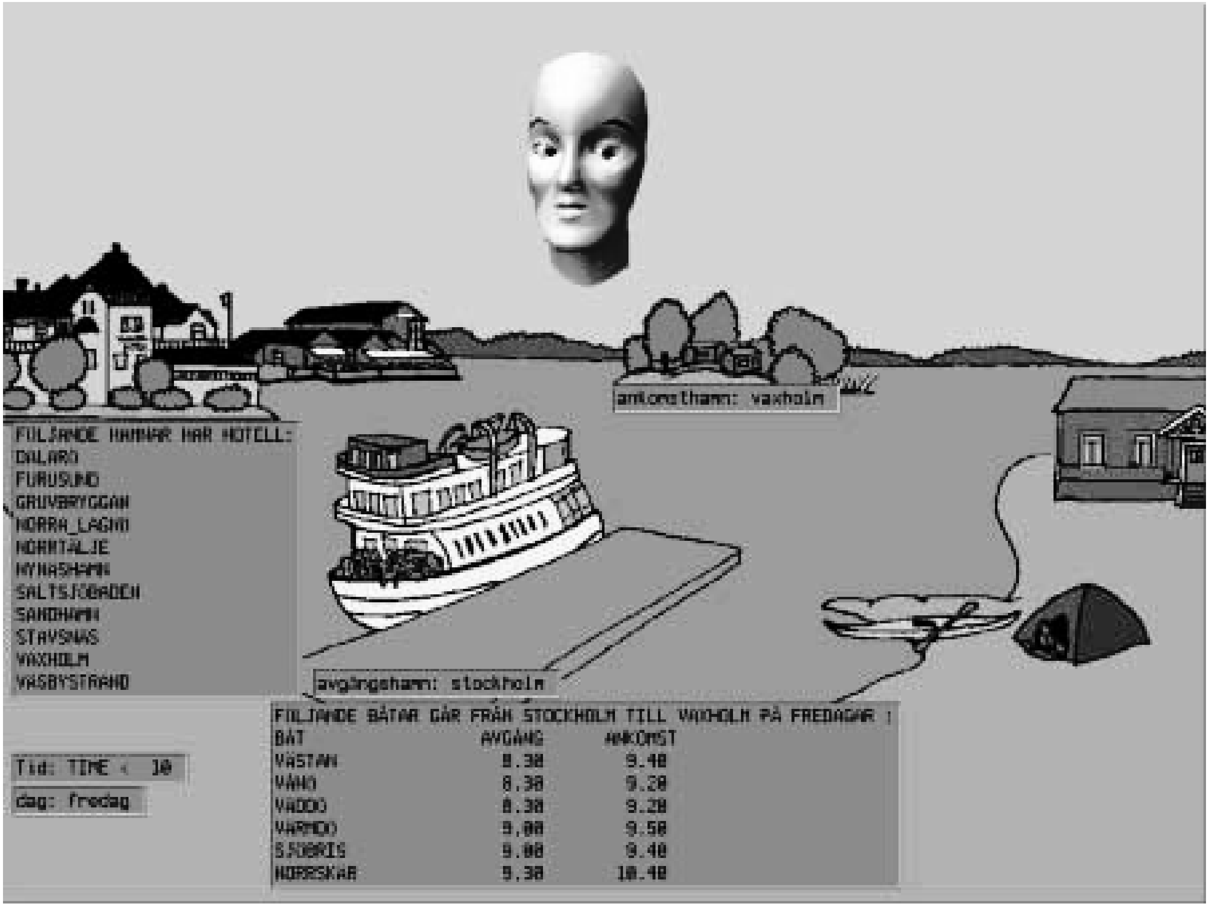

Figure 1. The graphical model of the WAXHOLM micro-world.

GRAPHICAL INTERFACE

The Waxholm system can be viewed as a micro-world, consisting of harbors with different facilities and with boats that you can take between them. The user gets graphic feedback in the form of tables complemented by speech synthesis. Up to now the subjects have been given a scenario with different numbers of subtasks to solve. A problem with this approach is that the subjects tend to use the same vocabulary as the text in the given scenario. We also observed that the user often did not get enough feedback to be able to decide if the system had the same interpretation of the dialogue as the user. To deal with these problems a graphical representation that visualizes the Waxholm micro-world is being implemented. An example is shown in Figure 1. One purpose of this is to give the subject an idea of what can be done with the system, without expressing it in words. Another purpose is that the interface continuously feeds back the information that the system has obtained from the parsing of the subject's utterance, such as time, departure port and so on. The interface is also meant to give a graphical view of the knowledge the subject has secured thus far, in the form of listings of hotels and so on.

For the speech-output component we have chosen our multi-lingual text-to-speech system (Carlson, Granström and Hunnicutt, 1991). The system is modified for this application. In dialog applications such as the Waxholm system we have a better base for prosodic modeling compared to ordinary text-to-speech, since, in such an environment, we will have access to much more information than if we used an unknown text as input to the speech synthesizer. The speech synthesis has recently been complemented with a face-synthesis module (Beskow, 1995). Both the visual and the speech synthesis are controlled by the same synthesis software.

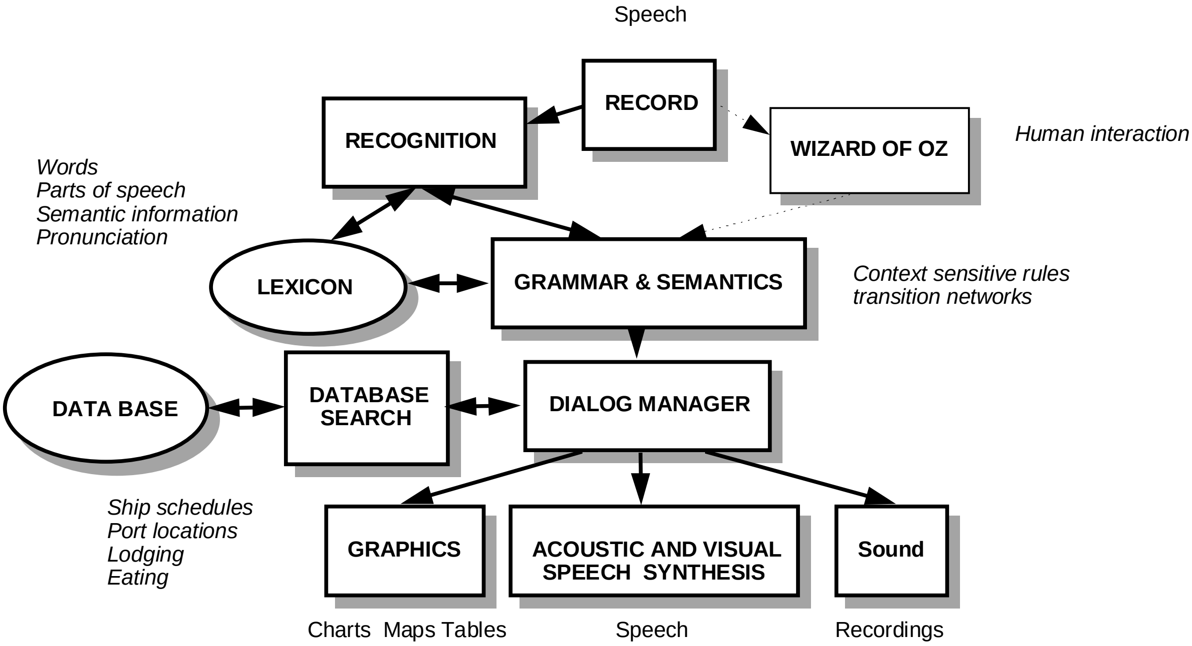

Figure 2. The modules in the Waxholm spoken dialog system.

NATURAL LANGUAGE MODELING

In this section we will give a short review of the natural language component, STINA. Some of our fundamental concepts were initially inspired by TINA, a parser developed at MIT, (Seneff , 1992.) STINA is knowledge based and contains a context-free grammar which is compiled into an ATN. (A detailed description of the parser can be found in Carlson and Hunnicutt, 1995.) Probabilities are assigned to each arc after training. These probabilities are primarily used to reduce search time and for hypothesis pruning.

The parsing is done in three steps. The first step makes use of broad categories such as nouns, while the following step expands these into more detailed solutions. The last step involves recalculation of hypothesis probabilities according to a multi-level N-gram model.

Domain dependent feature system

The feature system used in the parser plays an important role. Each lexical entry can have domain specific semantic features associated to it in addition to the basic syntactic features. The semantic features as used in STINA can be divided into two different classes, basic semantic features and function features. Basic features such as BOAT and PORT give a simple description of the semantic property of a word and are often domain specific. These features are hierarchically structured. In our domain we have specified that a PORT is part of an ISLAND, which is part of a REGION, which is part of a PLACE, which is part of the WORLD.

The second type of semantic features is the "function features." These features are not hierarchical. Typically they are associated with an action, such as TO_PLACE indicating the destination in an utterance regarding travel (Example 1). The function features are also node names in the parser. A verb can have function features set, allowing or disallowing a certain type of modifier to be part of a clause. The action itself in the TO_PLACE example has, of course, a broader scope than the traveling domain, and includes movements between any reference points. Thus, the node TO_PLACE is specified as a prepositional phrase starting with “to” and followed by any nominal expression. The scope of the phrase is changed according to the domain by training.

Example 1:
(TO_PLACE (“to”/TO “Waxholm”/noun))

The function features are powerful tools 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 or a prepositional phrase including a port name as a possible response from the user. This property of STINA gives the parser some of the advantages of a functional grammar parser.

Terminal node evaluation is primarily carried out on the grammatical features. If this basic constraint evaluation is accepted, the semantic features are also evaluated. The hierarchical structure has importance for the rule writing. During the unification process all semantic features which belong to the same semantic branch in the feature tree are considered. The whole tree of the lexical entry is moved into the hypothesis including the leaves on the feature tree. In our traveling domain a port name will keep its PORT feature even if only the PLACE is noted in the grammar. This has several advantages. The rules or terminal specifications do not have to be more specific than necessary and the domain knowledge can, to some extent, be part of the lexicon rather than the rules. This mechanism is extensively used in the sublanguage grammar for our application. In the next section we will see how the introduction of domain dependent terminal nodes is delayed during the parsing process.

General grammar to subgrammar

It has been an ambition in our work to create a general grammar which at least covers the type of dialog found in our domain. After an utterance initially has been parsed, we have a hypothesis in terms of grammar nodes and generic terminals such as nouns. In the next step the terminals are replaced by more domain specific labels. A domain specific list of possible terminals is processed by the parser during initialization and each such terminal is associated a generic terminal node. The domain specific nodes are typically constrained by domain specific semantic features in addition to the basic syntactic ones. In our case we have defined a number of terminals, such as port, hotel, boat and time-table. These are all part of the noun class and will replace the "noun" terminal whenever appropriate according to the semantic features of the lexical entry. In our application, then, the lexicon defines that there are nouns with a specific semantic feature, PORT, and is able to separate them from other nouns. In Example 1 the simple phrase is turned into the phrase "TO port" since Waxholm is a port and the terminal port is part of the noun class, Example 2.

Example 2:
(TO_PLACE ("to"/TO "Waxholm"/port))

With this approach we can formulate a general grammar and make it domain specific with the help of the feature system and lexical specifications.

The described method has some attractive side effects. Since the network specified by the ATN has generic terminals, the number of nodes and transitions are less than if the grammar were more specific. This makes the parsing faster since fewer hypotheses have to be evaluated. However, the probability calculation is less informative based on broad categories and has to be reconsidered. In our case this is done with the help of N-gram models.

N-gram models

It seems to be a general consensus that N-gram models, in the context of speech understanding, have at least as good predictive power as regular knowledge based grammars (Jelinek et al., 1992, and Stolcke 1995.) However, some research, such as the work by Seneff et al. (1995,) has shown, that a knowledge based parser including multilayered probabilities has some advantages. This is specially true for the following processing in the dialog system.

In STINA, smoothed N-gram models are used in addition to the regular transition probabilities. N-gram probabilities are added to the node probabilities after the domain specific node replacements have been performed and before a hypothesis is pushed on the probability ordered N-best stack. The N-gram probabilities include not only terminal node sequences but also phrase level heads. The work by Moore et al. (1995) has earlier shown the advantage in adding phrase heads in the N-gram calculation. In Example 3 the hypothesis score calculation includes for example:

p(boat | "TOP+SUBJ"),
p(verb | "SUBJ+boat+VP"),
p(TO | "VP+verb+TO_PLACE")
p(port | "VP+verb+TO_PLACE").

We have expanded the calculation to also include phrase level head node probabilities. However, they are based on phrase level head sequences.

p(SUBJ | "TOP")
p(VP | "TOP+SUBJ")
p(TO_PLACE | "TOP+SUBJ+VP").

Example 3:
(TOP (SUBJ "båten"/boat) (VP "går"/verb
(TO_PLACE ("till"/TO "Vaxholm"/port))))

As an additional example we find that the utterance “I want to go from X to Y” is more probable in our application than “I want to go to X from Y” as reflected in the node N-gram probabilities. Thus, this last step of hypothesis scoring is a powerful method to adjust the general grammar to the domain specific analysis that is needed. Certain phrases and phrase sequences will be well described in the N-gram statistics.

DIALOG MODELING

Two major ideas have been guiding the work on the dialog model. First, the dialog should be described by a grammar. Second, the dialog should be probabilistic. In our system, dialog building blocks are described by nodes. Each node has specifications concerning, for example, dialog action, constraint evaluation and system response. A graphical interface to the system presents the dialog grammar graphically. Both the syntax and the dialog can be modeled and edited graphically with this tool.

In the following description, we have used the term “topic” to describe what type of information a user is requesting or, in some cases, a special response from the system. In Table 2, some of the major topics are listed. Topic selection is accomplished based on probabilities calculated from user initiatives (Carlson, 1994; Carlson and Hunnicutt, 1995; Carlson, Hunnicutt and Gustafson, 1995). Lexical semantic information combined with semantic grammar nodes are used as factors in this calculation. The topic selection based on probabilities in our system has similarities with the effort at AT&T (Gorin, 1994; Gorin et al., 1994). A different approach, also based on training, has been presented by Kuhn and De Mori (1994) in their classification approach. A special session in the Eurospeech 1995 conference was devoted to word spotting including topic spotting based on keywords. The work by Nowel and Moore (1995) goes one step further exploring non-word based topic spotting.

The dialog component is controlled by hand-crafted context free rules which control all steps in the dialog including database search, speech and face synthesis, and other graphical feedback. The dialog component is probabilistic in terms of topic selection based on user initiatives.

Initially the system was designed based on intuition and discussions within the group. After the first subjects were recorded, we were able to base both the grammar and the dialog model on the gathered empirical data. The lexicon was expanded and the network probabilities were trained.

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 Waxholm? (Where is Waxholm located?)
EXIST Goal: to display the availability of lodging and dining possibilities. Example: Var finns det vandrarhem? (Where are there hostels?)
TRIP_MAP Goal: To present possible trip alternatives. Example: Var kan jag åka? (Where can I go?)
END_SCENARIO Goal: to end a dialog. Example: Tack. (Thank you.)
REPEAT Goal: To repeat the synthesis output. Example: Vad sa du? (What did you say?)
NO_UNDERSTANDING Goal: To take care of those cases when the system do not understand the user's intentions. Example: Jag heter Olle. (My name is Olle)
OUT_OF_DOMAIN Goal: to inform the user that the subject is out of the domain for the system. Example: Kan jag boka rum. (Can I book a room?)

Table 2. The main topics used in the Waxholm dialog model.

The semantic frame plays an important role in the dialog model. After the syntactic tree has been reduced to a semantic tree, a semantic frame is created with slots corresponding to attribute-value information taken from the tree. The semantic frame has a feature specification describing which features are used in the frame and which information might have been added to the frame from the dialog history. Each step in the dialog progress is associated to a new frame, with reference pointers to the history. The information that should be pushed forward is defined by semantic feature specifications in the dialog rules.

Dialog rules

Each predicted dialog topic is explored according a set of rules. These rules define which constraints have to be fulfilled and what action should be taken depending on the dialog history. A node can be a terminal node or the entry point to a network. In the case of a terminal, several node specific characteristics have to be defined in the rule system. These can divided into three basic groups: constraint evaluation, system questions and system actions.

The constraint evaluation is described in terms of features and in terms of 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. This whole process is handled by the feature passing between the dialog part and the grammar part of STINA.

If the response from the subject does not clarify the question, the robust parsing is temporarily disconnected so that informative error messages can be given to the user about lexical or syntactic problems. At the same time, a complete sentence is requested giving the dialog manager the possibility of evaluating whether the chosen topic is incorrect. This technique has been shown to be useful in helping subjects to recover from an error through rephrasing of their last input (Hunnicutt et al., 1992).

In most cases, the user *does* answer system questions, (Bertenstam et al., 1995b, 1995c). A positive response from the constraint evaluation opens the way for the selected action to take place, such as a database search or a graphic presentation of a map or a table.

Most of the terminal nodes have no direct input from the user. Rather they deal with small details such as making the synthetic face to look at a graphic object on the screen or to simply synthesize a message to the user. Thus, the dialog rules do not only model turns in an oral dialog, they cover all actions in the dialog system. The expanded grammar notation makes it possible to separate the dialog model from the system implementation.

A modification of the domain implies an addition of how to handle a new topic, but it is our ambition that the implementation and the training procedures should, as much as possible, be kept the same.

Topic selection

The decision about which topic 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.

Each semantic feature found in the syntactic and semantic analysis is considered in the form of a conditional probability to decide on the topic. 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 BOAT feature can be a strong indication for the TIME_TABLE topic but this can be contradicted by a HOTEL feature. The topic prediction has been trained using a labeled set of utterances taken from the Waxholm database. Only utterances indicating a topic (about 1200) have been included in this set. The probability is calculated according to: $p = (n+1)/(N+2)$, where N = number of times a feature can be a terminal node in the feature tree, and n = number of times a feature actually is a terminal node in a topic indicating utterance.

Introduction of a new topic

In this section we will give a simple example of how a new topic can be introduced. Suppose we want to create a topic called "out of domain." First a topic node is introduced in the rule system. Some new words probably need to be included in the lexicon and labeled 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 an output informing the user about the situation such as "We are not able to process things dealing with booking". Example sentences must be created that illustrate the problem and the dialog parser must be trained with these sentences labeled with the "out of domain" topic. Since the topic selection is done by a probabilistic approach that needs application-specific training, data collection is of great importance.

MAN-STINA INTERACTION

In the implementation of the parser and the dialog management, we have stressed an interactive development environment. It is possible to study the parsing and the dialog flow step by step when a graphic tree is built. It is even possible to use log files collected during Wizard of Oz experiments as scripts to repeat a specific dialog, including all graphic displays and acoustic outputs.

We have added a graphical interface to the system which presents each network graphically. Both the syntax and the dialog networks can be modeled and edited graphically with this tool. Earlier work on dialog modeling such as the Generic Dialog System Platform in the Danish dialog project (Larsen and Baekgaard, 1994) has been an inspiration for this expansion. This platform is based on a special tool in which the dialog flow can be described by a network of building blocks. These blocks can be edited graphically. The OASIS developed by the GTE Laboratories Incorporated (Zeigler and Mazor, 1994) should also be referred to in this context. It is based on dialog prototypes which include building blocks for the acquisition of factual information, for the verification of acquired information, and for reacquisition following a disconfirmation.

EVALUATION OF THE NLP AND DIALOG MODULES

Evaluation of the system has been performed using part of the Waxholm database. In this database, speech and text data was collected using the Waxholm system. Initially, a “Wizard of Oz” replaced the speech recognition module. A full report on the data collection and data analysis can be found in Bertenstam et al. (1995b and 1995c.)

The database was collected using preliminary versions of each module in the Waxholm system. This procedure has advantages and disadvantages for the contents of the database. System limitations will already from the beginning put constraints on the dialog, making it representative for a human-machine interaction. However, since the system was under development during the data collection, it was influenced by the system status at each recording time. After about half of the recording sessions, the system was reasonably stable, and the number of system "misunderstandings" had been reduced. As research on dialog systems develops, it becomes more important to develop new methods to evaluate human-machine interaction (Hirschman and Pao, 1993.)

Test material

The test material used in the experiments includes 68 subjects and 1900 dialog utterances containing 9200 words. The total recording time amounts to 2 hours and 16 minutes. The most frequent 200 words out of the total of 720 words cover 92 percent of the collected transcribed data. About 700 utterances are simple answers to system questions while the rest, 1200, can be regarded as user initiatives.

We can find a few examples of restarts in the database due to hesitations or mistakes on the semantic, grammatical or phonetic level. However, less than 3% of the utterances contain such disfluencies. Some of the restarts are exact repetitions of a word or a phrase. In some cases a preposition, a question word or a content word is changed. The average utterance length was 5.6 words. The average length of the first sentence in each scenario was 8.8 words.

Parser evaluation

The parser has been evaluated in several different ways. Most tests used a deleted estimation procedure. Using about 1700 sentences in the Waxholm database, 62 percent give a complete parse, whereas if we restrict the data to utterances containing user initiatives (about 1200), the result is reduced to 48 percent. This can be explained by the fact that the large number of responses to system questions typically have a very simple syntax.

If we exclude extralinguistic sounds such as lip smack, sigh and laughing in the test material based on dialog initiatives by the user, we increase the result to 60 percent complete parses. Sentences with incomplete parses are handled by the robust parsing component and frequently effect the desired system response.

The perplexity on the Waxholm material is about 26 using a trained grammar. If only utterances with complete parses are considered we get a perplexity of 23.

N-best resorting

The parser has also been evaluated in an N-best list resorting framework. Totally 290 N-best lists with about 10 alternatives each were generated, using an early version of the speech recognition module of the Waxholm system (Ström, 1995.) Since several of the utterances were answers to simple questions the utterance length only averaged about 5 words. The top choice using a bigram grammar as part of the recognition module gave a word accuracy of 76.0%. The mean worst and best possible accuracy in the lists were 48.0% and 86.1%. After resorting using the STINA parser the result improved to 78.6% corresponding to about 25% of the possible increase.

All topics		
Test material	N	% Error
woz input	1209	12.9
no extralinguistic sounds	1214	12.7
only complete parses	581	3.1
All topics excluding no "understanding"		
Test material	N	% Error
woz input	1154	8.8
no extralinguistic sounds	1159	8.5
only complete parses	580	2.9

Table 3. Results from the topic prediction experiments.

Evaluation of topic selection

We have performed a sequence of tests to evaluate the topic selection method. The evaluation has used one quarter of the material, about 300 utterances, as test material, and the rest as training material, about 900 utterances. This procedure has been repeated for all quarters and the reported results are the mean values from these four runs. The first result, 12.9% errors in Table 3, is based on the unprocessed labeled input transcription. The eight possible topics have a rather uneven distribution in the material with TIME_TABLE occurring 45% of the time. One of the topics, labeled "no understanding," is trained on a set of constructed utterances that are not possible to understand, even for a human. This topic is then used as a model for the system to give an appropriate "no understanding" system response. It should be noted that, in principle, these utterances still can have a reasonable parse. However, the topic prediction is certainly influenced by a poor parse. It seemed reasonable to exclude the "no understanding" prediction from the result since the system at least does not make an erroneous decision. The accuracy model in word recognition evaluation has the same underlying principle. By excluding 55 utterances, about 5% of the test corpus, predicted to be part of the "no understanding" topic, we reduce the error by about 4%.

In the next experiment, we excluded all extralinguistic sounds, about 700, in the input text. This will increase the number of complete parses with about 10% as discussed earlier. The prediction result was about the same compared to the first experiment.

The final experiment included only those utterances that gave a complete parse in the analysis. The errors were drastically reduced. We do not yet know if an increased grammatical coverage also will reduce the topic prediction errors.

SUMMARY

Lexical semantic information combined with the grammar rules describe the system constraints in our system. Thus, the choice of semantic features and terminal nodes will automatically turn the general grammar into a subgrammar based on the domain. The use of N-gram statistics improves the predictive power of the grammar on both terminal level and phrase structure. Topic prediction based on semantic features separates the surface form of an utterance from the intention of the subject. The dialog design can be data driven to some extent with the proposed method. The rule-based, and to some extent, probabilistic approach we are exploring makes the addition of new topics relatively easy. However, much manual work still remains to be done when an application domain should be changed.

ACKNOWLEDGMENT

We thank all the people in the Waxholm project group who have been engaged in data collection, data analysis and system design. The Waxholm group consists of staff and students at the Department of Speech, Music and Hearing, KTH. Most of the efforts are done part time. The members of the group in alphabetic order are: Johan Bertenstam, Jonas Beskow, 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. This work has been supported by grants from The Swedish National Language Technology Program.

REFERENCES

Aust H, Oerder M, Seide F and Steinbiss V (1994). Experience with the Philips Automatic Train Timetable Information System. In: Proc IEEE workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA94), 67-72.

Bertenstam J, Blomberg M, Carlson R, Elenius K, Granström B, Gustafson J, Hunnicutt S, Högberg J, Lindell R, Neovius L, Serpa-Leitao A de, Nord L and Ström N (1995a). The Waxholm system - a progress report. Proc. Spoken Dialog Systems, Vigsø, 81-84.

Bertenstam J, Blomberg M, Carlson R, Elenius K, Granström B, Gustafson J, Hunnicutt S, Högberg J, Lindell R, Neovius L, de Serpa-Leitao A, Nord L and Ström N (1995b). The Waxholm Application Data-Base. In: Proc Eurospeech '95, 4rd European Conference on Speech Communication and Technology, Madrid, 833-836.

Bertenstam J, Blomberg M, Carlson R, Elenius K, Granström B, Gustafson J, Hunnicutt S, Högberg J, Lindell R, Neovius L, Nord L, Serpa-Leitao A de and Ström N (1995c). Spoken dialog data collected in the Waxholm project. STL-QPSR, KTH, 1: 49-74. Dept. of Speech Comm. and Music Acoustics, KTH, Stockholm, 49-74.

Beskow J (1995). Rule-based visual speech synthesis. In: Proc Eurospeech '95, 4rd European Conference on Speech Communication and Technology, Madrid, pp. 299-302.

Blomberg M, Carlson R, Elenius K, Granström B, Gustafson J, Hunnicutt S, Lindell R and Neovius L (1993). An experimental dialog system: Waxholm. In: Proc Eurospeech '93, 3rd European Conference on Speech Communication and Technology, Berlin, 1867-1870.

Carlson R (1994). Recent developments in the experimental “Waxholm” dialog system. In: ARPA Human Language Technology Workshop, Princetown, New Jersey, 207-212.

Carlson R, Granström B, and Hunnicutt S (1991). Multilingual text-to-speech development and applications. In: Advances in speech, hearing and language processing (Ainsworth AW, ed.), London: JAI Press, UK.

Carlson R and Hunnicutt S (1995). The natural language component - STINA. In: STL-QPSR, . Dept. of Speech Comm. and Music Acoustics, KTH, Stockholm, 1: 29-48.

Carlson R, Hunnicutt S and Gustafson J (1995). Dialog management in the Waxholm system. In: Proc. ESCA/ETRW on Spoken Dialog Systems, Vigsø, Denmark, 137-140.

Clementino D and Fissore L (1993). A man-machine dialog system for access to train timetable information. In: Proc Eurospeech '93, 3rd European Conference on Speech Communication and Technology, Berlin, 1863-1866.

Dalsgaard P and Baekgaard A (1994). Spoken Language Dialog Systems. In: Proc in Artificial Intelligence, Infix. Presented at the CRIM/FORWISS workshop on 'Progress and Prospects of Speech Research and Technology, Munich.

Gerbino E and Danieli M (1993). Managing dialog in a continuous speech understanding system. In: Proc Eurospeech '93, 3rd European Conference on Speech Communication and Technology, Berlin, 1661-1664.

Glass J, Flammia G, Goodine D, Phillips M, Polifroni J, Sakai S, Seneff S and Zue V(1995). Multilingual spoken-language understanding in the MIT Voyager System. In: Speech Communication. Vol. 17:1-2:1-18.

Gorin A (1994). Semantic associations, acoustic metrics and adaptive language acquisition. In: Proc ICSLP, International Conference on Spoken Language Processing, Yokohama, 79-82.

Gorin A, Hanek H, Rose R, and Miller L (1994). Automatic call routing in a telecommunications network. In: Proc IEEE workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA94), 137-140.

Hirschman L and Pao C (1993). The cost of errors in a spoken language system. In: Proc Eurospeech '93, 3rd European Conference on Speech Communication and Technology, Berlin, 1419-1422.

Hunnicutt S, Hirschman L, Polifroni J, and Seneff S (1992). Analysis of the effectiveness of system error messages in a human-machine travel planning task. In: Proc ICSLP International Conference on Spoken Language Processing, Alberta, Canada, pp. 197-200.

Jelinek F, Mercer R., and Roukos R (1992). Principles of lexical language modeling for speech recognition. In Advances in Speech Signal Processing Eds. Furui S and Sondhi M. Marcel Dekker, 651-699.

Kuhn R and De Mori R (1994). Recent results in automatic learning rules for semantic interpretation. In: Proc ICSLP, International Conference on Spoken Language Processing, Yokohama, 75-78.

Larsen LB and Baekgaard A (1994). Rapid prototyping of a dialog system using a generic dialog development platform. In: Proc ICSLP International Conference on Spoken Language Processing, Yokohama, pp. 919-922.

Moore R, Appelt D, Dowlding J, Gawon M and Moran D (1995). Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS. In Proc ARPA Workshop on Human Language Technology.

Nowel P and Moore R (1995). The application of dynamic programming techniques to non-word based topic spotting. In: Proc Eurospeech '95, 4rd European Conference on Speech Communication and Technology, Madrid, 1355-1358.

Oerder M and Aust H (1994). A real-time prototype of an automatic inquiry system. In: Proc ICSLP International Conference on Spoken Language Processing, Yokohama, 703-706.

Peckham J (1993). A new generation of spoken dialog systems: results and lessons from the SUNDIAL project. In: Proc Eurospeech '93, 3rd European Conference on Speech Communication and Technology, Berlin, 33-40.

Seneff S (1992). “TINA: A natural language system for spoken language applications.” Computational Linguistics, 18:1:61-66.

Seneff S, McCandless M and Zue V (1995). Integrating Natural Language into the Word Graph Search for Simultaneous Speech Recognition and Understanding. In: Proc Eurospeech '95, 4rd European Conference on Speech Communication and Technology, Madrid, 1781-1784.

Shirai K and Furui S (1995). Spoken Dialog. In: (Shirai and Furui, eds.) Special issue of Speech Communication, 15(3-4).

Stolcke A (1995). Combining N-grams and SCFGs in Speech Language Models. In Proc IEEE Automatic Speech Recognition Workshop, Snowbird USA.

Ström N (1994). Optimizing the lexical representation to improve A* lexical search. In: STL-QPSR, . Dept. of Speech Comm. and Music Acoustics, KTH, Stockholm, 2-3: 113-124.

Zeigler B and Mazor B (1994). Dialog design for a speech interactive automation system. In: Proc IEEE workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA94), pp. 113-116.