AN EDUCATIONAL DIALOGUE SYSTEM WITH A USER CONTROLLABLE DIALOGUE MANAGER

Joakim Gustafson[1], Patrik Elmberg[2], Rolf Carlson[1] and Arne Jönsson[2]

[1] Department of Speech, Music and Hearing, KTH, Sweden
[2] Department of Computer and Information Science, Linköping University, Sweden

ABSTRACT

We have developed an educational environment for a modular spoken dialogue system. The aim of the environment is to provide students, with different backgrounds, means to understand the behaviour of spoken dialogue systems. Focus in this paper is on dialogue and dialogue management. The dialogue is recorded in a dialogue tree whose nodes are dialogue objects. The dialogue objects model the constituents of the dialogue and consist of parameters for modelling dialogue structure, focus structure and a process description describing the actions of the dialogue system. Various dialogue system behaviours can be achieved by modifying these parameters. This is done using the educational environment, which is interactive and facilitates examination, expansion and modification of the dialogue object parameters and hence the system. The educational system has been used in a number of courses at various universities in Sweden.

1. INTRODUCTION

The speech group at KTH has developed a modular spoken dialogue system that gives students from diverse backgrounds an understanding of the modules used in a spoken dialogue system [1][2]. In a joint research project the dialogue group at Linköping University and the group at KTH has integrated a highly flexible dialogue manager [3] into the educational dialogue system. The students can test the system themselves and are able to examine each module in detail. They are also able to extend and develop the functionality of the system. The goal is to increase their understanding of the problems and issues involved in developing and using spoken dialogue systems. In a separate presentation at this conference a web based educational program is presented for teaching speech technology by hands-on-experiments [4].

In the dialogue system, called GULAN, students are presented with a simple spoken dialogue application for searching in the web-based "Swedish Yellow Pages" on selected categories regarding facilities in Stockholm. This domain has several advantages, since it either can be very restricted for specific experiments with limited vocabulary or can be expanded to a very general and open domain.

In the educational environment students can use the system as it is but also modify it in several respects, such as dialogue behaviour, grammar, speech generation and speech recognition. For example, on the basic level, new words can be added to the system. These will be automatically transcribed with the help of a text-to-speech system. After listening to the transcription the students can if necessary correct it and investigate new alternative pronunciations. In addition to the phonetic transcription, the lexical entries also have semantic labels for the domain description and syntactic labels for the recognition grammar.

2. ARCHITECTURE

The dialogue system has been under development for a number of years [1]. During this process modules in the system have been refined or replaced by completely new ones. The basic modules in the system are speech recognition [5], speech synthesis, parsing [6], dialogue management [7], web-database-search, and an interactive map. The modules are implemented in different programming languages, but have been provided with an interface written in the Tcl language [8]. This speech technology toolkit makes it easy to create new applications quickly and easily on a number of platforms, currently PC (Win NT/95) and Unix machines (HP-UX, Linux, Sun Solaris, SGI IRIX).

An architecture for communication between clients and servers on different computers at different locations has also been developed. (For a detailed description see [9].) The central server, the Broker, handles the communication between clients and servers over the Internet. All communication within the broker system is in text form to ensure portability and aid in debugging with the included debugging tool. Binary data, such as speech, is sent over separate private connections. This architecture allows us keep the speech technology servers in-house where they can run on powerful machines, and be updated at any time. The approach makes it very easy to develop new modules at other research groups, and it makes it possible to collect speech data regardless of where the applications are used.

3. THE DIALOGUE SYSTEM IN USE

During the last two semesters 1997/1998 the first versions of the instructional environment have been used in five different courses by four different departments at three universities in Sweden, see Table 1. About 200 last-year Masters students participated in these classes. The students worked in groups of two and were given a list of tasks that where to be carried out.

- The first task was to use the dialogue application in order to determine its capabilities and limitations.
- The next task was to test the speech recognition module stand alone, with the explicit purpose that the students should gain some insight into the limitations of current HMM based

speech recognition technology, regarding for example, regarding noise, speaking style and out of vocabulary words.

- The main assignment was to add new fields from the Yellow Pages, new street names from the map; and new words or phrases to the system. All new words in the lexicon had to be labelled with appropriate syntactic and semantic tags, and correct transcriptions. Students also had to extend the example-based grammar with new constructs.
- In the text generation module, students had to insert additional response templates to handle the new facilities. This included experimenting with different prosodic patterns in the sentences.
- Finally, the students where required to demonstrate the system and show that it worked according to the new specifications.

Overall, the students were very satisfied with the system and rated it four on a five point scale in the course evaluation. The lab environment, together with the underlying toolkit, was an important aid in giving students an understanding of spoken language technology. The main criticism by the students was that they wanted to be able to make greater changes to the system and to go deeper into some of its modules. In all these exercises a simple dialogue manager was used. It was quite clear that a more elaborate dialogue manager would add a new stimulating and valuable dimension to the assignment. The next section presents a new dialogue manager, which will be used in future courses.

University	**Course**	**Students**
KTH/nada	Advanced Graphics and Interaction	80 Masters Students in Computer Science
KTH/nada	Language Technology	20 Masters Students in Computer Science
KTH/tmh	Speech Technology	25 Masters Students in Computer Science
Linköping University	Speech Technology	13 Masters Students in Computer Science and Cognitive Science
Uppsala University	Language Technology	20 Masters Students in Computational Linguistics

Table 1. The courses that used GULAN the 1998 semester.

4. DIALOGUE MANAGEMENT

The dialogue manager utilised in the GULAN-system is based on a dialogue model originally developed for written interaction, LINLIN [3] and verified for spoken interaction [7]. This model assumes that dialogues can be represented with a simple initiative-response scheme. The dialogue is further structured into three different classes; the entire dialogue D, discourse segments IR, and the actual speech acts, initiatives I or responses R. The scheme only accepts units consisting of an initiative followed by a response or embeddings of such units in higher IR-units, e.g. (I R), or successive and recursive embeddings such as (I (I R) R), (I (I R) (I R) R), or (I (I (I R) R) R) etc. All moves must belong to some discourse segment, and no segments with the structure (I I R) or (I R R) are allowed. However, users need not to respond to an initiative, e.g. ((I) (I R)). Thus, the dialogue structure forms a dialogue tree, figure 1.

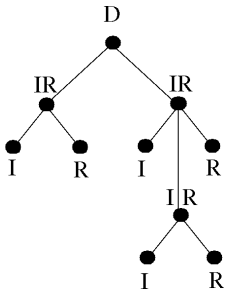

Figure 1. A dialogue structure drawn as dialogue tree

What makes the dialogue model especially well suited for educational systems is the use of dialogue objects. These correspond to the three classes above and each dialogue object consists of three components; dialogue structure, focus structure and a process description.

The dialogue structure models information, such as, initiator, responder, illocutionary type and topic. This is modelled in a dialogue grammar. The dialogue grammar controls the creation of the dialogue tree.

The focus structure records entities mentioned in the discourse. These entities pertain to domain objects and related properties providing the dialogue manager with information, thus allowing a user to refer to them in the course of interaction. Considering the GULAN domain, a typical domain object is *restaurant* and related properties could be *opening hours*, *type of food* etc. A common user initiative in these types of applications, i.e. simple information retrieval, is asking for the value of a property of a domain object.

The process description models a stereotypical dialogue behaviour based on inspection of the dialogue and focus structure parameters. This can for instance be under what circumstances to initiate a clarification sub dialogue and how to integrate the response from such a clarification request with previous information. Also, the process description specifies what expectations the dialogue manager will have for the next utterance.

4.1. Changing the behaviour

Thus, different dialogue behaviours can be illustrated by modifying the dialogue objects. This is facilitated by providing the students with a set of primitives corresponding to focus handling, interpretation, and presentation. If e.g. focused objects are appended to the list of objects already modelled in the focus structure, we will have a different dialogue behaviour than if the new objects entirely replace existing objects or just the ones corresponding to the same class. In the lab environment a general principle for focus handling can be set from a menu. Also, individual focus handling is possible for each system response, see below.

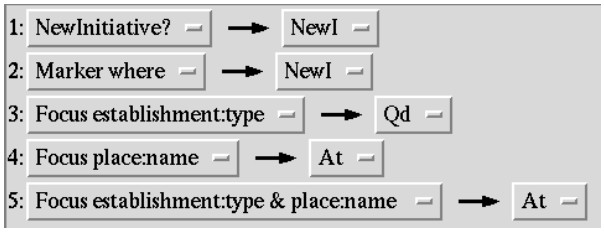

Figure 2. Rule editing window, with the student specified rules used in section 4.2.

When interpreting an utterance, primitives are used to inspect the dialogue and focus structures of the dialogue objects. These primitives form a small set of rules. Figure 2 above shows the student's view of the rule editing window. It contains five numbered rules, each with the activation condition on the left hand side and the resulting action to the right. The rules are presented as buttons, which when activated yields a menu with available primitives and system behaviours respectively.

The condition is described using primitives and arguments, corresponding to the dialogue objects' focus parameters. In GULAN the arguments to the primitives refer to additional semantic information about the domain objects.

The action describes the system's behaviour, e.g. responding with an answer or initiating a clarification request. When specifying the system response, the student can change how the system presents its results to the user (speech synthesis, visual information etc). This can be used to illustrate how the user is influenced by the system's response and also how it affects the overall quality of the dialogue. An editing window for this pops up when using the right mouse button on the right hand side of a rule, i.e. the button to the right of the arrow in Figure 2.

During the interaction the students can monitor what is happening. The dialogue tree is drawn on the screen, and by stepping through the process descriptions, the students are able to follow the change of focus structure in the dialogue objects.

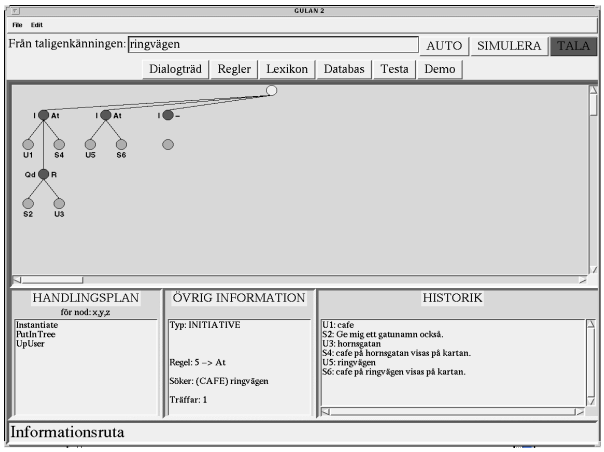

Figure 3. An overview of the environment for the dialogue management.

Figure 3 above shows a screenshot from the interface to the educational system in action. In the centre, the dialogue tree is drawn. Each node (dialogue object) can be inspected with a mouse click. As a convenience, the process description for the current dialogue object is shown at the bottom left, and a history list of the dialogue in plain text at the bottom right. In between there is a window where the interpretation process is monitored, e.g. which rule was applied and how successful the database search was.

4.2. An illustrative example

One important feature of the educational system is the possibility for students to model the behaviour of the dialogue manager. Figure 2 is an example of such a model. In the following we will illustrate how the manager will process a simple example. A corresponding dialogue tree is generated during the interaction (see Figure 3). The dialogue below is an English translation, with the user U and system S taking turns:

U1: Where can I find a coffee shop?

S2: I need a street name too.

U3: Hornsgatan.

S4: Coffee shops on Hornsgatan are shown on the map.

U5: ……and on Ringvägen?

S6: Coffee shops on Ringvägen are shown on the map.

In U1, the user is asking for a coffee shop. Having coffee shop in focus, rule #3 (see Figure 2) states that only establishment information is not enough and therefore initiates a clarification request, **Qd**. This results in the response S2, and the system sets its expectations to **place:name** waiting for a user response. The dialogue manager creates a new user move node, the unattached, lower circle in the left part of Figure 4.

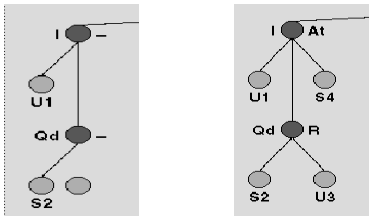

Figure 4 Two snapshots of the dialogue tree being built.

The user answers the clarification request, in U3, by providing a street; *Hornsgatan*. Now, as *Hornsgatan* fulfils the **place:name** criteria, the system can connect U3 in the tree, integrating the information up to the initiating node, the top node in the right part of Figure 4. This dialogue object invokes rule #5 since both a type of establishment and a name of a place are in focus. After a database search, the system responds with an answer, **At**, in S4.

Next the user provides another street name, *Ringvägen*, in U5. As the dialogue object already has an establishment type, Coffee shops, in focus, the new street, *Ringvägen*, replaces *Hornsgatan* and the system can successfully access the database and provide the new information, S6.

The "replace-focus"-strategy used in the example above is reasonable for street focus shifts, but, students can also change the focus strategy. Compare the following system responses, S8a and S8b, to user question U7 in the following continuation of the dialogue:

U7: Restaurants.

S8a: Restaurants on Ringvägen are shown on the map.

S8b: Coffee shops and restaurants on Ringvägen are shown on the map.

The response in S8a is one possible response to U7 using the replace-strategy, but it could also be appropriate to append the new type of establishment to the previous and respond as in S8b. This is a reasonable strategy if, for instance, the information can be conveniently presented to the user in a table (cf. [3]) or, as in this case, on a map. What is more important here, however, is that the educational system allows the students to easily modify the dialogue systems behaviour to investigate various dialogue systems.

5. FUTURE WORK

Several issues will be addressed in the next phase of the project. From the educational point of view, several facilities will be added to further improve the laborative environment, e.g. enabling undo/redo-functions and giving support for novice students writing grammars. New modules will also be added for example multimodal synthesis [10], prosodic analysis, dynamic lexica and dialogue dependent speech recognition [11] and speech synthesis.

A dialogue system is currently on display as part of the activities celebrating Stockholm as the Cultural Capital of Europe. In this system several domains has been possible to combine, thanks to the modular functionality of the architecture. Each domain has it own dialogue manager and an example based topic spotter is used to relay the user utterances. In this system the multimodal agent “August” is presenting different tasks such as taking the viewers on a trip through the Department of Speech, Music and Hearing, giving street information (with the help of the educational system GULAN) and also presenting short excerpts from the production of the Swedish author, August Strindberg, when waiting for someone to talk to. Experiences from these new experiments will give valuable feedback for the next generation of GULAN.

6. SWEDISH DIALOGUE SYSTEMS

The development of the educational system presented in this paper is part of the research project Swedish Dialogue Systems, supported by the Swedish Language Technology Program. The aim is to bring together research groups in Sweden who have competence in speech analysis and synthesis, dialogue structure and the interpretation of natural language and move them towards the creation of a common environment, which will allow for the creation of generic dialogue systems and the investigation of fundamental issues in dialogue technology. In this project we focus on the higher system levels: dialogue management, robust utterance interpretation, including use of prosodic information, and effective utterance planning. In order to ensure flexibility and ecological validity of dialogue systems we will also create a platform for multimodal spoken language corpora which will contain spoken language audio and video recordings with standardised transcriptions.

7. ACKNOWLEDGMENTS

The project was supported in part by the Swedish Language Technology Program (NUTEK/HSFR). We are indebted to Zhulia Ayani for work on the dialogue manager.

8. REFERENCES

1 Sjölander K. and Gustafson J. “An Integrated System for Teaching Spoken Dialogue Systems Technology,” Proc. Eurospeech 97, Rhodes, Greece, 1997.

2 Carlson R., Granström B., Gustafson J., Levin E. and Sjölander K. “Hands-on speech technology on the web", Proc. ELSNET in Wonderland, Soesterberg, The Netherlands, 1998.

3 Jönsson, A. “A Model for Habitable and Efficient Dialogue Management for Natural Language Interaction,” Natural Language Engineering, Vol: 3, No: 2/3, pp: 103-122, 1997.

4 Sjölander K., Beskow J., Gustafson J., Carlson R. and Granström B. “Web based educational tools for speech technology,” Proc. ICSLP 98, Sydney, Australia, 1998.

5 Ström, N. *Automatic continuous speech recognition with rapid speaker adaptation for human/machine interaction*, Dr. Thesis, KTH, Stockholm, 1997.

6 Carlson, R., “The Dialog Component in the Waxholm System,” Proc. ICSLP´96, Philadelphia, USA, 1996.

7 Jönsson, A. “A Model for Dialogue Management for Human Computer Interaction”, Proceedings of ISSD'96, Philadelphia, 1996.

8 Sjölander, K., “The Snack Sound Visualization Module,” http://www.speech.kth.se/SNACK/ 1997

9 Lewin, E., “The Broker Architecture at TMH,” http://www.speech.kth.se/proj/broker/, 1997.

10 Beskow J. “Animation of Talking Agents,” In Proceedings of AVSP'97, ESCA Workshop on Audio-Visual Speech Processing, Rhodes, Greece, 1997 .

11 Gustafson, J., Larsson, A., Carlson, R. and Hellman, K., “How do System Questions Influence Lexical Choices in User Answers?” Proc. Eurospeech 97, Rhodes, Greece, 1997.