EVALUATING THE DIALOGUE COMPONENT IN THE GULAN EDUCATIONAL SYSTEM

Pernilla Qvarfordt and Arne Jönsson

Department of Computer and Information Science
Linköping University, SE-581 83, LINKÖPING, SWEDEN
perqv@ida.liu.se, arnjo@ida.liu.se

ABSTRACT

In this paper we present results on the learning effects of using the GULAN educational system to understand spoken dialogue systems. The investigation is restricted to the dialogue management component, which uses a subset of the LINLIN dialogue manager. The results are based on an evaluation of questionnaires given to the students and of tutors assessment of the students knowledge before and after using the educational system. The study shows that the students knowledge on focus and dialogue structure for dialogue systems improved after using the system. We also gained insights on further development of the system.

1. Introduction

The GULAN system is a software package for teaching spoken dialogue systems technology [1]. The domain used in GULAN is the Yellow pages, but the current lexicon only includes information on restaurants and cinemas in Stockholm. Its aim is to provide students with means to study the various components of a spoken dialogue system. GULAN has been used for a couple of years at various universities and by different student groups in Sweden [2]. One component is a customizable dialogue manager, which allows students to study the behaviour of a dialogue system for various dialogue strategies.

2. The dialogue model

The dialogue model utilised in GULAN is based on the LINLIN dialogue manager [3]. This dialogue manager models dialogue in two structures; one concerns modeling dialogue structure, i.e. the structure of the dialogue as it proceeds in the interaction, and the other models focus structure, i.e. entities being discussed in the interaction. Information on dialogue structure and focus structure is modeled in dialogue objects.

The dialogue is controlled by a dialogue grammar with three descriptive levels, the dialogue (D), a recursive initiative-response (IR) level, and the move. Typical moves include question Q, and answer, A. Dialogue structure in the LINLIN-model also distinguishes between various topical domains. A topical domain reflects what information source the dialogue manager needs to consult in order to understand a user request. LINLIN utilises three topical domains: the background system (t), information on the systems capabilities (s), and information on the ongoing dialogue (d). To illustrate this, consider a system with information on places to eat, such as restaurants and cafées. Such a system can not provide information on the user request *What computer stores are there on Odengatan?* If the background system had been consulted, the response would be *There are no computer stores on Odengatan*, i.e. the user's presupposition that the database contains information on computer stores would not be addressed. A more proper response should inform the user that the system does not have information on computer stores, i.e. information on the systems capabilities[1].

Focus structure is modeled in parameters corresponding to information in the background system. For information seeking applications we have found that a user initiative is often a request for database information; users specify a database object, or a set of objects, and ask for the value of a property of that object or set of objects. In LINLIN this is modeled with one parameter pertaining to the database object and another to properties related to that object. The dialogue objects are recorded in a dialogue tree, where local focusing principles states how information from the previous segment, i.e. dialogue object, is copied to the current dialogue object. This dialogue object is then updated with information provided in the user initiative and the response from the background system.

3. Student tasks

The dialogue manager in GULAN only utilises a subset of the features of the LINLIN dialogue manager, and the specification of the dialogue manager's behaviour has been simplified in order to facilitate understanding of the basic tasks of a dialogue manager. The task for the students is to implement different focus and dialogue structuring strategies and discuss their implications for a dialogue system.

For studying dialogue structure the task for the students

[1]This has, however, not been utilised in the GULAN system that is evaluated in this paper.

is to write a dialogue grammar that can engage a user in a dialogue with the background system and discuss implications for the dialogue based on how the system handles different dialogue situations, such as when and how to initiate a clarification request.

Initially there is only one rule in the grammar which is able to handle regular information seeking requests of the form: Qt1 $\rightarrow$ At1 corresponding to a LINLIN-rule of the form: QtAt $\rightarrow$ Qt At. This rule states that a question, Q, asking for information found in the database, t, should have an answer with database information, At. The students define the behaviour in terms of how the question, Qt1, should be specified in order to activate the rule, At1. For instance, to handle the request *What restaurants are there on Odengatan?* Qt1 is specified as Qt1:Focus (FACILITY)CLASS(PLACE)INSTANCE meaning that the dialogue object contains a facility class, such as restaurant, and an instance of a place, such as Odengatan. The response is specified as At1: Answer up. The response provided from the action Answer is further specified as a canned text depending on how many instances that were found in the database. For example, if the database responds with one information item, the answer from the system is specified as: At1: (1) FACILITY on STREET is shown on the map. If clarification is needed, the students have to write another rule, Qt1 $\rightarrow$ Qd1[2], where the Qd further specifies what information to request from the user, e.g. a street, which is also used by the dialogue manager to ensure that the response from the user adhere to this.

The task for the students when studying focus structure is to modify the local focus behaviour, i.e. how to copy information from the previous IR-unit to the current IR-unit. This can be carried out in three ways: AddToFocus appends the information from the user move to the information copied from the previous IR-unit to the new, ChangeFocus replaces the information copied from the previous IR-unit and CmpAndChange changes some of the old information. Consider, the following interaction:

U1: Show me the restaurants on Valhallavägen
S2: Restaurants on Valhallavägen are
shown on the map
U3: Cinemas
S4: Cinemas on Valhallavägen are shown on the map

In this dialogue the focused object property[3] (restaurants) of the object (Valhallavägen) is replaced with cinemas using the CmpAndChange focusing principle. If the focus copying principle instead had been AddToFocus, the system would respond:

S4b: Cinemas and restaurants on Valhallavägen
are shown on the map

Finally, if the principle is ChangeFocus, no information is copied from the previous IR-unit and the system would need to ask for missing information. Again consider the interaction above. With ChangeFocus we would have the following continuation after S3:

S4c: You must provide a place, e.g. a street

The place information from the previous IR-unit, i.e. Valhallavägen, is changed to include only information from the user and as no information on place is provided, a clarification request is needed. The students not only change the focusing principles, but must also construct clarification requests, for instance, providing examples on what a facility means as in S4c.

4. The study

The aim of the study was to investigate if the students gained knowledge about dialogue management by using GULAN. In order to do so, assessments of students knowledge before and after using GULAN was collected. The study was conducted on a class of 27 students forming 13 groups taking an AI course at a Cognitive Science programme. The students had a lecture on the LINLIN dialogue manager and an article presenting the dialogue model [3]. They were then confronted with the system in a lab assignment which included a short introduction to the system.

Before starting to use the system the students had to fill in a questionnaire (the pre-questionnaire) on their understanding of dialogue and focus structure. The students filled in the pre-questionnaire before they read the short introduction to the system. After finishing the lab assignment, they also had to fill in a new questionnaire (the post-questionnaire). In parallel to this the two tutors ranked the knowledge of the students according to their responses on the pre-questionnaire and the performance on the lab assignment.

In the pre-questionnaire the 13 lab groups had to describe what they knew about dialogue and focus structure. They also had to rank their knowledge on a 1-5 scale. The questionnaire was filled in by each lab group.

The lab assignment was constructed so that the students had to try out the various focus structures, and to write a dialogue grammar, as discussed above. The students worked with the lab assignments in groups, two or three students in each. Each lab assignment group had to write a report that was handed over to the tutors for assessment. Of the 13 groups 12 finished the report.

The aim of the post-questionnaire was to measure various aspects of the students knowledge on dialogue and focus structure, as well as the general level of knowledge, i.e.

[2]In LINLIN this corresponds to modifying the QtAt to QtAt $\rightarrow$ Qt QdAd At which in turn requires another rule of the form QdAd where information is taken from the user, i.e. the ongoing dialogue. However, this is also simplified in the lab system.

[3]In GULAN we do not distinguish between Objects and Properties

Table 1: Tutors and students assessment of the overall knowledge of dialogue structure. The grades were 1-5.

	Median	Mean	Std. deviation
Tutor			
Pre-quest.	1	1.29	.582
Post-quest.	3	2.96	.964
Student			
Pre-quest.	1	1.25	.620
Post-quest.	3	2.67	.612

Table 2: Tutors and students assessment of the overall knowledge of focus structure. The grades were 1-5.

	Median	Mean	Std. deviation
Tutor			
Pre-quest.	1	1.37	.608
Post-quest.	3.5	3.38	.829
Student			
Pre-quest.	1	1.42	.900
Post-quest.	4	3.83	.661

if the students understood the dialogue/focus structure, understood how to change it, and understood the consequences of the changes. The aspects were measured using a semantic differential 7-grade scale. Assessments of the overall knowledge of dialogue and focus structure were collected using a 5-grade scale, the same as in the pre-questionnaire. It was also possible to give written comments on the understanding of dialogue and focus structure. Other measures collected include how much time a student spent on solving the tasks. The students filled in the post-questionnaire individually. A total of 24 students answered the questionnaire, which means that 3 students participated in a lab assignment group, but did not answer the post-questionnaire.

Students' ratings of their knowledge before and after the lab assignment were compared. In order to do so the median of the two to three student ratings from the post-questionnaire in each lab assignment group were compared with the ratings from the pre-questionnaire. Since not all participants in the lab assignment groups answered the post-questionnaire, the comparison could only be carried out for 9 groups.

The two tutors rated the knowledge according to the pre-questionnaire independently on a scale 1-5. They also rated the lab reports in the same manner as the students in the post-questionnaire. Each tutor rated all groups, and the final rating for each group was the sum of both tutors' ratings.

5. Results

Most of the lab assignment groups spent 3-4 lecture hours using GULAN. Only two groups spent more time with the system. No effect on the amount of time spent with the system could be noticed on the ranking of the assignment, mainly because so few groups spent more than four hours using the system. All results were gained using Wilcoxon Signed Rank test.

5.1. Dialogue structure

The grades from the tutors' assessment of the students knowledge on dialogue structure showed that the students had significantly higher grades after using GULAN ($Z=-2.944$, $p<.01$, two-tailed). The results also showed that the students believed that their knowledge on dialogue structure had improved. Using Wilcoxon Signed Rank test the rankings on their knowledge after using the system were significantly higher ($Z=-2.692$, $p<.01$, two-tailed) for the dialogue structure. The median and mean for the tutors' and students' assessment is shown in Table 1. There was no significant difference between the students' and the tutors' ranking of the overall knowledge of dialogue structure in neither the pre- nor the post-test.

Table 3: Comparison between assessment of various aspects of dialogue and focus structure.

Understanding of	Z-value	probability
Tutor		
focus/dialogue	-2.388	p<.05
how to change	-2.339	p<.05
consequence of change	-2.430	p<.05
Students		
focus/dialogue structure	-2.256	<.05
how to change	-3.137	p<.01
consequences of change	-3.561	p<.01

5.2. Focus structure

The grades from the tutors' assessment of the students knowledge on focus structure showed that the students had significantly higher grades after using GULAN ($Z=-2.944$, $p<.01$, two-tailed). The results also showed that the students believed that their knowledge on focus structure had improved. Using Wilcoxon Signed Rank test the rankings on their knowledge after using the system were significantly higher ($Z=-2.716$, $p<.01$, two-tailed) for the focus structure. The median and mean for the tutors' and students' assessment is shown in Table 2. There was no significant difference between the students' and the tutors' ranking of the overall knowledge of focus structure in neither the pre- nor the post-test.

5.3. Comparison between dialogue and focus structure

The comparison between dialogue and focus structure was made from three aspects; understanding the dia-

Table 4: Tutors assessment of students knowledge on the various aspects of dialogue and focus structure. The grades were given on a scale 0-6, and the two tutors' grades was summarised.

	Median	Mean	Std. dev.
Dialogue			
usage of structure	7.5	7.58	2.940
how to change	8	7.92	2.640
consequence of change	7.5	7.42	2.710
Focus			
usage of structure	9.5	9.00	1.910
how to change	10	9.50	1.780
consequences of change	9.5	9.17	2.040

Table 5: Students' assessment of their knowledge on the various aspects of dialogue and focus structure. The grades were given on a scale 0-6.

	Median	Mean	Std. dev.
Dialogue			
usage of structure	3.5	3.46	1.690
how to change	3	3.17	1.990
consequence of change	2.5	2.63	1.790
Focus			
usage of structure	4	4.33	1.400
how to change	5.5	4.71	1.520
consequences of change	5.5	4.37	1.500

logue/focus structure, understanding how to change it, and understanding the consequences of the changes. The results showed that the students thought that it was easier to understand the aspects on focus structure than on dialogue structure. Table 3 shows the results for these aspects. The median and mean for the tutors' assessment of the various aspects are shown in Table 4 and students' assessment are shown in Table 5.

The overall grades from the tutors' assessment of the students showed that the students had significantly higher grades on focus structure than dialogue structure ($Z=-2.134$, $p<.05$, two-tailed). The students overall ranking on their knowledge of dialogue and focus structure showed the same tendency ($Z=-2.716$, $p<.01$, two-tailed).

5.4. Comments from students

The students comments on dialogue structure were that it was hard to use. One reason, according to the students, is that there were many sources of information to keep track of, for example the dialogue grammar, the dialogue tree, and the lexicon.

Most of the students said they believed that the focus structure was easier to modify. Some wanted to write rules, just as with the dialogue structure. In this way the focus structure could be changed depending on what kind of question the user had asked the system.

6. Discussion

Students knowledge on dialogue and focus structure improved using the GULAN dialogue component. However, their knowledge on dialogue and focus structure before starting to work with the assignment was very low. One reason for this is that not all students had read the required paper before doing the assignment. It is still an open question if students with better pre-knowledge would have improved their knowledge to the same extent. What the study shows is that for courses, such as AI, which includes many different topics, the education system helps students understand properties of dialogue systems.

Students thought that it was easier to understand focus structure than dialogue structure. This might be due to the fact that the simplifications made to the dialogue component of GULAN did not allow enough dialogue structure modifications in order to understand its implications for the interaction. Further refinements of the system will modify the dialogue component to more resemble the dialogue manager of LINLIN. This includes, for instance, allowing the students to also study the use of a system information database, as discussed above, and allow for more varied focus structure modifications, as also proposed by the students. This has the benefit that the educational system corresponds better with the course material.

The evaluation has investigated a subset of the LINLIN dialogue manager and more specifically the notions of focus and dialogue structure. Although these concepts are not general, dialogue systems need means to control the interaction and means to record entities being discussed during the conversation and in that respect the results are general.

Acknowledgments

This work is supported by The Swedish Transport & Communications Research Board (KFB) and the joint Research Program for Language Technology (HSFR/NUTEK).

7. REFERENCES

1. Rolf Carlson, Björn Granström, Joakim Gustafson, Erland Levin, and Kåre Sjölander. An integrated system for teaching spoken dialogue systems technology. In *Proceedings of Eurospeech'97, Rhodes, Greece*, 1997.
2. Joakim Gustafson, Patrik Elmberg, Rolf Carlson, and Arne Jönsson. An educational dialogue system with a user controllable dialogue manager. In *Proceedings of ICSLP'98, Sydney, Australia*, 1998.
3. Arne Jönsson. A model for habitable and efficient dialogue management for natural language interaction. *Natural Language Engineering*, 3(2/3):103–122, 1997.