Dialogue Models for Bilingual Human-Computer Interaction in a City Information System*

Jana Schwarz[1] and Václav Matoušek[2]

[1] Institute of Slavistics, Technical University of Dresden, Germany
Jana.Schwarz2@mailbox.tu-dresden.de
[2] Department of Computer Science and Engineering,
University of West Bohemia in Plzeň, Czech Republic
matousek@kiv.zcu.cz

Abstract. In this paper we report on some special problems arising during the building of a bilingual dialogue system, concerning especially parsing and dialogue modeling. We describe an unique automated formalization technique, with help of which authentic dialogues recorded in city information centres can be transformed into a special internal representation, closed to the symbolism of pseudo-flowcharts. The dialogue representation consists of many simple dialogue steps, each of them represents a very small part of the dialogue being described by 2 – 3 sentences. The use of the developed technique requires the preassignment of dialogue acts to the corresponding problem domain, preselection of typical user utterances and their careful transcription into grammatically correct defined sentences, machine comparison of transcribed user's utterances with preselected domain oriented sentences and recognition of the concrete dialogue step. The developed formalization technique enables the generation of an effective dialogue management, further implemented in a standard city information system communicating with the user in two different languages – in this case in Czech and German ones. The second prestigious advantage of the presented technique is the possibility of an automatic generation of the grammar, formally describing the subset of utterances (sentences) from the given domain; the grammar is later used in the step of linguistic analysis of user utterances.

1 Introduction

Effective dialogue management is a key issue in speech-based interfaces to all information systems since it can ensure a cooperative interaction with the user. Cooperativeness requires techniques which allow the user to efficiently access information and also techniques which compensate for limitations in system knowledge and speech technology. Our guiding principle is that dialogue is a joint activity in which user input is interpreted as instructions how to update the system's model of the evolving dialogue, and that system output transparently reflects the state of the model. The model itself encodes techniques for contextual semantic interpretation, and dialogue strategies to handle clarification, confirmation and failure-repair in manner appropriate to dialogue

* The work presented in this paper was supported by the Bilateral German – Czech Research Cooperation Project CZE-00-034

progression. The authentic (real) human-human dialogues recorded in city information centres of 19 Czech towns were used as the initial material to develope the first version of the dialogue system. The selected dialogues were transcribed with help of German GAT Transcription System into the written form and further linguistically analyzed. By analyzing human-human and several human-computer dialogues (simulated by students of the Multimedia Systems branch), we find that while users expect reduced linguistic competence from a system (and reflect this in a simplification of their own linguistic behavior), they still expect the system to retain many of the characteristics of a human service agent, whilst compensating for its own performance limitations.

Some of the main dialogue characteristics with system include global structure (the dialogue has an opening, a body and a closing, withal in the body, the system takes responsibility for obtaining information), mixed initiative (the user generally takes the initiative, but the system can take the initiative to confirm information has been correctly understood, obtain information necessary for database access and, if the dialogue deteriorates, to constrain the form and content of user utterances), over-informativeness (users may provide) over-informative answers), contextual interpretation (user input may only provide partial and ambiguous information whose interpretation needs to be estabilished in the discourse context) and failure-repair (users expect the system to repair failures which may arise from performance limitations, especially speech recognition, as well as the limitations in system knowledge). A minimal requirement is that the system plays the role of a co-operative agent so that the resulting interaction is comfortable, comprehensive and comprehensible to the user. From the user's point of view, whether the interaction is co-operative or not is judged solely on the basis of what the system says. Even if difficulties arise in the interaction, the system should still provide a response which does not lead to dialogue failure. To achieve this, both interpretation of user utterances and production of system utterances must be informed by past and current states of the interaction. Co-operative dialogue management, therefore, requires the construction and maintenance of an interactional model, i.e. a model which specifies the layers of structure which can be distinguished in dialogue interactions. We distinguish linguistic structure, attentional (or discourse) structure, and intentional structure. Intentional structure is further differentiated into dialogue structure and task one [5].

Our approach realized in a dialogue manager where each structural layer is represented in a separate model and each model, together with maintenance and update routines, was based on the use of dialogue grammars. When the dialogue manager is ready to process user input, the linguistic interface calls the parser with a set of predictions, and communicates the result of recognition and parsing to the dialogue module. The parser result is either a semantic representation of the user turn, or an error message. After semantic, task and dialogue processing, a set of goals is selected as the dialogue continuation. The linguistic interface passes these to a linguistic generator for linguistic processing and synthesis.

2 Dialogue Type and Sources

Our goal is to build a dialogue system for a city information system, that would allow users (tourists and guests as well as citizens of the town) to get information on cultural events in the city and surroundings. In order to bear in mind various types of cultural events and institutions, we prefer a multiple frame-based system, that will dynamically combine a certain number of frames (e.g. museums, galeries, and exhibitions). The system we report on is under construction.

The system should be able to allow mixed-initiative interaction, because it was found out that in real touristic interaction people do not have only one information need. Usually they start with some field of interest and continue asking for further information – for instance: "Where is the museum ? Is here only one museum?" – "How long is it open?" and so on. Furthermore, mixed-initiative systems are considered to be the most user-friendly, because they take into consideration the multifariousness of authentic human-human interaction. They meet the needs of more complex tasks that have to be handled by the dialogue system. Nevertheless, parts of the dialogue will be restricted and directed by the system. Our dialogue manager will be modelled after authentic information-seeking dialogues in Czech tourist information centres.

In our case, we have to deal with a practical, medium complex task (to find and provide exhausting information on museums, galeries, and exhibitions), which has to be solved collaboratively. After a greeting turn by the system, which is explaining the main features of the system, too, the user has to formulate his information need. Afterwards the system has to decide whether it is necessary to clarify or ask for additional information in order to answer the question. The following steps are chosen according to the actual dialogue state and the user's actions.

Based on the fact, that we are designing a dialogue system in a Czech town visited by numerous German tourists, we decided to enable the system to give information in German, too. This task will of course have several effects on our work.

3 Demands on the Dialogue System

Starting from the goals and needs formulated so far, we have to deal with the following special problems:

- Powerful speech recognizers for the two languages are needed.
- We need a grammar for language parsing in Czech and German. Since our first work is concentrated on Czech, it is necessary to test the transferability of our rule-based grammar.
- Our goal is to build a dialogue manager which reflects the interactional model of human-human interaction and the turn-taking system at the information desk, because we want our system to be maximally cooperative and effective.
- It is necessary to create a context-handling module (including dialogue history as well as presuppositions and other knowledge in the given world), dictionaries with referring and synonymous expressions and an interpreter of illocutionary forces of user-utterances (their possible intentions in the given context).
- The grammars used for the creation of training corpora and language parsing have to be modulated in order to fit the needs of utterance-generating on the part of the system.

In the considerations given below we will focus on grammar modelling and the use of authentic dialogue models for the building of our dialogue system.

4 Language Parsing

In former publications, we reported on our work with the corpus of authentic interaction in touristic information centres [6]. We created a rule-based grammar for Czech information-seeking utterances and used it for the automatic generation of training sentences. This process had two effects: First, we realised problems and mistakes in our grammar and had so the chance to correct them early. Secondly, we created a big amount of training material useful for the speech recognizer.

This grammar reflects insights into spoken Czech (and phenomena in every spoken language) like elliptical utterances, ”strange” word order, colloquial vocabulary and morphology, syntactic fragments and so on. It takes into account the functioning of utterances in the given context (i.e. their illocutionary forces) and probabilities of occurrence and combination of certain rules and types.

It is understandably not sentence-, but utterance-based. According to the terminology of conversation analysis it would be better to name this items ”turn construction unit” (TCU). Each TCU has its concrete function in this context and can be combined with other TCUs. One or more TCUs form a turn, simultaneously providing so-called transition relevant places (TRP) [8]. These TRPs are crucial points – on the one hand, for the identification of the beginning of barge-ins by the user, on the other hand, for planning of barge-ins by the system without risking of user frustration. Furthermore, the methods of conversation and interactional analysis provide us the opportunity of handling fragments of user-utterances (examples see [6]).

For the purpose of testing the compatibility of our grammar with a German dialogue manager at the Regensburg University (see [1]), we transcribed our grammar into Java TM Speech Grammar Format [4] and expanded it with respect to other user activities and system activities. A very small part of this grammar looks as follows:

```
public <confirm> = (  (<yes> {$M} | <no> {$M}) )
                   {$M} ;
           <yes> = ( ano | no | jo | nono | ”ano chci” )
                   {dm_marker = yes};
           <no>  = ( ne | nene | nenene | kdepak | ”ne nechci” |
                   nechci )  {dm_marker = no};
```

Trying to ”translate” the Czech grammar into German turned out to be quite complicated, that is why we did it in fact only for a small part of the grammar in order to test it. It is inevitable to write a new German grammar for our system on the basis of authentic dialogue material using therefore parts of existing and efficient grammars.

The difficulties associated with bilingual performing of information giving are the following:
German and Czech, although being both indoeuropean languages and typologically related, show quite a lot of relevant differences. They differ with regard to word order (especially separable verbs in German, relatively free word order in Czech with restrictions for clitics and thematically accentuated parts of speech), definiteness, and degree of inflection (the case system in Czech is much more detailled, and congruence is therefore crucial).

First tests have shown that it is possible to translate the rules, templats and terminals fo our grammar, but almost every rule has to be corrected with respect to the mentioned problems (insertion of articles, change of cases, subsuming of non-terminals etc.). This method does not seem to be effective for a more sophisticated grammar, and it is quite sure, that it cannot cover all the possible utterances in this context for German.

We will have to implement an authentic-based grammar (even if it was from another task-field, in this case it is possible to expand it).

5 Dialogue Modeling

We are using the results and cognitions of conversation analysis and interactional linguistics [7] in order to construct and further improve our dialogue manager. It has to mirror the mode humans interact together, their manner of turn-taking (who takes initiative in which cases, in which state-of-communication and how; the sequencing of turns in real conversation and the conditional relevance of re-actions after certain actions: query-answer; ask for repeat-repeat etc.).

Information seeking dialogues consist of the following parts (subdialogues), each of them using special and common rules: start/domain (greeting, help, choice of the domain), keyword (concrete information need: address, opening hours ...), confirm (confirmation of user-needs), end. Furthermore, there are additional language models for handling of problems (misunderstandings because of acoustic troubles, bad intention-recognition, referential problems).

6 Formalization of the Dialogue

The transcribed dialogues were (for the time being manually) transformed into the relatively large set of flowchart diagrams created by means of the dialogue functional specification developed within the framework of the European Joint Project ACCeSS [1]. The flowcharts fully describe the decomposition of the dialogue strategy to simple dialogue steps; each dialogue step is represented by the separate flowchart diagram. To illustrate this technique, let us show the following (partially formalized) short dialogue (asking information about the visit to the Pilsen Brewery Museum):

Dobrý den. Mluvíte s informačním systémem města Plzně. Můžeme vám poskytnout informace o kultuře a turistice v našem městě a okolí. Z které oblasti si přejete informace? `<start>`

Dobrý den. Můžu se zeptat jak je to pivovarský muzeum otevřený ?
`<domain1 = museum><keyword1 = opening hours>`

Přejete si informace o pivovarském muzeu v Plzni ? `<confirm_domain1>`

Ano, chci. `<yes>`

Pivovarské muzeum sídlí v centru města Plzně, ve Veleslavínově ulici číslo šest. Pivovarské muzeum je otevřené každý den od desíti do osmnácti hodin. `<inform1><inform2>`

Hmmm. `<nothing>`

Potřebujete další informace ? `<offer>`

Je třeba tam objednat lístky pro skupinu ? `<keyword2>`

Chcete jít do pivovarského muzea se skupinou ? `<confirm_kw2>`

No a potřebuju vědět, jestli je musím rezervovat. `<yes>`

Pro rezervaci lístků do pivovarského muzea zavolejte prosím na číslo sedm nula šest dva osm osm osm. `<inform3>`

Jo díky a dokdy jste řekl, že je otevřený ?	`<thanks><keyword3>`
Chcete vědět otevírací hodiny pivovarského muzea ?	`<confirm_kw3>`
Ano.	`<yes>`
Pivovarské muzeum je otevřené každý den od desíti do osmnácti hodin.	`<inform2>`
Aha, dobře, děkuji, na shledanou.	`<accept>`
Přejeme vám příjemný pobyt v našem městě, na shledanou.	`<end>`

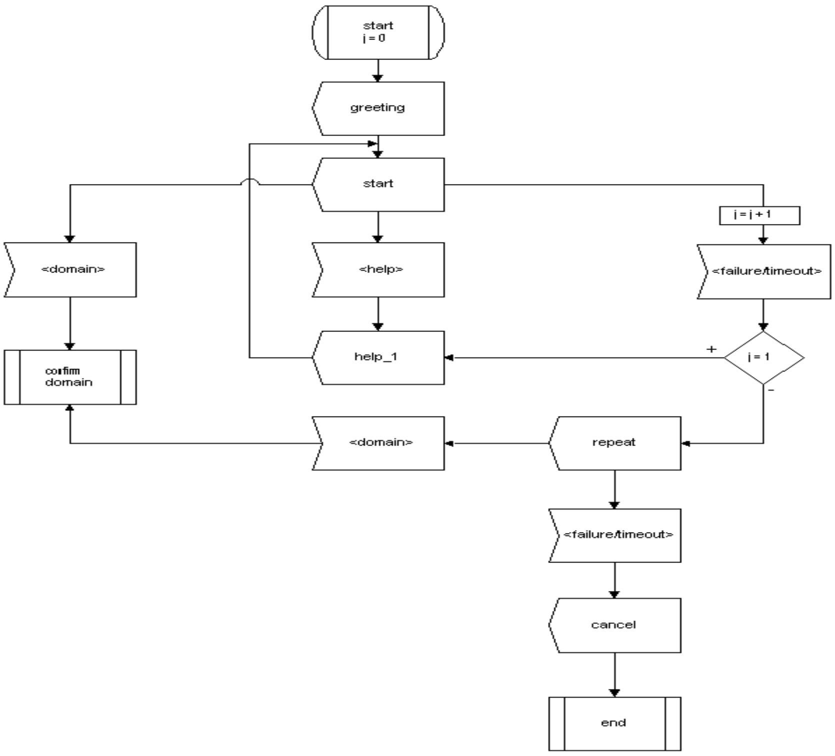

Fig. 1. Example of the formal description – a main dialogue path

The main dialogue task (flowchart of the highest level of the dialogue) as the illustrative example of the use of this technique is shown in the Fig. 1. Additionaly, the grammar of possible user's utterances was generated; the part of the grammar generated for the above presented example is shown below:

```
#JSGF V1.0;  /* JSGF grammar for Czech Information Centre */
grammar cic.museums;

/****************************************************
Subgrammar Start/Domain - activated at the beginning
****************************************************/
public <start> =
        ( <domain> | <problem> ) ;
public <confirm> =
        ( <yes> | <no> ) ;
```

```
/***********************************************************
Subgrammar Information - user reacts after information-giving.
***********************************************************/
public <accept> =
        ( <accept> | <nothing> ) ;
public <thanks_bye> =
        ( <thanks> <bye>  ) ;
public <domain_keyword> =
        ( <domain> | <keyword> ) ;
public <no_thanks> =
        ( <no> | <thanks> ) ;
public <confirm> =
        ( <yes> | <no> ) ;
/******************************************
  Rules domain, keyword
******************************************/
<domain, keyword> =
        ( (<INTENT>, (<GALL_SFN> | (<INDET_SNN>, <MUSEUM_SNN>)))
        | ([<GREET>,] [<POLITE>,] [<INTENT>,] ((["ta"] <EXHB_SFN>)
        | (["to"] <MUSEUM_SNN>) | (["ta"] <GALERY_SFN>)) "je"
          <INTERROG_LOC>)
        | (["a"] <INTENT>, [<HESIT>,] (<MUSEUM_SNN> | <GALL_SFN>)
          "nebo" <ALTERNAT> "ne\v{z}" "to" <MUSEUM_SMN> "je" "tady"
          <ALTERNAT>, <OPEN_SNN>)
        | ("je" "tady" [<POLITE>], ["n\v{e}kde"], (([<INDET_SNN>],
          <MUSEUM_SNN>)
        | (<GREET>, [<HESIT>,] <INTENT>, <INTERROG_LOC> "je"
           "vstup do" "toho" <MUSEUM_SNG>)
        | (<GREET>, <INTENT>, [<HESIT>] "je" "te\v{d}" <INTOWN>,
          <INDET_SFN>, <EXHB_SFN>)
                 *  *  *
        | (<POLITE> "jak" ("se dostanu" | "se dostanem") "do"
          <MUSEUM_SNG>)
        | (<TIME_LIMIT>, <OPEN_SM>, <MUSEUM_SMN>)    ) ;
<ALTERNAT> =
        ("e\v{s}t\v{e} n\v{e}co"
        | "e\v{s}t\v{e} n\v{e}co zaj\'{\i}mav\'{y}ho") ;
<CONDIT> =
        ( "nejsme tady autem pot\v{r}ebovali sme se kdy\v{z} tak
          maxim\'{a}ln\v{e} dopravit autobusem nebo vlakem") ;
<DAY_SFA> =
        ( "sobotu" | "ned\v{e}li | "st\v{r}edu" ) ;
<EXHB_SFN> =
        ( "v\'{y}stava" | "expozice" ) ;
<EXHB_SFA> =
        ( "v\'{y}stavu" | "expozici" ) ;
                .  .  .
```

The generated grammar, formally describing the subset of utterances (sentences) from the selected problem domain, is later used in the step of the linguistic analysis of user utterances, i.e. it is used as the base for partially automated implementation of the linguistic analysis module of the City Information Dialogue System.

7 Conclusion

Some problems arising during the building of a bilingual dialogue system, special parsing and dialogue modeling as well as the unique automated formalization technique, with help of which authentic dialogues recorded in city information centres are transformed into a special internal representation were presented in this paper. The use of the developed technique requires the preassignment of dialogue acts to the corresponding problem domain, preselection of typical user utterances and their careful transcription into grammatically correct defined sentences, machine comparison of transcribed user's utterances with preselected domain oriented sentences and recognition of the concrete dialogue step. But, the recorded dialogues must unfortunately be transcribed by hand into the written form, because the real recordings show usually bad quality and cannot be recognized and understood by the machine.

The developed formalization technique enables the generation of an effective dialogue management, further implemented in a standard city information system communicating with the user in Czech and/or German language. The second prestigious advantage of the presented technique is the possibility of an automatic generation of the grammar (formally describing the subset of user's utterances), which is later used in the step of linguistic analysis of user utterances. Both formalisms were used for the development and implementation of a simple experimental information dialogue system offering the information about several sightseeings and museums of the city of Plzeň.

References

1. Brey T., Hanrieder G., Heisterkamp P., Hitzenberger L., Penner M., Regel-Brietzmann P.: *Automated Call Center Through Speech Understanding System (ACCeSS) – Funktionale Spezifikation der Applikation*. Forschungsbericht zum Projekt LE 1802, Regensburg, Ulm, Juli 1999
2. Allen J., Byron D., Dzikovska M., Ferguson G., Galescu L., Stent A.: *Towards conversational human-computer interaction*. In: AI Magazine, 2001 (pdf-file at: www.cs.rochester.edu/research/cisd/pubs/).
3. Belina F., Hogrefe D.: *The CCITT-Specification and Description Language SDL.* In: Computer Networks and ISDN Systems 16 (1988/89), pp. 311 – 341.
4. http://java.sun.com/products/java-media/speech/forDevelopers/JSGF/
5. McGlashan S.: *Towards Multimodal Dialogue Management*. http://www.sics.se/∼scott/papers/twlt96/
6. Schwarz J., Matoušek V.: *Creation of a Corpus of Training Sentences Based on Automated Dialogue Analysis*. In: Matoušek V., Mautner P., Mouček R., Taušer K. (eds.): *Text, Speech and Dialogue*. Proc. of the 4th International Conference on Text, Speech and Dialogue in Železná Ruda (Czech Republic), September 2001. LNAI Series No. 2166, Springer Verlag, Berlin, Heidelberg, Sept. 2001, pp. 418 – 426.
7. Selting M., Couper-Kuhlen E.: *Argumente für die Entwicklung einer 'interaktionalen Linguistik'*. In: Gespraechsforschung: Online-Zeitschrift zur verbalen Interaktion 1(2000), pp. 76 – 95 (www.gespraechsforschung-ozs.de).
8. Selting M.: *TCUs and TRPs : The Construction of Units in Conversational Talk*. Universität Konstanz, InLiSt Nr. 4, 1998.