27

Developing a Dialogue System that Interacts with a User in Estonian

HALDUR ÕIM AND MARE KOIT

27.1 Introduction

There are many spoken dialogue applications in different languages available in the world: flight reservation systems worked out in USA within the DARPA programme, flight and train schedule systems developed in Europe within the SUNDIAL programme, the Verbmobil meeting agreement system in Germany, a help desk and bus schedule system developed within the Interact project in Finland, etc (McTear 2004).

No such system is available for Estonian so far. The analysis of actual human-human dialogues is needed in order to find out the variants of universal norms and rules that are typically used in particular language and culture. To get the empirical material, we are collecting Estonian spoken dialogues. The Estonian Dialogue Corpus (EDiC) presently includes about 600 spoken human-human dialogues. A typology of dialogue acts has been worked out and is used for annotating the corpus (Hennoste, Rääbis 2004; Gerassimenko et al. 2004, Hennoste et. al. 2003). The typology is based on the conversation analysis (CA) approach. Dialogue acts are divided into two big groups (1) acts that form adjacency pairs (AP) where the first part requires a certain second part (e.g. questions and answers) and (2) non-AP acts (e.g. acknowledgement).

In this paper, we shall analyse calls for information to find out methods and ways used by people for ordering and giving information, and to model them in a DS. Let us call these methods communicative strategies (cf. Jokinen 1996).

Inquiries into Words, Constraints and Contexts.
Antti Arppe et al. (Eds.)

27.2 Dialogue System as Conversation Agent

A conversation agent is a program that consists of six (interacting) modules (cf. Koit, Õim 1998):

$$\mathrm{DS} = \{\mathrm{PL}, \mathrm{TS}, \mathrm{DM}, \mathrm{INT}, \mathrm{GEN}, \mathrm{LP}\},$$

where PL – planner, TS – task solver, DM – dialogue manager, INT – interpreter, GEN – generator, LP – linguistic processor. PL directs the work of both DM and TS, whereby DM controls the communication process and TS solves domain-related tasks. The task of INT is to make the semantic analysis of a partner's utterances and that of GEN is to generate semantic representations of agent's own contributions. LP carries out linguistic analysis and generation. The conversation agent uses a knowledge base KB in its work. In our model, the KB consists of four components: $\mathrm{KB} = (\mathrm{KB_W}, \mathrm{KB_L}, \mathrm{KB_D}, \mathrm{KB_A})$, where $\mathrm{KB_W}$ contains world knowledge, $\mathrm{KB_L}$ linguistic knowledge, $\mathrm{KB_D}$ knowledge about dialogue and $\mathrm{KB_A}$ knowledge about interacting agents. $\mathrm{KB_A}$ has two parts: the knowledge of DS about itself and a partner model – the knowledge about a 'standard' user. A necessary precondition of for communication is existence of shared knowledge of interacting agents (e.g. a common picture of the world, a common language of interaction).

27.2.1 Frames of Dialogue Acts

The DS must be able to recognize a user's acts and generate its own responding acts. The full processing cycle of a dialogue act pair can be represented as follows:

speech recognition $\Rightarrow$ text analysis $\Rightarrow$ task solving $\Rightarrow$
text generation $\Rightarrow$ speech synthesis

The communicating agents exchange acts that express their goals. DS as a cooperative partner must take over the user's goal (in our case, the goal is to get information) and try to fulfil it. Therefore, the dialogue knowledge $\mathrm{KB_D}$ of a DS must include descriptions of dialogue acts that make it possible to infer user's goals. Dialogue acts can be represented as frames having the slots SETTING, GOAL, PLOT and CONSEQUENCE (cf. Saluveer, Õim 1985). The slot SETTING gives preconditions of the dialogue act, including the author's beliefs about the addressee (as a part of the partner model) that can be true or false. An unexpected reaction by the partner signals that a belief was wrong. For example, a user asks for a bus timetable supposing that the DS knows it but actually the data base includes only flight information. If the SETTING is not satisfied then the speaker initiates a subdialogue – asks a specifying question or initiates a repair.

Let us consider the frame of 'closed yes/no question'[1] (the idea is taken from Saluveer, Õim 1985; cf. Bunt 1999; Jurafsky, Martin 2000). The following notations are used: S speaker (author of the act), H hearer (addressee), p proposition (true or false, e.g. *This is a direct bus*). Both the user and the DS perform the roles of S and H alternately.

QUF:CLOSED_YES/NO
SETTING:
S has a wish to know whether p (or not-p)
S believes that H knows whether p
GOAL: H knows that S has a wish to know whether p
PLOT: S informs H that S has a wish to know whether p
CONSEQUENCE: H knows that S has a wish to know whether p

Example: *Is this a direct bus?*

27.2.2 User Model

For a dialogue system, a user (client, C) is a conversation agent like itself. In its work, DS supposes that C has analogous six processing modules and four knowledge bases as the DS itself, and that the intersection of its knowledge bases with those of the user is not empty (otherwise, the interaction would be impossible).

Let us consider the KB_A component of the knowledge base. It includes the knowledge of DS (1) about itself, and (2) about a 'standard' client – his/her beliefs, desires, intentions, and algorithms that are used to generate plans. In the case of information dialogues, a client's beliefs, desires and intentions are related to ordering and getting some information. When asking a question, C believes that DS has the needed information, and his/her intention is to get this information. When analysing a question, DS recognises C's beliefs and intention, and tries to satisfy his/her goal, i.e. to provide him/her the asked information. Therefore, a BDI model which operates with agent's beliefs, desires and intentions can be implemented here (cf. Allen 1995; Koit, Õim 1998; Koit, Õim 2004).

Every question or directive sets up a (new) goal that is reached if a requested answer is received. If C's goals are unsatisfied after (s)he got an answer, then (s)he initiates a clarification subdialogue, asking a new, specifying question. The dialogue manager must keep accounts of C's beliefs and goals. A suitable data structure is a stack. Every question/directive (the first part of

[1] Frame names (equal to dialogue act tokens in our typology of dialogue acts) are originally in Estonian. Every token consists of two parts separated by a colon: the first two letters form an abbreviation of the act group name (e.g. QU = question). The third letter is only used for AP acts: the first (F) or the second (S) part of an AP act. The second part of a token is the full name of the act.

an AP) adds a new goal (subgoal) into the stack, and every answer or fulfilling a directive (the second part of the AP) may delete the upper goal.

27.3 Information-Sharing Strategies in Estonian Spoken Dialogues

27.3.1 Overview of Empirical Material

20 institutional dialogues (calls for information) were chosen from EDiC where a client (C) orders information and an information provider (P) provides them. The calls are short – the average length of a dialogue is 13 utterances. The total number of utterances is 275 and the number of words – about 1,000. A typical call consists of three parts: a conventional beginning, main information part, and a conventional ending. The kernel of the information part is a question – answer (or directive – grant) AP: a question is asked (or a request made) and an answer (or grant) is obtained. Still, subdialogues can occur after a question and/or answer: an adjusting/specifying question is asked and answered, or a repair for solving a communication problem is initiated and performed. The kernel can be repeated; more than one question can be asked and answered. In the analysed dialogues, C asks for a phone number in most cases (Table 1).

Table 1. What is being asked for

Client's goal	Number of dialogues
Phone number	16
Bus time	2
Film in cinema	1
Start of street	1
Address	1

Typically, C has only one goal (17 dialogues from 20), e.g. to obtain a phone number. The goal is reached after the answer is received, and then the conversation can be finished.

In the remaining 3 dialogues, C has more than one goal. In the first of them, C asks for the phone number of one travel bureau and then of another. These two goals can be considered as subgoals of a general goal – to take a trip (but it is beyond this dialogue). In the second dialogue, C similarly asks two questions, both of which are about a bus departure time. Supposedly, C intends to take a bus. In the third dialogue, C's first four questions (which film, the start time today evening and tomorrow morning, the price of a ticket) point to his/her intention to go to the cinema. The last question concerns a phone number and is not connected with going to the cinema.

27.3.2 Dialogue Acts that Set Up Goals

All of the analysed dialogues have a standard beginning part – P responds the call by saying the name of the company (e.g. *Estmar information*[2]), introducing himself/herself (e.g. *Leenu is hearing*) and greeting (*good morning*). Typically, C responds to the greeting and immediately requests information. Every question or request sets up a goal. In a cooperative conversation, P will share C's goal and assist C in reaching it.

A limited number of dialogue acts are used to express the goal (Table 2): request (e.g. *give me the teachers room of the Karlova school*), an indirect speech act which we call open yes/no question (*could you tell me the departure time of the bus to Tallinn*), or alternative question (*where does the Aleksandri Street begin—at the town centre or at the other end*). An advance note sometimes precedes a request (*I have a question*), or an additional information follows (*[the ticket office of the theatre Vanemuine,] such a place where tickets can be bought, please*). Some new goals are set up in such cases where P is not able to fulfill the request and offers substituting information (e.g. a phone number where C can get information).

In many dialogues, C starts his/her request with a cue phrase which precisely determines the following dialogue act: a request (*I wanted to know*), open yes/no question (*could you tell me*). Such phrases provide good features for automatic recognition of the dialogue act type (pragmatic analysis) and its meaning (semantic analysis).

P does not always succeed in giving a sufficient answer in the analysed dialogues. C obtains the requested information only in half of the cases, and substituting information in seven dialogues. (S)he does not get any information in three dialogues (e.g. the requested phone number is missing in the data base), therefore his/her goal will not be reached.

Table 2. Dialogue acts used by clients

[2] The examples are translated from Estonian.

Client's dialogue act	Typical phrases	Number of cases
request	*I'd like to know, please*	13
open yes/no question	*could you tell, is it possible to know*	6
accept of an offer/request	-	4
advance note + open yes/no/ alternative question	*I have such a question, one more question*	2
wh-question	*please tell me*	1
request + additional information	-	1

27.3.3 Communicative Strategies Used by Client

Calls for information form a simple dialogue type where the client has only one certain question in most cases. We found 27 questions/requests of C in analysed 20 dialogues.

P recognized C's goal immediately in 16 cases (Table 3) and either provided the requested information or informed C that it was missing. The information part of a dialogue consists of one adjacency pair of dialogue acts. In the remaining cases, C did not formulate his/her question precisely enough, and a subdialogue was started in the ensuing process. The initiator of a subdialogue is either P or C. In one dialogue, C reformulated his request three times, and in another dialogue, specified the answer (C: *Is it near the department store?* P: *Farther away, to the Lille hill.* C: *Is Lille the street which goes from the department store?*).

Table 3. How a client orders information

Strategy	Number of requests/questions
Request/question that does not need adjusting	16
Request/question that needs adjusting	
(a) client initiates adjusting	5
(b) information provider initiates adjusting	4
Request/question that needs reformulation	2

27.3.4 Communicative Strategies Used by the Information Provider

In a typical case, P gives the asked information either immediately or after adjusting (in 20 cases out of 27). If the requested information is missing then P either offers substituting information, or behaving non-cooperatively, does not offer anything (Table 4).

Table 4. How an information provider gives information

Strategy	Number of answers
The needed information exists and is provided immediately	12
The needed information exists and is provided after adjusting initiated by the client	5
The needed information exists and is provided after adjusting initiated by the information provider	4
The needed information does not exist; the provider offers a substitution	4
The needed information does not exist; the provider does not offer a substitution	2

In seven dialogues out of 20, P initiates an information-sharing subdialogue before answering. The subdialogue always consists of one AP of dialogue acts (cf. Hennoste et al. 2005): P's alternative question or wh-question followed by C's giving information, or P's question which offers an answer (sometimes clarification) followed by C's agreement. An information-sharing subdialogue explains which information C needs, and helps him/her to reach the goal. C reaches the original goal in three cases and gets substituting information in three cases. The answer turns out to be wrong in one case (the phone number in another town).

Information providers are specially trained to tell phone numbers. In the analysed calls, phone numbers consist of three, five or six digits. In case of three digits, all the digits are spelled out in sequence which the client acknowledges (*mhmh*). A number of five digits is spelled out in two parts – two and three digits separated by a micropause. C either repeats all the digits, or the last three ones, and P confirms (*yes*). A number of six digits similarly is given in two parts – the first three and the last three digits. C's response de-

pends on the length of pause between the two parts. In case of a long pause, C either repeats all the first three digits, or acknowledges them (*yes*). C always repeats all the last three digits. Sometimes (s)he adds the word *yes?* waiting for P's confirmation.

27.4 Information Provider as a Conversation Agent

The DS which performs the role of information provider implements a formal grammar for dialogue management (Figure 1). The grammar is based on APs of dialogue acts. When requesting information, a client uses the first part of an AP: question (QUF) or directive (DIF). Dialogue act names are terminals of the grammar (capital letters are used in act names).

The DS uses a stack to keep shared goals. C's request or question sets up a goal which goes to the stack. If DS needs additional information for answering then it initiates an information-sharing subdialogue by asking an adjusting question. The question asked sets up a subgoal of the original goal and goes to the stack onto the original goal. When the answer is obtained then the goal will be removed from the stack. If the stack is empty then all the goals have been achieved (Table 5).

Table 5. Example of using a goal stack

Utterance	**Dialogue act**	**Goal stack**
/---/		
C: *tell me please the phone number of the dentist Vigoroovit*	DIF:REQUEST	
P: *where the dentist is located*	QUF:WH-QUESTION	Phone number
C: *2, Tuglase Street*	QUS:GIVING_INFORMATION	Address Phone number
P:		~~Address~~ Phone number
/---/		

The described ideas and results of corpus analysis are implemented only partly at the moment. A DS is being worked out (author Margus Treumuth)

Notation	Number of z's
$(z)^+$	one or more
[z]	zero or one
$(z)^*$	zero or more

information_dialogue ::= beginning main_part ending
beginning ::= [RIF:INTRODUCTION] RIF:GREETING
RIS:GREETING
ending ::= RIF:THANKING RIS:PLEASE [RIF:GOODBYE
RIS:GOODBYE]
main_part ::= (ordering_information (ordering_information
giving_information)* (giving_information)$^+$
(ordering_information giving_information)*)$^+$
ordering_information ::= Questions_first | Directives_first |
(advance_note)* ordering_information
giving_information ::= Questions_second | Directives_second |
giving_information (additional_information)*
advance_note ::= SA:ADVANCE_NOTE
additional_information ::= AI:SPECIFICATION | AI:ASSESSMENT
Questions_first ::= QUF:CLOSED_YES/NO | QUF:OPEN_YES/NO |
QUF: ALTERNATIVE | QUF:WH-QUESTION |
QUF:OFFERING_ANSWER
Directives_first ::= DIF:REQUEST | DIF: PROPOSAL | DIF:OFFER
Questions_second ::= QUS:YES | QUS:NO | QUS:AGREEING_NO |
QUS:ALTERNATIVE:ONE | QUS:ALTERNATIVE:BOTH |
QUS:ALTERNATIVE:THIRD_CHOICE |
QUS:ALTERNATIVE:NEGATIVE |
QUS:GIVING_INFORMATION |
QUS:MISSING_INFORMATION | QUS:REFUSAL |
QUS:POSTPONEMENT
Directives_second ::= DIS:GIVING_INFORMATION |
DIS:MISSING_INFORMATION | DIS:REFUSAL |
DIS:AGREEING | DIS:DISAGREEING |
DIS:RESTRICTED_AGREEING | DIS:POSTPONEMENT

Figure 1. Grammar of a simple information dialogue

which interacts with the user in Estonian and gives information about the flights departing from the Tallinn airport. The user inserts his/her question on a web page (http://www.ut.ee/~treumuth/) in the form of a written sentence or phrase in Estonian, and gets an answer in form of text and/or synthesized speech. The world knowledge base KB_W contains information of flight times and destinations. The linguistic processor LP performs a morphological analysis of the user's utterances in order to find out the cue words, and uses ready-made sentence templates with some word forms generated by the morphological synthesis to compile the answers. The text-to-speech module is integrated into the DS.

27.5 Conclusion

We have analysed spoken human-human dialogues in Estonian with the aim of investigating how people request and receive information. Some information-sharing strategies used by clients and information providers have been established. DS that performs the role of an information provider is a conversation agent which consists of various functional blocks and uses various knowledge bases in its work. The dialogue management block uses a formal grammar of dialogue acts. The grammar expresses the idea of adjacency pairs of dialogue acts – one of fundamental ideas of conversation analysis. Every question or request of a client (the first part of an AP) is expecting an answer (the second part of the corresponding AP). Every question and request sets up a new goal or subgoal. DS as a cooperative partner shares client's goals. A stack is used for these shared goals. Every satisfactory answer removes a goal from the stack. For the DS, a user is a conversation agent similar to itself. Beliefs, desires and intentions of a user must be taken in account in order to give him/her the needed information. This work is still in progress. Our further work will concentrate on finding out of more detailed communicative strategies and on formal definitions of more dialogue acts that make it possible the automatic recognition of user goals in a cooperative dialogue system.

Acknowledgement

This work is supported by Estonian Science Foundation (grants 5685, 5534).

References

Allen, J. 1995. *Natural Language Understanding*. 2nd ed. The Benjamin/Cummings Publ. Comp., Inc.

Bunt, H. 1999. Dynamic Interpretation and Dialogue Theory. *The Structure of Multimodal Dialogue I.*, ed. M.M. Taylor, F. Neel, and D.G. Bouwhuis, 139–166. Philadelphia/Amsterdam: John Benjamins.

Hennoste, T., O. Gerassimenko, R. Kasterpalu, M. Koit, A. Rääbis, K. Strandson, M. Valdisoo. 2005. Information-Sharing and Correction in Estonian Information Dialogues: Corpus Analysis. *Proc. of the Second Baltic Conference on Human Language Technologies*. 249–254. Tallinn.

Gerassimenko, O., T. Hennoste, M. Koit, A. Rääbis, K. Strandson, M. Valdisoo, E. Vutt. 2004. Annotated Dialogue Corpus as a Language Resource: An Experience of Building the Estonian Dialogue Corpus. *Proc. of the First Baltic conference "Human Language Technologies. The Baltic perspective"*, 150–155. Riga.

Hennoste, T., M. Koit, A. Rääbis, K. Strandson, M. Valdisoo, E. Vutt. 2003. Directives in Estonian Information Dialogues. *Text, Speech and Dialogue. 6th International Conference TSD 2003*, ed. V. Matousek, P. Mautner, 406–411. Springer.

Hennoste, T., A. Rääbis. 2004. *Dialoogiaktid eesti infodialoogides: tipoloogia ja analüüs*. Tartu: TÜ.

Jokinen, K. 1996. Cooperative Response Planning in CDM: Reasoning about Communicative Strategies. *TWLT11. Dialogue Management in Natural Language Systems*, ed. S. LuperFoy, A. Nijholt, G. Veldhuijzen van Zanten, 159–168. Enschede: Universiteit Twente.

Koit, M., H. Õim. 2004. Argumentation in the Agreement Negotiation Process: A Model that Involves Natural Reasoning. *Proc. of the Workshop W12 on Computational Models of Natural Argument. 16th European Conference on Artificial Intelligence*, 53–56. Valencia, Spain.

Koit, M., H. Õim. 1998. Developing a Model of Dialog Strategy. *Text, Speech and Dialogue. International Conference TSD 1998*, 387–390. Springer.

Saluveer, M., H. Õim. 1985. Frames in linguistic descriptions. *Quaderni di Semantica* 6(2):282–292.

McTear, M.F. 2004. *Spoken Dialogue Technology. Toward the Conversational User Interface*. Springer.