Interaction with the Computer: How to Model Argumentation?*

Mare Koit
Institute of Computer Science
University of Tartu
mare.koit@ut.ee

Abstract

The paper considers communication in a natural language between two participants, *A* and *B*, where *A* has a communicative goal that his/her partner, *B*, will make a decision to do an action *D*. *A* has to argue the usefulness, pleasantness, etc. of doing *D* for *B*, in order to guide *B´*s reasoning into right direction. A computational model of argumentation is developed which includes reasoning.

1 Introduction

There are many dialogue systems (DS) that interact with a user in a spoken natural language and help him/her to solve practical problems, e.g. to book flights, to get information about bus or train timetables, to detect computer faults, etc. [McTear, 2004; Möller, 2004] Usually, such tasks do not include argumentation. Rather practical dialogue is implemented in such systems [Allen *et al.*, 2000]. On the other hand, there are tasks and situations where not only information seeking, but also argumentation is required. Therefore, modeling of argumentation process is one of our aims.

A computational model of dialogue is described in [Koit and Õim, 2004]. The agreement negotiation process is modeled where one participant is trying to influence his/her partner to agree to do an action. Argumentation is used by one participant to direct the reasoning of the partner. This type of dialogue constitutes one kind of so-called agreement negotiation dialogues [Yuan *et al.,* 2002]. Such a dialogue can be considered, on a more general level, as rational behaviour of conversation agents which is based on beliefs, desires and intentions of agents, at the same time being restricted by their resources [Allen, 1994; Jokinen, 1996; Webber, 2001].

Estonian spoken dialogues taken from the Estonian Dialogue Corpus (EDiC)[1] have been analyzed in [Koit, 2005, 2006, 2007] with the purpose to verify an argumentation model - calls of clients to travel agencies where a travel agent is using arguments for booking a trip by a client, and calls from an educational company to different institutions where a salesclerk argues for taking a training course by a client. The analyzed travel dialogues turn out to be mostly information seeking dialogues, a travel agent typically does not argue for a certain trip. Salesclerks, to the contrary, try to persuade clients stressing the usefulness of a course. Still, clients seldom agree to take a course. In most cases, a decision will be postponed. A client may develop collaboration with a salesclerk, looking together for arguments for taking a course, or be antagonistic, looking for counter-arguments. The study of actual spoken dialogues has demonstrated the need to refine the argumentation model.

In this paper, we will develop the model considered in [Koit and Õim, 2004]. The paper has the following structure. Section 2 gives an overview of the model. Partner influencing methods as well as ways of reasoning are considered. We will make links between the partner influencing and reasoning. Section 3 discusses implementation of the model. Section 4 gives a summary and plans for the future work.

2 Modeling the Communication Process

Let us consider conversation between two agents - *A* and *B* - in a natural language. In the goal base of one participant (let it be *A*) a certain goal G^A related to *B´*s activities gets activated and triggers in *A* a reasoning process. In constructing his/her first turn *A* must plan the dialogue acts (DA) and determine their verbal form as a turn r_1. This turn triggers a reasoning process in *B* where two types of procedures should be distinguished: the interpretation of *A´*s turn and the generation of his/her response r_2. *B´*s response triggers in *A* the same kind of reasoning cycle in the course of which s/he has to evaluate how the realization of his/her goal G^A has proceeded, and depending on this s/he may activate a new sub-goal of G^A, and the cycle is repeated: *A* builds a new turn r_3. Dialogue comes to an end, when *A* has reached his/her goal or abandoned it.

2.1 Reasoning Model

After *A* has expressed his/her desire to *B* that *B* will decide to do *D*, then *B* can respond with agreement or rejection, depending on the result of his/her reasoning. Rejection can be (but not necessarily) supported with an argument. These arguments can be used as giving information about the reasoning process that brought *B* to the given decision.

The general principles of our reasoning model are analogous to the BDI model [Allen, 1994] but it has some specific traits which we consider important.

First, along with desires we also consider other kinds of

* The support of he Estonian Science Foundation and the Estonian Ministry of Education and Research is acknowledged (projects No 7503, SF0180078s08, and EKKTT09-52).

[1] http://www.cs.ut.ee/~koit/Dialoog/EDiC

motivational inputs for creating the intention of an action in an actor (if the actor considers the action useful to him/her or s/he is forced to do it independent of his/her immediate wish - e.g. s/he is forced to do it by some obligation or by threats of the partner).

Second, we have worked out a detailed model of reasoning which leads to the emergence of the intention (goal) in an actor to do the action in question, or to refuse to do it. Our reasoning model is based on the studies in the commonsense conception of how the human mind works in such situations, cf. [D´Andrade, 1987] since in natural communication people depart from this conception, not from any scientific one.

In our model we try to reflect the main types of determinants that motivate humans to act. Thus the strategy used depends on which determinant is chosen as the target of influence.

The reasoning model consists of two functionally linked parts: 1) a model of human motivational sphere; 2) reasoning schemes. In the motivational sphere three basic factors that regulate reasoning of a subject concerning *D* are differentiated. First, subject may wish to do *D*, if pleasant aspects of *D* for him/her overweight unpleasant ones; second, subject may find reasonable to do *D*, if *D* is needed to reach some higher goal, and useful aspects of *D* overweight harmful ones; and third, subject can be in a situation where s/he must (is obliged) to do *D* - if not doing *D* will lead to some kind of punishment. These factors are called WISH-, NEEDED- and MUST-factors, respectively.

Resources of the subject concerning *D* are any internal and external circumstances which create the possibility to perform *D* and which are under the control of the reasoning subject.

The values of the dimension obligatory/prohibited are in a sense absolute: something is obligatory or not, prohibited or not. On the other hand, the dimensions pleasant/unpleasant, useful/harmful have a scalar character: something is pleasant or useful, unpleasant or harmful to a certain degree. For simplicity´s sake, it is supposed that these aspects have numerical values and that in the process of reasoning (weighing the pro- and counter-factors) these values can be summed up. Still, in reality people do not operate with numbers. On the other hand, existence of certain scales also in human everyday reasoning is apparent.

The model of motivational sphere of a subject can be represented by the following vector of weights:
$\boldsymbol{w}$ = (w(resources), w(pleasant), w(unpleasant), w(useful), w(harmful), w(obligatory), w(prohibited), w(punishment-for-doing-a-prohibited-action), w(punishment-for-not-doing an obligatory action)). In the description, w(pleasant), etc. means weights of pleasant, etc. aspects of *D*.

The aspects of *D* can be divided into positive and negative ones. The positive aspects are pleasantness and usefulness of *D* but also the punishment for not doing *D* if it is obligatory. The negative aspects are unpleasantness, harmfulness but also the punishment for doing *D* if it is prohibited.

The second part of the reasoning model consists of reasoning schemes that supposedly regulate human action-oriented reasoning. A reasoning scheme represents steps that the agent goes through in his/her reasoning process; these consist in computing and comparing the weights of different aspects of *D*; and the result is the decision to do or not to do *D*.

How does the reasoning itself proceed? It depends on the determinant which triggers it (WISH, NEEDED or MUST). In addition, a reasoning model, as a naïve theory of mind, includes some principles such as: People want pleasant states and do not want the unpleasant ones, People prefer more pleasant states to less pleasant ones.

As an example, let us present a reasoning procedure which is triggered by NEEDED-determinant, that is, if the subject believes that it would be useful (needed) to do *D* in order to reach some of the activated goals.

```
// Reasoning triggered by NEEDED-determinant
Presupposition: w(useful) > w(harmful) //
1.  Are there enough resources for doing D?
2.  If not then do not do D.
3.  Is w(pleasant) > w(unpleasant)?
4.  If not then go to 10.
5.  Is D prohibited?
6.  If not then do D.
7.  Is w(pleasant)+w(useful) > w(unpleasant)
+  w(harmful)  +  w(punishment-for-doing-a-
prohibited-action)?
8.  If yes then do D.
9.  Otherwise do not do D.
10. Is D obligatory?
11. If not then do not do D.
12. Is w(pleasant)+w(useful) + w(punishment-
for-not-doing-an-obligatory-action)>
w(unpleasant)+ w(harmful)?
13. If yes then do D.
14. Otherwise do not do D.
```

In the case of other input determinants (WISH and MUST) the general structure of the algorithm is analogous, but there are differences in certain steps.

2.2 Communicative Strategies and Tactics

A communicative strategy is an algorithm used by a participant for achieving his/her goal in interaction. An agent can realize a communicative strategy by means of several communicative tactics.

There is one relevant aspect of human-human communication which is relatively well studied in pragmatics of human communication and which we have included in our model as the concept of communicative space. Communicative space is defined by a number of coordinates that characterize the relationships of participants in a communicative encounter. Communication can be collaborative or confrontational, personal or impersonal; it can be characterized by the social distance between participants; by the modality (friendly, ironic, hostile, etc.) and by intensity (peaceful, vehement, etc.). Just as in case of motivations of human behaviour,

people have an intuitive, "naïve theory" of these coordinates.

In our model the choice of communicative tactics depends on the "point" of the communicative space in which the participants place themselves. The values of the coordinates are again given in the form of numerical values.

A communicative strategy for A can be presented as the following algorithm.

```
// Communicative strategy //
1.  Choose a communicative tactic.
2.  Choose an initial point in the communicative space.
3.  Implement the tactic to generate an utterance: inform the partner of the communicative goal (agreeing to do an action D).
4.  Did the partner agree to do D? If yes then finish (the communicative goal has been achieved).
5.  Give up? If yes then finish (the communicative goal has not been achieved).
6.  Change the point in the communicative space? If yes then choose a new point.
7.  Change the communicative tactic? If yes then choose a new tactic.
8.  Implement the tactic to generate an argument.
9.  Go to the step 4.
```

The participant *A* can realize his/her communicative strategy in different ways (using different arguments for): stress pleasant aspects of *D* (i.e. entice *B*), stress usefulness of *D* for *B* (i.e. persuade *B*), stress punishment for not doing *D* if it is obligatory (threaten *B*). We call communicative tactics these concrete ways of realization of a communicative strategy. Communicative tactics are ways of argumentation. The participant *A*, trying to direct *B*´s reasoning to the positive decision (to do *D*), proposes various arguments for doing *D* while *B*, when opposing, proposes counter-arguments.

There exist three tactics for *A* in our model which are connected with three reasoning procedures (WISH, NEEDED, MUST). By tactics of *enticing* the reasoning procedure WISH, by tactics of *persuadin*g the procedure NEEDED and by tactics of *threatening* the procedure MUST will be tried to trigger in the partner.

The participant *A* when implementing a communicative strategy uses a partner model - a vector w^{AB} - which includes his/her imagination about weights of the aspects of the action *D*. The more *A* knows about *B* the more similar is the vector w^{AB} with the vector w^{B} of the motivational sphere of the partner *B*. We can suppose that *A* has sets of statements for influencing the weights of different aspects of *D* for the partner *B*: $\{st^{A}_{i}\text{-}asp_j,\ i=1, \ldots, n^{A}_{asp\,j};\ j=1, \ldots, n\}$ where asp_j is the *j*-th aspect of *D* and *n* is the number of different aspects. All the statements have their weights as well.

If *A* uses a statement $st^{A}_{i}\text{-}asp_j$ then s/he can increase/decrease the weight of that aspect in the following way:

$w(asp_j) := w(asp_j) + w(st^{A}_{i}\text{-}asp_j)$ if asp_j is positive, and

$w(asp_j) := w(asp_j) - w(st^{A}_{i}\text{-}asp_j)$ if asp_j is negative

where $w(st^{A}_{i}\text{-}asp_j)$ is the weight of the statement $st^{A}_{i}\text{-}asp_j$.

For illustration, let us present a schematic description of the tactic of persuasion, based on the reasoning procedure NEEDED. The general idea underlying this tactic is that *A* proposes statements for usefulness of *D* trying to keep the weight of usefulness for B high enough and the possible negative values of other aspects brought out by *B* low enough so that the sum of positive and negative aspects of *D* would bring *B* to the decision to do *D*. *B* may verbalize is rejection to do *D* bringing out a certain statement about a certain aspect of *D* (e.g. if *B* says *I do not have enough time* then s/he indicates that resources are missing for doing *D*). On the other hand, *B* may only say *No, I don´t do*, without any argument. Still, we can suppose that *B* has a set of statements $\{st^{B}_{i}\text{-}asp_j,\ i=1, \ldots, n^{B}_{asp\,j};\ j=1, \ldots, n\}$ for indicating the aspect which weight caused his/her rejection where asp_j is the *j*-th aspect of *D* and *n* is the number of aspects. We suppose that *A* when persuading uses each statement only once.

```
//Persuasion: A persuades B to do D//
while B is rejecting AND A is not giving up do
 case B´s answer of
   stB-resources //no resources//:
if there are statements stA-resources then present a statement stA-resourcesi in order to point at the possibility to gain the resources, at the same time showing that the cost of gaining these resources is lower than the weight of the usefulness of D.
//The expected result:
 wB(resources):=wB(recourses)+w(stA-resourcesi)//
  else exit //there are no more statements, give up//
   stB-harmful //much harm//:
if there are statements stA-harmful then present a statement stA-harmfuli to decrease the value of harmfulness in comparison with the weight of usefulness
// The expected result:
 wB(harmful):=wB(harmful)-w(stA-harmfuli)//
   stB-unpleasant //much unpleasant//:
if there are statements stA-unpleasant then present a statement stA-unpleasanti in order to downgrade the unpleasant aspects of D as compared to the useful aspects of D
// The expected result:
 wB(unpleasant):=wB(unpleasant)-w(stA-unpleasanti)//
   stB-punishment-for-doing-a-prohibited-action //D is prohibited and the punishment is great//:
if there are statements stA-punishment-for-doing-a-prohibited-action then present a statement stA-punishment-for-doing-a-prohibited-actioni in order to downgrade the weight of punishment as compared to the usefulness of D
```

//The expected result:
w^B(punishment-for-doing-a-prohibited-action):=w^B(punishment-for-doing-a-prohibited-action) - w(st^A-punishment-for-doing-a-prohibited-action$_i$)//

st^B-pleasant //little pleasant//:

if there are statements *st^A-pleasant* **then** present a statement st^A-pleasant$_i$ in order to stress pleasantness

else if there are statements *st^A-unpleasant* THEN present a statement st^A-unpleasant$_i$ in order to downgrade unpleasantness

st^B-obligatory //not obligatory; in such a case, B´s reasoning finished on the step 11, see above//

if there are statements *st^A-pleasant* **then** present a statement st^A-pleasant$_i$ in order to stress the pleasant aspects of D

else if there are statements *st^A-unpleasant* **then** present a statement st^A-unpleasant$_i$ in order to downgrade the unpleasant aspects of D

end case

if there are statements *st^A-useful* **then** present a statement st^A-useful$_i$ in order to stress usefulness

//The expected result:
w^B(useful):=w^B(useful)+w(st^A-useful$_i$)//

else exit //give up//.

3 Discussion

When communicating in a natural language, where *A* tries to influence *B* in order to bring him/her to a decision, *A* uses several statements to increase the weights of the positive aspects and to decrease the weights of the negative aspects of the action under consideration.

If *B* indicates a certain aspect which does not allow him/her to do *D* then *A* simply can choose a statement for attacking this aspect until there are statements at his/her disposal. When reasoning, *B* can make his/her negative decision on steps 2, 9, 11, or 14. For example, if *B* says that resources are missing then s/he arrived to step 2. If *A* points that resources can be obtained then *B* has to start his/her reasoning again from the beginning (step 1). If *B* indicates the harmfulness of *D* then s/he finished his/her reasoning on the step 9 or (in the case if *D* is obligatory) on the step 14. Receiving a statement from *A* that decreases the weight of harmfulness *B* has to go back to the step 7 or 12, respectively.

If *B* does not indicate a certain reason of rejection then *A* only can stress the usefulness.

Let us consider an example from the EDiC where a travel agent (*A*) calls back to a client (*B*) and offers a trip to a water centre[2]. *A* represents several arguments trying to persuade the client to take the trip.

A: .hh ee teie 'helistasite meile Abetsee 'reisidesse?
you have called our ABC travel agency
B: [jaa?]
yes
A: [ja] tundsite uvi meie Tartu reisi Aura 'keskuse vastu.
and were interested in a trip to Tartu, the Aura centre
.hh tähendab meil on nüüd niimodi=et me oleme pla'neerinud küll teisel no'vembril,
we have now planned a trip for November, 2.
.hh et kas 'teil oleks nagu huvitud
would you be interested in
see 'reis maksab meil 'kakssada=viiskend 'krooni.
this trip costs two hundred and fifty kroon
B: ja 'm:is sinna 'alla nagu 'käib.
and what does it include? //B needs more information; A will give it//
A: 'sinna läheb sis sisse meil üks 'tunniajane ekskurs'joon mööda Tartu 'linna see on ['Toome]mägi=ja=ja kõik=se 'ülikool=ja [niuke] kõik.
it includes a tour in Tartu for one hour, there are Dome hill and the university, and such all
=ja:=ja siis=ee veepargi 'pilet.
and then a ticket to the water park
.hh ee 'Tallinast Sakalast väljutakse ommikul kell: 'kaheksa [ja]
the departure is at 8 a.m. at Sakala in Tallinn
B: [jaa?]
yes
'tagasijõudmine on no õhtul ütleme tav- tavaliselt jõuavad nad kell kaheksa ka 'tagasi.
the arrival is in the evening, let's say, usually at 8 p.m.
//B does not make a decision therefore the usefulness of the trip is not big enough//
/.../
A: =et=on: näiteks 'üks laps, 'kaks last ja kaks 'täiskasvanut
if there are 1 child, 2 children and 2 adults, for example //A offers a discount, i.e. she tries to increase the usefulness of the trip for B//
B: jah=
yes
A: =siis meil on pere'soodustus ja see on 'kakssada kakskend krooni per 'inimene sel 'juhul.
then we have a family discount and it is two hundred and twenty kroon per person in this case
B: mhmh,
uhuh
/.../
B: siis=ee kas: 'on ka nagu mingeid ma=i=tea 'toidu kuidas öelda 'pause või selliseid.
then, are there some breaks, food breaks //B is looking for a counter-argument but A can avert it//
A: ee toidupause no 'kindlasti tehakse 'üks paus 'Adaveres.
food breaks will be made, one break will be in Adavere
/.../
B: ma olen 'kuulnud=et siin on 'nädalavahetustel vist üsna [nagu pikad 'järjekorrad.]

[2] Transcription of conversation analysis is used.

I have heard that there are quite long queues on weekends
//B still represents a counter-argument but A averts it//
A: [ee seal on üle'üldse] väga suured prob'leemid ja väga pikad 'järjekorrad [ütlen] teile 'ausalt
yes there are big problems, very long queues
'bussi'juht seisab 'järje'korras [ja] 'elavast järjekorrast võtab võtab meile
the bus driver stands in a queue and takes (tickets) for us
aga teised teevad sel ajal:] giidiga ekskurs'jooni=ja bussi-juht seisab sis=seal 'järjekorras lihtsalt.
the others make a guided tour this time and the bus driver simply stands in a queue
/.../
A: .hh aga mm kui=te=nüd 'soovite, ma panen teid 'kirja.
if you want, I register you //A repeats the offer but B refuses to make a decision//
B: ee ma vel 'mõtlen.
I´ll think more.

4 Conclusion

We have worked on different aspects of developing a model of dialogue, including its computer realization in the lines of BDI model. The main specific traits of our model are: 1) taking into account the "naïve" common-sense reasoning as the basis of dialogue, 2) modeling dialogues where the initiator´s goal is to impose the partner to do a certain action. As a practical realization of the model we have in view a computer program which we call "communication trainer". An experimental dialogue system is implemented which in interaction with a user can play the role of both *A* or *B*. At the moment the computer operates with semantic representations of linguistic input/output only, the surface linguistic part of interaction is provided in the form of a list of ready-made and classified utterances (statements) which are used both by the computer and user. In the future, we will go over to processing of natural language input and output. When considering generation of argumentative utterances in Estonian, we plan to implement the ideas represented in [Elhadad, 1995] and [Piwek and Deemter, 2007].

We are continuing our work in the following directions: 1) refining the reasoning model, 2) specifying resistance strategies, 3) specifying descriptions of actions using frames, 4) developing linguistic knowledge, 5) analysis of human-human dialogues in the EDiC in order to verify the model.

References

[Allen, 1994] James Allen. *Natural Language Understanding.* Second Edition. The Benjamins/Cummings Publ. Co. Redwood City etc., 1994.

[Allen *et al*., 2001] J. Allen, D.K. Byron, M. Dzikovska, G. Ferguson, L. Galescu, A. Stent. Towards conversational human-computer interaction. In *AI Magazine*, 2001. http://www.cs.rochester.edu/research/cisd/pubs/2001/allen-et-al-aimag2001.pdf

[D´Andrade, 1987] R. D´Andrade. A Folk Model of the Mind. In D. Holland and A. Quinn, eds. *Cultural models of Language and thought,* 112-148. Cambridge University Press, London, 1987.

[Elhadad, 1995] Michael Elhadad. Generating coherent argumentative paragraphs. In *Journal of Pragmatics,* 24: 189-220. North Holland, Elsevier.

[Jokinen, 1996] Kristiina Jokinen, Rational Agency. In M. Fehling, ed. *Rational Agency: Concepts, Theories, Models, and Applications, Proc. of the AAAI Fall Symposium,* 89-93. MIT, Boston, 1996.

[Koit, 2007] Mare Koit. Collaboration and Antagonism in Estonian Spoken Dialogues: Corpus Analysis. In *Proc. of the Workshop on Computational Models of Natural Argument. 20th International Joint Conference on Artificial Intelligence*, 31-34. Hyderabad, India, 2007.

[Koit, 2006] Mare Koit. Argumentation in Negotiation Dialogues: Analysis of the Estonian Dialogue Corpus. In *Proc. of the Workshop W2 on Computational Models of Natural Argument. 17th European Conference on Artificial Intelligence*, 173-176. Riva del Garda, Italy, 2006.

[Koit, 2005] Mare Koit. Argumentation in Institutional Dialogues: Corpus Analysis. In Chris Reed, ed. *IJCAI-05 Workshop on Computational Models of Natural Argument. Working Notes,* 80-83, Edinburgh, Scotland, 2005.

[Koit and Õim, 2004] Mare Koit and Haldur Õim. Argumentation in the Agreement Negotiation Process: A Model that Involves Natural Reasoning. In *Proc. of the Workshop W12 on Computational Models of Natural Argument. 16th European Conference on Artificial Intelligence*, 53-56. Valencia, Spain, 2004.

[Möller, 2004] Stephan Möller. *Quality of telephone-based spoken dialogue systems.* Springer, New York, 2004.

[McTear, 2004] Michael F. McTear *Spoken Dialogue Technology: Toward the Conversational User Interface.* Springer, London, 2004.

[Piwek and Deemter, 2007] Paul Piwek and Kees van Deemter. Generating under Global Constraints: the Case of Scripted Dialogue. In *Journal of Research on Language and Computation*, 5(2):237-263, 2007.

[Yuan *et al.*, 2002] T. Yuan and D. Moore and Alec Grierson. Human-Computer Debate: a Computational Dialectics Approach. In *Proc. of CMNA-02*, 2002. http://www.csc.liv.ac.uk/~floriana/CMNA/-YuanMooreGrierson.pdf

[Webber, 2001] B. Webber. Computational Perspectives on Discourse and Dialogue. In D. Schiffrin, D. Tannen, H. Hamilton, eds. *The Handbook of Discourse Analysis,* 798-816. Blackwell Publishers Ltd., 2001.