Argument and Computation, 2015
Vol. 6, No. 2, 101–129, http://dx.doi.org/10.1080/19462166.2014.915233

A computational model of argumentation in agreement negotiation processes

Mare Koit[a*] and Haldur Õim[b]

[a]*Institute of Computer Science, University of Tartu, Tartu, Estonia;* [b]*Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia*

(*Received 17 May 2013; accepted 3 April 2014*)

The paper describes a computational model that we have implemented in an experimental dialogue system (DS). Communication in a natural language between two participants A and B is considered, where A has a communicative goal that his/her partner B will make a decision to perform an action D. A argues the usefulness, pleasantness, etc. of D (including its consequences), in order to guide B's reasoning in a desirable direction. A computational model of argumentation is developed, which includes reasoning. Our model is based on the studies in the common-sense conception of how the human mind works in such situations. Theoretical considerations are followed by an analysis of Estonian spoken human–human dialogues. First, calls of clients to travel agencies are studied where a travel agent could use various arguments in order to persuade a client to book a trip. The analysis demonstrates that clients are primarily looking for information; argumentation occurs in a small number of dialogues. Secondly, calls of sales clerks of an educational company to different organisations are analysed where training courses are offered. The sales clerks, unlike the travel agents, try to persuade clients, stressing the usefulness of a course. Finally, face-to-face conversations are studied where one participant is arguing for an action by the partner. The corpus analysis shows that our model covers simple situations occurred in actual dialogues. Therefore, we have chosen a limited application of the model in a DS – communication trainer. The computer can optionally perform A's or B's role and the interaction with the user in Estonian follows norms and rules of human–human communication.

Keywords: formal models of argumentation; argumentation in human–computer interaction; argumentation support; conversational agent; reasoning model; communicative strategy; communicative tactics

1. Introduction

There are many dialogue systems (DSs) that interact with a user in a natural language and help him/her to solve practical problems, e.g. to book flights, to receive information about bus or train timetables, to detect computer faults, etc. (Jokinen & McTear, 2009). Usually, such tasks do not include argumentation. Rather practical dialogue (Allen, Ferguson, & Stent, 2001) is implemented in these systems. On the other hand, there are tasks and situations where not only information retrieval but also argumentation are required. Argumentation in the general sense is a large and complicated research field of human interaction with its specific units, structures and rules, the concrete inventory of which depends on the interaction domain, participants and their goals.

This paper describes a computational model of agreement negotiation processes, which involves natural reasoning. More detailed treatment of the theoretical framework of our approach will be given in the next section, but a preliminary delimitation of the senses in which we use the concepts 'agreement negotiation' and 'natural reasoning' would already be appropriate here.

[*] Corresponding author. Email: mare.koit@ut.ee

First, the general type of interaction we are dealing with represents a kind of directive interaction where the goal of one participant, *A*, is to get another participant, *B*, to carry out a certain action *D*. In this interaction situation, the concepts of agreement and negotiation have quite specific contents, as compared with the generally accepted ones (e.g. Rahwan et al., 2004). We want to stress two points. (1) 'Agreement' means here primarily '*B* agrees/does not agree to do *D*', that is, *B*'s agreement concerns the wish of *A* that *B* would do *D*, not some general standpoint of *A* concerning the situation or problem under discussion and (2) the interaction (dialogue) between *A* and *B* can be considered as negotiation because we are studying such interactions (dialogues) only where *B*'s decision to do or not *D* is preceded by a discussion where different aspects of 'doing *D*' (*D* as an acceptable goal, different kinds of resources, sub-actions, etc. needed to carry out *D*) are considered by the participants, that is, it is preceded by a process of argumentation in the sense of natural human interaction.

Second, when dealing with argumentation, we will concentrate on the process of reasoning as its central part. Argumentation is used by communication participants to direct reasoning of the partner. And here again, the specific context determines the choice of the focus of our interests. We want to model a 'naïve' theory of reasoning, a 'theory' that people themselves use when they are interacting with other people and trying to predict and influence their decisions. Naive in the sense that it is not scientific, that is, not constructed according to any system of principles accepted in scientific discourse, but nevertheless is based on certain principles which, taken together, form a conceptual whole that can be characterised as a 'theory'. In psychology, the corresponding approach is known as Theory Theory, which says that people use a 'naïve' theory of psychology to infer the mental states of other people and understand or predict their future behaviour (e.g. Carruthers & Smith, 1996). In this sense, the conception of Theory Theory constitutes a subtopic of the Theory of Mind, a general research area of scientific psychology, which deals with human mental states and processes, including those involved in communication. We use the term 'natural reasoning' in the context of this general approach.

Departing from these theoretical considerations, one of our aims has been to investigate actual dialogues, not artificially constructed ones. Analysis of human–human dialogues can provide information about their structure and linguistic features for developing DSs, which interact with a user in a natural language. For this paper, three sub-corpora of the Estonian Dialogue Corpus (EDiC) were analysed. EDiC includes mainly information-seeking dialogues, that is why it was hard to find argumentation for and against doing an action. However, as a result of the corpus analysis, typical sequences of dialogue acts (DAs) were found in human–human spoken dialogues that form agreement negotiations and reflect reasoning of the participant who has to make a decision about an action.

We have implemented the model in an experimental DS. We have planned one of the practical applications of the DS as a participant in communication training sessions – communication trainer. Here the DS can establish certain restrictions on argument types, on the order in the use of arguments and counter-arguments, etc. But let us stress that our interest lies in the development of such DSs as are able to interact with human users by following not only the formal rules of exchanging information but also the 'rules' (regularities) of human reasoning, specifically in the process of argumentation. The training system described represents just one of the implementations by using which the adequacy of our theoretical approach can be checked.

The main goal of the paper is to introduce our formal model of argumentation, which is based on a naive theory of reasoning, to justify the model on actual human–human argumentation dialogues and implement a simple DS, which includes the model of reasoning and makes it possible to demonstrate how the model will work.

The paper has the following structure. Section 2 describes the theoretical background of the study and gives an overview of the related research. Section 3 introduces the model of a conversational agent. Partner-influencing methods as well as ways of reasoning are included in the model. We will make links between the reasoning and partner influencing. In Section 4, three sub-corpora of EDiC are analysed, in order to justify the model. Section 5 discusses the implementation of the model in a simple DS – a communication trainer. Section 6 gives a summary and plans for further work.

2. Theoretical background and related work

Communication is the only means that agents can use to influence the actions of other agents and thereby to increase their knowledge about their communication partners (Wells & Reed, 2006). There are certain methods that people use in the construction of communication. These methods grant the purposeful, result-producing and coherent nature of communication. Reasoning-influencing strategies are one such method. They exemplify ways of influencing and directing the reasoning and decision-making processes of partners in communication – such as asking, luring, persuading, insisting, demanding, threatening, etc. Since the use of reasoning-influencing strategies critically depends on the selective manipulation of the partner's wishes, emotions, assessments, beliefs, reasoning patterns, etc., the application of such a strategy necessarily presupposes from its author the ability to operate with the model of partner as a functional whole: the ability to predict 'what leads to what' in a partner's inner functioning, using it as the basis of explication of the 'model of the human mind' in general.

This explication should represent the intuitive, naive model of functioning of the human mind. It should reflect knowledge which human beings have and use about their partners in everyday communication. At the same time, this model should be more than a simple collection of beliefs. It should form a basis for drawing inferences from concrete data in concrete situations, for making predictions concerning the possible outcomes of the acts of influence exercised on the partner. In this sense, it may be called a 'theory' humans use in their everyday life, a 'naïve' theory. In theoretical psychology dealing with human thinking, there is a subfield which aims at explicating the nature of this 'naïve theory'. This is the Theory Theory approach we referred to in the Introduction. In the context of this paper, it is reasonable to differentiate *two lines of research* which, although ultimately aiming at the same goal, are doing it on different conceptual and formal levels and use different data and methods.

The first approach may be called cognitivistic because its origins lie in the interdisciplinary research field called cognitive science which, among other disciplines, originally included artificial intelligence and, specifically, the explication of human-language-understanding ability. At present, one of the hottest topics in this specific area is the role of reasoning in human social interaction, and the role of argumentation in this context in particular. For instance, why and how do humans reason at all before making decisions in social interaction.

For a good overview, we refer to the paper of Mercier and Sperber (2011). According to the format of such papers in the *Behavioral and Brain Sciences* journal, almost half of the paper consists of comments on the authors' text made by other researchers, and thus one can get a relatively adequate picture of the state of the art in the field. One of the main conclusions important for the topic of this paper can be summarised as follows: 'reasoning in social interaction necessarily involves, for a participating agent, reasoning about the predictable reasoning processes of other participating agents in the context of the topic under consideration'. In other words, a naive theory of reasoning.

In the broader context, it should be added that the discussion about such naive theories of reasoning and argumentation has transcended the borders of psychology and become a natural part of neighbouring disciplines that deal with human communication in the frames of the 'cognitive paradigm'. First of all, it involves (linguistic) semantics and pragmatics (e.g. Jackendoff, 2002; Ortony, 2003; Paradis, Hudson, & Magnusson, 2013).

But even more important is that other areas related to human social interaction have also been and are involved in the discussion, starting with those works of cultural anthropology from which the concept of ‘folk theory’ originated (D’Andrade, 1987) and ending with contemporary studies on the ‘collective wisdom’ (Landemore & Elster, 2012).

The second approach we have in view is the one directly related to creating DSs which interact with human users but, at the same time, being computer systems, have in their focus the regularities of human reasoning. That is, if a dialogue at hand does not constitute a simple question and answer exchange, the computer must understand what kind of reasoning processes started the additional questions or investigations in the partner, and be able to give them adequate answers or explanations if needed. The main point is how this should be realised in a computer program.

A good overview of the context in which we place ourselves can be found in the paper (Rahwan et al., 2004), as pointed out in the Introduction. In particular, we consider critically important the basic list of requirements to an argumentation-based negotiation (ABN) agent noted in the paper (p. 358): an ABN agent needs to be able to (1) evaluate incoming arguments and update its mental state accordingly; (2) generate candidate outgoing arguments and (3) select an argument from a set of available arguments. We hope that the description of our model demonstrates at least attempts to satisfy these requirements.

Below we give a short overview of some more concrete views concerning the modelling of human–computer interaction the description of which should make clearer the conceptual borderlines of the area into which we ourselves place our model described in the following sections.

Researchers of communication are studying formal dialogue games – interactions between agents, where each agent ‘moves’ by making utterances, according to a given set of rules. Dialogue games have been used to study human–computer interaction to explain sequences of human utterances and model complex human reasoning (Levin & Moore, 1978; McBurney & Parsons, 2009; Yuan, Moore, Reed, Ravenscroft & Maudet, 2011). Dialogue games are useful because they provide stereotypic patterns of dialogue. Loui (1998) characterises negotiation as an argumentative game. Hadjinikolis, Modgil, Black, Mcburney, and Luck (2012) provide an argumentation-based framework for persuasion dialogues, based on a logical conception of arguments that an agent may undertake in order to strategise over the choice of moves to make in a dialogue game, based on its model of its opponents.

Understanding and simulating behaviour of human agents, as presented in text, is an important problem to be solved in decision-making and decision support tasks (Morge & Mancarella, 2012). On the one hand, argument structures can be learnt from previous experience with these agents. Another approach is based on the assessment of quality and consistency of argumentation of agents (Chesñevar, Maguitman, & Loui, 2000).

Principles and criteria that characterise patterns of inference have to be formalised in order to create a logical model of common-sense reasoning. Dębowska, Łoziński, and Reed (2009) investigate the links between the philosophical and linguistic study of human reasoning and argumentation expressed in language, and on the other hand the formal, logical accounts of argument structures. Typically, argumentation takes place on the basis of incomplete information. Classical logic is inadequate since it behaves monotonically. Argumentation is a non-monotonic logic of reasoning with incomplete and uncertain knowledge (Bench-Capon & Dunne, 2007). Loui (1998) investigates the appropriateness of formal dialectics as a basis for non-monotonic reasoning and defeasible reasoning.

For the current state of argumentation theory, see, for example, Bench-Capon and Dunne (2007), Besnard and Hunter (2008) and Chesñevar et al. (2000).

Argumentation theorists Douglas Walton and Erik Krabbe propose a taxonomy of six types of dialogue: (1) persuasion, (2) negotiation, (3) inquiry, (4) deliberation, (5) information-seeking dialogue and (6) eristic dialogue. These types are characterised by their main goal, initial situation and participants' aims. Each type of dialogue has its own distinctive rules and goals, its permitted types of move, and its conventions for managing the commitments incurred by the participants as a result of the moves they make. Each type of dialogue they add exhibits a normative model, an enveloping structure that can aid us in evaluating the argumentative and other moves contained in it (Walton & Krabbe, 1995, pp. 8–9). Boella, Hulstijn, and van der Torre (2004) consider persuasion dialogues – the dialogues in which one agent is trying to influence the behaviour of another agent. There are three ways of influencing behaviour, each corresponding to the manipulation of one of the mental attitudes: by issuing a command, by convincing and by suggestion. These strategies have different requirements regarding the social setting of the dialogue. For example, for commands an authority relationship between agents is necessary. One agent models the other agent and tries to predict its response, given the other agent's model of itself (Boella et al., 2004).

Negotiation is a process where each party tries to gain an advantage for itself by the end of the process. Negotiation is intended to achieve compromise. Rahwan and Larson (2011) explore the relationships between mechanism design and formal logic, particularly in the design of logical inference procedures when knowledge is shared among multiple participants. Tohmé (2002) analyses the dynamics of beliefs and decision, in order to determine conditions on the agents that allow them to reach agreements. When one person initiates communication with another s/he mainly proceeds from the fact that the partner is a human being who feels, reasons and has wishes and plans, on the one hand, like every human being and, on the other, as this individual person. In order to be able to foresee what processes will be triggered in the partner after a DA (move), the agent must know the inner workings of the partner's psychological mechanisms. When aiming at a certain goal in communication, the agent must know how to direct the functioning of these mechanisms in order to bring about the intended result in the partner. Because of this, we have modelled the reasoning processes that people supposedly go through when working out a decision whether to perform an action or not. In a model of conversational agent, it is necessary to represent cognitive states as well as processes. One of the most well-known models of this type is the belief-desire-intention (BDI) model of human practical reasoning developed as a way of explaining future-directed intention (Allen, 1995; Boella & van der Torre, 2003; Bratman, 1987; Grosz & Sidner, 1986). This model is based on the Theory Theory.

3. Modelling 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 DAs and determine their verbal form as a turn tr_1. This turn triggers a reasoning process in B where two procedures are distinguished: the interpretation of A's turn and the generation of his/her response tr_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 realisation of his/her goal G^A has proceeded, and depending on this s/he may activate a new sub-goal of the initial goal G^A, and the cycle is repeated: A builds the next turn tr_3, etc. The dialogue comes to an end when A has reached his/her goal or abandoned it, or the participants have agreed to continue the interaction in the future.

3.1. *Model of conversational agent*

In our model, a conversational agent is a program that consists of six (interacting) modules (cf. Koit & Õim, 2000, 2004; for its further elaborations, see e.g. Koit, 2011):

$$(\mathrm{PL}, \mathrm{PS}, \mathrm{DM}, \mathrm{INT}, \mathrm{GEN}, \mathrm{LP}),$$

where PL is planner, PS is problem solver, DM is dialogue manager, INT is interpreter, GEN is generator and LP is linguistic processor. PL directs the work of both DM and PS, where DM controls communication process and PS solves domain-related tasks. The task of INT is to make semantic analysis of partner's utterances and that of GEN is to generate semantic representations of agent's own contributions. LP carries out linguistic analysis (starting from speech recognition) and generation (ending with speech synthesis). Conversational agent is using goal base GB and knowledge base KB in its work. Knowledge base consists of four components

$$\mathrm{KB} = (\mathrm{KB}_\mathrm{W}, \mathrm{KB}_\mathrm{L}, \mathrm{KB}_\mathrm{D}, \mathrm{KB}_\mathrm{S}),$$

where KB_W contains knowledge of the domain where the action under discussion belongs to, KB_L – linguistic knowledge, KB_D – knowledge about dialogue, and KB_S – knowledge about interacting subjects. KB_D contains definitions of DAs (declarative knowledge) and algorithms that are applied to reach communicative goals – communicative strategies and tactics (procedural knowledge). KB_S contains knowledge about evaluative dispositions of participants towards the action(s) (e.g. what do they consider as pleasant or unpleasant, useful or harmful), and, on the other hand, algorithms that are used to generate plans for acting on the world.

A necessary precondition of interaction is existence of common knowledge of agents:

$$\mathrm{KB}_\mathrm{L}^\mathrm{A} \cap \mathrm{KB}_\mathrm{L}^\mathrm{B} \neq \varnothing, \quad \mathrm{KB}_\mathrm{W}^\mathrm{A} \cap \mathrm{KB}_\mathrm{W}^\mathrm{B} \neq \varnothing, \quad \mathrm{KB}_\mathrm{D}^\mathrm{A} \cap \mathrm{KB}_\mathrm{D}^\mathrm{B} \neq \varnothing, \quad \mathrm{KB}_\mathrm{S}^\mathrm{AB} \cap \mathrm{KB}_\mathrm{S}^\mathrm{B} \neq \varnothing,$$
$$\mathrm{KB}_\mathrm{S}^\mathrm{BA} \cap \mathrm{KB}_\mathrm{S}^\mathrm{A} \neq \varnothing.$$

3.2. *Argumentation that involves reasoning*

After *A* has expressed his/her goal that the partner *B* would decide to do *D*, *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 but it has some specific traits that we consider important.

First – especially in the case of *B*'s reasoning – along with desires we also consider other kinds of motivational inputs for creating the intention of an action in an actor (e.g. whether 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). In other words, we are trying to analyse the concept of 'desire' used in the BDI model in fact as a cover concept for motivation to do something into some more detailed kinds, and this just so far as we are able to differentiate (and model formally) the effects that these factors will have on the following reasoning process of the agent.

Second, starting from this general idea, we have worked out a model of reasoning which leads to the emergence of the intention (goal) in an actor to do the action in question, or refuse to do it. Our reasoning model is based on the studies in the common-sense conception of how the human mind works in such situations (cf. D'Andrade, 1987). We suppose that in natural (everyday or even institutional) communication people start, as a rule, from this conception, not from any consciously chosen scientific one. We want to model a 'naïve' theory of reasoning, a 'theory' that people themselves use when they are interacting with other people and trying to predict and influence their decisions.

In our implementation, the reasoning model consists of two functionally linked parts: (1) a model of human motivational sphere and (2) reasoning algorithms.

3.2.1. *Model of reasoning subject*

In the motivational sphere, three basic factors that regulate reasoning of a subject concerning *D* can be differentiated. First, the subject may wish to do *D*, if pleasant aspects of *D* for him/her outweigh unpleasant ones; second, the subject may find it reasonable to do *D*, if *D* is needed to reach some higher goal, and useful aspects of *D* outweigh harmful ones; and third, the subject can be in a situation where s/he must (is obliged) to do *D* – where not doing *D* will lead to some kind of punishment. We call these factors *wish*-, *needed*- and *must*-factors, respectively. These factors should be taken as an approximation of the real constituents of the reasoning process. As we can see in the analysis of real-life dialogues (see Section 3), it is always not easy to differentiate, on the level of concrete DAs, between them (thus, for instance, *needed* as well as *must* can be said to include, or rely on, the *wish*-factor in some way, not to speak of the exact connections between these factors with emotions, such as joy or fear, for instance). Nevertheless we think that it is important to make these kinds of differentiation; quite another question is, how deep we will be able to go in using them in modelling the human reasoning, e.g. in the argumentation dialogue.

Thus, according to our present model, the model of motivational sphere of a subject can be represented by the following vector of 'weights':

$$\boldsymbol{w} = (w(\mathit{resources}), w(\mathit{pleasant}), w(\mathit{unpleasant}), w(\mathit{useful}), w(\mathit{harmful}), w(\mathit{is\text{-}obligatory}),$$
$$w(\mathit{is\text{-}prohibited}), w(\mathit{punishment\text{-}do}), w(\mathit{punishment\text{-}do\text{-}not})).$$

In the description, $w(\mathit{pleasant})$, $w(\mathit{unpleasant})$, $w(\mathit{useful})$, $w(\mathit{harmful})$ mean weight of pleasant, unpleasant, useful, harmful aspects of *D* (and/or its consequences), $w(\mathit{punishment\text{-}do})$ – weight of punishment for doing *D* if it is prohibited and $w(\mathit{punishment\text{-}do\text{-}not})$ – weight of punishment for not doing *D* if it is obligatory. Resources of the subject concerning *D* constitute any kinds of internal and external circumstances which create the possibility to perform *D* and which are under the control of the reasoning subject.

Let us briefly discuss the categories that will be used in the formulation of the reasoning algorithm:

(1) pleasant/unpleasant – these categories represent primary subjective-emotional evaluations of *D*, which are anchored in the sensual system of the subject
(2) useful/harmful – these are prototypical rational evaluations, i.e. they are based on certain beliefs or certain knowledge of the subject, and there are certain criteria for making the corresponding judgements. These criteria are connected, first of all, with the goals of the subject – an aspect of *D* is useful if it helps the subject to achieve some goal and, correspondingly, an aspect of *D* is harmful if it prevents the subject from achieving some goal
(3) obligatory/prohibited – these are also rational evaluations, but they are based on either the knowledge of certain social norms or some directive communicative act of a person who is in the position (has the power) to exercise his/her will upon the reasoning subject. Obligations and prohibitions are related to the concept of punishment, which is an action taken by some other subject because the reasoning subject has not followed the corresponding obligations or prohibitions
(4) resources of the subject concerning *D* constitute any kinds of internal (psychical and/or physical) and external circumstances that create the possibility to perform *D*.

However, we are aware that these motivational factors are not independent of each other. Thus, a useful outcome of an action is in some sense also pleasant for the subject, punishment is unpleasant (and can be harmful) for the punished person, etc., but we will not go into these details in our present model.

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 here that these aspects have numerical values (weights) and that in the process of reasoning (when weighing the pro- and counter-arguments) these values can be compared and summed up in a certain way. Here $w(resources) = 1$ if the subject has resources necessary to do D (otherwise 0); $w(is\text{-}obligatory) = 1$ if D is obligatory for the reasoning subject (otherwise 0); $w(is\text{-}prohibited) = 1$ if D is prohibited (otherwise 0). The values of other weights are non-negative natural numbers. Still, in reality people do not operate with numbers. On the other hand, the existence of certain scales in human everyday reasoning is also apparent. For instance, for the characterisation of pleasant and unpleasant aspects of some action there are specific words: enticing, delightful, enjoyable, attractive, acceptable, unattractive, displeasing, repulsive, etc. Each of these adjectives can be expressed quantitatively. In the simplest case, the numeric value of an aspect (e.g. pleasantness) can be taken equal to the number of different arguments supporting this aspect (e.g. if (1) a person likes to visit cinema and (2) s/he likes history then the weight of pleasantness of the action 'to visit cinema in order to see an historical film' is 2 for him/her).

If there are many actions $D_1, \ldots, D_n$ instead of one action D, then similar components must be added into the vector of weights for all these actions.

Knowledge base about interacting subjects, $\mathrm{KB_S}$, contains the vectors $\boldsymbol{w}^X$ (subjective evaluations of all subjects X of all possible actions) as well as vectors $\boldsymbol{w}^{XY}$ (X's beliefs concerning Y's evaluations). The vector $\boldsymbol{w}^{XY}$ is used by X as a partner (Y) model.

3.2.2. *Reasoning algorithms*

The second part of the reasoning model consists of reasoning algorithms that regulate (as we suppose) human action-oriented reasoning. A reasoning algorithm 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 D or not.

According to our model, the reasoning process directed at an action D of a subject can be triggered by three main types of factors: *wish*-, *needed*- or *must*-factors. The reasoning process consists of a sequence of steps where such aspects participate as resources of the reasoning subject for doing D, positive aspects of D or its consequences (pleasantness, usefulness, also punishment for not doing D if it is obligatory) and negative aspects (unpleasantness, harmfulness, punishment for doing D if it is prohibited).

How does the reasoning itself proceed? It depends on the factor which triggers it. In addition, a reasoning model, as a naive theory of mind, includes some principles which represent the interactions between factors and the causal connection between factors and the decision taken. For instance, the principles fix such concrete preferences as:

- people want pleasant states and do not want unpleasant ones
- people prefer more pleasant states to less pleasant ones
- the more pleasant the imagined future state, the more intensively a person strives for it.

In addition, there are also more concrete preference rules, e.g.:

- if D has been found pleasant (and the subject wishes to do it) then the subject checks the *needed*- and *must*-factors first from the point of view of their possible negative values ('what harmful consequences would D have?')

- if the sum of the values of the inner (*wish*- and *needed*-)factors and the value of the external (*must*-)factor appear equal in a situation (i.e. there arises a conflict) then the decision suggested by the inner factors is preferred.

We do not go into details concerning these principles here. Instead, we refer to (Davies & Stone, 1995).

There are three reasoning procedures in our model (*wish, needed, must*), which depend on the factor that triggers the reasoning. Each procedure represents steps that a subject goes through in the reasoning process (computing and comparing weights of different aspects of D), and the result is the decision to do D or not. As an example, let us present a reasoning procedure *wish* which is triggered by *wish*-factor, that is, the subject believes that it would be pleasant to do D, i.e. $w(pleasant) > w(unpleasant)$ (Jackson structured programming (JSP) diagram in Figure 1).

In the case of other input factors (*needed, must*), the general structure of the algorithm is analogous, but there are differences in concrete steps.

The reasoning model is connected with the general model of conversational agent in the following way. First, the planner PL makes use of reasoning schemes and second, the $\mathrm{KB_S}$ contains the vector $\boldsymbol{w}^A$ (A's subjective evaluations of possible actions) as well as vectors w^{AY} (A's beliefs concerning Y's evaluations, where Y denotes agent(s) A may communicate with). The vectors $\boldsymbol{w}^{AY}$ are used as partner(s) model(s).

When comparing our model with the BDI model, beliefs are represented by knowledge of the conversational agent with reliability less than 1; desires are generated by the vector of weights $\boldsymbol{w}^A$; and intentions correspond to goals in GB. In addition to desires, from the weights vector we can also derive some parameters of the motivational sphere that are not explicitly conveyed by the basic BDI model: needs, obligations and prohibitions.

Some wishes or needs can be stronger than others: if $w(pleasant_{Di}) - w(unpleasants_{Dj}) > w(pleasant_{Dj}) - w(unpleasant_{Dj})$, then the subject's wish to do an action D_i is stronger than that to do another action D_j. In the same way, some obligations (prohibitions) can be stronger than others, depending on the weight of the corresponding punishment.

3.3. *Communicative strategies and tactics*

A communicative strategy is an algorithm, which is used by a communication participant to achieve his/her communicative goal. About the general concept of the communicative strategy, see, for example, Heritage (1991).

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 is the concept of communicative space. For a general description of the approach, see, for example, Brown and Levinson (1987). Communicative space is defined by a number of coordinates that characterise the relationships of participants in a communicative encounter. Communication can be collaborative or confrontational (that is represented by one axis of the communicative space), personal or impersonal (another axis); it can be characterised by the social distance between participants (the third axis). The next axes represent the modality of communication (friendly, ironic, hostile, etc.), the intensity (peaceful, vehement, etc.), etc. Together, these dimensions bring the social aspect of communication into the model (cf. Boella et al., 2004). Just as in the case of motivations of human behaviour, people have an intuitive, 'naïve theory' of these coordinates, too. The values of the coordinates can be expressed by specific words (such as friendly, ironic, and so on) as in the case of pleasant, useful, etc. aspects of an action (Section 2.2.1). Instead, we use numerical values as approximations in our model.

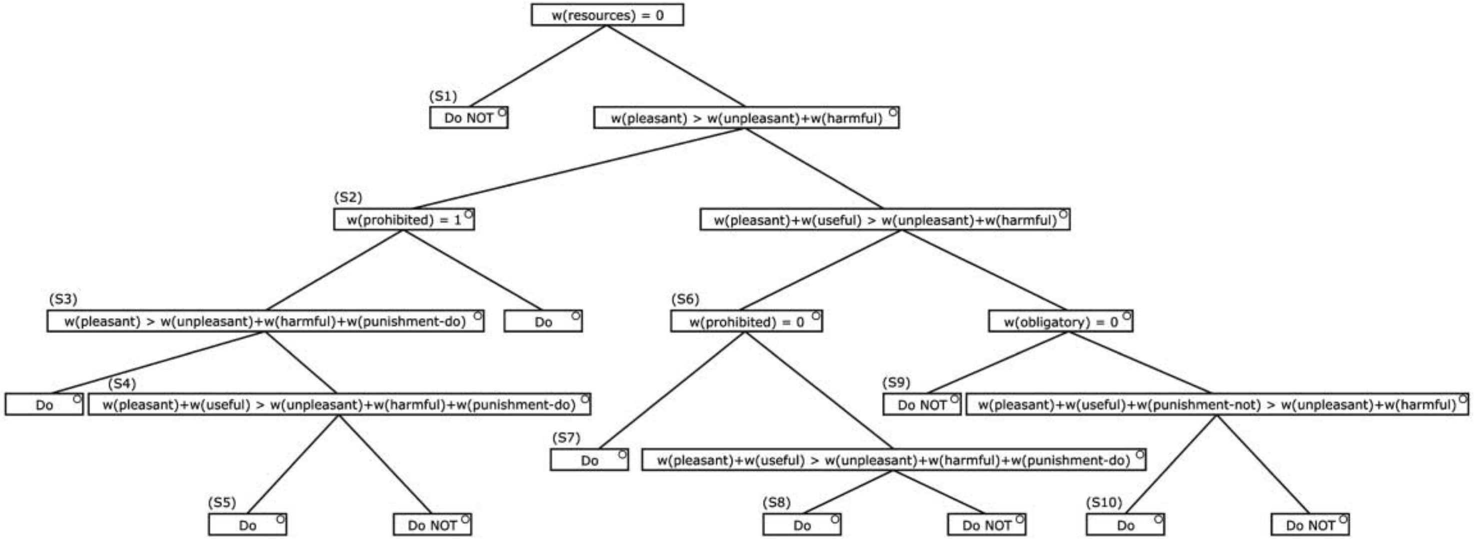

w(resources) = 0
(S1)
Do NOT
w(pleasant) > w(unpleasant)+w(harmful)
(S2)
w(prohibited) = 1
w(pleasant)+w(useful) > w(unpleasant)+w(harmful)
(S3)
w(pleasant) > w(unpleasant)+w(harmful)+w(punishment-do)
Do
(S6)
w(prohibited) = 0
w(obligatory) = 0
Do
(S4)
w(pleasant)+w(useful) > w(unpleasant)+w(harmful)+w(punishment-do)
(S9)
Do NOT
w(pleasant)+w(useful)+w(punishment-not) > w(unpleasant)+w(harmful)
(S7)
Do
w(pleasant)+w(useful) > w(unpleasant)+w(harmful)+w(punishment-do)
(S5)
Do
Do NOT
(S8)
Do
Do NOT
(S10)
Do
Do NOT

Figure 1. Reasoning procedure *wish*. Precondition: $w(pleasant) > w(unpleasant)$.

A communicative strategy can be implemented in various ways, e.g. the participant *A* can use different arguments, in order to influence his/her partner *B* to do *D*: stress pleasantness of *D* (i.e. *entice B*), stress its usefulness (*persuade B*, i.e. *suggest B* to do *D*), or stress punishment for not doing *D* if it is obligatory (*threaten B*). We call these concrete ways of realisation of a communicative strategy communicative tactics. In our model, the choice of communicative tactics depends on the 'point' in the communicative space where the participants believe to find themselves.

A communicative strategy for *A* can be presented as the following algorithm (Figure 2).

The general strategy used by *A* may be called reasoning-influencing strategy since it is based on the model of reasoning we described in the previous section. The aim of the strategy is to influence the reasoning process of the partner in such a way that its output would match the goal of the *A*.

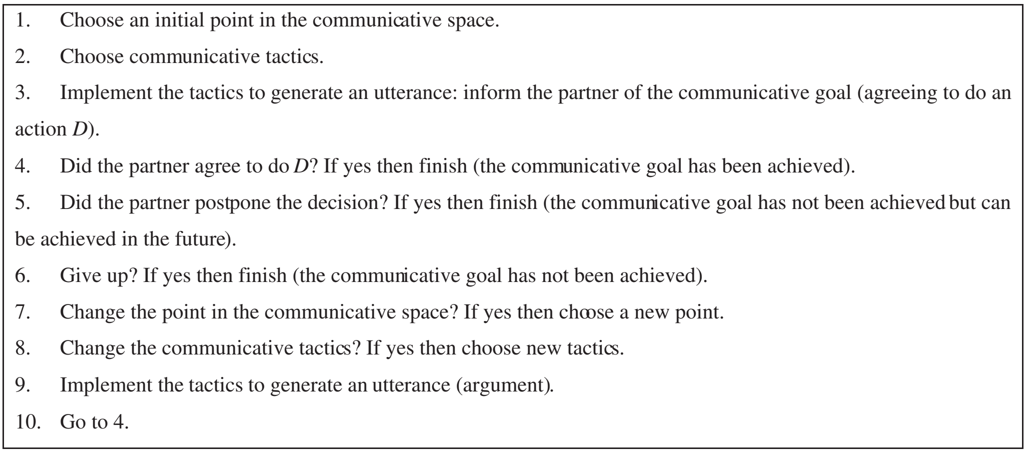
The strategy may be realised in different forms depending on the steps in the reasoning process of the partner through which *A* is trying to influence this process. The dialogue proceeds in the following way.

- *A* informs the partner *B* of the communicative goal, formulating it, e.g. as a request or proposal.
- *B* goes through the reasoning process – whether to do *D* or not – and makes a decision, of which s/he informs *A*.

At this point there appears a need to model *B*'s reasoning process and incorporate this model into the dialogue model. When the reasoning process has started, the subject considers various positive and negative aspects of *D*. If the positive aspects weigh more, the subject will make the decision to do *D*, otherwise s/he will make the decision not to do *D*.

- If *B*'s decision is negative, *A* will try, during the following process of interaction, to influence *B*'s reasoning in such a way that ultimately s/he would make a positive decision.

Three communicative tactics exist for *A* in our model, which are related to the three reasoning procedures (*wish, needed, must*). The reasoning procedure *wish* may be triggered in the partner by tactics of *enticing*, the procedure *needed* by tactics of *persuading* (*suggesting*), and the procedure



1. Choose an initial point in the communicative space.
2. Choose communicative tactics.
3. Implement the tactics 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. Did the partner postpone the decision? If yes then finish (the communicative goal has not been achieved but can be achieved in the future).
6. Give up? If yes then finish (the communicative goal has not been achieved).
7. Change the point in the communicative space? If yes then choose a new point.
8. Change the communicative tactics? If yes then choose new tactics.
9. Implement the tactics to generate an utterance (argument).
10. Go to 4.

Figure 2. Communicative strategy.

1. Determine the aspect of D brought out by the partner and choose an argument for attacking it (if there are arguments).
2. If the aspect brought out by the partner was not the title aspect and if there are no arguments for attacking it, or the partner did not bring out any aspect, then choose an argument for stressing the title aspect
3. If there are no arguments for stressing the title aspect then fail.

Figure 3. Communicative tactics.

needed by tactics of *threatening*. These tactics are similar to the persuasion strategies considered in (Boella et al., 2004): convincing, suggestion and a command, respectively.

Participant A when implementing a communicative strategy uses a partner model – a vector $\boldsymbol{w}^{AB}$ – 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 $\boldsymbol{w}^{AB}$ with the actual vector $\boldsymbol{w}^{B}$ of the motivational sphere of the partner B. We suppose here that A has (can generate) several sets of statements for changing the weights of different aspects of D for the partner $B : \{st_i^A - \mathrm{asp}_j, i = 1, \dots, n^A_{\mathrm{asp}_j}; j = 1, \dots, n\}$ where asp_j is the j-th aspect of D and n is the number of different aspects. All the statements have their (numerical) weights as well.

If A uses a statement $st_i^A - \mathrm{asp}_j$, then s/he can increase/decrease the weight of that aspect in the following way:

$$w(\mathrm{asp}_j) := w(\mathrm{asp}_j) + w(st_i^A - \mathrm{asp}_j)\text{if } \mathrm{asp}_j \text{ is positive and}$$

$$w(\mathrm{asp}_j) := w(\mathrm{asp}_j) - w(st_i^A - \mathrm{asp}_j)\text{if } \mathrm{asp}_j \text{ is negative,}$$

where $w(st_i^A - \mathrm{asp}_j)$ is the weight of the statement $st_i^A - \mathrm{asp}_j$.

Every tactic has its own 'title aspect' – the most important aspect of the action D for that tactic. The title aspects are: pleasantness of D for enticing, usefulness for persuading and punishment of not doing (an obligatory action) D for threatening. While enticing, A first of all tries to increase the pleasantness of D for the partner B; while persuading, A tries to increase the usefulness of D, and while threatening A stresses punishment for not doing D. Using the notion of the title aspect, all the tactics can be represented by the following algorithm (Figure 3).

Therefore, we consider every tactic used by a participant as an algorithm for choosing (generating) his/her next utterance (argument) as a reaction to the partner's refusal to perform the action. A participant can follow the same tactics consistently until there are arguments (see Figure 2).

A detailed description of the tactic of *persuasion* can be found in (Koit, Roosmaa, & Õim, 2009). The tactics of *enticing* and *threatening* can be represented in an analogous way.

We will go back to some details of the model in Section 4.

4. Corpus analysis

How do people argue for or against an action? When aiming at a certain goal in communication, the agent must know how to direct the functioning of the partner's psychological mechanisms in one or another direction in order to bring about the intended result in the partner. This knowledge forms a necessary part of human communicative competence. A model of human being as a partner in communication should explicate, in particular, people's intuitive knowledge about the inner mental structures of human beings, a naive, or folk theory. Because of that, the aim of the corpus analysis is to justify our model on Estonian spoken human–human dialogues taken from the dialogue corpus EDiC (Hennoste et al., 2008). This is not verification in the mathematical sense. However, arguments and counter-arguments expressed in a natural language appear in the actual negotiation dialogues and demonstrate, even though indirectly, how the participants are reasoning before making a decision. DAs are annotated in our corpus by using a special annotation typology (Hennoste et al., 2008) which is based on conversation analysis (Hutchby & Wooffitt, 1998). Still, the major part of the typology coincides with well-known typologies (e.g. DAMSL). Three sub-corpora were chosen for analysis, two of them consisting of institutional and one of everyday dialogues: (1) calls of clients to travel agencies, (2) calls of sales clerks of an educational company where training courses are offered for other companies and (3) face-to-face conversations between acquaintances. However, it was hard to find suitable empirical material for our study because the corpus contains mostly information-seeking dialogues. We did not record, transliterate nor annotate new dialogues especially for the current analysis but used the existing corpus.

4.1. *Travel agency*

First, 36 dialogues were taken from the EDiC where clients call travel agencies supposedly with a goal to travel somewhere. A travel agent is interested in booking a trip by the client, therefore, we may expect that the agents try to influence clients in such a way that they decide to book the trip immediately, during the current conversation. Among the considered communicative tactics (Section 2.3), threatening can be excluded because a travel agent as an official person is obligated to communicate cooperatively, impersonally, friendly and peacefully (i.e. to stay in a neutral, fixed point of the communicative space). An agent can only persuade or entice a client.

Argumentation has only been found in four dialogues. It turned out that the clients mainly want to get information, not to book a trip. Moreover, no persuasion by a travel agent finishes with the client agreeing to book a trip.

In Dialogue 1 (in the Appendix), participant *A* (a travel agent) presents several arguments (giving several pieces of information) trying to persuade the client *B* to take the trip but the client does not make a decision. In Dialogue 2 (in the Appendix), *A* is similarly proposing various arguments for the usefulness of booking the trip.

In the analysed travel dialogues, a typical sequence of DAs that form argumentation is as follows (the subsequence marked by → can be repeated):

$$\begin{array}{rl} A: & \text{Proposal} \\ \rightarrow A: & \text{Giving Information} \\ \rightarrow B: & \text{Continuer} \\ B: & \text{Deferral} \end{array}$$

It should be mentioned that the used DA typology does not include a special act Argument. In the travel dialogues, arguments of travel agents are represented by giving information (the DA Giving Information). A client does not propose any counter-arguments but allows a travel agent to give all his arguments only signalling that she is waiting for further information (the DA Continuer, e.g. ‘I see’, ‘uhuh’, etc.). After that, she informs the agent about her decision (typically, deferral like utterance 16 in Dialogue 1 in the Appendix).

The subsequence marked by → can be considered as an information-sharing sub-dialogue (Chu-Carroll & Charberry, 1998), which is initiated by *A* to give further information to *B* in such a way that *B* after reasoning could make an informed decision about the proposal.

4.2. *Educational company*

Second, a closed part of the EDiC has been analysed, consisting of 44 calls where sales clerks in an educational company offer different courses of their company (language, secretary training, bookkeeping, etc.) to managers of other companies. The dialogues have been put into a secret list, according to an agreement with the educational company (only fragments of dialogues may be published). Fourteen dialogues out of 44 have been excluded from this study (the needed person is not present, the number the clerk is calling is wrong, the recording breaks off). Like a travel agent, a sales clerk can only persuade or entice a client to take a course; threatening is excluded. Various arguments are used by the clerks to indicate the usefulness or pleasantness of the courses. However, the decisions of clients will be postponed in most cases. It is not surprising because the recordings belong to the initial period of negotiations.

The analysed 30 dialogues can be divided into two groups: 1) *A* and *B* are communicating for the first time (six dialogues) or 2) they have been in contact previously (24 dialogues).

4.2.1. *First call*

A typical dialogue starts with *A*'s introduction and a question whether *B* does know the educational company. Then a short overview of the company is given (e.g. we are an international company, we are dealing with sale, marketing). All the statements can be considered also as arguments for taking a training course. Then a proposal is made by *A* to *B* to take some courses. *A* points to the activities of *B*'s organisation, which demonstrates that he has previous knowledge about her institution (e.g. your firm is dealing with retail and wholesale therefore you could be interested in our courses). If *B* does not make a decision, then *A* asks *B* to tell more about her institution in order to get more information (arguments) for the necessity of the courses for *B* and offers them again[1]:

***A*:** `ja no Ti- Tiritamm pakub just nüd ka sellist sellist`
`koolitust et kuidas kuidas neid (0.5) mm kliente nüd`

1. **`and Tiritamm offers just such a training how [to find]`**
`customers`
`(1.8)`
`leida eks=ole, oma turgu`

2. **`to find, yes, [to increase] your own market`**
`(1.5)`
`e suurendada. ja (0.8) ja (0.5) ja samas ka see et=et`
`kuidas neid püsikliente 'hoida (1.0) kas e (...) suhtlemist`
`et. kuidas teiele tundub kas ned teemad võiksid teile huvi`
`pakkuda?`

3. **`to increase, and how to keep regular customers. how do you`**
`think - are you interested in these topics?`

All the dialogues end with an agreement to keep in touch (*A* promises to send information materials to *B*, to call *B* later, etc.), and *B* does not accept or reject taking a course but postpones her decision. That can be considered as a good result for *A* – his arguments have been reasonable. *B* needs more time before making the final decision.

4.2.2. *Later calls*

B agrees to take a course only in one conversation, she agrees with reservations in two dialogues and does not agree in one dialogue. In the remaining dialogues, *A* and *B* come to the agreement to keep in touch as in the case of the first communication, i.e. *B* postpones the decision.

A always starts the conversation with referring to a previous communication (*we communicated in November, I sent catalogues to you*):

> ***A*:** `'kevadel rääkisime natuke 'pikemalt sin (.) 'viimati. (.) et e (.) kudas teil 'läheb ka? (.)`
> **`we talked in the spring for quite a long time last time, how do you do?`**

It is significant that the introductory part is quite long in the dialogues. *A* behaves very politely, friendly and sometimes familiarly (this holds especially for male clerks):

> ***A*:** `mt (.) kuidas on elu 'vahepeal läinud, kõik kenad 'reisid on 'seljataha jäänud.`
> **`how did you do meanwhile, all the nice trips are finished?`**

In this way, *A* prepares a suitable background for his proposal and herewith makes a refusal more difficult for *B*.

In the main part of a dialogue, *A* presents various arguments for the usability of the courses for *B*'s institution and meanwhile collects new information about *B* by asking questions in order to learn more and get new arguments for doing *D*:

> ***A*:** `ee küsiks nüd 'seda et=et ta on (.) noh põhimõtselt möeldud ütleme mt (.) e 'juhtidele ja 'spetsialistidele ütleme kes 'vastutavad 'rahvusvaheliste kontaktide 'arendamise eest.`
> **1. `I'd like to ask that, is it designed for managers in general and for the specialists who are responsible for development of international contacts`**
> ***B*:** `mhmh.`
> **2. `uhuh`**
> ***A*:** `a kas teil on 'rahvusvahelisi 'suhteid,`
> **3. `but do you have international relations?`**
> ***B*:** `mm=`
> **4. `uhuh`**

A typical sequence of DAs that form argumentation includes more DAs as compared with travel dialogues (DAs in brackets '['...']' can be missed; subsequences marked with → can be repeated):

A : Giving Information
Proposal
→ *A* : Giving Information
→ *B* : Continuer
[→ *A* : Question]
[→ *B* : Giving Information]
B : Accept/Reject/Deferral

Proposal will be made by *A* after an introductory part – giving information. In the main part, *A*'s arguments are presented by giving information, *B* responds using continuer as in the travel dialogues (*uhuh*, *I see*, etc.). In addition, *A* can ask questions in the main part of a dialogue.

We can conclude that in institutional dialogues, *A* (an official) presents his arguments by giving information but the partner *B* (a customer) usually does not give counter-arguments which explicitly show how does her reasoning process proceed. The outcome of *A*'s influencing process (*B*'s decision) becomes clear only at the end of interaction.

4.3. *Face-to-face conversations*

Finally, nine face-to-face dialogues were analysed where the participants know each other well. There are parts in the conversations where performing a certain action is under consideration and arguments for and against are presented by the participants. Here we can hope to find all the communicative tactics considered in our model. Actually, enticing and persuading were used but not threatening.

In Dialogue 3 (in the Appendix), *A* tries to achieve *B*'s agreement to publish his personal data at the website of an institute but *B* does not want to publicise his relations with the institute.

The utterances of the participants form adjacency pairs of opinions/statements and agreements/disagreements as known in conversation analysis (Hutchby & Wooffitt, 1998), which can be interpreted as arguments and counter-arguments. *A* implements the tactics of persuasion, stressing the usefulness of the action (publishing *B*'s data). *B*, trying to postpone the decision, points out less usefulness or more harmfulness in most cases. *B* does not make a decision during the conversation; therefore, *A*does not achieve his communicative goal.

The general structure of the analysed face-to-face dialogues can be represented by the following DAs (a subsequence marked with → where the exchange of arguments/counter-arguments takes place can be repeated several times):

$$\begin{array}{rl} A: & \text{Proposal/Request} \\ \to B: & \text{Reject} + \text{Assertion} \\ \to A: & \text{Reject} + \text{Assertion} \\ B: & \text{Accept/Reject/Deferral} \end{array}$$

We can summarise that in the analysed everyday dialogues, partner influencing is different as compared with institutional dialogues – here *B*'s reasoning process is explicated by giving counter-arguments (assertions). In such a case, *A* has more information for choosing a new argument (new assertion) on the next step of negotiation in order to direct *B*'s reasoning to a (desirable) positive decision.

The corpus analysis demonstrates that the actual human–human spoken conversations can be more complicated than the dialogues which can carry out the conversational agent modelled in Section 2. For that reason, we have chosen a limited task for implementation of the model in an experimental DS – 'a communication trainer', which can train the user to influence his/her partner (computer) consistently using certain communicative tactics. The DS can establish certain restrictions on argument types, on the order in the usage of arguments and counter-arguments, etc. (cf. Bringsjord et al., 2008; Yuan, Moore & Grierson, 2008). The DS can alternately play two roles: (1) of the participant *A* who is influencing the reasoning of user *B* in order to achieve *B*'s decision to do an action *D* or (2) the role of the participant *B* who is rejecting arguments for doing the action *D* proposed by user *A*. In the first case, the DS does not deviate from the selected tactics but follows them in a systematic way. In the second case, the DS does not deviate from the selected reasoning procedure.

Early attempts to produce a general purpose automated reasoning system were widely seen as a dead end. Special purpose reasoners for smaller, tractable problems, such as theorem proving, have superseded them (Morris, Tarassenko, & Kenward, 2005). Scheuer, Loll, Pinkwart, and McLaren (2009) review the way in which argumentation has been supported and taught to students using computer-based systems. Studies have indicated that argumentation systems can be beneficial for students, and there are research results that can guide system designers and teachers as they implement and use argumentation systems. They conclude that both on the technology and educational psychology sides a number of research challenges remain to be addressed in order to make real progress in understanding how to design, implement and use educational argumentation software.

5. Implementation: a conversational agent

Four kinds of dialogue management architectures are most common (Ginzburg & Fernandez, 2010). The earliest and also one of the most sophisticated models of conversational agent behaviour is based on the use of planning techniques (Allen, 1995). The two simplest and most commercially developed architectures are finite-state and frame-based (Jurafsky & Martin, 2008). The most powerful are information-state dialogue managers (Traum & Larsson, 2003). The information state may include aspects of dialogue state and also beliefs, desires, intentions, etc. of dialogue participants.

In our application, we have chosen the information-state approach. An experimental DS is implemented (in programming language Java), which in (text-based) interaction with a user in Estonian can optionally play the role of *A* or *B* (Koit, 2012). Our aim is, first of all, to demonstrate how reasoning can be influenced and modelled in dialogue, therefore, not all the modules listed in Section 3.1 are implemented in our DS (for example, the DS uses only predefined set of sentences and does not make language generation).

In the first version of the DS, both the participants only operate with semantic representations of linguistic input/output, the surface linguistic part of interaction is provided in the form of a list of ready-made sentences in Estonian, which are used both by the computer and the user. The user can choose his/her sentences from a menu. The sentences are only classified semantically according to their possible functions and contributions in a dialogue (e.g. the sentences used by *A* to increase the usefulness of the action, the sentences used by *B* to indicate harmfulness of the action, etc.). Still, the files of Estonian sentences can easily be substituted with their translations and interaction can take place in another language. The weight of each sentence (argument) is equal to 1.

In the second version, there are ready-made sentences only for the computer. The user can put in free text which will be interpreted by the computer using keywords or phrases in order to classify the user texts semantically. The sentences chosen by the computer as well as texts put in by the user can have different weights. Speech recognition and speech synthesis are not included.

5.1. *Representation of information states*

Let us consider the situation where the computer (conversational agent) plays *A*'s role. The most important component of an information state of the conversational agent is the partner model, which is changing during the interaction.

There are two parts of an information state of a conversational agent – private (information accessible only for the agent) and shared (accessible for both participants). The private part consists of the following components (Koit, 2011):

- current partner model (vector w^{AB} of weights – A's picture of B). In the beginning of interaction, A can create the vector randomly or take into account pre-knowledge about B, if it is available
- tactic t_i^A chosen by A for influencing B (can be fixed randomly in the beginning of interaction)
- reasoning procedure r_j which A is trying to trigger in B and bring to a positive decision (it is determined by the chosen tactics, e.g. when enticing, A triggers the reasoning procedure WISH in B)
- stack of aspects of D under consideration. In the beginning, A puts the 'title' aspect of the chosen tactics into the stack (e.g. pleasantness when A is enticing B)
- set of DAs DA $= \{d_1^A, d_2^A, \ldots, d_n^A\}$. There are following DAs for A: proposal, statements for increasing or decreasing weights of different aspects of D for B, etc.
- (finite) set of utterances for increasing or decreasing the weights ('arguments for/against') $U = \{u_{i1}^A, u_{i2}^A, \ldots, u_{iki}^A\}$. Every utterance has its own weight – numerical value (some arguments weigh more than others): $V = \{v_{i1}^A, v_{i2}^A, \ldots, v_{iki}^A\}$ where v_{ij}^A is the value of u_{ij}^A, respectively ($j = 1, \ldots, k_i$).. Every utterance can be chosen by A only once. Therefore, A has to abandon its initial goal if there are no more arguments remained.

The shared part of an information state contains

- set of reasoning models $R = \{r_1, \ldots, r_k\}$
- set of tactics $T = \{t_1, t_2, \ldots, t_p\}$
- dialogue history – the utterances together with participants' signs and DAs $p_1 : u_1[d_1], p_2 : u_2[d_2], \ldots, p_i : u_i[d_i]$, where $p_1 = A; p_2$, etc. – A or B.

5.2. *Update rules*

There are different categories of update rules which will be used for moving from the current information state into the next one:

I. rules used by A in order to interpret B's turns and
II. rules used by A in order to generate its own turns
    (1) for the case if the title aspect of the used tactics is located on top of the goal stack (e.g. if the tactic is enticement then the title aspect is pleasantness)
    (2) for the case if another aspect is located over the title aspect of the tactics used (e.g. if A is trying to increase the usefulness of D for B but B argues for unpleasantness, then unpleasantness lies over the usefulness)
    (3) for the case if there are no more utterances for continuation of the current tactics (and new tactics should be chosen if possible)
    (4) for the case if A has to abandon its goal
    (5) for the case if B has made the positive decision, and therefore A has reached the goal.

Let us consider two examples of update rules. The first one (1) belongs to the category I and the second one (2) to the II-2.

(1) `IF the used tactic is enticement (i.e. its title aspect` *`pleasantness`* `lies in the stack) AND B's last utterance was about` *`harmfulness`* `of doing D THEN` **`put`** *`harmfulness`* `into the stack (over` *`pleasantness`*`).`

(2) `IF the used tactic is enticement (i.e. its title aspect` *`pleasantness`* `lies at the bottom of the stack) AND` *`harmfulness`* `lies on the top of the stack (i.e. B said that`

```
doing D is too harm) AND there are utterances for
decreasing the harmfulness by x units AND reasoning
triggered by the wish-factor on the changed partner model
gives a decision 'do D' THEN choose this utterance (and the
corresponding dialogue act) AND eject harmfulness from
the goal stack.
```

There are special rules for updating the initial information state.

5.3. *Examples of interactions with the computer*

When playing A's role, the computer chooses tactics (of enticing, persuading or threatening) and generates (randomly) a model of the partner, according to which the corresponding reasoning procedure (*wish, needed* or *must*) yields a positive decision, i.e. the computer 'optimistically' presupposes that the user can be influenced in this way. A dialogue begins by an expression of the communicative goal (this is the first utterance u_1 of the computer). If the user refuses (after his/her reasoning by implementing a normal human reasoning, which we are trying to model here), the computer determines (on the basis of the user's utterance u_2) the aspect of D the weight of which does not match the reality and changes this weight in the user model so that a new model will give a negative result as before but it is an extreme case: if we increased this weight by one unit (in the case of positive aspects of D) or decreased it by one unit (in the case of negative ones) we should get a positive decision. (For simplicity, we suppose here that each argument will change the corresponding weight in the user model exactly by one unit.) On the basis of a valid reasoning procedure (and tactics) the computer chooses a (counter-)argument u_3 from the set of sentences for increasing/decreasing this weight in the partner model. A reasoning procedure based on the new model will yield a positive decision. Therefore, the computer supposes, that its next utterance will bring the user to agreement to do the action. Now the user must enter (or choose from a menu) his/her next utterance (e.g. a new argument against the action), and the process continues in a similar way (see Example 1). Each argument or counter-argument can be chosen only once. A similar approach was used in Loui's Skeletal Model (Loui, 1998).

Example 1 Let us suppose that the computer has chosen the tactic of enticing and has generated the following user model (cf. Section 3.2.1):

$$\begin{aligned}
&w^{AB} = (w(are - resources) = 1, w(pleasant) = 9, w(unpleasant) = 7, w(useful) = 5,\\
&w(harmful) = 0, w(obligatory) = 1, w(prohibited) = 0, w(punishment - do) = 0,\\
&w(punishment - do - not) = 1) = (1, 9, 7, 5, 0, 1, 0, 0, 1).
\end{aligned}$$

The reasoning procedure *wish* yields a positive decision on this model because $w(\textit{are-resources}) = 1$, $w(\textit{pleasant}) > w(\textit{unpleasant}) + w(\textit{harmful})$, $w(\textit{pleasant}) + w(\textit{useful}) > w(\textit{unpleasant}) + w(\textit{harmful})$, $w(\textit{prohibited}) = 0$ (cf. Figure 1 in Section 3.2.2). Let us suppose that the user chose a refusing sentence from a menu, indicating that $w(\textit{harm})$ should be corrected. There are three possible negative outcomes when applying the procedure *wish* (Figure 1). In our case, $w(\textit{obligatory}) = 1$, therefore we get $w(\textit{harmful}) \geq w(\textit{pleasant}) - w(\textit{unpleasant}) + w(\textit{useful}) + w(\textit{punishment-do-not}) = 8$. Thus, in the corrected model $w(\textit{harmful}) = 8$. The procedure *wish* will yield a negative decision as before but if we decreased (choosing the next utterance) the value of $w(\textit{harmful})$ by 1 we should reach a positive decision soon.

If there is more than one possible non-empty domain of allowed values for correcting a weight we shall choose the domain with the greatest lower barrier (for negative aspects of *D*) or with the least upper barrier (for positive aspects), i.e. the worst case.

Example 2 is an excerpt of an enticing dialogue where the goal of the conversational agent is to reach the partner's decision to travel to Venice (*A* – computer, *B* – user, ready-made sentences were used by both the computer and the user; the user makes his/her options from a menu).

Let us suppose that the computer starts with the user model (*1, 9, 7, 5, 0, 1, 0, 0, 1*) considered in Example 1 and tries to entice the user. The reasoning procedure *wish* gives a positive decision on this model.

Example 2 (originally in Estonian)

`A: Would you like to travel to Venice? Our firm needs to conclude a contract there.`

`B: Why me?` //Refusal without any argument. The pleasantness is not so high as supposed by the computer, therefore the user model has to be corrected. The new model will be (*1, 8, 7, 5, 0, 1, 0, 1*)//

`A: You look very smart, this is important for making contracts.` //The computer supposes that its next argument will lead to the model (*1, 9, 7, 5, 0, 1, 0, 1*) where the procedure *wish* gives a positive decision//

`B: Why do I suit better than Mary?` //Refusal. The pleasantness has to be corrected once more//

`A: You have a talent for making such contracts.`

/- - -/

`B: When I am abroad my husband will be unfaithful.` //Refusal: the harmfulness is greater than that supposed by the computer//

`A: Sorry, I could not convince you.` //The computer has no more arguments for decreasing the harmfulness nor for increasing the pleasantness. It does not change its tactics (e.g. does not go over to persuading or threatening) and gives up)//

Example 3 demonstrates in more detail how the partner model is used in an interaction (cf. Koit, 2012). The communicative goal of the conversational agent is to reach the partner's decision to do an action *D*, which is 'to become a vegetarian' (*A* – computer, *B* – user, ready-made sentences are used by the computer, the user puts in free text).

Example 3 *A* will implement the tactic of persuasion and generates a partner model: $w^{AB} = (w(\mathit{resources}) = 1$, $w(\mathit{pleasant}) = 5$, $w(\mathit{unpleasant}) = 4$, $w(\mathit{useful}) = 6$, $w(\mathit{harmful}) = 2$, $w(\mathit{obligatory}) = 0$, $w(\mathit{prohibited}) = 0$, $w(\mathit{punishment\text{-}do}) = 0$, $w(\mathit{punishment\text{-}not}) = 0)$.

The reasoning procedure *needed* yields a positive decision in this model (to do *D*).

The initial information state of *A* is as follows.

Private part

- initial partner model $\mathbf{w}^{AB} = (1, 5, 4, 6, 2, 0, 0, 0, 0)$
- the tactic chosen by *A* – persuasion
- *A* tries to trigger the reasoning procedure *needed* in *B*, the pre-conditions are fulfilled: $w(\mathit{resources}) = 1$; $w(\mathit{useful}) > w(\mathit{harmful})$
- stack contains usefulness – the title aspect of persuasion tactic
  - set of DAs at *A*'s disposal: {proposal, statements for increasing or decreasing weights of different aspects of *D* for *B*, etc.}

- set of utterances for expressing the DAs at *A*'s disposal, together with their values: {*Doctors argue for vegetarian food* – value 1, *Meat contains much cholesterol* – value 1, etc.}.

The shared part of the initial information state contains

- the reasoning procedures *wish, needed* and *must*
- the communicative tactics of enticement, persuasion and threatening
- dialogue history – an empty set.

The dialogue (originally in Estonian) is generated jointly by the computer (*A*) and the user (*B*). *A*'s first DA is proposal:

A: *Would you like to become a vegetarian?*

B: *I like animal food.* //Refusal: the user pointed out the unpleasantness of the action, therefore the computer must increase the weight of unpleasantness in the user model in such a way that the reasoning procedure *needed* will give the negative decision as before but the computer's next argument will increase the weight of usefulness by one unit and the reasoning procedure *needed* will give the positive decision again. The new model will be (*1,5,**9**, 6, 2, 0, 0, 0, 0*)//

A: *Doctors argue for vegetarian food.* //The user model will be: (*1,5,9,**7**,2, 0, 0, 0, 0*)//

B: *Meat contains many useful ingredients.* //Refusal: the usefulness is lower than expected by the computer, the new partner model will be (*1, 5, 9,**6**,2, 0, 0, 0, 0*)//

A: *Vegetarian food contains less harmful transcendental fats.* //A new argument for increasing the usefulness will change the model again: (*1, 5, 9,**7**,2, 0, 0, 0, 0*)//

etc.

5.4. *Experiments and evaluation*

Six voluntary users have carried out a number of experiments with the DS. Let us consider here the case when the computer plays *A*'s role. When starting a dialogue the computer randomly generates a user model. At the beginning, we have set only one restriction: we required that the initial model should satisfy the presumption(s) underlying the corresponding reasoning procedure. Thus, for enticing $w(pleasant) > w(unpleasant)$, for persuading $w(useful) > w(harmful)$ and for threatening $w(obligatory) = 1$. But the experiments have shown that such an initial model has given relatively bad results. 65% of the dialogues were hopeless for *A* because the weights of negative aspects had reached such a level compared with the positive aspects that it was impossible to try to reach a partner model where the reasoning would yield a positive decision.

The situation improved considerably when we added another restriction to the initial model: we required that the chosen reasoning procedure should aim at getting a positive decision in this model. In real life, this restriction is also meaningful: while making a proposal or request we suppose that our partner will agree and only when counter-arguments are put forward shall we try to refute them.

In dialogues, two kinds of 'special cases' occurred:

(1) 'a dead point'. This is a situation when after exchanging two pairs of turns (i.e. after utterances ..., $u_i^A, u_i^B, u_{i+1}^A, u_{i+1}^B$) the partner model has remained the same and applying a reasoning procedure to it gives a negative decision in this model. In that case the computer either gives up and finishes the dialogue or changes its tactics.

(2) 'an endless dialogue'. This is a situation where after two pairs of turns the partner model is constant as before but a valid reasoning procedure gives a positive decision in this model.

In this case, the dialogue will continue until one of the participants runs out of utterances. (We suppose that each utterance can be used only once.)

The process of correcting the user model works effectively and yields new models only if the user points out some different aspects of *D* in his/her consecutive expressions. A model will remain constant if the user repeats one aspect over and over again. For example, if the user time and again indicates the unpleasantness of the action then a new value for *w*(*unpleasant*) will be computed every time but it will be the same as the other weights in the model do not change. In our implemented model, this problem was solved as follows:

(a) if it reaches 'a dead point' in two consecutive computations, the computer gives up and terminates the dialogue
(b) if it is 'an endless dialogue', the computer continues until it runs out of utterances.

Often a decision can be negative if a user points out less pleasantness of *D* in an enticing dialogue or less usefulness of *D* in a persuading dialogue. The reason is that after computing a new value for *w*(*pleasant*) or *w*(*useful*) this value can be less than *w*(*unpleasant*) or *w*(*harmful*), respectively, and the supposition of valid reasoning procedure will not be satisfied any more.

On this ground, the following tactics could be recommended to the (malicious) user in case of failing dialogues:

(1) in dialogues of enticing and persuading, point out less pleasantness and/or usefulness of *D*;
(2) choose an aspect of *D* and consistently use utterances for this aspect. If the computer has in its file less utterances of this class than the user then the computer will give up and
(3) if a user model exists already where the reasoning procedure gives a negative decision, then point out the unpleasantness and/or harmfulness of *D*.

In reality, a user does not know whether his/her model in the computer changed or not. This gives the computer an additional chance to continue a dialogue.

The first two tactics can rather simply be eliminated by the programmer:

(1) for all classes of user utterances (if s/he can only use the predetermined sentences from a menu), make the corresponding computer classes of utterances at least as powerful. For example, if the user has 10 utterances for pointing out the harmfulness of *D* then the computer could have at least 10 utterances for decreasing the harmfulness
(2) make the computer's class of utterances for increasing the pleasantness (or usefulness) at least as much as the user classes of utterances for pointing out the insufficiency of pleasantness (usefulness) and asking for additional information about pleasantness (usefulness)
(3) make the computer class of utterances for increasing the resources at least as big as the user has for pointing out the insufficiency of resources and for requesting additional information about resources.

On the other hand, a user can support a dialogue with quite simple means. S/he could indicate the insufficiency of resources (independent of computer tactics), insufficiency of usefulness (if the computer is enticing or threatening) and insufficiency of pleasantness (if the computer is persuading or threatening).

The generated dialogues are not quite coherent because the computer uses predefined sentences (e.g. User: 'I don't have proper dresses'. Computer: 'You won't be alone – there are three other people with you'). This makes the development of a linguistic processor important for real applications.

In the second version of the software tool, a database is used for identifying different key words and phrases in the user input (the input is checked against regular expressions). The database also includes an index of answer files and links to suitable answers as well as files corresponding to different communicative tactics containing various arguments to present to the user.

The use of unrestricted natural language text as input is both an advantage and a disadvantage for the application, as it helps in creating more natural dialogues, but at the same time if the database is compiled poorly, the conversation can become unnatural in a few moves.

In general, the experiments give a reason to believe that such software could be beneficial for training argumentation skills in a certain domain. Still, we had to implement natural language processing, i.e. language analysis and generation before to do real training with users in order to justify this assumption. So far, we have used two different scenarios. In the first scenario, the goal of *A* is to convince the partner *B* to make a decision about a business trip (see Example 2). In the second scenario, a decision about vegetarianism will be made (see Example 3). To include new scenarios, new sets of sentences (arguments) should be compiled and their weights assigned. In the case of free user input, key words and phrases should be determined that help the computer to relate the user arguments to different aspects of the action under consideration.

6. Summary, conclusions and future work

Our main aim is to model argumentation in agreement negotiation processes. We investigate the conversations where the goal of one partner, *A*, is to get another partner, *B*, to carry out a certain action *D*. Because of this, we consider the reasoning processes that people supposedly go through when working out a decision whether to perform an action or not. We state that people construct folk theories, or naive theories, for the important fields of their experience and that they rely on these theories when acting inside of these domains. They include knowledge, beliefs and image concerning the corresponding domains, but also certain principles and concrete rules that form the basis of operating with these mental structures involved.

The general principles of our reasoning model are analogous to the BDI model but it has some specific traits, which we consider important. First, along with desires we also consider other kinds of motivational inputs for creating the intention of an action in an actor. Secondly, we have worked out a 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 consists of a model of human motivational sphere and of reasoning algorithms.

The interaction proceeds in the following way. (1) *A* informs partner *B* of the communicative goal (to do *D*). (2) *B* goes through the reasoning process and makes a decision, of which s/he informs *A*. In the reasoning process, the subject considers various positive and negative aspects of *D*. If the positive aspects weigh more, the subject will make the decision to do *D*, otherwise s/he will make the opposite decision – not to do *D*. (3) If *B*'s decision is negative, *A* will try, during the following process of interaction, to influence *B*'s reasoning in such a way that ultimately *B* would make a positive decision. There exist three reasoning procedures for *B* in our model (initiated by *B*'s wish, need or obligation to do *D*, respectively) and three communicative tactics, i.e. ways of achieving the communicative goal for *A*, related to the reasoning procedures (respectively, enticement, persuasion and threatening).

We justify our argumentation model in two ways; first of it being the analysis of real human–human dialogues. Arguments and counter-arguments used by humans in negotiation dialogues demonstrate, even though indirectly, how the participants are reasoning before making a decision. Three types of dialogues taken from the EDiC were analysed. First, calls to travel agencies were studied in order to find out communicative tactics used by travel agents to persuade clients to book a trip. Second, we analysed calls of sales clerks of an educational company to other institutions where various training courses were offered to clients. Third, we studied face-to-face conversations between acquaintances. We were looking for sequences of DAs (moves) used by participants to express their arguments and counter-arguments for/against doing an action. We found persuasion and enticement in the dialogues. We can claim that in main lines, the process runs in accordance with the procedures that represent tactics of persuasion and enticement in our model.

The second way to justify our model is its implementation in an experimental DS – the communication trainer. When interacting, the computer uses ready-made Estonian sentences while the user can choose his/her utterances from a menu (in the first version of software) or put in free text in Estonian (in the second version). The computer can optionally play two roles: (1) of participant *A* who is influencing the reasoning of user *B* in order to achieve *B*'s decision to do an action *D* or (2) of participant *B* who is rejecting arguments for doing action *D* proposed by user *A*. In the first case, the DS does not deviate from the fixed communicative tactics but follows them consistently. In the second case, the DS does not deviate from the selected reasoning procedure.

The goal of the paper is to introduce our agreement negotiation model, check its validity on our existing dialogue corpus and implement a simple DS, which includes the reasoning model as the central module that is based on a naive theory of reasoning. An equal development of other modules in the DS has not been our current aim.

The novelty of our approach, as we see it, lies in the starting point of our reasoning model – it is based on studies in the common-sense conception of how the human mind works in such situations; we suppose that in everyday or even institutional communication people start, as a rule, from this conception, not from any consciously chosen scientific one. The reasoning model, as a naive theory of mind, includes some principles which represent the relations between different aspects of the action under consideration and the decision taken.

We are continuing our work in the following directions: (1) refining the reasoning model by means of incorporating the ideas presented in the literature on argumentation and improving the formalism, (2) extending the EDiC by new agreement negotiation dialogues and their analysis in order to justify and develop the model, (3) developing a linguistic processor in order to make it possible for the computer to react adequately to the user input and generate semantically coherent dialogues in Estonian, and (4) evaluating the strategies in the implemented DS using the metrics proposed, for example, in Danieli and Gerbino (1995).

Acknowledgement

We thank our students for implementing and testing the model.

Funding

This work is supported by the European Regional Development Fund through the Estonian Centre of Excellence in Computer Science (EXCS) and the Estonian Research Council (projects 8558 and 9124).

Notes

1. In the examples, transcription of conversation analysis is used (Hutchby, Wooffitt 1998). Utterances are numerated.

References

Allen, J. (1995). *Natural language understanding* (2nd ed.). Redwood City, CA: Benjamin/Cummings.

Allen, J., Ferguson, G., & Stent, A. (2001). An architecture for more realistic conversational systems. In *IUI '01: Proceedings of the 6th international conference on intelligent user interfaces* (pp. 1–8). Santa Fe, NM: ACM Press.

Bench-Capon, T.J.M., & Dunne, P.E. (2007). Argumentation in artificial intelligence. *Artificial Intelligence, 171*, 619–641.

Besnard, P., & Hunter, A. (2008). *Elements of argumentation*. Cambridge, MA: MIT Press.

Boella, G., Hulstijn, J., & van der Torre, L. (2004). *Persuasion strategies in dialogue*. Proceedings of the ECAI workshop on computational models of natural argument (CMNA'04), Valencia.

Boella, G., & van der Torre, L. (2003). *BDI and BOID argumentation*. Proceedings of CMNA-03. The 3rd workshop on computational models of natural argument at IJCAI-2003, Acapulco. Retrieved from www.cmna.info

Bratman, M.E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press (reissued by CSLI 1999).

Bringsjord, S., Taylor, J., Shilliday, A., Clark, M., Arkoudas, K., & Khemlani, S. (2008). *Slate: An argument-centered intelligent assistant to human reasoners*. Proceedings of CMNA-08, Patras, Greece. Retrieved from http://www.cmna.info/CMNA8/

Brown, P., & Levinson, S. (1987). Universals in language usage: Politeness phenomena. In E. Goody (Ed.), *Questions and politeness: Strategies in social interaction* (pp. 56–289). Cambridge: Cambridge University Press.

Carruthers, P., & Smith, P.K. (Eds.). (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.

Chesñevar, C., Maguitman, A., & Loui, R. (2000). Logical models of argument. *ACM Computing Surveys, 32*(4), 337–383.

Chu-Carroll, J., & Carberry, S. (1998). Collaborative response generation in planning dialogues. *Computational Linguistics, 24*(3), 355–400.

D'Andrade, R. (1987). A folk model of the mind. In D. Holland and A. Quinn (Eds.), *Cultural models of language and thought* (pp. 112–148). Cambridge: Cambridge University Press.

Danieli, M., & Gerbino, E. (1995). *Metrics for evaluating dialogue strategies in a spoken language system*. Proceedings of the 1995 AAAI spring symposium on empirical methods in discourse interpretation and generation, Palo Alto, CA.

Davies, M., & Stone, T. (1995). *Folk psychology: The theory of mind debate*. Oxford: Blackwell.

Dębowska, K., Łoziński, P., & Reed, C. (2009). Building bridges between everyday argument and formal representations of reasoning. *Studies in Logic, Grammar and Rhetoric, 16*(29), 95–135.

Ginzburg, J., & Fernández, R. (2010). Computational models of dialogue. In A. Clark, C. Fox, and S. Lappin (Eds.), *The handbook of computational linguistics and natural language processing* (pp. 429–481). Chichester: Wiley-Blackwell.

Grosz, B., & Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*(3), 175–204.

Hadjinikolis, C., Modgil, S., Black, E., Mcburney, P., & Luck, M. (2012). *Investigating strategic considerations in persuasion dialogue games* (Vol. 241: STAIRS 2012, pp. 137–148). Frontiers in artificial intelligence and applications. IOS Press. doi:10.3233/978-1-61499-096-3-137

Hennoste, T., Gerassimenko, O., Kasterpalu, R., Koit, M., Rääbis, A., & Strandson, K. (2008, May 28–30). *From human communication to intelligent user interfaces: Corpora of spoken Estonian*. Proceedings of the sixth international language resources and evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco. Retrieved from www.lrec-conf.org/proceedings/lrec2008/

Heritage, J. (1991). Intention, meaning, and strategy: Observations on constraints on interaction analysis. *Cognitive Sciences, 24*, 311–332.

Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis. Principles, practices and applications*. Cambridge: Polity Press.

Jackendoff, R. (2002). *Foundations of language. brain, meaning, grammar, evolution*. New York: Oxford University Press.

Jokinen, K., & McTear, M.F. (2009). *Spoken dialogue systems*. Princeton, NJ: Morgan & Claypool.

Jurafsky, D., & Martin, J.H. (2008). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics*. Upper Saddle River, NJ: Prentice-Hall.

Koit, M. (2011, October 26–29). Conversational agent in argumentation: Updating of information states. In J. Filipe and J.L.G. Dietz (Eds.), *Proceedings of the international conference on knowledge engineering and ontology development: KEOD 2011* (*international conference on knowledge engineering and ontology development*) (pp. 375–378). Paris: SciTEC.

Koit, M. (2012, August 27). Developing software for training argumentation skills. In F. Grasso, N. Green, & C. Reed (Eds.), *Proceedings of CMNA-2012: The 12th workshop on computational models of natural argument at ECAI-2012* (pp. 11–15), Montpellier. Retrieved from www.cmna.info/CMNA12/

Koit, M., & Õim, H. (2000). Reasoning in interaction: A model of dialogue. In E. Wehrli (Ed.), *TALN 2000. 7th conference on automatic natural language processing* (pp. 217–224), Lausanne.

Koit, M., & Õim, H. (2004). Argumentation in the agreement negotiation process: A model that involves natural reasoning. In F. Grasso, C. Reed, & G. Carenini (Eds.), *Proceedings of the workshop W12 on computational models of natural argument. 16th European conference on artificial intelligence* (pp. 53–56), Valencia.

Koit, M., Roosmaa, T., & Õim, H. (2009, October 6–8). Knowledge representation for human-machine interaction. In J.L.G. Dietz (Ed.), *Proceedings of the international conference on knowledge engineering and ontology development: International conference on knowledge engineering and ontology development, Madeira* (*Portugal*) (pp. 396–399). Portugal: INSTICC.

Landemore, H., & Elster, J. (Eds.). (2012). *Collective wisdom. Principles and mechanisms*. Cambridge: Cambridge University Press.

Levin, J.A., & Moore, J.A. (1978). Dialogue-games: Metacommunications structures for natural language interaction. *Cognitive Science, 1*(4), 395–420.

Loui, R.P. (1998). Process and policy: Resource-bounded non-demonstrative reasoning. *Computational Intelligence, 14*, 1–38.

McBurney, P., & Parsons, S. (2009). Dialogue games for agent argumentation. In I. Rahwan & G. Simari (Eds.), *Argumentation in artificial intelligence* (pp. 261–280). Berlin: Springer.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–111.

Morge, M., & Mancarella, P. (2012). *Arguing over goals for negotiation: Adopting an assumption-based argumentation decision support system*. Dordrecht: Springer Science + Business Media. doi:10.1007/s10726-012-9324-4

Morris, R.M.G., Tarassenko, L., & Kenward, M. (2005). *Cognitive systems – information processing meets brain science*. Cambridge: Academic Press – Psychology.

Ortony, A. (Ed.). (2003). *Metaphor and thought*. Cambridge: Cambridge University Press.

Paradis, C., Hudson, J., & Magnusson, U. (Eds.). (2013). *The construal of spatial meaning: Windows into conceptual space*. Oxford: Oxford University Press.

Rahwan, I., & Larson, K. (2011). Logical mechanism design. *The Knowledge Engineering Review, 26*(1), 61–69.

Rahwan, I., Ramchurn, S.D., Jennings, N.R., McBurney, P., Parsons, S., & Sonenberg, L. (2004). Argumentation-based negotiation. *The Knowledge Engineering Review, 18*(4), 343–375.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B.M. (2009). Computer-supported argumentation: A review of the state of the art. *Computer-Supported Collaborative Learning*. doi:10.1007/s11412-009-9080-x

Tohmé, F. (2002). Negotiation and defeasible decision making. *Theory and Decision, 53*(4), 289–311.

Traum, D., & Larsson, S. (2003). The information state approach to dialogue management. In J. van Kuppevelt & R. Smith (Eds.). *Current and new directions in discourse and dialogue* (pp. 325–353). Dordrecht: Kluwer.

Walton, D.N., & Krabbe, E.C.W. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. Albany, NY: SUNY Press.

Wells, S., & Reed , C. (2006). *Testing formal dialectic. Argumentation in multi-agent systems*. Lecture Notes in Computer Science, Vol. 4049, pp. 74–87. Utrecht: Springer.

Yuan, T., Moore, D., & Grierson, A. (2008). A human-computer dialogue system for educational debate, a computational dialectics approach. *International Journal of Artificial Intelligence in Education, 18*(1), 3–26.

Yuan, T., Moore, D., Reed, C., Ravenscroft, A., & Maudet, N. (2011). Informal logic dialogue games in human-computer dialogue. *Knowledge Engineering Review, 26*(2), 159–174.

Appendix. Examples of human–human dialogues from the EDiC

In the examples, transcription of conversation analysis is used (Hennoste et al. 2008). Utterances are numerated. Slashes // are used for comments.

Dialogue 1 A travel agent (*A*) calls back to a client (*B*) and offers a trip to a water park.

A: *.hh tähendab meil on nüüd niimodi = et me oleme pla’neerinud küll teisel no’vembril,*

1. we have now planned a trip for 2 November

/- - -/

see ’reis maksab meil ’kakssada = viiskend ’krooni.

2. this trip costs two hundred and fifty kroon

B: *ja ’m:is sinna ’alla nagu ’käib.*

3. 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.*

4. 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.

5. and then a ticket to the water park

.hh ee ’Tallinast Sakalast väljutakse ommikul kell: ’kaheksa [ja]

6. the departure is at 8a.m. at Sakala in Tallinn

B: *[jaa?]*

7. *yes*

/- - -/

//*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*

8. 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 =*

9. yes

A: *= siis meil on pere’soodustus ja see on ’kakssada kakskend krooni per ’inimene sel ’juhul.*

10. then we have a family discount and it is two hundred and twenty kroon per person in this case

B: *mhmh,*

11. uhuh

/- - -/

B: *ma olen ’kuulnud = et siin on ’nädalavahetustel vist üsna [nagu pikad ’järjekorrad.]*

12. I have heard that there are quite long queues on weekends //*B* still presents a counter-argument but *A* averts it//

A: *[ee seal on üle’üldse] väga suured prob’leemid ja väga pikad ’järjekorrad [ütlen] teile ’ausalt*

13. yes there are big problems, very long queues

’bussi’juht seisab ’järje’korras [ja] ’elavast järjekorrast võtab võtab meile

14. **the bus driver stands in a queue and takes (tickets) for us**

/- - -/

A: *.hh aga mm kui = te = nüd 'soovite, ma panen teid 'kirja.*

15. **if you want, I register you now**//*A* repeats the proposal but *B* postpones the decision//

B: *ee ma vel 'mõtlen.*

16. **I$^{\infty}$ll think more.**

Dialogue 2 A travel agent *A* is proposing various arguments for usefulness of booking the trip by the client *B*.

/- - -/

A: tändab on praegu välja pakkuda (.) küllaltki 'soodne variant on 'Finnääriga lend = on.

1. **an advantageous variant can be offered to fly with Finnair**

B: jah?

2. **yes**

/- - -/

A: peab mõtlema selle 'hinna = üle praegu on 'kolm = tuhat üks'sada tuleks nagu koos 'täksidega 'kokku.

3. **one must think about this price it is three thousand one hundred with taxes**

(.) see 'hind.

4. **this price**

(.)

B: ah[ah]

5. **aha**

/- - -/

A: ['praegu mo'mendil] 'kohti 'on.

6. **there are free places at the moment**

(0.5)

B: [< 'selge. >]

7. **clear**

/- - -/

Dialogue 3 *A* tries to achieve *B*'s agreement to publish his personal data at the website of an institute but *B* does not want to publicise his relations with the institute.

/- - -/

B: a:ga ma mõtlen oopis 'teist asja,

1. **but I mean another thing**

[võib]olla ma = ei::: 'pane sinna 'ültse mingeid andmeid.

2. **maybe I don't put any data there**//Refusal//

/- - -/

A: aga [kõik 'teised] 'panevad.

3. **but all others will put** //Disagreement + a statement for increasing the weight of usefulness: it is useful to be like other people//

/- - -/

B: jaahhhhh ((ohates))

4. **yes ((sighing))**

ei = no ma = i ma = i 'näe nagu suurt 'vajadust 'miks: 'miks see peaks seal isegi 'olema.

5. **no, I don't see any reason why it should be there** //Refusal + a statement for decreasing the weight of usefulness//

(0.3) ausalt. = hh =

6. **honestly**

A: = noh ma = i = tea esiteks on see .hhhhhh nagu nimodi=hhhh .hhhhh 'otsustatud = et = hh et nõukogu 'tuleb = ja: = ja q noh (1.0) .hhhhhhhhh eeeee ikkagi noh, see on ka 'enda tutvustamise = hhhhhhh 'mõttes kogu = se .hhhhhh > Kaerajaani instituudi kodu'lehekülg. = hhh < .hhh

7. **well, I don't know, first, it was decided that the council will come and and well still well it is for presentation, the website of the Kaerajaani Institute** //A statement for increasing the weight of usefulness//

B: ei seda 'küll =

8. **no, indeed**
   aga ma mõtlen mõtlen noh igasuguseit meili'aadresse = ja
   'vär[ke = ja]
9. **but I mean, well, every e-mail addresses and such things** //Refusal + a statement for increasing the weight of harmfulness: personal data will be published//