Quignard, M. & Baker, M.J. (1997). Modelling Argumentation and Belief Revision in Agent Interactions. *Proceedings of "European Conference on Cognitive Science"*, Manchester (UK), 1997.

MODELLING ARGUMENTATION AND BELIEF REVISION IN AGENT INTERACTIONS

MATTHIEU QUIGNARD and MICHAEL BAKER

Centre National de la Recherche Scientifique (CNRS) / Université Lyon 2,
UMR GRIC (Groupe de Recherches sur les Interactions Communicatives),
Equipe COAST, ENS de Lyon, 46 allée d'Italie, 69364 Lyon cedex 07, France.
Tel. +33 (0)4 72 72 85 38, Fax. +33 (0)4 72 72 80 80.
Email. {Matthieu.Quignard@ens-lyon.fr, Michael.Baker@ens-lyon.fr}

Abstract : We describe a model for resolution of belief conflicts by argumentation in agent interactions, based on aspects of dialogic logic, speech act theory and belief systems research. This is part of a long term project on the relations between argumentation and cognitive change in collaborative problem-solving interactions. A set of communicative acts for argumentative interaction is defined, based on viewing them as multifunctional with respect to argumentation and interaction control. We concentrate on defining different intra- and inter-personal conflict situations, in terms of agents' mental states, and specify the communicative acts that are relevant within each of them. An illustrative example of an argumentation sequence is presented and discussed in relation to the model. Our main claims are that the dialectical and rhetorical dimensions of argumentation need to be situated within a more general interactional framework, and that resolution of belief conflicts can give rise to increased coherence in agents' epistemic states.

Keywords: argumentation, belief, dialogue, speech acts, agent models.

1. Introduction

Research on cooperative or collaborative learning has up to recently been dominated by the question as to whether Vygotskyan cooperation is more potentially productive for learning than neo-Piagetian conflict. However, it has been recognised that these two phenomena, and associated theoretical approaches, should not necessarily be opposed. Thus a number of authors (e.g. [18], [2], [4]) have claimed that it is not the incidence of verbal conflicts *per se* which is important for cognitive change, but rather the way in which conflicts are *cooperatively resolved*, in sequences of *argumentation dialogue*. This view is associated with the general and growing recognition of the importance of studying interactive *processes* [23], [20], in addition to relating initial conditions of the learning situation to outcomes.

The research that we describe here is situated within these tendencies, and aims ultimately to understand the possible relations between features of argumentation and possible cognitive effects in individuals. Here we present the first part of our project, which concerns the elaboration of a computational model[1] for argumentative interaction between artificial agents, at the level of exchange of speech acts (natural language understanding and generation are not adressed). We also present an illustrative example of an interaction that could be generated by the model, based on a corpus of verbal interactions between students solving problems of qualitative modelling (energy) in physics [11]. In a later stage of our project, a similar corpus will be used to validate the model.

Our model takes into account two main dimensions of argumentation [19] as a form of *interaction*[2] : *dialectical* and *rhetorical*. Along the dialectical dimension, argumentation is viewed as a (verbal) *game*, played using certain allowable *moves*, according to certain *ground rules*, and oriented towards a determinate *outcome* (i.e. designation of a "winner" and a "loser"). Along the rhetorical dimension, argumentation is principally viewed as the attempt to *persuade* listeners to accept the speaker's point of view, by verbal means that are evaluated in terms of their *efficacity*, sometimes at the expense of

---

[1] The model described previously is implemented in CLOS (Common Lisp Object System). This enabled us to easily integrate an existing JTMS in Lisp and to define generic classes of arguing agents.

[2] Argumentation may, of course, also be viewed as a structure of reasoning in textual or monologual discourse [25], as a phenomenon of scientific and logical reasoning, or even as the basis for a general semantic theory (c.f. [1], [22]).

rationality. More generally, we view it as essentially concerned with *cognitive effects* on agents within an *actional* perspective on language.

These two dimensions are taken into account by synthesising aspects of the following three areas of research : *dialogic logic* [5], *speech act theory* (e.g. [6], [7], [8]), and *belief systems* research (e.g. [12], [9], [15]). Dialogic logic provides the starting point, in terms of a basic dialectical framework that includes formal definitions of "conflicts of avowed opinions", and rules that govern what moves can and must be made in specific contexts. However, the fact that it is based on propositional logic, with no reference to arguing agents' mental states (beliefs, intentions, …), means that it is ill-suited to modelling cognitive effects on agents. We have therefore replaced the purely syntactic operators of dialogic logic with speech act operators based on pertinence conditions in, and effects on agents' mental states. An important aspect of our model is that speech acts are viewed as multifunctional (see §3.1 below). We have used a justification-based truth-maintenance system ("JTMS" - [12]) as a basis for representing agents' beliefs in propositions that are relevant to the argumentation

According to our model, an argumentative interaction is fundamentally one which 'resolved' by agents' *externalisations* of *revisions* in their beliefs. *Increased coherence* in agents' beliefs therefore constitutes the type of cognitive change which we take into account here, in relation to argumentation.

In the remainder of the paper we describe the theoretical basis of the model in more detail, then the internal structure and dynamics of the model itself. This is followed by discussion of a detailed example of argumentation generation, in relation to an example taken from an interaction corpus. In conclusion we discuss limitations of the model and further work.

2. Theoretical framework

Our model is based on the basic *dialectical framework* of dialogic logic, rather than on the details of the language itself, its proof and model theories. We therefore present the basic components of the dialectical framework, then describe how they are 'recast' in our model, drawing on other areas of cognitive science research.

The main theoretical underpinnings and components of dialogic logic are as follows :

— *Principle of (Verbal) Externalization of Dialectics*. "Whether a certain move is permissible shall depend on what has been said, and not on intentions, beliefs, etc." (see [5], p. 60).

— *Conflict definitions*. A "conflict of avowed opinions" is defined by a quadruplet <C,T,B,A> ; A and B are language users, T the thesis, C a set of concessions (commonly agreed statements). Concessions must be established prior to debating ; they can not be modified. Conflicts may be "simple" (a single thesis is debated) or "mixed" (more than one thesis is debated).

— *Allowable statements*. In turn, each participant may make a single statement, corresponding to a propositional content and a speech act operator (assertion U, hypothetical assertion (?)U, question U?, exclamation U!).

— *Dialectical roles*. With respect to a thesis T, the participants A and B are assigned to dialectical roles of *proponent* or *opponent*, which determine their "statemental dialogue attitudes" *(pro* or *contra)* with respect to their future declarations (e.g proponent is always *pro* with respect to own statements ; opponent is always *contra* the proponent's). Roles can not change during a debate.

— *Attacking and defending moves*. The allowable statements may be used to make attack or defense moves, under the restriction of certain rules. The moves are termed "structural" when they depend on the propositional content[3] of U, otherwise, they are "general" when they appeal to general rules of the debate itself (e.g the defense *Ipse Dixisti !* - "You said so yourself!", used when a participant attacks one of his own previous statements).

— *Argumentation outcomes*. The dialogue continues until *exhaustion of rights* (to continue debating) of one of the participants : either the participant has no further legal defenses at his disposition, or else has committed an infraction of a general rule of the debate. The participant whose rights are exhausted is said to have "lost" the debate with respect to T (and externalises this), the other is said to have "won" (successfully defended T). In a mixed confict with several 'chains', the winner of the most recent chain has "won" the whole debate.

The most basic change that we introduce to the above-described dialectical framework concerns the *principle of externalization of dialectics*. Once this principle is modified, other modifications follow in turn. We thus assume precisely that dialectical principles are *relativised* to individual agents : what

[3] For example, in dialogic logic an "attacking" move on the proposition "A ∧ B" is defined on the basis of propositional logic as "A?" or "B?" ("how do you defend A?"). Using a JTMS the attack could correspond to "I do not believe that A" ("A" is OUT of the speaking agent's set of beliefs).

counts as an attack or defense is decided from the point of view of an agent's own mental states, as is the case with the outcome of the debate itself. Fundamentally, a debate will be resolved in our model when the *beliefs* of one of the participating agents are *revised* so that interpersonal conflicts (i.e. mutually believed sets of logically inconsistent propositions) are no longer believed to exist. Of course, agents can and do "externalise" their mental states using speech acts, to the extent that there is adequation between agents' models of each other. Such revisions lead to an augmentation of the coherence of agents' beliefs.

The theoretical principles on which our model is based may be summarised as follows, in direct comparison to those underlying dialogic logic :

— *Principle of relativisation of dialectics*. Whether a certain move shall be allowable is decided from the point of view of an agent's own cognitive states ; agents' moves must be consistent with their own utterances and mental states (sincerity).

— *Conflict definitions*. A belief conflict between two agents, A and B, obtains when an agent A believes that A and B mutually believe a set of inconsistent propositions ("no-good") ; the set may be distributed across agents *(inter-personal)* or be believed by a single agent *(intra-personal)*. Specific mental states of conflict are defined according to the possession of information about propositions, and the nature of their justifications, if any.

— *Allowable communicative acts* ("CA"s). These are defined *simultaneously* along two dimensions : (1) task / control CAs [6]. Here the former operate on the level of argumentation itself, and the latter are used to manage the argumentative interaction (e.g. opening, closing, turn-taking, ...). (2) type of speech act, in accordance with Vanderveken's [26] taxonomy (assertives, directives, commissives, declaratives, expressives). Finally, we view *all* of these CAs as potentially *multifunctional*, on the task and control levels (e.g. an "assert" CA can open a new argumentation sequence, whilst controlling turn-taking, is assertive on the belief level).

— *Dialectical roles*. Since agents' belief can evolve in and as a result of the dialogue, such roles are no longer fixed.

— *Attacking and defending moves*. We replace structural attacks and defenses with relations between agents' propositional attitudes.

— *Argumentation outcomes*. Whether the argumentation is *closed* or not depends on mutual beliefs concerning revisions of agents' attitudes and their expression in CAs. It is possible that a conflict can *not* be resolved in this way.

In summary, what we retain from dialogic logic is the general theoretical framework, as resolution of a verbally expressed conflict governed by rules that are specifically *dialectical*. We situate the dialectical framework within an *interactional* and *cognitive* one.

3. Description of the model

Here we present a more precise description of our model, from the following points of view : agent interaction modelling, agents' cognitive mechanisms and the argumentation model.

3.1 Agent interaction model

Dialogue is the interactional framework in which the agents evolve. They interact only by sending messages : communicative acts (see below). In order to potentially model the communication channel, we use an intermediate agent, the "reporter", which can alter communication by retention of information and record the interaction history.

The purpose of the model is not only to resolve belief conflicts, but also to manage an interaction in which such resolution can take place. Certain elementary rules of dialogue must therefore be respected (turn taking, opening and closing of dialogue). Thus the communicative acts incorporated in the model have this dual function : (1) argumentational function - expressing opinions, questioning, justifying, revising, … ; (2) dialogical function - alternating, opening and closing the debate, … Communicative acts will also play a number of roles in the model : they produce effects on beliefs but also on the continuation of the dialogue itself. This relates to the *multifunctionality of acts* mentioned previously. For example, even before being a request for justifications, a communicative act such as "A : REQ-JUSTIF **p1**" corresponds also to the taking of a turn that *commits* the speaker to be silent on the next turn.

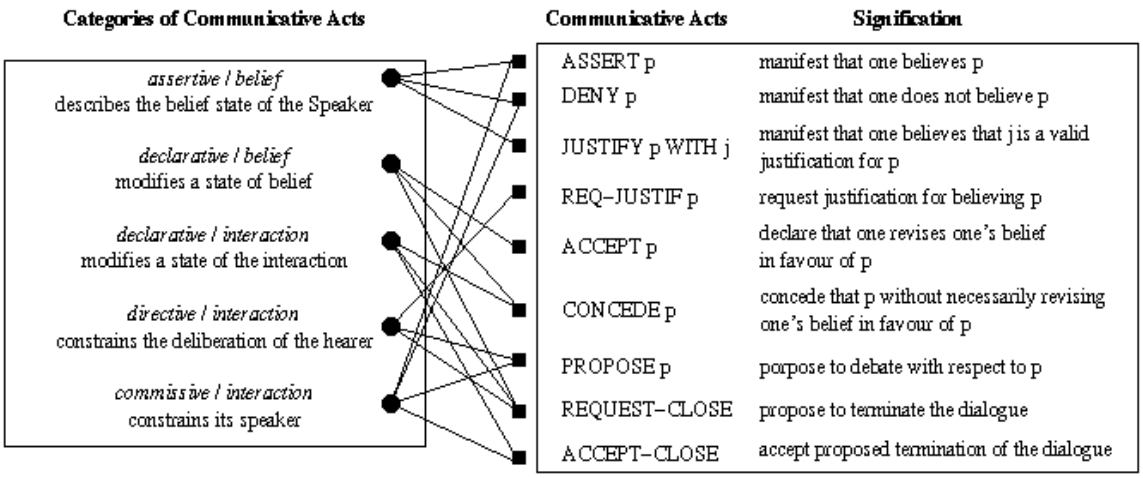

Figure 1 : Multifunctionality of Communicative Acts

We describe this multifunctionality, at the conceptual level, in terms of Searle and Vanderveken's [24] [26] taxonomy of speech acts (see §2 above). Given the dual nature of the interaction context (beliefs and dialogue), there are theoretically eight[4] ways of intervening on the context, of which we only retain the five most important. **Figure 1** shows the functional categories of communicative acts, with the main effects on the context (interaction and belief). The multifunctionality of acts is shown by the fact that a given act can be classified under several different categories - for example, ASSERT can express an agent's belief state (assertive) and commit the agent (commissive).

3.2 Modelling cognitive agents

Agents possess two types of representations : the dialogue history and beliefs (their own and of their interlocutors), relative to a domain (subject of debate).

The dialogue history is a memory of previous moves. The simplest version is a list containing the number of the act in the argumentation, the communicative act expressed and its arguments.

Beliefs are represented as JTMS networks [12], or nodes representing propositions[5]. When a node is "active" (IN) this signifies the belief of the system in the proposition. Each agent possesses a representation of his own beliefs as well as of those of his interlocutor. After each argumentation move agents update both sets of beliefs and decide what is the current belief state of each. Subsequently, agents execute the most relevant communicative act. Cognitive mechanisms in the next section explains in details generation of relevant moves.

Figure 2 shows the dynamic description of agents, including the phenomena of belief revision and argumentation that the system models. Mechanisms operate in parallel in each agent, even though one will not 'speak'.

After perception of messages itself, that include recording them in the dialogue history, agents must understand the messages by situating them in the dialogue context (dialogical understanding) and in the task context (here, epistemic understanding). The comprehension mechanism draws on the current context, represented by beliefs and the dialogue history (ie the current *expectations*).

Once the message is understood, the context is updated : the argumentation has advanced by one step. Agents will therefore have to modify their dialogue state and beliefs. The message is analysed according to the five categories of CAs as described above.

---

[4] We do not require the (purely) expressive dimension of speech acts in our model. This reduces the number of speech acts dimensions to 4, and thus the number of categories to a maximum of eight.

[5] In a little-cited later section of his 1979 paper, entitled "Dialectical Arguments" Doyle had in fact hinted at the possibility of using the JTMS as an underlying basis for argumentation. Despite the fact that an ATMS [9] has been argued to incorporate a more efficient revision algorithm, we have nevertheless chosen to work with a JTMS here since its explicit representation of justifications for and against propositions corresponds more naturally to defenses and attacks in argumentational interactions.

The agent now enters the phase of argumentation analysis : it will select all moves it is allowed to do, with respect to its updated beliefs representations (see the next section).

In the deliberation phase, only the most relevant move is chosen from the set of permissible moves. Evaluating relevance is difficult since it varies with the specific goals and behaviours of agents. Attitudes and strategies weight the relevance of moves, that are then filtered by the interaction constrains (directive or commissive). Afterwards, the reaction message can be formulated and sent.

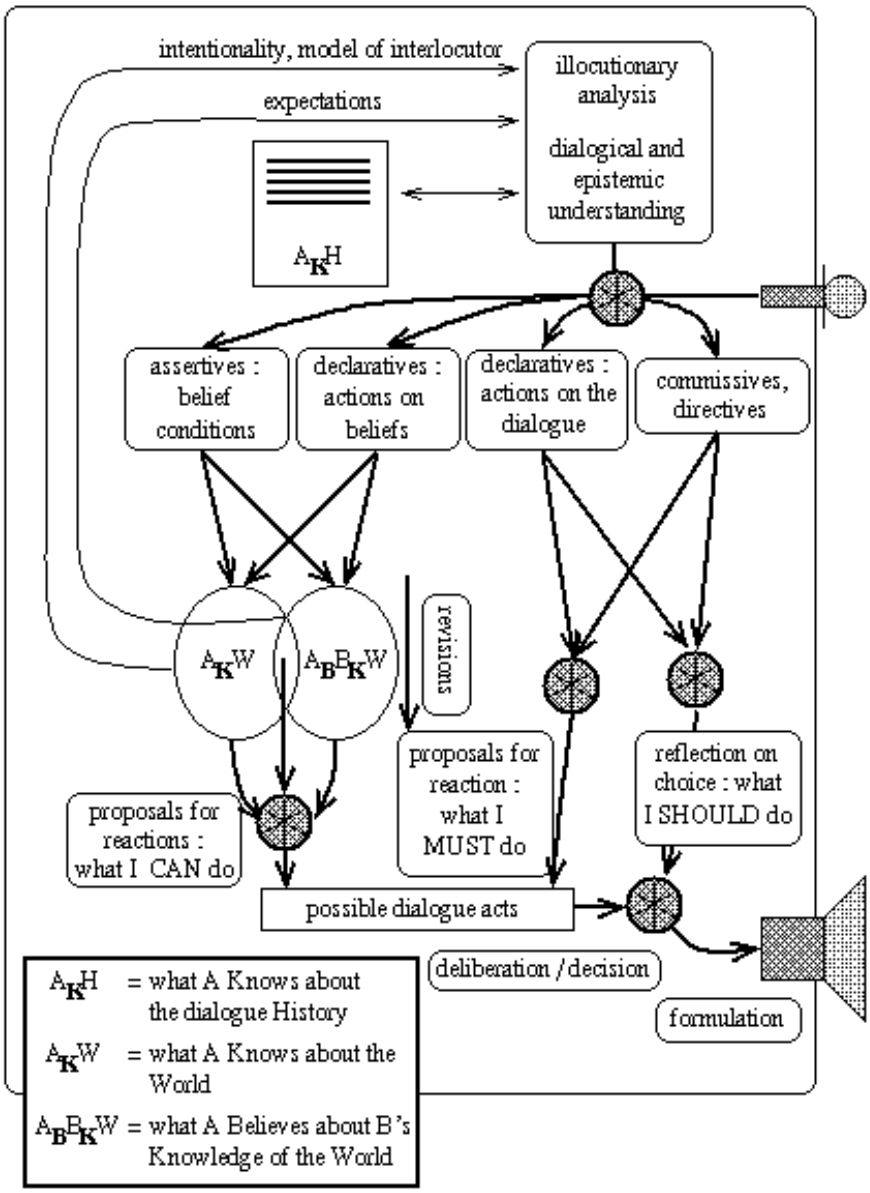

Figure 2. Internal structure of agents.

3.3 The argumentation model

Once the argumentation context has changed, the new state must be analysed, principally in terms of beliefs. In an argumentation phase with respect to a proposition **p**, the debate is analysed along 3 specific directions, each corresponding to a particular orientation of the 'language game'. These analyses are not performed within the totality of the network but rather within a *focus space* restricted to **p**, its justifications and all the nodes that are directly incompatible with it (NO-GOODs). The agent analyses its personal beliefs with respect to **p** and all the propositions **q** that are incompatible with **p**. The aim of this analysis is to remove internal contradictions, maintaining coherence of communicated propositions.

The representations on which these analyses are based are summarised in Tables 1, 2 and 3. They represent the different possible mental states of an agent, with the communicative acts that are relevant

in them, with respect to two incompatible propositions, p and q, for the following three possible types of conflict :

— C_1: *Intrapersonal* conflict, within an agent **A**'s own beliefs (Table 1) ;

— C_2: *Intrapersonal* conflict, within the beliefs of an agent **B**, as represented by **A** (Table 2) ;

— C_3: *Interpersonal* conflict, between **A**'s own beliefs and those that it attributes to **B** (Table 3).

Table 1. Intrapersonal conflict within an agent A's own beliefs, with possible CAs.

	p	**node does not exist**	**node IN by hypothesis**	**node IN, justified**	**node OUT**
q	**p & q** incompatible	no knowledge of p	believes that p	believes that p	does not believe that p
node does not exist	no known contradiction with p	knows nothing about p **CLOSE**	belief$_1$ in p **PROPOSE**, **ASSERT**, **JUSTIFY** p	belief$_3$ in p **PROPOSE**, **ASSERT**, **JUSTIFY** p	non-belief$_1$ in p **REQ-JUSTIF** p
node IN by hypothesis	believes that ¬p because of q	knows nothing about p **CLOSE**	contradiction$_1$ **RETRACT** p or q	contradiction$_2$ **RETRACT** q	non-belief$_2$ in p **REQ-JUSTIF** p, **DENY** p, **PROPOSE** q
node IN, justified	believes that ¬p because of q	knows nothing about p **CLOSE**	contradiction$_2$ **RETRACT** p	contradiction$_3$ **REQ-JUSTIF** p	non-belief$_3$ in p **REQ-JUSTIF** p, **DENY** p, **PROPOSE** q
node OUT	does not believe that ¬p	knows nothing about p **CLOSE**	belief$_2$ in p **PROPOSE**, **ASSERT**, **JUSTIFY** p	belief$_4$ in p **PROPOSE**, **ASSERT**, **JUSTIFY** p	non-belief$_4$ in p **REQ-JUSTIF** p, **CLOSE**

Table 1 shows different possible belief states of an agent with respect to two propositions, **p** and **q** that form a no-good (ie incompatible), with the communicative acts that are *possible* in each state. For example, in the case of A believes that **p** with justification for it, and believes that **q** is not the case by hypothesis, (and **p** and **q** are linked together with a No-Good), A is a particular state of belief to p (belief4). Within each agent, we assume there are 4 levels of belief to **p**, 4 levels for disbelief and 3 levels of contradiction (internal conflict). In his state of belief, the relevant moves for A are PROPOSE, ASSERT and JUSTIFY **p**.

The analysis of belief states assumes existence of another agent : the agent analyses the beliefs that he attributes to others in order attack or, in the case of 'cooperative' argumentations, to help the other resolve his contradictions, and maybe the main conflict of opinions. Table 2 shows the belief states of the *hearer*, from a *speaker's point of view*, and the CAs that the *speaker* can perform in each belief state.

Table 2. Intrapersonal conflict within the beliefs of an agent B, as represented by an agent A, with communicative acts that A believes B could perform.

	p	**node does not exist**	**node IN by hypothesis**	**node IN, justified**	**node OUT**
q	**p & q** incompatible	has no knowledge of p	believes that p	believes that p	does not believe that p
node does not exist	no known contradiction with p	knows nothing about p	no conflict	no conflict	no conflict
node IN by hypothesis	believes that ¬p because of q	no conflict	contradiction$_1$ **PROPOSE** q	contradiction$_3$ **PROPOSE** q	no conflict
node IN, justified	believes that ¬p because of q	no conflict	contradiction$_2$ **PROPOSE** q	contradiction$_4$ **PROPOSE** q	no conflict
node OUT	does not believe that ¬p	no conflict	no conflict	no conflict	no conflict

Finally, in the analysis of an *interpersonal conflict*, agents analyse their own beliefs in comparison with those that they attribute to others in order to resolve the conflict that opposes them. Table 3 shows the possible communicative acts for a speaker in such comparative belief states.

Table 3. Interpersonal conflict, between an agent A's own beliefs and those that it attributes to B, with possible communicative acts for the speaker A.

A's beliefs with respect to p **B's beliefs (according to A) with respect to p**		**node does not exist** has no knowledge of p	**node IN by hypothesis** believes that p	**node IN, justified** believes that p	**node OUT** does not believe that p
node does not exist	has no knowledge of p	(impossible case !)	**PROPOSE** p **JUSTIFY** p	**PROPOSE** p **JUSTIFY** p	no conflict
node IN by hypothesis	believes that p	**ACCEPT, CONCEDE, CLOSE**	no conflict	no conflict	**REQ-JUSTIF** p
node IN, justified	believes that p	**ACCEPT, CONCEDE, CLOSE**	no conflict	no conflict	**REQ-JUSTIF** p
node OUT	does not believe that p	no conflict	**ASSERT** p **JUSTIFY** p	**ASSERT** p **JUSTIFY** p	no conflict

4. An example

We now describe an example of an argumentation that the model could generate[6]. The example involves two agents having perfect knowledge of each other. The system is run by the initialisation of the dialogue histories of each agent : at instant 0, A has uttered "PROPOSE p1". The initial belief states are shown in Figure 3.

In order to make the example more comprehensible, we have instantiated the propositions on the basis of analysis of reasons for and against a thesis, taken from a real example [4]. Note that this is *not* at this stage of our research viewed as a validation of the model, but simply an illustration.

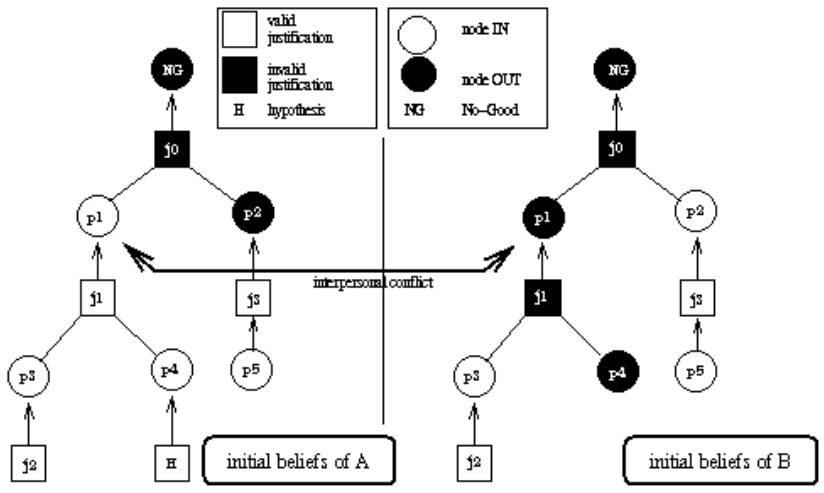

Figure 3. Initial belief states of agents in the example argumentation interaction.

Briefly, the students' task is to construct a qualitative model for energy storage, transformation and transfer (an "energy chain") for an experimental situation where a bulb is connected to a battery by two wires (see [11]). The lower part of Figure 4 gives the correct solution, for reference and the upper part, the conflicting solutions. An interpersonal conflict arises concerning the direction of the energy transfers between battery (reservoir) and bulb (transformer). Student A believes (and claims) that there is an energy transfer also from the bulb to the battery ((s)he confuses electrical

6 Here it is hand-generated from Tables 1, 2 and 3 since the current state of the implementation does not permit such a complete generation

current with energy), whereas student B claims that two transfers go from battery to bulb. In the real interaction case, B does in fact win the verbal conflict, using the conclusive counter-argument that energy can not go back to the battery, otherwise the bulb would never go out (!).

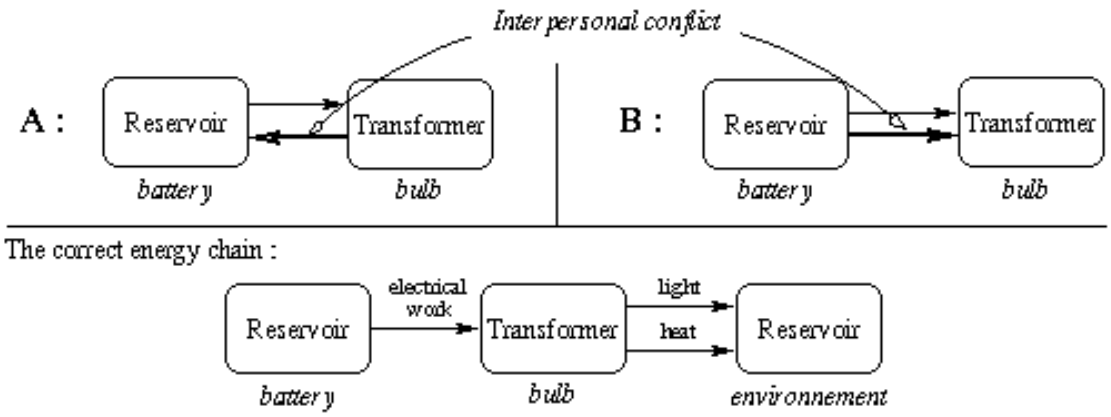

Figure 4. Conflicting energy chains of students A and B, with correct solution for reference.

Table 4 shows the argumentation from instant 0 to its closure. The first column numbers interventions, the next two show CAs of A and B with their instantiated propositions, the next their interpretation, and the final two the beliefs that underlie generation of the CAs.

Table 4. Dialogue history of argumentation, as recuperated by the reporter.

N	**A** CAs **B**		Interpretation of CAs	**A** Beliefs **B**	
0	PROPOSE **p1** *There is an energy transfer from the bulb to the battery*		*assertive* of belief in **p1** *directive* : assert/deny **p1** *declarative* of control : focus on **p1**	p1(in,out)	p1(out,in)
1		DENY **p1** *I don't think so !*	*assertive* of non-belief in **p1**		p1(out,in)
2	REQ-JUSTIF **¬p1** *Why <don't you think so>?*		*directive* : justify **¬p1**	¬p1(out,in)	
3		JUSTIFY **p2** *There is no energy transfert from the bulb to the battery.*	*assertive* of belief in **p2**		p2(in,out)
4	REQ-JUSTIF **p2** *Why <no transfert from the bulb to the battery>?*		*directive* : justify **p2** *declarative* of control : focus on **p2**	p2(out,in)	
5		JUSTIFY : **j3** () *The bulb will eventually go out.*	*assertive* : **j3** justifies **p2**		p3(in,0)
6	ACCEPT **p2** *You're right*		*declarative* of belief in **p2**	p3(in,0)	
7		REQ-JUSTIF **p1** *Why did you put a transfert from the bulb to the battery ?*	*directive* : justify **p1**		p1(out,in)
8	JUSTIFY : **j1**(p3, p4) *A complete energy chain must start and end with a reservoir, and the electrical current brings energy back to battery*		*assertif* : **j1** justify **p1**	p3(in,0) p4(in,0)	

9		REQ-JUSTIF **p4** *Why does the electrical current bring back energy to the battery ?*	*directive* : justify **p4** *declarative* of control : focus on **p4**		p4(out,0)
10	JUSTIFY **p4** : (H) *Maybe yes...*		*assertive* : H justifies **p2**	p4(H,0)	
11		DENY **p4** *I don't think so !*	*assertive* of non-belief in **p4**		p4(out,0)
12	RETRACT **p4** *Ok, you may be right.*		*declarative* of non-belief in **p4** *declarative* of closing on **p4**	p4(out,0)	
13		REQUEST-CLOSE *So we agree, don't we ?*	*directive* : closure	p1(out,in)	p1(out,in)
14	ACCEPT-CLOSE *Yes, we do.*		*declarative* of control : closure	p1(out,in)	p1(out,in)

Rather than attempting a detailed move-by-move account of the generation of the example (it can be checked from Figure 1 and Tables 1, 2 and 3), we shall concentrate on the basic structural features of the interaction.

The example is interesting in terms of the relative complexity of the argumentation generated, given the equally restricted conflict situation. It is in fact at the same time a *simple* conflict (a single proposition/thesis debated : p1) and a *mixed* one (p1 and ¬p1 are debated).

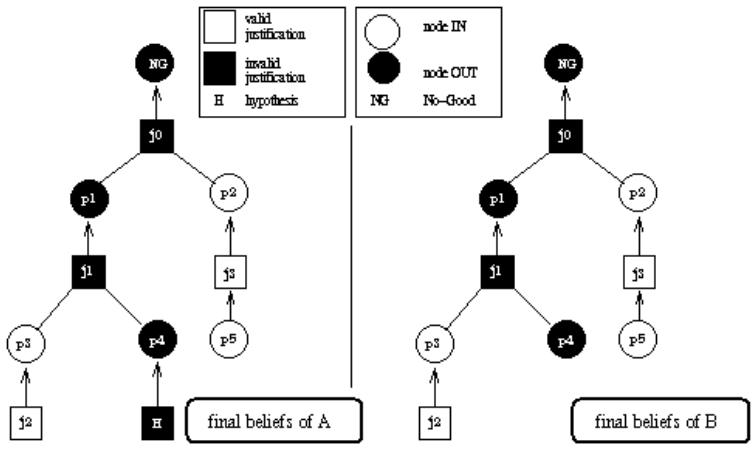

Figure 5. Final belief states of agents in the example argumentation interaction.

The conflict is resolved when agreement is reached (i.e. mutual belief that there is no longer a belief conflict), via the opening of debates on "local theses" (p4 in this example). The main and local theses debated are shown in Figure 6.

Below we summarise a structural description of the argumentation in the example, that brings out the different phases of the conflict resolution.

1. *Opening of the belief conflict (0-2)* : A asserts p1, B denies it (in fact s/he has reasons to believe that p1 is not the case, i.e. ¬p1). A attacks ¬p1 shifts the debate to the acceptability of ¬p1 (is it possible to accept q given that p1 and q are incompatible)[7].

2. *Resolution of debate on local thesis ¬p1 (2-7)* : B justifies that ¬p1 is acceptable given p2. A requests justifications since A thinks that p2 can not be maintained. B justifies with p5 in which A believes. A accepts the argument and revises his/her related beliefs : henceforth A believes that p2. His NO-GOOD therefore detects a contradiction. There is thus a double conflict : an interpersonal one with respect to p1 as well as an interpersonal one with respect to p1 and p2.

3. *Resolution of debate on local thesis p1 (7-9)* : since it is B' turn, s/he reopens the debate on p1 (it is still the main thesis of the conflict). B requests justifications for accepting p1. In addition, this request is made (cf. Table 2) in the context of a **cooperative attitude**, where B helps A resolve his/her internal conflict. A justifies p1 with p2 and p4. B attacks p4, opening a local debate with thesis p4.

4. *Resolution of debate on local thesis p4 (9-12)* : at this point neither of the participants has a (hard-wired) justification. A holds p4 by hypothesis. B does not believe that p4. There is no other way out for A than to retract his/her hypothesis.

7 In fact, this attacking phase is not yet allowed by the system, since it would require a more sophisticated representation of NO-GOODs themselves than is provided by the JTMS.

5. *Closure of the belief conflict (12-14)* : Since A has retracted his hypothesis, the debate on p4 is closed by propagation of belief statuses. A no longer maintains that p4, nor that p1. A and B have resolved their conflict. B requests closure, which A ratifies.

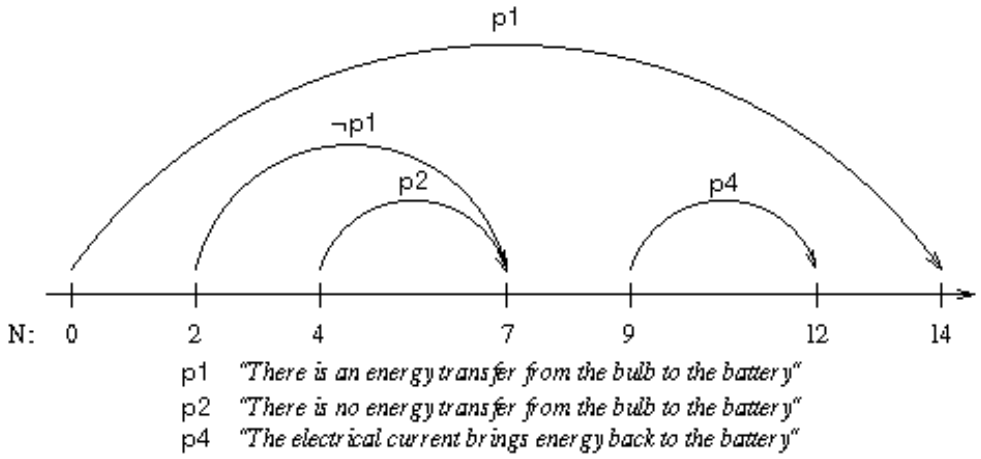

Figure 6. Main theses and local-theses debated in the example.

5. Conclusions and further work

We have described a model for argumentation in agent interactions that shows how intra- and inter-personal belief conflicts may be resolved, concentrating on the relations between agents' cognitive states and their choice of relevant argumentation moves. The model has been developed as part of our longer term research goal of exploring the relations between argumentation as a form of verbal interaction and certain types of cognitive change. Our basic claim is that such argumentations may lead to increased coherence in agents' beliefs, in a way that is analytically and empirically distinct from who wins or loses the argumentation game [10].

One basic issue that we have not yet fully addressed concerns *deliberation* , i.e. an effective procedure for evaluating the relevance of argumentation communicative acts at a given stage of the interaction. Fundamentally, this depends on the choice of *argumentation style* for a given interaction, since argumentations can be *cooperative* as well as *adversarial* [27]. Supporting such styles would require a set of higher-level attitudes of the interaction itself [14], in addition to a more sophisticated dialogue history.

At present the analysis of conflict resolution, concentrating purely on the attitude of belief, remains too simplistic. As Van Eemeren and Grootendorst [13] argue, other intermediary attitudes in argumentation need to be taken into account, such as conviction and acceptance. In addition, use of attentional constraints on belief needs to be more fully explored (see e.g. [16], [17]), since this in fact creates the initial belief incoherences that argumentation can remove. At present our model assumes perfect communication ; in the longer term we intend to introduce understanding of illocutionary forces into it, since this is important for modelling certain types of argumentation, such as those where participants 'argue past each other'.

Acknowledgements

We would like to thank the CNRS and Andrée Tiberghien, director of the COAST research team, for providing a framework within which this research could be carried out. The research described here is based on the DEA and PhD research in Cognitive Science of M. Quignard [21], directed by M. Baker and J. Caelen.

References

[1] Anscrombe, J.-C. & Ducrot, O. (1983) *L'argumentation dans la langue,* Bruxelles : Mardaga.

[2] Baker, M.J. (in press). Argumentation et co-construction des connaissances. *Interaction et Cognitions*. Paris : L'Harmattan.

[3] Baker, M.J. & Lund, K. (1996) Flexibly structuring the interaction in a CSCL environment. In *Proceedings of European Conference on Artificial Intelligence and Education, October 1996, Lisbon, Portugal*. Brna P., Paiva A. & Self J. (Eds) : pp. 401-407

[4] Baker, M.J. (1996). Argumentation and Cognitive Change in Collaborative Problem-Solving Dialogues. COAST Research report CR-13/96, COAST Research Team, ENSL, Lyon.

[5] Barth, E.M. & Krabbe, E.C.W. (1982). *From Axiom to Dialogue : A philosophical study of logics and argumentation*. Berlin : Walter de Gruyter.
[6] Bunt, H.C., (1989). Information Dialogue as Communicative Action in relation to Partner Modelling and Information Processing. In *The Structure of Multimodal Dialogue*. M.M. Taylor, F. Néel & D.G. Bouwhuis
[7] Bunt, H.C., (1994) Dialogue Controle Functions and Interaction Design. In *Dialogue and Instruction : Modeling Interaction in Intellignet Tutoring Systems*. Beun R.-J., Baker M. & Reiner M. (Eds), NATO ASI Series F, Vol. 142. Berlin : Springer Verlag : pp. 197-214
[8] Cohen, P.R., Morgan, J.& Pollack, M.E. (eds) (1990) Intentions in Communication. Cambridge, Mass: MIT Press.
[9] deKleer, J. (1986). An assumption-based TMS ; "Extending the ATMS" & "Problem-solving with the ATMS", *Artificial Intelligence* 28(2), 127-162, 163-196, 197-225.
[10] Dennett, D.C. (1981) *Brainstroms : Philosophical Essays on Mind and Psychology*. Brighton (UK) : Harvester Press.
[11] Devi, R., Tiberghien, A., Baker, M. & Brna, P. (1996). Modelling students' construction of energy models in physics. *Instructional Science* 24, 259-293.
[12] Doyle, J., (1979). A Truth-Maintenance System. Artificial Intelligence 12(3) : pp. 231-272.
[13] Eemeren, F. (van) & Grootendorst, R. (1984) *Speech Acts in Argumentative Discussions*. Dordrecht-Holland : Foris Publications.
[14] Galliers, J.R. (1992) Autonomous belief revision and communication. In *Belief Revision*. (Ed.) Gardenförs, P. Cambridge Tracts in Theoretical Computer Science 29, Cambridge University Press
[15] Gardenförs, P. (1992) Belief Revision : An Introduction. In *Belief Revision*. Cambridge Tracts in Theoretical Computer Science 29, Cambridge University Press : pp. 1-28.
[16] Grosz, B.J., (1981). Focusing and description in natural language dialogues. In *Elements of discourse understanding*. Joshi A.K., Webber B.L. & Sag I.A. (eds), Cambridge U.P. : pp. 84-105.
[17] Huang, X. (1992). Inconsistent Beliefs, Attention, and Student Modelling. *Journal of Artificial Intelligence in Education* 3(4), 417-428.
[18] Mavarech, Z.R. & Light, P.H. (1992) Peer-based interaction at the computer : looking backward, looking forward. *Learning and Instruction,* 2, pp. 275-280.
[19] Plantin, C. (1990). *Essais sur l'Argumentation*. Paris : Editions Kimé.
[20] Pontecorvo, C. (ed.) (1993). *Cognition and Instruction*, 11, 3 & 4. Special Issue : Discourse and Shared Reasoning
[21] Quignard, M. (1995). *Argumentation et révision de croyances : modélisation de dialogues entre élèves en situation de résolution de problèmes*. DEA en Sciences Cognitives de Grenoble. (CNRS-COAST, ENS de Lyon).
[22] Raccah, P.-Y. (1991) Argumentation in lexicon. In *Sprache - Kommunikation - Informatik, Akten des 26. Linguistischen Kollokiums, Poznan*. Tübingen : Max Niemeyer Verlag.
[23] Resnick, L.B., Levine, J.M. & Teasley, S.D. (1991). *Perspectives on Socially Shared Cognition*. Washington DC : American Psychological Association.
[24] Searle, J. (1969). Speech Acts. Cambridge : Cambridge University Press.
[25] Toulmin, S. (1958) The Use of Argument. Cambridge : Cambridge University Press.
[26] Vanderveken, D. (1990). *Meaning and Speech Acts. Volume 1 : Principles of Language Use*. Cambridge : Cambridge University Press.
[27] Walton, D.N. (1989) *Informal logic : a handbook for critical argumentation*. Cambridge : Cambridge University Press.