KNOWLEDGE ACQUISITION MODELING THROUGH DIALOGUE BETWEEN COGNITIVE AGENTS

Keywords: Intelligent agents dialogue, tutoring systems , logical reasoning.

Abstract: The work described in this paper tackles learning and communication between cognitive artificial agents. Focus is on dialogue as the only way for agents to acquire knowledge, as it often happens in natural situations. Since this restriction has scarcely been studied as such in artificial intelligence (AI), until now, this research aims at providing a dialogue model devoted to knowledge acquisition. It allows two agents, in a 'teacher' - 'student' relationship, to exchange information with a learning incentive (on behalf of the 'student'). The article first defines the nature of the addressed agents, the types of relation they maintain, and the structure and contents of their knowledge base. It continues by describing the different aims of learning, their realization and the solutions provided for problems encountered by agents. A general architecture is then established and a comment on an a part of the theory implementation is given. Conclusion is about the achievements carried out and the potential improvement of this work.

1 Introduction

This research aims at defining a set of algorithms for knowledge acquisition through dialogue between artificial cognitive agents. By *cognitive agents* we mean entities possessing knowledge as well as acquisition and derivation modes. In other words, they are able to capture knowledge externally, and to process and modify it through reasoning. Moreover, agents are characterized by one or several goals. As an artificial intelligence (AI) entity, each agent owns a knowledge base and attempts to make it evolve either by environment observation (reactivity) or by derivation modes (inductive or deductive reasoning). However, human beings as natural cognitive agents favor dialogue as another mean for knowledge revision. This mean leads each agent to consider any fellow agent as a knowledge source. The source is "triggered" through questioning, and information is acquired, from the answer, as an external possible hypothesis. This naturally is the anchor of a revision based process, where the hypothesis is subject to confrontation with the inner knowledge source of the requiring agent. Thus, it drives the latter to proceed to derivation (by reasoning). The feedback commonly observed in natural dialogue is analyzed as such; the knowledge source could be addressed in order to test wether the acquisition process has succeeded. It is a sort of confirmation process. In a nutshell, this is what happens in *tutored learning*.

As a technique, tutored learning has been a fruitful resource of inspiration to researchers in both humman-machine dialogue systems, and intelligent tutoring systems. This is briefly outsketched in next section, as the related literature upon which this works partly relies. The other underlying trend is that of belief and knowledge revision, and the inner mechanisms related to knowledge update. This type of research has been thoroughly discussed in knowledge representation (KR) and AI literature.

The particularity of this paper is to bridge both fields within a single research, assuming that human behavior does that so commonly that it is generally overlooked. People tend to focus on either communication habillities or reasoning capabilities, forgetting that language is a medium for both activities: communication and thought. If dialogue systems are nowadays emerging as possible scripts for knowledge revision, most papers detail game dialogues and not language-based learning activities. Thus, we believe that learning through a language-based or language-inspired dialogue is an interesting track within the AI field of cognitive agents cooperation as well as in KR processes involved with revision and explanation.

To simulate learning through dialogue, many examples could have been taken. In order to stress upon the revision process, and not be lost in the meanders of the communication process, we have chosen a socratic dialogue ('teacher' - 'student') where knowledge is presented exclusively by means of a question-answer mode of interaction. This is a simple way of checking both acceptance and revision. Knowledge contained in an answer is as an "external" fact presented to the requiring agent, and thus actions a revision process within its knowledge base. The 'student' agent owns belief revision mechanisms and all axioms leading to formal reasoning.

Since we simplified communication pragmatics to highlight knowledge revision, we also wished not to overload dialogue with language intrinsic ambiguity (i.e. pure "natural language" problems). Therefore we designed a "skeleton protocol": message data will be exchanged in first-order logic. This has been decided only to be able to focus on dialogue, seen as a knowledge-related process, particular features; we attempt, the best we can, to make the dialogue situation as close to a natural human-human dialogue between a 'teacher' and a 'student', as possible. We assume that agents use a common formalism concerning terms, predicates and functions. Nevertheless, the 'student' agent may not have predicates (or functions) given by the 'teacher' agent and so can question it on this subject before revising its base. In this paper, we try to show how dialogue initiates reasoning, which leads to an increase as well as a revision of the 'student' knowledge base according to hypothesises we simulate and revision mechanisms we define.

After offering an overview of related literature in next section, we define, in section three, the theoretical framework in which we have placed our model. In section four we sketch the general achitecture of the system and the first experiments lead to test the model. Last we discuss the results obtained as well as the model features, hinting at further developments to achieve further goals.

2 Dialogue and Learning: A Brief Description of Related Literature

Several papers deal with human learning via dialogue (Draper and Anderson, 1991). Those related to computer devices, within intelligent tutoring systems literature, usually rely on human-machine dialogue models (Baker, 1994; Cook, 2000), mostly about cooperation and negotiation as striking features in knowledge acquisition. However, for artificial agents only, the very few papers about communication as an acquisition mode are in the framework of noncognitive environment like robots (Asoh et al., 1996) or noncognitive software agents. It seems that, in artificial systems, learning is often realized without dialogue.

2.1 Learning without Dialogue.

There are many kind of learning methods for symbolic agents like reinforcement learning, supervised learning (sometimes using communication as in (Mataric, 1997)), without speaking about neural networks models that are very far from our domain. This type of learning prepares agents for typical situations, whereas, a natural situation in which dialogue influences knowledge acquisition, has a great chance to be unique and not very predictable (Ravenscroft and Pilkington, 2000).

2.2 Dialogue Models.

Most dialogue models in computer science (namely in AI) are based on *intentions* (Allen and Perrault, 1980; Cohen and Levesque, 1992), rely on the Speech Act Theory (Austin, 1975; Searle, 1969), to define dialogue as a succession of planned communicative actions modifying implicated agents' mental state, thus emphasizing the importance of *plans* (Pollack, 1998). When agents are in a knowledge acquisition or transfer situation, they have goals: teach or learn a set of knowledge chunks. However, they do not have predetermined plans: they react *step by step*, according to the interlocutor's answer. This is why an *opportunistic* model of linguistic actions is better than a planning model. A new trend about dialogue situations as learning situations has lately appeared in literature, but is mostly confined to game strategies (Amgoud and Prade, 2003). Therefore, dialogue is more an interaction process than a real "logos" (the greek root in the word dialogue, i.e. related to discourse and language) process, and appears as an inappropriated term to designate the described interaction.

2.3 Negotiation Models Related to Reasoning and Revision.

Another approach, closer in spirit to what has been achieved by (Baker, 1994) about negotiation as a process of increasing or revising knowledge, is trying to emerge in the AI and KR literature. Researchers such as (Parsons et al., 1998; Wooldridge and Parsons, 2000) have been considering negotiation as a qualitative process in belief revision. Moreover, others such as (Zhang et al., 2004) defend the capabilities of negotiation, as a mutual belief revision process. This clearly comes close to assumptions rising from human behavior, and akin to those we try to develop in this paper. However, the situation described, that of a consensus achieved as an equilibrium after an initial set of demands and offers, is not exactly the type of learning-teaching situation we aim at. In a negotiation, both parties have an interest. In a tutored learning situation, the teacher plays an altruistic role, and has nothing to gain from a cognitive or an economic point of view, at least on a first level appreciation. Nevertheless, this "altruistic" attitude, related to cooperation as a whole, has proven to be an optimized process of knowledge spreading in human communities, and thus benefited both seeker and provider, in a rather complex feedback process. Therefore, it is interesting to transpose it to artificial agents communities, knowing that a 'teacher' agent might become in its turn, a 'student' in a domain where it lacks knowledge.

2.4 Some Features of Tutored Learning Situations.

2.4.1 Dialogue Peculiarities.

Even reduced to socratic dialogue, a tutored learning situation implies a *finalized* dialogue (aiming at carrying out a task) as well as *secondary* exchanges (precision, explanation, confirmation and reformulation requests can take place to validate a question or an answer). Therefore, speech acts appear as crucial elements in the interaction process. We have chosen to assign functional roles (FR) to speech acts since this method, described in (Sabah et al., 1998), helps unpredictable situations modeling, whereas the Speech Act Theory (i) assigns multiple illocutionary values to the same speech act thus maintaining ambiguity ; (ii) is more efficient *a posteriori* than *a priori* ; (iii) relies on verbs interpretation by human-based pragmatics, and therefore is difficult to transform into a reliable computational model. The FR theory is closer to an adaptive computational model since it tries to compute an exchange as an adjusment between locutors mental states. We have adapted this method, originally designed for human-machine dialogue, to artificial agents.

2.4.2 Reasoning

Reasoning, from a learning point of view, is a knowledge derivation mode, included in agent functionalities, or offered by the 'teacher' agent. Reasoning modifies the recipient agent state, through a set of reasoning steps. *Learning* is considered as the result of a reasoning procedure over new facts or predicates, that ends up in engulfing them in the agent knowledge base. Thus, inspired from human behavior, the described model acknowledges for three types of reasoning: deduction, induction and abduction. Currently, our system uses inductive and deductive mechanisms. Abduction is not investigated as such, since we consider *dialogue as an abductive bootstrap technique* which, by presenting new knowledge, enables knowledge addition or retraction and therefore leads to knowledge revision (Josephson and Josephson, 1994; Pagnucco, 1996).

2.5 Peculiarity of our Approach: Simplification Due to Artificial Agents.

Although our system is heavily inspired from dialogue between humans and from human-machine dialogue systems, it differs from them with respect to the following items:

- Natural language is not used as such and a formal-based language is prefered, in the tradition of languages such as KIF, that are thoroughly employed in artificial agents communication. These formal languages prevent problems that rise from the ambiguity intrinsic to natural language.
- When one of the agents is human, then his/her knowledge is opaque not only to his/her interlocutor (here, the system) but also to the designer of the system. Therefore, the designer must build, in his system, a series of "guessing" strategies, that do not necessarily fathom the interlocutor's state of mind, and might lead to failure in dialogue. Whereas, when both agents are artificial, they are both transparent to the designer, if not to each other. Thus, the designer embeds, in both, tools for communication that are adapted to their knowledge level. The designer might check, at any moment, the state variables of both agents, a thing he or she cannot do with a human.

These two restrictions tend to simplify the problem, and more, to stick to the real core of the task, i.e., controlling acquisition through interaction.

3 The Theoretical Framework

3.1 Agents Frame

Our environment focuses on a situation where two cognitive artificial agents are present, and their sole interaction is through dialogue. During this relationship, an agent will play the role of a 'teacher' and the other will momentarily act as a 'student'. We assume they will keep this status during the dialogue session. Nevertheless, role assignation is temporary because it depends on the task to achieve and on each agent's skills. The 'teacher' agent must have the required skill to teach to the 'student' agent, i.e., *to offer unknown and true knowledge, necessary for the 'student' to perform a given task*. Conventionally, 'student' and 'teacher' terms will be used to refer, respectively, to the agents acting as such. The 'teacher' aims at 'freely' offering a set of predetermined knowledge to the 'student'. This, naturally subsumes that agents cooperate. Thereby, no erroneous data will be exchanged and agents will attempt, using all means they can, to satisfy their interlocutor's expectancy. Nevertheless, as in a natural situation, the 'student' could be not really self-motivated and by this way making harder the 'teacher's task. For instance, the 'student' could provide indefinite data to the 'teacher'.

3.2 Knowledge Base Properties

3.2.1 First-Order Logic

Each agent owns a knowledge base (KB), structured in first-order logic, with functions, so the knowledge unit is a **formula**.

First-order logic has been prefered to propositional logic,or description logic, because of the expressive power of predicates, and the existence of functions was necessary to the nature of our first test corpus, which was in physics (teaching laws of mechanics). However, functions have been abandoned because of intrinsic difficulties, and we changed the corpus into a basic science corpus. Since quantifiers not being tested, the traps related to them in first-order logic have been avoided. So, first-order logic here mostly appears because FR modeling, introducing particular predicates (functional roles), has driven us to use this level of expressivity.

3.2.2 Basic Assumptions about True and False

The 'student' can make mistakes, i.e., possess *wrong* knowledge. From an axiomatic point of view, if an agent acts as a 'teacher' in relation to a given knowledge set, then the 'student' will consider as true every item provided by the 'teacher'.

3.2.3 Conventions

Each KB is manually initiated, however, its update will be automatic, thanks to 'learning' and reasoning abilities. In order to simplify modeling, we only use formulas such as (P), $(P \rightarrow Q)$ and $(P \leftrightarrow Q)$. (P) and (Q) are predicates conjunctions (or their negation) of type $(p(A))$ or $(p(X))$ (or $(not(p(A)))$ or $(not(p(X)))$), where $A = \{a_1, a_2, \ldots, a_n\}$ is a set of terms and $X = \{x_1, x_2, \ldots, x_n\}$ a set of variables. For simplification sake, we note P and Q such predicates conjunctions. Universal quantification is implicit for each formula having at least one variable. We consider that, to initiate learning (from the 'student' position), the 'teacher' has to rely on the 'student's previous knowledge. This constraint imitates humans' learning methods. Therefore, before performing a tutored learning dialogue, agents must have a part of their knowledge identical (called *basic common knowledge*). The 'teacher' will be able to teach new knowledge by using the 'student''s already known one. However, our agents do not 'physically' share any knowledge (their KBs are independent).

3.2.4 Connexity as a KB Fundamental Property.

During learning, each agent will attempt to make its KB as "connex" as possible.

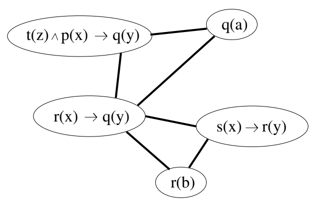
Definition 1. A KB is *connex* when its associated graph is connex. A graph G_Γ is associated to a KB Γ as such:
Each formula is a node. An edge is created between each couple of formulas having the same premise or the same conclusion or when the premise of one equals the conclusion of the other. For the moment, variables and terms are not taken into account in premise or conclusion comparison [1]. Thus, in a connex KB, every knowledge element is linked to every other, the path between them being more or less long. As the dialogic situation must be as close as possible to a natural situation, **agents' KBs are not totally connex**: a human agent can often, but not always, link two items of knowledge, haphazardly taken.

Examples:
A connex KB: $\Gamma_1 = \{t(z) \wedge p(x) \rightarrow q(y), r(x) \rightarrow q(y), s(x) \rightarrow r(y), q(a), r(b)\}$
A non connex KB: $\Gamma_2 = \{t(z) \wedge p(x) \rightarrow q(y), r(x) \rightarrow q(y), s(x) \rightarrow u(y), q(a), u(b)\}$
Figures 1(a) and 1(b) respectively represent graphs associated to Γ_1 and Γ_2.

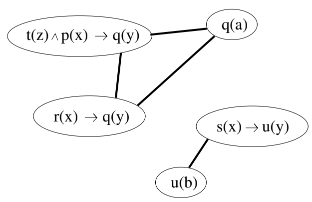
Definition 2. A connex component (or just component) is a connex subset of formulas in a KB.

Theorem 1. Let A, B and C be three connex formulas sets. If $A \cup B$ and $B \cup C$ are connex then $A \cup B \cup C$ is connex.

[1] An abductive reasoning mechanism is contemplated as a possible mean to compare a constant fact $q(a)$ with a predicate with a variable $q(y)$. We only consider the result of a succeeding abduction.



(a) Γ_1 Associated Graph



(b) Γ_2 Associated Graph

Figure 1: KB Associated Graphs

Proof. Let us assume that $A \cup B$ and $B \cup C$ is connex and G_A, G_B and G_C are graphs respectively associated to A, B and C. According to definition 1: $A \cup B$ connex is equivalent to $G_A \cup G_B$ connex. Also, $B \cup C$ connex is equivalent to $G_B \cup G_C$ connex. And according to connex graph properties: $G_A \cup G_B$ connex and $G_B \cup G_C$ connex implies $G_A \cup G_B \cup G_C$ connex. So $A \cup B \cup C$ is connex.

The agent situation we envisage is such that agents will not attempt to increase the number of their connex components. However, there will be some cases where the 'student' will be forced to do so. Fortunately, in some other cases, learning new knowledge may link two connex components into a new larger one, decreasing the components total number (according to theorem 1).

3.3 Dialogue: Using Functional Roles (FR)

A dialogue session is the image of a *lesson*. A lesson is performed either because the presumed 'student' has been asking for an important piece of information (not limited to a simple yes-or-no question or 'where do I find something...' questions), or because the 'teacher' finds him/herself in a situation where he/she has to transmit his/her knowledge. Transposed to artificial agents situation, both cases are available. In those, the assigned 'teacher' must know what knowledge to teach to the 'student': therefore a lesson has to be planified. It is then composed of several elements, each of them contained, in our framework, in a logic formula. In our model for artificial agents, the teaching agent provides each formula to the 'student'. However, before that, the teacher waits for the 'student''s understanding (or misunderstanding) message of the last formula. If the 'student' doesn't understand or is not at ease, it can just inform its interlocutor of the misunderstanding or, requests a particular information bit.

The FR theory, that models exchanges in this dialogue, provides a role attachment to each utterance. Both agents, when receiving a message, know its role and come up with an adequate answer. In our framework, knowledge about roles is possible, because opacity, natural to human situation, is absent. A particular clause is expressed and conveyed to the interlocutor. This clause denotes the *dialogic role* of the formula to be transmitted. At the same time, this clause provides an indication about the *formula evaluation.* We assign the type *knowledge* to universal or existential general logical formulas. This type is neutral from the evaluation point of view, that is, knowledge might either be true or false, as a formula. Whereas, we assign the type *information* to constant-uttering answer (or question): i.e., which value is 'true', 'false', or 'unknown'. Knowledge and Information, when exchanged, might be stamped as 'inevitably true' or 'possibly false' or, and this takes us out of the first-order logic language, 'unknown', i.e. not evaluable to both locutors. What makes the evaluation possible, is a complex result of three components: the dialogue situation, the agent role, and the conveyed functional role.

To illustrate this, let us detail the main FR types used in the our tutored learning dialogue.

1. give-knowledge. Used to teach a knowledge and introduce an exchange.
   Argument's general form: $(P \rightarrow Q)$ or $(P \leftrightarrow Q)$.
   Example:
   $give - knowledge(cat(x) \rightarrow mortal(x))$:
   "Cats are mortal."
   When uttered by the teacher, a give-knowledge argument has to be evaluated as true by the 'student' (see 'FR Interpretation Axioms' section).

2. **askfor/give-information** **(**boolean evaluation case):
   askfor-information.
   Argument's general form: either (P), or $(P \rightarrow Q)$, or $(P \leftrightarrow Q)$, with or without variable.
   Examples:
   $askfor - information(cat(Folley))$:
   "Is Folley a cat?"
   $askfor - information(cat(x) \rightarrow mortal(x))$:
   "Are cats mortal ?"
   When conveyed to the interlocutor this function

bids him/her to answer. In a very cooperative framework as the one we need to install between artificial agents, the teacher agent is compelled to answer with a give-information utterance, which gives the interpretation of the formula $cat(x) \rightarrow mortal(x)$ according to the teacher.
give-information.
Argument's general form: either $(True)$, or $(False)$, or $(Unknown)$.
Example:
$give - information(true)$: "Yes."

3. give-explanation (predicate case).
Argument's general form: either $(p(x) \leftrightarrow P)$, or $(Unknown)$.
Example:
$give - explanation(cat(x) \leftrightarrow$
$(animal(x)\, and\, pet(x)\,))$: "A cat is a pet animal."
A give-explanation formula is provided as an answer to a question of the type : "what is X ?" or "Why/how is X related to Y ?" . In other words, when a student has no value to a predicate, or cannot relate it to another, the student asks the teacher to provide the links between the unknown element and other possibly known predicates. The situation can be triggered by an askfor-explanation clause taking as an argument the unknown predicate or formula. The student expects the teacher to provide a formula in which known predicates are related to the unknown one. By this process, the student might augment its KB while increasing its connexity. A "Why/how is X related to Y ?" question is about connexity, and a "what is X ?" question increases KB elements through KB connexity.

4. say-(dis)satisfaction: tells the other agent that the last provided data has (has not) been well understood. This is a meta-evaluation clause, since it has no direct argument, but leads to the evaluation of the interaction (and not of the formula). say-(dis)satisfaction is particular to dialogue modeling (most linguistic and pyscholinguistic theories account for interaction evaluation), and is very useful in checking dialogue feedback.

There are some FR we do not detail here ($askfor - knowledge$, $askfor/give - example$, $askfor/give - precision$, $askfor/give - reformulation$) likewise some specific uses like the type $askfor/give - information$ in the case of an evaluation by a function. So FR are dialogic clauses leading to the interpretation of exchanged formulas. A functional role of the "$ask - for$" kind implies one or a series of clauses of the "$give$" type, with the possibility of using another "$ask - for$" type if there is a misunderstanding. This case will bring about a clause without argument: "$say - dissatisfaction$". Only "$ask - for$" type roles will lead to interpretative axioms. Other ones are behavioral startings.

3.4 Tutored Learning

3.4.1 Axioms

Our reasoning system is hypothetical-deductive, so it allows belief revision and dialogue is the mean by which this revision is performed. Two groups of axioms are defined: fundamental axioms of the system and those corresponding to the FR interpretation in the system. Each knowledge chunk of each agent is seen as an assumption.

Fundamental axioms. Our system revision axioms include the hypothetisation axiom, hypothesis addition and retraction, implication addition, implication retraction or modus ponens and the *reductio ad absurdum rule.*

Let Γ be the 'student's assumptions finite set. The knowlede acquisition mode is represented by addition or substraction "deducted" by the fraction bar symbol. Generalised to the ensemblist implication, this symbol means that in premise (numerator) there is an ensemblist implication and in conclusion (denominator) there is another ensemblist implication, deductible from the previous one, whose objective is to make the knowledge set evolving. System revision axioms (taken from (Manna, 1974)):

- The hypothetisation axiom:
$$\Gamma, A \Rightarrow A \qquad (1)$$
if the agent knows an assumption A, then it can deduce it from its own system.
- The assumption addition:
$$\frac{\Gamma \Rightarrow B}{\Gamma, A \Rightarrow B} \qquad (2)$$
if the agent can deduce B, then it will be able to deduce it from any superset of its own system.
- The assumption retraction:
$$\frac{\Gamma, A \Rightarrow B \ and\ \Gamma, \neg A \Rightarrow B}{\Gamma \Rightarrow B} \qquad (3)$$
if the validity of an assumption A of the system doesn't influe on the assumptions B deductible from this system, A must be removed. To allow A influing on B, the assumptions set Γ and A must be (but this is not sufficient) connex.
- The implication addition:
$$\frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \supset B} \qquad (4)$$
if B is deductible from an assumptions set and from an assumption A, then the rule $A \supset B$ is deductible

from the system. The connexity notion is present here as we need the fact that B is deductible from A to be able to add the rule $A \supset B$, which means that a path between A and B must be present.

- The implication retraction or modus ponens :

$$\frac{\Gamma \Rightarrow A \ and \ \Gamma \Rightarrow A \supset B}{\Gamma \Rightarrow B} \tag{5}$$

if A is deductible from the system and if A is the premise of a system deductible rule, then the conclusion of this rule is direclty deductible from the system.

- The rule called *reductio ad absurdum* :

$$\frac{\Gamma, A \Rightarrow B \ and \ \Gamma, A \Rightarrow \neg B}{\Gamma \Rightarrow \neg A} \tag{6}$$

if B can be deductible AND falsifiable from the system including A, then A is falsified. This axiom introduce the conflict management dealed lately.

FR interpretation axioms. Interpretation axioms are not in the first order since they introduce clauses and multiple values (like the "unknown" one). Our syntax will be in the first order, but the interpretation is not monotonous.

- $give - knowledge(A) \Rightarrow A \vdash T$;
any knowledge supplied by the teacher is considered as true.
- $give - information(A) \equiv A \epsilon [T, F, U]$;
any supplied information is a formula interpretable in a multi-valued space.
- $give - explanation(A) \equiv$
$(give - information(P), A \leftrightarrow P)$;
any explanation consists in supplying a right formula, equivalent to the formula A that has to be explained.

With T for True, F for False and U for Unknown.

3.4.2 Tutored Learning Situations

Learning can have several goals like enriching the KB with new data, increasing the KB connexity, widening the predicates base, understanding why some formulas imply others. We mainly focus on the first one because of its importance. In order to learn, the 'student' must first understand received data. By *understanding*, we mean "not increasing the KB components number": the 'student' understands a data that is linked to at least one component of its KB. By definition, we consider that a 'student' agent *knows* a formula if it owns it. If the taught data is not linked to any component, the 'student' have to inform the 'teacher' of its misunderstanding.

3.5 Dialogue Strategy

There are several dialogue strategies depending on the goals chosen by the learner. In this paper, being limited in scope because of the experiment needs, we consider one goal: enriching the KB with new data while maintaining connexity as best as possible. This is the commonest and the most natural goal, the one that seems to appear most frequently. Thus we suggest an appropriate common strategy: solving a misunderstanding problem by choosing adequate questions and answers. We have adopted a technique inspired from the socratic teaching method.
For each predicate p_i to be taught, the 'teacher' knows another one p_j linked with p_i by an implication or an equivalence F. Therefore, to ensure that the 'student' understands p_i thanks to p_j, he/she (or it, in the case of an artificial teacher) will have to ask the 'student' if the latter knows p_j. If the 'student' knows it, then the 'teacher' only has to give F to the 'student'. Otherwise, the 'teacher' will find another formula that explains p_j and so on.

Here is an example of such misunderstanding and its solving:

— 'teacher' - $give - knowledge(human(x) \rightarrow mortal(x))$
; "Humans are mortal."
; the 'student' doesn't know these two predicates, nevertheless it knows the following predicates $(animal(x))$, $(intelligent(x))$ and $(can - speak(x))$.
— 'student' - $say - dissatisfaction()$; "I don't understand."
; then the 'teacher' tries to explain to the 'student' the formula's premise: what is a human.
— 'teacher' - $ask - for - information(animal(x))$; "Do you know what is an animal ?"
— 'student' - $give - information(True)$; "Yes."
— 'teacher' - $ask - for - information(intelligent(x))$; "Do you know what is "being intelligent ?"
— 'student' - $give - information(True)$; "Yes."
— 'teacher' - $ask - for - information(can - speak(x))$; "Do you know what is "being able to speak ?"
— 'student' - $give - information(True)$; "Yes."
— 'teacher' - $give - explanation((animal(x)$ $and\, intelligent(x)\, and\, can - speak(x)) \leftrightarrow human(x))$
; "A human is an intelligent animal that can speak."
— 'student' - $say - satisfaction()$; "I understand."

As the 'student' has understood what is a human, it can know learn the knowledge $(human(x) \rightarrow mortal(x))$ by just adding it to its KB, the connexity is preserved. Once data is understood, the 'student' may realize that some bits are contradictory with its KB, leading to a *conflict*.

Conflict Management. We have studied several types of conflict, those related to implications as well as those related to facts. In this paper, we will only present the first one, which typically takes place when the 'student' has a formula $(P \rightarrow Q)$ and attempts to learn a formula $(P \rightarrow not(Q))$. The solution, for the 'student', is removing $(P \rightarrow Q)$ from its KB and adding $(P \rightarrow not(Q))$. It acts so because this is **'teacher''s knowledge** (thus true) and so it gets the upper hand on the 'student' one (first axiom). However, the conflict could be hidden if the 'student' has the next formulas: $(P_1 \rightarrow P_2)$, $(P_2 \rightarrow P_3)$, ..., $(P_{n-1} \rightarrow P_n)$ and attempts to learn $(P_1 \rightarrow not(P_n))$: the 'student' has an equivalent to the formula $(P_1 \rightarrow P_n)$. Instead of using a baseline solution consisting in removing all the series of implications, we opted for a more flexible one which attempts to look for a wrong implication and only remove this one. Indeed, removing one implication is sufficient to solve the conflict. The 'student' will then attempt to validate each implication with the 'teacher' through an "$askfor - information$" request. As soon as a wrong implication is found, the 'student' removes it and safely adds the new one. However, if none of the implications is neither validated nor rejected by the 'teacher', the 'student' will be forced to remove all the series before adding the new one to be sure to end up the conflict. We have studied other implication conflict types that are even less easily detectable, but we have not a sufficient space to detail them.

4 System Architecture and Implementation

The theoretical approach of section 3 has been specified and partially implemented. The specification is general, the implementation contains some of its elements presented in section 4.2.

4.1 Architecture

The figure 2 displays the main architecture elements of our tutored learning system. It is composed of five main structures: the 'teacher', the 'student', the FR, the strategies and the 'World'. 'Teacher' and 'Student' are agents.
The FR are a shared knowledge base about dialogic clauses, to which both agents have access. The strategies are meta-rules of behavior that help both teacher and student to achieve satisfaction (positive meta-evaluation) and thus to end the dialogue with success (an ending with failure is possible, since a repeated negative meta-evaluation might appear. Then, the 'teacher' ends up the dialogue, because not correct evolution of the situation is observed). The 'World' is a sharable knowledge base, a pool of predicates available to agents.
Each agent has a KB, a model of itself and of its interlocutor. a model of the interlocutor is what the agent knows that its interlocutor knows. The 'teacher' agent mostly checks its interlocutor model and updates it, by asking the student questions, when it (the teacher) needs to explain something to the student. The model the student has of its teacher is that all what the teacher says is true, and thus a $give - knowledge(P)$ clause is equivalent to $P\ is\ true$ (first axiom). Naturally, each agent when shifting from a role to another, in a different situation, modifies its interlocutor's model according to its present role. It can freely

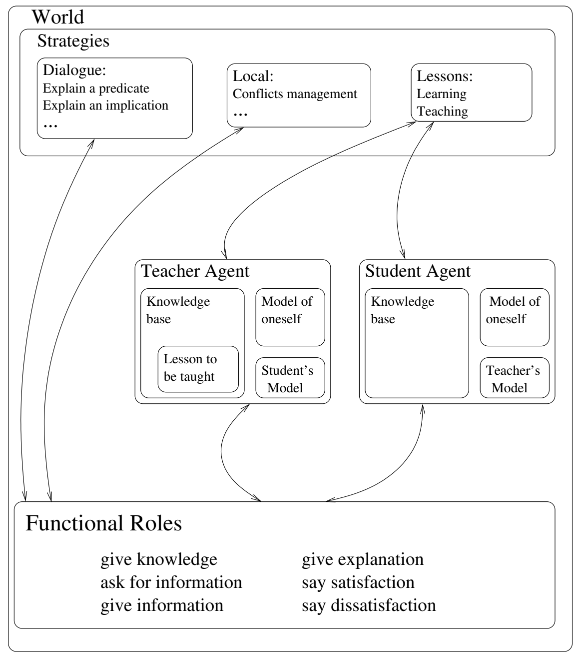


Figure 2: The tutored learning system through dialogue between artificial cognitive agents

update them in order to make them evolve. It has an access to strategies, for learning and teaching lessons and to all the FR rules (seen in section 3.3).

4.2 Implementation

We have implemented a Java[2] program to test conflict solving. This program is a basic prototype aiming at getting experimental results of a part of our theory. Each agent is an instance of the Thread Java class and has a name, a knowledge base (KB) and a pointer to

- a World class (the environment at which it belongs),
- its possible teacher, student and interlocutor agents,
- the strategies class,
- the fonctional roles (FR) class.

[2] http://java.sun.com/

The KB is made of two types of simple objects: facts (a predicate name and a term) and implications (two predicate names, two variables and a direction). An exchanged message is one entry of the KB (a fact or an implication) plus a FR type. Strategies are methods defining sequential actions to perform in order to accomplish the specific asked strategy. They can be used directly by the agent (for lesson teaching for example) or they can be called by their FR component (for predicate explanation implied by a "say-dissatisfaction" for example). The FR component is a switch that routes the agent message to the good FR method according to its FR type. Each FR method uses the adequate strategies to satisfy agents.

We have tested our program with different KB and lessons. Here is an example dealing about the implication conflict. The basic common knowledge ($\Gamma(agent)$) of our two agents are:
For the 'teacher':
$\Gamma(T) = \{human(x) \rightarrow mortal(x),$
$mortal(x) \rightarrow live(x),$
$can - be - killed(x) \rightarrow live(x),$
$live(x) \rightarrow can - repoduce(x)\}$
For the 'student':
$\Gamma(S) = \{human(x) \rightarrow can - be - killed(x),$
$can - be - killed(x) \rightarrow live(x),$
$not(can - reproduce(x)) : -live(x).\}$
The lesson is:
$\delta(T) = \{human(x) \rightarrow mortal(x),$
$human(x) \rightarrow live(x),$
$human(x) \rightarrow can - reproduce(x)\}$

A natural constraint we have set is: $\delta(T) \subset \Gamma(T)$. Which means the 'teacher' can only teach data it already knows. Then the 'teacher' will teach each formula to the 'student', waiting a satisfaction, disatisfaction or some questions between each dialogue utterance, here is this dialogue:

— 'teacher' - $give - knowledge(human(x) \rightarrow mortal(x))$
; this formula is learned easily by the 'student'.
— 'student' - $say - satisfaction()$
— 'teacher' - $give - knowledge(human(x) \rightarrow live(x))$
; the 'student' already knows this formula (deducted from its KB), then the 'student' understand but doesn't learn this formula.
— 'student' - $say - satisfaction()$
— 'teacher' - $give - knowledge(human(x) \rightarrow can - reproduce(x))$
; this new data is in contradiction with $\Gamma(S)$, then the 'student' will have to locate its wrong knowledge(s) among those which generate the problem, i.e. among $human(x) \rightarrow can - be - killed(x)$, $can - be - killed(x) \rightarrow live(x)$ et $live(x) \rightarrow not(can - reproduce(x))$
— 'student' - $ask - for - information(human(x) \rightarrow can - be - killed(x))$
— 'teacher' - $give - information(Unkown)$
; the 'teacher' doesn't know if a human can be killed, the 'student' conserve this formula yet.
— 'student' - $ask - for - information(can - be - killed(x) \rightarrow live(x))$
— 'teacher' - $give - information(True)$
— 'student' - $ask - for - information(live(x) \rightarrow not(can - reproduce(x))$
— 'teacher' - $give - information(False)$
; this data is in contradiction with the 'teacher's KB which will then assess it as false.
; thus the 'student' removes this data from its KB.
; its conflict solved, the 'student' can know learn the knowledge $(human(x) \rightarrow can - reproduce(x))$ safely.
— 'student' - $say - satisfaction()$
; it closes the dialogue by informing the 'teacher' of its satisfaction.

The new 'student's KB is then:
$\Gamma'(E) = \{human(x) \rightarrow can - be - killed(x),$
$can - be - killed(x) \rightarrow live(x),$
$human(x) \rightarrow animal(x),$
$human(x) \rightarrow mortal(x),$
$human(x) \rightarrow can - reproduce(x)\}$

Running the process shows that the 'student' has detected a conflict between its KB and a new data provided by the 'teacher'. It then asks the 'teacher' to validate some potentially conflictual knowledge and finally removes the wrong implication.

5 Conclusion

Our system allows artificial cognitive agents, in a tutored learning situation between a 'teacher' and a 'student', to acquire new knowledge through sole dialogue. This is possible, because language is both a communication and a knowledge representation mean. Each 'message' is then a knowledge chunk presented to a learner, not from its environment as a general framework, but from a particular other agent that possesses this knowledge. Since cooperation, and thus truthfullness, are assumed, then the knowledge chunk provided is evaluated as true. The learning agent has then to integrate it into its own KB. To do this, the agent uses dialogue as a feedback mean to maintain the coherence of its knowledge base. The study of such a constrained situation has led us to define a notion of *connexity* for a knowledge base (KB), allowing to assess the connection level between each element of knowledge of an agent and so to give it a new goal: increasing its KB connexity. As the dialogue situation in highly impredictable and may follow no previous plan, we have adopted the functional (FR) role theory to easily model dialogical exchanges. Agents use strategies to learn new knowledge and solve conflicts between external and internal data. (Angluin, 1988) tackles the problem of identifying an unknown subset of hypothesis among a set by queries to an oracle. Our work differs mainly in

- the communication mean: we use imbricated dialogues instead of queries;
- the learning's aim: our agents aim at learning new

formulas and increasing their KB connexity instead of identifying hypothesises.

This work is a first approach in learning by dialogue for cognitive artificial agents. Its aim is to define a set of requirements for an advanced communication. Some paths could be explored like enriching the KB content by new formula types or defining new dialogue strategies. Last, this type of learning could be used in complement with others that rely on interaction with environment, in order to multiply knowledge sources.

REFERENCES

Allen, J. and Perrault, R. (1980). Analyzing intention in utterances. *Artificial Intelligence*, 15(3):11–18.

Amgoud, L. and Prade, H. (2003). A possibilistic logic modeling of autonomous agents negotiation. In *EPIA*, pages 360–365.

Angluin, D. (1988). Queries and concept learning. *Machine Learning*, 2(4):319–342.

Asoh, H., Motomura, Y., Hara, I., Akaho, S., Hayamizu, S., and Matsui, T. (1996). Acquiring a probabilistic map with dialogue-based learning. *Hexmoor, H. and Meeden, L., editors, ROBOLEARN '96: An Intern. Workshop on Learning for Autonomous Robots*, pages 11–18.

Austin, J. (1975). *How to Do Things with Words*. ed. J. O. Urmson and Marina Sbis, Mass: Harvard University Press, Cambridge.

Baker, M. (1994). A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5(2):199–254.

Cohen, P. and Levesque, H. (1992). *Intentions in Communication*. Rational Interaction as the Basis for Communication. Bradford Books, MIT Press, seconde dition, chap. 12, Bradford.

Cook, J. (2000). Cooperative problem-seeking dialogues in learning. *G. Gauthier, C. Frasson and K. VanLehn (Eds.) Intelligent Tutoring Systems: 5th International Conference, ITS 2000 Montreal, Canada*, 1839:615–624.

Draper, S. and Anderson, A. (1991). The significance of dialogue in learning and observing learning. *Computers and Education*, 17(1):93–107.

Josephson, J. and Josephson, S. (1994). *Abductive Inference, Computation, Philosophy, Technology*. Cambridge University Press, New York.

Manna, Z. (1974). *Mathematical Theory of Computation*. International Student Edition. McGraw Hill Computer Science Series.

Mataric, M. (1997). Using communication to reduce locality in multi-robot learning. In *AAAI-97*, pages 643–648, Menlo Park. CA: AAAI Press.

Pagnucco, M. (1996). *The Role of Abductive Reasoning Within the Process of Belief Revision*. PhD thesis, University of Sydney.

Parsons, S., Sierra, C., and Jennings, N. (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292.

Pollack, M. E. (1998). Plan generation, plan management, and the design of computational agents. In *Proceedings of the 3rd International Conference on Multi-Agent Systems*, pages 643–648, Paris, France.

Ravenscroft, A. and Pilkington, R. (2000). Investigation by design: Developing dialogue models to support reasoning and conceptual change. *International Journal of Artificial Intelligence in Education*, 11(1):273–298.

Sabah, G., Ferret, O., Prince, V., Vilnat, A., Vosniadou, S., Dimitrakopoulo, A., and Papademetriou, E. (1998). What dialogue analysis can tell about teacher strategies related to representational change. In *Modelling Changes in Understanding: Case Studies in Physical Reasoning*, Oxford. Cambridge University Press.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge.

Wooldridge, M. and Parsons, S. (2000). Languages for negotiation. In *Proceedings of ECAI2000*, pages 393–400.

Zhang, D., Foo, N., Meyer, T., and Kwok, R. (2004). Negotiation as mutual belief revision. In *Proceedings of AAAI 2004*.