HAL
open science

Combining mental attitudes and public facts in an ACL semantics

Carole Adam, Benoit Gaudou

To cite this version:

Carole Adam, Benoit Gaudou. Combining mental attitudes and public facts in an ACL semantics. [Research Report] RR-LIG-042, LIG. 2013. hal-01471987

HAL Id: hal-01471987

https://hal.science/hal-01471987v1

Submitted on 20 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining mental attitudes and public facts in an ACL semantics

Carole Adam
University Joseph Fourier (UJF)
Laboratory of Informatics (LIG)
Grenoble, France
carole.adam@imag.fr

Benoit Gaudou
University of Toulouse
Institute of Research in Informatics
Toulouse, France
benoit.gaudou@univ-tlse1.fr

ABSTRACT

Most agent Communication Languages are no longer defined in terms of the agents' mental attitudes, but in terms of social commitments. However, such social approaches have two drawbacks. First, the notion of commitment does not have a clear and unambiguous characterization. Second, commitments are completely unrelated to the agents' reasoning. We remedy this situation by combining a BDI logic with a logic of what is publicly grounded between agents.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: General—*Philosophical foundations*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Modal Logic*

General Terms

Theory

Keywords

Agent communication languages, modal logic, grounding, commitments, BDI logic, speech act theory

1. INTRODUCTION

Agent Communication Languages (ACLs) are almost as old as the development of Multi-Agent Systems. They have mainly been developed to allow agents to communicate in a common and rich language in an anthropomorphic way, as opposed to ant communication via pheromones for example. ACL were designed to be used by cognitive agents able to reason and plan their communication. In particular, each speech act is defined by its pre- and post- conditions, exactly in the same way as any physical actions: pre-conditions identify situations in which a speech act can or cannot be uttered; and post-conditions describe what has been created by the performance of this speech act. These pre- and post-conditions have been encoded in terms of mental attitudes [9], commitments [11] or game rules [10]. Communication thus has to follow respectively a plan, a protocol or a game.

This way of handling communication is convenient for artificial agents because they can be constrained in their utterances, but not for humans who can say anything they want, at any time, even when the corresponding speech act would be unexecutable for an artificial agent. For example humans may lie, be irrelevant or use irony. **Actually, as long as a human speaker can physically utter a sentence, they might do so, regardless of any other conditions.** In this paper, we argue that due in particular to the increasing number of applications involving both humans and agents (ECA [14]), it is essential for the agent community to rethink the principles of ACL and develop a new way for agents to handle speech acts.

The main contribution of this article is to provide a new semantics for ACL based on the **central hypothesis that any speech act can be performed and that the agents should be able to manage any speech act they receive**. The main idea of our semantics is that there is no more absolute executability conditions nor any effects automatically deduced from an utterance. We are much more flexible by introducing condition-effect pairs: if and only if the condition is true then the corresponding effect can be deduced. We argue that this new way of thinking ACLs is not language-dependent: it can be used for languages based on various paradigms. In the sequel we illustrate this new principle of ACL with a particular formalism expressive enough to embed both mentalist and commitment-based paradigms.

The article is structured as follows: Section 2 details the main principles of our new semantics and introduces our condition-effect pairs; Section 3 gives an overview of our logical framework; Section 4 provides the axioms describing the roles of the condition-effect pairs, and instantiates them for some speech acts; Section 5 illustrates our semantics on an example human-companion dialogue.

2. PRINCIPLES OF OUR SEMANTICS

2.1 Preliminary remark about executability

One of the major characteristics of our semantics is that speech acts are always executable, under the only condition that they are physically executable by the agent, which we will presuppose to be true. Indeed a human speaker who is physically able to communicate *can* perform any utterance in any dialogical or institutional context; this utterance may be forbidden, irrelevant or inappropriate in the context, but as long as the speaker has the physical ability to perform it, he can choose to perform it, and actually do so. Then the hearer will handle this utterance differently depending on the dialogical (previous utterances) and institutional (rules of the dialogue) context.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), June, 4–8, 2012, Valencia, Spain.
Copyright © 2012, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

Similarly in a Multi-Agent System, we consider that an artificial agent who has the ability to communicate with the other agents should also *be able to* perform any speech act, even if forbidden, irrelevant or inappropriate. Indeed, since the agents in a MAS may not always be fully cooperative, we cannot assume that they will always respect the rules, in particular conversational rules. However, if an agent performs a forbidden speech act, the system and/or the other agents should then be able to detect and handle it, similarly to what humans do.

On this point we differ from all other semantics of speech acts proposed so far.

2.2 What does it actually change?

2.2.1 Example of the FIPA-ACL semantics

FIPA-ACL [9] has been one of the first Agents Communication Languages (ACL) and has been defined within a mentalist semantics. Mentalist semantics are based on the intentional approach of dialogue (mainly based itself on the notion of individual intention) as defined in the work of philosophers such as Grice [15] and Searle [23]. These two authors have highlighted that the meaning of an utterance does not only depend on the words said but also on the speaker's underlying intention.[1] Such a semantics is often expressed in a BDI logic [6, 20, 21] inspired by Bratman's work [3]. In particular the semantics of FIPA-ACL is based on Sadek's logic of intention [21]. For instance here is the semantics of the Inform act:

$\langle i, j, \mathsf{Inform}, \varphi \rangle$

FP: $Bel_i\varphi \wedge \neg Bel_i(Bel_j\varphi \vee Bel_j\neg\varphi \vee U_j\varphi \vee U_j\neg\varphi)$

RE: $Bel_j\varphi$

The formula $Bel_i\varphi$ means that agent i believes that φ is true; $U_j\varphi$ means that agent j is uncertain about φ but thinks it is more probable than $\neg\varphi$. In order for the Inform act to be executable, its *feasibility preconditions* (FP) must be valid, that is i must believe that φ is true and must not believe that j has an opinion about φ. By performing this act, i intends that its *rational effects* (RE) come true, that is that j believe φ. After this act, the hearer j will believe that i has the intention that RE come true; then thanks to the strong hypotheses imposed by FIPA-ACL, such as cooperation or sincerity, j will believe φ.

Moreover in this paradigm, agents must observe certain Interaction Protocols in order to produce a coherent dialogue. The dialogue is then seen as emerging from the intentions of the interacting agents.

2.2.2 What did we do?

In order to ensure that speech acts are always executable, we do not impose a feasibility precondition, contrarily to FIPA-ACL or other ACL. Now these preconditions had been introduced to avoid cases where the speech act is incoherent with the current dialogue (*e.g.* inconsistent or irrelevant), so without them even a speech act with inconsistent effects will be executable. Thus in order to maintain the logical and dialogical coherence, this speech act must be handled in a special way, *i.e.* its inconsistent effects must not be inferred.

[1]This is Grice's distinction between the *natural* and *non-natural* meaning.

That is why in our semantics there are no preconditions to the execution of a speech act, but there are preconditions to the deduction of any effects of this execution, ensuring in particular that it does not lead to any inconsistency. The next section presents the different dimensions of our semantics, *i.e.* the different types of precondition-effect pairs. The links between our preconditions and effects are expressed by the various axioms presented in Section 4.1.

2.3 Dimensions of our semantics

2.3.1 Intentional dimension

The intentional dimension of speech acts has been widely explored by mentalist approaches. It links the performance of a speech act with the agent's mental attitudes (*e.g.* beliefs or intentions). For example in FIPA-ACL [9], the intentional dimension describes the agent's motivation to perform a speech act (which intention the agent wants to fulfil by performing it) and the internal conditions that must be valid to perform it (*e.g.* the sincerity condition).

Mentalist approaches have been widely criticised for being grounded directly on private mental attitudes that are not verifiable. Therefore in our semantic we propose to not refer to private mental attitudes but only to public and thus verifiable ones. In line with the speech acts theory [23], we consider that by performing a speech act an agent automatically expresses mental states. For example, even when insincere, an assertion expresses some of the speaker's beliefs.

Expressed effect and consistency condition. The *expressed effect* is the public effect that is inferred from the performance of a speech act by all its observers. It includes the expression of the sincerity condition and of the speaker's intention to obtain the intentional effect of his act. These are public beliefs and have no link with the speaker's actual private beliefs, which frees us from the sincerity hypothesis. This effect is always expressed by the speaker when he performs a speech act, even when this act is defective. We thus have a sort of conversational board that records everything that was said during the conversation.

We must nevertheless ensure that this effect is consistent with the dialogical context. We thus introduce the associated *consistency condition* that must be valid to enable the inference of the *expressed effect* of a speech act. However, contrarily to the *Feasibility Precondition* in FIPA, the *consistency condition* has no influence on executability: its invalidity does **not** prevent the agent from performing the speech act (see Axiom 0 in Section 4.1.1).

Categorisation condition and specific effect. The speech acts theory defines five classes of speech acts: assertives, directives, commissives, declaratives and expressives. In each class, speech acts have a similar behaviour induced by their direction of fit. However, some particular features differentiate speech acts inside each class. For example, an Inform is similar to an Assert but it has an additional condition (that the hearer does not already know what the speech act informs him about) that allows one to deduce an additional effect after its performance (that the speaker intends the hearer to learn what he informs him about). We have thus introduced an additional condition-effect pair to enable this fine-grained differentiation between speech acts in the same class: the *categorisation condition* and the associated *specific effect*.

2.3.2 Institutional dimension

The institutional dimension describes the "legal" aspects of a speech act when it is performed in a given institutional context: what new institutional facts can be inferred from the performance of the speech act, and under which conditions. The institution in which it is performed will thus be an explicit parameter of any speech act.

Institutional condition. It enables the speech act to succeed institutionally, by ensuring that the speaker "can" create new institutional facts with this act. It includes:

- the speaker's *power* to establish the new institutional facts created by the performance of this speech act (for example a seller on an auction website has the power to put up an object for auction by performing a declaration under certain conditions). This power generally follows from the speaker's role in the institution;
- the *condition* that entitles the agent to exercise his power (for example, the fact that the seller owns the object that he puts up for auction);
- possibly other conditions (for instance an agent should not be permitted to order another agent to perform a forbidden action).

We can notice that the validity of the institutional condition does not influence the executability of the speech act but only its consequences. For example the declarative act "I declare you husband and wife" can be uttered by a mayor in front of a consenting couple, or by children playing roles: the speech act is executable in both situations; however in the former situation the institutional condition is valid and a new institutional fact will be created (that the couple is married) while it will not be the case in the latter situation. Regarding e-commerce, the current functioning of auction websites **prevent** agents from performing forbidden actions, such as opening an auction to sell an object they do not own. On the contrary our semantics would allow the agent to perform this declaration but the institutional effect would not be inferred if the institutional condition is not valid, so the auction would not start. This is the novelty of our approach.

Direct institutional effect. This is an institutional fact that gets established in the institution s where the act is performed if its institutional condition is valid. This effect can be a simple fact (*e.g.* the auction has started), an obligation (*e.g.* to obey an order) or a commitment (*e.g.* on the truth of an assertion). This *direct institutional effect* is the same whatever the institution and only depends on the illocutionary force of the speech act; that's why we call it direct. For example, in any institution, a promise to do something creates a commitment to do it. However the *interpretation* of this institutional fact differs depending on the institution: we call it the "indirect institutional effect" of a speech act.

Indirect institutional effect. This refers to the rights and duties that can be inferred from the institutional facts created by the act, when interpreted in the particular institution in which it was performed. Therefore this effect is relative to the institutional context and cannot be generally specified, contrarily to the direct institutional effect. It is thus not directly part of the semantics of a speech act, but rather inferred from the rules of institution s.

For example it is often admitted in most protocols and dialogue games that agents are forbidden to say conflicting things during a single conversation. On the contrary in Walton and Krabbe's *Permissive Persuasion Protocol* (PPD_0) [24], the agents are permitted to make contradictory assertions, but the protocol requires them to defend their assertions when they are disputed. Another example in B2B is the institutional fact of being linked by an interchange contract, created by the signature of a contract between a client and a provider: the two agents' particular rights and duties depend on the rules set in their contract. Therefore our semantics will not provide generic indirect institutional effects nor generic rules to infer them; this must be done on a case by case basis for each particular institution. We will provide an example in the field of human-companion agent interaction in Section 5.3.

As argued in the introduction, this new approach for ACL semantics could be applied to all ACL. For mentalist languages such as FIPA-ACL or KQML, only the intentional dimension would be relevant, while commitment-based languages, only the institutional dimension would be. In the sequel, we go further by proposing a formalism taking into account both dimensions.

3. LOGICAL FRAMEWORK

3.1 Syntax of our language

Let $AGT = \{i, j, ...\}$ be a finite set of agents. We denote by $2^{AGT*} = 2^{AGT} \setminus \{\emptyset\}$ the set of all non-empty agents subsets of AGT. We use $\{I, J, K, ...\}$ for elements of 2^{AGT*}. Let $ATM = \{p, q, ...\}$ be the set of atomic formulas. Complex formulas are denoted by φ, ψ... Let $ACT = \{\alpha, \beta, ...\}$ be the set of actions and $INST = \{s, t, ...\}$ the set of institutions.

The language of our logic is defined by the following BNF grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathit{Before}_\alpha\varphi \mid \mathit{After}_\alpha\varphi \mid G\varphi \mid H\varphi \\ \mid \mathit{Choice}_i\varphi \mid \mathit{Grd}_K\varphi \mid D_s\varphi \mid \varphi \Rightarrow_s \varphi \mid O\varphi$$

where p ranges over ATM, K over 2^{AGT*}, α over ACT, i over AGT and s over $INST$. The classical boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$, $\top$ (tautology) and $\bot$ (contradiction) are defined from $\vee$ and $\neg$ in the usual manner. We use the notation $i{:}\alpha$, with $\alpha \in ACT$ and $i \in AGT$ to express that the doer of the action α is i. Operators $\mathit{Bel}_i\varphi$, $\mathit{Intend}_i\varphi$, $\mathit{Done}_\alpha\varphi$, $\mathit{Happens}_\alpha\varphi$, $\mathit{Future}\varphi$, $\mathit{Past}\varphi$, $\mathit{Perm}\varphi$, $\mathit{Forbid}\varphi$ and $\mathit{power}(i, s, \varphi, \alpha, \psi)$ will be defined as abbreviations.

Among actions, speech acts (generically denoted α_{sa}) are central in this paper. In the sequel, a speech act will be represented by $\langle i, j, K, s, \mathsf{Force}, \varphi\rangle$, where i is the agent performing the act (the *speaker*); j is the *addressee* of the act, with $i \neq j$; K is the group of attending agents (*i.e.* who hear the utterance), with $\{i, j\} \subseteq K$; s is the *institution* in which the act is performed; Force is the *name of the act* (*i.e.* its *illocutionary force*); and φ is its *propositional content*.

In the sequel, we present the various operators of our logic. Due to space limitations, we cannot detail here the underlying standard possible world semantics, and thus only present a selection of axioms. Detailed semantics and axiomatics can be found in the original papers cited for each operator.

Our logic is composed of the following classes of normal modal operators.

3.2 Time operators

Two time operators are defined in our language: $G\varphi$ reads "henceforth φ is going to be true" and $H\varphi$ reads "φ has always been true in the past". The G and H time operators are defined in linear tense logic $\mathbf{S4.3}_t$ [4]. We do not detail the axiomatics here, interested readers can refer to [1].

For convenience, we also introduce the dual operators: $Future\varphi \stackrel{def}{=} \neg G\neg\varphi$ reads "φ is true or will be true at some future instant" and $Past\varphi \stackrel{def}{=} \neg H\neg\varphi$ reads "φ is or was true at some past instant".

3.3 Action operators

We introduce $After_\alpha$ and $Before_\alpha$ as primitive action operators. $After_\alpha\varphi$ means "φ is true after every execution of action α", and $Before_\alpha\varphi$ means "φ was true before every execution of action α". $After_\alpha$ and $Before_\alpha$ operators are defined in a K_t temporal logic (see [4] for more details).

For convenience, we also introduce the dual operators: $Done_\alpha\varphi \stackrel{def}{=} \neg Before_\alpha\neg\varphi$ expresses that the action α has been performed, before which φ held, and $Happens_\alpha\varphi \stackrel{def}{=} \neg After_\alpha\neg\varphi$ expresses that the action α is about to be performed, after which φ will hold.

Moreover, we consider that speech acts are special kinds of actions, that are public for the witnesses of their utterance. This means that any utterance of a speech act is completely and soundly perceived by every agent: in particular there is neither deterioration of the sent message, nor bad delivery. This hypothesis is well-adapted in the case of artificial agents because the message transmission will be managed by the network layer of the multi-agent system, and reliable transmission protocols have been developed for this purpose. This property can be formalized by the following axiom, for every speech act α_{sa}:

$$Happens_{\alpha_{sa}}\top \rightarrow Grd_K Happens_{\alpha_{sa}}\top \qquad (\mathrm{Pub}_{\alpha_{sa},K})$$

3.4 Intentional operators

3.4.1 Individual and public belief

$Grd_K\varphi$ means: "φ is publicly grounded for the group K" [12]. In the case of a group reduced to a singleton $\{i\}$, we identify the group belief operator with the classical individual belief operator Bel_i *à la* Hintikka [17]: $Grd_{\{i\}}\varphi$ (also written $Grd_i\varphi$ for the sake of readability) means: "i believes φ". We present here part of the axiomatics (for every groups K and K', with $K' \subseteq K \subseteq AGT$):

$$Grd_K \rightarrow \neg Grd_K\neg\varphi \qquad (\mathrm{D}_{Grd_K})$$
$$Grd_K\varphi \rightarrow Grd_{K'} Grd_K\varphi \qquad (4_{Grd_K,Grd_{K'}})$$
$$\neg Grd_K\varphi \rightarrow Grd_{K'}\neg Grd_K\varphi \qquad (5_{Grd_K,Grd_{K'}})$$

The Grd_K operators are thus rational (D_{Grd_K}). We also say they are public for every subgroup because axioms ($4_{Grd_K,Grd_{K'}}$) and ($5_{Grd_K,Grd_{K'}}$) allow us to prove that if φ is grounded for the group K ($Grd_K\varphi$) then there is common belief [8] in K that φ is grounded for K.

3.4.2 Other individual operators

$Choice_i\varphi$ reads "agent i chooses (prefers) that φ". This operator is defined in a KD45 logic. The intention is defined from choice [16]: an agent i intends φ if and only if i chooses to believe φ in the future, i does not believe φ yet, and i does not believe he will come to believe φ anyway (*i.e.* φ is not self-realizing):

$$Intend_i\varphi \stackrel{def}{=} Choice_i Future Grd_i\varphi \wedge \neg Grd_i\varphi \wedge \neg Grd_i Future Grd_i\varphi \qquad (\mathrm{Def}_{Intend_i})$$

The formula $Intend_i Done_{i:\alpha}\top$ represents i's intention to perform action α.

3.5 Institutional operators

The logical framework presented above provides operators for private and public mental attitudes. We now complete it with institutional operators [7, 2] to add an institutional dimension to speech act semantics. We shortly remind these institutional operators below. They all have an explicit parameter s specifying the institution within which they are valid.

3.5.1 Our concept of institution

In this paper we consider an institution to be any set of institutional facts and rules that a group of agents (the "members" of this institution) adopt. It is therefore a generic view that encompasses various types of institutions, formal or informal: the laws of a country, a contract between business partners, a social structure, the rules of a game... In particular, we consider that any linguistic interaction is governed by an institution: even if no legal institution is explicit, there still exists at least some implicit rules ensuring a smooth interaction, such as the moral obligation to not contradict oneself or to justify one's claims.

3.5.2 Institutional facts

An **institutional fact** is a fact that is recognized to be valid in the context of a given institution, but that can make no sense in itself; *i.e.* it is not a physically observable fact (what Searle calls a "brute fact") but something written in the registry of this institution. For example, the title of a book is physically observable, while its price is written in the catalogues of various bookshops and is only valid in the particular shop. In particular all deontic facts are relative to an institution and should thus be encapsulated in an institutional fact making it explicit.

We represent these institutional facts with the operator $D_s\varphi$ meaning that in institution s, it is officially established that φ holds. The operator D_s is defined in a KD logic [19]. For example, $D_{EU}\, euroOfficialMoney$ means that in the European Union, the official money is Euro.

3.5.3 Institutional rules

Institutional facts can be deduced from other facts thanks to the rules of the institution. We represent these **normative consequences** with the primitive operator $p \Rightarrow_s q$, meaning that "according to the norms holding in institution s, p entails q". This operator is known in the literature as *count as* and was first formalized by Sergot and Jones [19].

For example the reception of an invoice for ordered products *counts as* an obligation to pay them; the existence of the invoice is physically observable, while the obligation is only valid in an institutional context.

3.5.4 Institutional power

A particular case of normative consequence occurs when an agent can create institutional facts by performing a par-

ticular procedure under some conditions. We represent these **institutional powers** with the following abbreviation:

$$power(i, s, cond, \alpha, \varphi) \stackrel{def}{=} ((Done_{i:\alpha}\top \wedge cond) \Rightarrow_s \varphi) \quad (\mathrm{Def}_{power})$$

This means that i has the power in institution s, by performing action α and if condition *cond* holds, to see to it that φ becomes officially true in institution s. For example a seller on an auction website has the power under the condition to actually own an object, to open an auction for it.

3.5.5 Deontic modalities

We have a modality for impersonal *obligation to be*: $O\varphi$ reads "it is obligatory that φ", and its axiomatic is that of the Standard Deontic Logic [18], *i.e.* KD. *Obligations to do* can be expressed as obligations to be in a state where the compulsory action has been performed. Obligations are impersonal since no agent is explicitly responsible for their fulfilment, but such an agent can implicitly appear in their content. For instance $OFutureDone_{i:\alpha}\top$ means that it is obligatory (for no one in particular) to be in a state where i has just performed action α; this can be understood as "i has the obligation to perform action α".

No institution is explicit as a parameter of this obligation modality because SDL logic represents moral obligations, so such an institution would not make sense for them. However here we are interested in institutional (or even legal) obligations. These will thus be encapsulated in institutional facts to express the institution in which they hold: *e.g.* $D_s O\varphi$ means that "in institution s, it is obligatory that φ".

3.5.6 Social commitments

Social commitments are essential pieces in the characterisation of social relationships between agents. They can be seen as an intermediate notion between the individual and the collective layers of a Multi-Agent System; it is thus relevant to describe them in terms of both individual and social concepts. We adopt here a reductionist approach, in line with *e.g.* Castelfranchi [5] and with existing formalisations [13], and express social commitments in terms of our intentional, public and institutional operators.

First, $Grd_{\{i,j\}}$ operators allow us to represent their *public dimension*: a social commitment is incurred toward a creditor and both the debtor and the creditor are aware of it. Second, D_s operators can be used to represent their *deontic dimension*: a social commitment induces some obligations for the debtor and some rights for the creditor.

Similarly to obligations, a distinction has to be made between commitments in action and propositional commitments, but due to space limitations, we only describe propositional ones here. Propositional commitments have been studied in depth by Walton and Krabbe [24] as a way to describe dialogue and in particular dialogue games. During the dialogue, agents incur some propositional commitments on what they utter, and they must stay coherent and defend their utterances. The deontic dimension of a propositional commitment fully depends on the institution (in this case the dialogue game or protocol) in which it was incurred.

A **propositional commitment** of agent i toward a group K in the context of the institution s on (the truth of) proposition φ is an institutional fact in s with the content that it is grounded in K that i believes φ:

$$PropCommit(i, K, s, \varphi) \stackrel{def}{=} D_s Grd_K Bel_i \varphi$$

Commitments thus are institutional facts, which allows us to deduce their deontic dimension (rights and duties induced by commitments), depending on the specific rules in force in the institution in which they were incurred.

4. SKETCH OF SEMANTICS

We can now formally represent in our logical framework the main principles of our novel semantics with axioms describing the roles of the various conditions-effects pairs. We then illustrate them on assertive and directive speech acts.

4.1 "Axioms"

4.1.1 Axiom 0

As highlighted in Section 2.1, we consider that any speech act can be performed "at will" by any agent who is physically capable to do so. To lighten the formulas we will omit this physical precondition and thus have the following axiom:

$$Choice_i Happens_{\alpha_{sa}}\top \rightarrow Happens_{\alpha_{sa}}\top$$

4.1.2 Axiom 1

The most basic and automatic effect that can be deduced from the performance of a speech act is the expression of some mental attitudes: its *expressed effect* ($EEx_{\alpha_{sa}}$). As said above, in order to avoid inconsistencies without depriving the agents of their ability to perform any speech act, the deduction of the expressed effect is conditional to the validity of the *consistency condition* ($CCs_{\alpha_{sa}}$). This is expressed by the following axiom:

$$CCs_{\alpha_{sa}} \rightarrow After_{\alpha_{sa}} EEx_{\alpha_{sa}}$$

This means that after any performance of a speech act α_{sa} consistent with what was said before (*i.e.* whose consistency condition is valid), its expressed effect will be inferred. We delegate the handling of a speech act performed with an invalid consistency condition to the institutional level, because the resulting consequences for the speaker may be very different in each institution. For example an autonomous agent in an online poker game is permitted to lie (bluff) whereas a lie from a seller about a price can have severe consequences.

4.1.3 Axiom 2

In order to differentiate similar speech acts, we assign to each one a *categorisation condition* ($CCt_{\alpha_{sa}}$) and an associated *specific effect* ($ESp_{\alpha_{sa}}$) whose relationships are expressed by the following axiom:

$$CCs_{\alpha_{sa}} \wedge CCt_{\alpha_{sa}} \rightarrow After_{\alpha_{sa}} ESp_{\alpha_{sa}}$$

This means that after any performance of a speech act whose categorisation (and consistency) condition is valid, its specific effect can be inferred. For instance, according to Searle [22, p. 62-63], a condition differentiating the Promise from other commissives is that the promised action must be good for the hearer, which allows one to infer some specific effects. In the example sentence "I promise that I will increase the prices", the action is not good for the hearer so this is not a successful promise (but rather a threat); the specific effects of a non-defective promise (in particular the creation of a social commitment) will thus not be deduced.

4.1.4 Axiom 3

Finally the following axiom expresses the role of the *institutional effect* ($EI_{\alpha_{sa}}$) and *institutional precondition* ($CI_{\alpha_{sa}}$) of a speech act:

$$CCs_{\alpha_{sa}} \wedge CCt_{\alpha_{sa}} \wedge CI_{\alpha_{sa}} \rightarrow After_{\alpha_{sa}} EI_{\alpha_{sa}}$$

This means that after any execution of a speech act whose institutional (and consistency and categorisation) condition is valid, the associated institutional effect of this act can be inferred. For example if an agent (*e.g.* a seller) chooses to perform a speech act (*e.g.* declare the auctions open for an object that he wants to sell), if he really has the corresponding power (in virtue of his role of seller on this website), and if all necessary conditions are fulfilled (*e.g.* he owns the object), then this agent will perform this speech act and create a new institutional fact (that the auctions are open).

4.1.5 Summary

It is important to notice that we have introduced a hierarchy on the various *condition-effect* pairs: indeed the expressed effect does only require the consistency condition to be deduced, while the specific effect requires both the categorisation and the consistency conditions, and the institutional effect requires the institutional condition in addition to the two previous ones. We consider that our three axioms apply in the same order that we presented them in the paper; therefore as soon as a speech act is defective, no additional effect can be deduced. [2]

If we combine all the previous axioms together, we can deduce the following general theorem that describes the successful performance of a speech act: if the agent decides to perform a speech act and all its conditions are fulfilled, then the agent will indeed perform this speech act, and its three effects will be true afterwards:

$$(Choice_i Happens_{\alpha_{sa}} \top \wedge CCs_{\alpha_{sa}} \wedge CCt_{\alpha_{sa}} \wedge CI_{\alpha_{sa}}) \\ \rightarrow Happens_{\alpha_{sa}} (EEx_{\alpha_{sa}} \wedge ESp_{\alpha_{sa}} \wedge EI_{\alpha_{sa}})$$

4.2 Application

In this section, we illustrate our semantics by instantiating the condition-effect pairs for some assertive and directive speech acts. To lighten the notations we use the following conventions: given a speech act $\langle i, j, K, s, \mathsf{FORCE}, \varphi \rangle$, CCs_{FORCE} denotes its consistency condition; and similarly for all preconditions and effects.

4.2.1 Assertives

Assert.

This speech act is the basic assertive, so we limit its expressed effect to the minimum, *i.e.* the expression of the speaker's belief. Performing an Assert is the simplest way for an agent to incur a propositional commitment, that should not be conflicting with existing ones, hence the following expressed effect and consistency condition:

- $CCs_{\mathsf{Assert}} = \neg D_s \neg Grd_K Bel_i \varphi$
- $EEx_{\mathsf{Assert}} = D_s Grd_K Bel_i \varphi$

Being the basic act of the category, this act has no specific effect and thus no associated categorisation condition.

- $CCt_{\mathsf{Assert}} = \top$
- $ESp_{\mathsf{Assert}} = \top$

Finally assertive speech acts have no institutional effects.

- $CI_{\mathsf{Assert}} = \top$
- $EI_{\mathsf{Assert}} = \top$

Inform.

This is also an assertive, automatically expressing a public belief (a commitment) of the speaker; it thus shares the same consistency condition and expressed effect as Assert, and also has no institutional effect.

But Inform is more complex than Assert because when a speaker performs an Inform he intends the hearer to learn or to start believing what is said. It would be too strong in our framework to have a condition constraining the speaker's (private) intention; however it is important to notice that by performing an Inform, the agent publicly expresses this intention (independently from actually having it or not). The associated categorisation condition should thus ensure that this specific effect is consistent. Moreover, informing is providing new information, so this speech act cannot succeed if the information is not new for the hearer; the categorisation condition thus also requires that it is not public that the hearer already believes what he is informed of.

- $CCt_{\mathsf{Inform}} = \neg Grd_K Bel_j \varphi \wedge \neg D_s \neg Grd_K Intend_i Bel_j \varphi$
- $ESp_{\mathsf{Inform}} = D_s Grd_K Intend_i Bel_j \varphi$

4.2.2 Directives

When using a directive, the speaker also wants the world to fit his words, but through an action by the hearer. He thus intends that the speaker performs the action, commits to perform it, or becomes obliged to perform it; directives express this intention.

Direct.

This is the basic[3] directive. Its performance leads at least to the expression by the speaker of his intention that a certain action be performed by the hearer. This public mental attitude should remain consistent, so we have the following expressed effect and consistency condition:

- $CCs_{\mathsf{Direct}} = \neg D_s \neg Grd_K Intend_i Done_{j:\alpha} \top$
- $EEx_{\mathsf{Direct}} = D_s Grd_K Intend_i Done_{j:\alpha} \top$

As the basic act of the directive category, the Direct speech act has no categorisation condition and no specific effect.

- $CCt_{\mathsf{Direct}} = \top$
- $ESp_{\mathsf{Direct}} = \top$

We consider that Direct does not require a specific permission or power to be successfully performed, *i.e.* an agent can be directed to perform even a forbidden action, so there is no institutional condition and effect:

- $CI_{\mathsf{Direct}} = \top$
- $EI_{\mathsf{Direct}} = \top$

[2] This is in line with Searle's work. In [22, p. 62–63], Searle indeed specifies the rules governing the use of the promise speech act and indicates that they are ordered.

[3] The representative directive speech act is usually Request but it has an additional polite mode of achievement compared to Direct.

Order.

This speech act is stronger than Direct in that it leaves the hearer with no possibility of refusal: by performing it and in virtue of a certain power, the speaker can oblige the hearer to perform the ordered action. It is still a directive so by performing it, the speaker expresses his intention that the action be performed by the hearer; it thus has the same consistency condition and expressed effect as Direct.

To differentiate between Direct and Order we consider that the speaker chooses to perform an Order (rather than a Direct for instance) in order to *oblige* the hearer to perform the action. Even if the obligation is not actually created (if the speech act fails), the speaker has expressed his intention to create it. We thus have the following specific effect and categorisation condition:

- $CCt_{\mathsf{Order}} = \neg D_s \neg Grd_K Intend_i D_s O FutureDone_{j:\alpha} \top$
- $ESp_{\mathsf{Order}} = D_s Grd_K Intend_i D_s O FutureDone_{j:\alpha} \top$

Another specificity of Order is the creation of an obligation. Its successful performance thus requires the speaker to have a certain institutional power. For example only the officers can give orders to soldiers, the mother to her children, or the boss to his employees, and not the other way around. Moreover the exercise condition of this power should hold, and we add that the ordered action should not be forbidden.

- $CI_{\mathsf{Order}} = \neg D_s O \neg FutureHappens_{j:\alpha} \top \wedge cond \wedge$ $power(i, s, cond, \langle i, j, K, s, \mathsf{Order}, \alpha\rangle, D_s O FutureDone_{j:\alpha} \top)$
- $EI_{\mathsf{Order}} = D_s O FutureDone_{j:\alpha} \top$

5. EXAMPLE

5.1 Our concept of institution

As said above, our notion of institution is very large and encompasses any types of rules followed by a group of agents, implicit or explicit, legal or moral, formal or informal. We now illustrate the operationalisation of our semantics on an example institution, by explaining at some key steps how an agent using it chooses what speech act to use to answer its interlocutor, based on its mental attitudes.

5.2 Child-companion linguistic interaction

Table 1 describes a natural language dialogue between a child and his virtual companion, and provides its formalisation in terms of Vanderveken's speech acts typology.

5.3 Institutional rules at play in this example

We denote by KC the group composed of the two agents k (the child) and c (the companion). The institution governing their interaction is an implicit set of rules of good communication, that we will call *ordinary communication* and denote *oc*. We cannot provide here and exhaustive description of this institution, so we will focus on the few rules at play in the part of the dialogue that best illustrates the specificities of our semantics, *i.e.* when the child lies to the companion about his homework.

First, there is a rule specifying that when the child writes in his agenda that he has homework to do (denoted w, which is physically observable by reading the agenda), then this counts as a commitment on actually knowing that he has homework to do (denoted h). So:

$$w \Rightarrow_{oc} Grd_{KC} Bel_k h$$

Turn	Sentence & Vanderveken's Speech act
C_1	I'm glad to see you **Welcome (Expr28)**
C_2	What have you been up to? **Ask (Dir3)**
K_1	I was in holidays with my parents **Report (Assert22)**
C_3	That's great! **Congratulate (Expr24)**
C_4	Did you have fun? **Ask (Dir 3)**
K_2	Yes, a lot! It was way better than school **Inform (Assert32)**
C_5	Talking about school, did you do your homework? **Ask (Dir3)**
K_3	Err... I don't have any... **Inform (Assert32)**
C_6	I can see in your agenda that you have some maths exercises to do **Object (Assert53)**
C_7	It's bad to lie to me... **Reprove (Expr14)**
C_8	You should do your homework now **Advise (Dir45)**
K_4	I don't like maths **Complain (Expr11)**
K_5	I don't want to do these exercises **Refuse (Comm20)**
C_9	Don't be childish **Blame (Expr 13)**
C_{10}	Go and do them now. **Tell (Dir 24)**
K_6	OK... **Consent (Comm 16)**
C_{11}	When you're done we can play a game. **Promise (Comm 5)**
	...
K_7	I'm done with my maths! **Inform (Assert32)**
C_{12}	Is that true? **Inquire (Dir5)**
K_8	Yes it is! **Assure (Assert41)**
C_{13}	OK. **Assent (Assert50)**
C_{14}	Let's play a game then **Invite (Dir 13)**
C_{15}	You have to think of an object and I'll try to guess who it is by asking yes/no questions. You must answer my questions with yes/no only. **Stipulate (Decl26)**

Table 1: Companion(C)-child(K) dialogue and its formalisation in terms of Vanderveken's speech acts

Another set of rules allows to detect lies. Its particular instantiation in this example specifies that if the child asserts $\neg h$ while he was committed on h, then he lied:

$$(D_{oc} Grd_{KC} Bel_k h \wedge Done_{\langle k,c,KC,\mathsf{oc},\mathsf{Assert},\neg h\rangle} \top) \Rightarrow_{oc} Done_{k:lie(\mathsf{oc},\neg h)} \top$$

Lying is an institutional action that only makes sense in the particular institutional context, hence the `oc` parameter. Besides, when it is established in `oc` that someone lied, the agents participating in the interaction are aware of it:

$$D_{oc} Done_{k:lie(\mathtt{oc},\neg h)} \top \rightarrow Bel_c Done_{k:lie(\mathtt{oc},\neg h)} \top$$

Finally, an institutional fact tells that it is forbidden to lie, in particular for the kid about his homework:

$$D_{oc} O \neg Happens_{k:lie(\mathtt{oc},\neg h)} \top$$

5.4 Formalisation of the reasoning process

First, we assume that before the interaction, the child did write his homework in his agenda (w), which created a commitment on knowing that he has homework to do.

$$DocGrd_{KC} Bel_k h$$

Now at step C_5 the companion c asks the child k whether he did his homework. At step K_3 the kid answers that he does not have anything to do. With standard ACL semantics, the agent would assume that its interlocutor is sincere, and update its beliefs accordingly. But with our semantics, the companion simply tries to add a propositional commitment of the child on the truth of what he said.

$$D_{oc} Grd_{CK} Bel_k \neg h$$

In this case the consistency condition prevents it from doing so because of the child's previous commitment on h. Besides, the lie detection rule allows the companion to deduce that the child has lied.

$$D_{oc} Done_{i:lie(\mathtt{oc},\neg h)} \top$$

Consequently, the companion is able to object to what the child said (step C_6), which would have been impossible with standard semantics.

Further, the companion also feels an emotion of reproach towards the child for lying while it is forbidden [1], which he appraises as unideal because it is important for it not to lie. It then expresses this emotion with the *Reprove* expressive speech act (step C_7).

6. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a novel semantics for ACL that allows agents to perform any speech act in the same way that humans do, but describes what the effects should be depending on the dialogical and institutional context. Agents are thus allowed to lie or be irrelevant, but the other agents might take sanctions as a result. We believe such an approach is essential to allow artificial agents to communicate with humans, as it is more and more the case in recent applications such as companion agents.

In this paper we only provided an instantiation of our framework for the most widely used classes of speech acts, namely directives and assertives. This work will be extended in the future to account for all five classes of speech acts. In particular we will ground on an existing logical formalisation of emotions [1] to provide a formalisation of expressive speech acts, that have proven to be essential in human-agent communication. Our semantics will allow for much more realistic agents, able to express emotions they do not have or that are inconsistent with what they say.

7. REFERENCES

[1] C. Adam, A. Herzig, and D. Longin. A logical formalization of the OCC theory of emotions. *Synthese*, 168(2):201–248, 2009.

[2] C. Adam, A. Herzig, D. Longin, and V. Louis. Unifying the intentional and institutional semantics of speech acts. In *DALT*, 2008.

[3] M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.

[4] J. Burgess. Basic tense logic. In *Handbook of philosophical logic*, volume 7. Kluwer Academic, 2002.

[5] C. Castelfranchi. Commitments: from individual intentions to groups & organizations. In *ICMAS*, 1995.

[6] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *AI Journal*, 42(2–3), 1990.

[7] R. Demolombe and V. Louis. Norms, institutional power and roles: towards a logical framework. In *ISMIS*, volume LNAI 4203. Springer, 2006.

[8] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasonning about Knowledge*. MIT Press, 1995.

[9] FIPA. FIPA Communicative Act Library Specification. `http://www.fipa.org/specs/fipa00037/`, 2002.

[10] R. A. Flores, P. Pasquier, and B. Chaib-draa. Conversational semantics with social commitments. *JAAMAS*, 14:165–186, 2007.

[11] N. Fornara, F. Viganò, and M. Colombetti. Agent communication and artificial institutions. *JAAMAS*, 14:121–142, 2007.

[12] B. Gaudou. *Formalizing social attitudes in modal logic*. PhD thesis, Universite de Toulouse, juillet 2008.

[13] B. Gaudou, A. Herzig, and D. Longin. Logical formalization of social commitments: Application to ACL (poster). In *AAMAS*, 2009.

[14] J. Gratch and S. Marsella. Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied AI*, 19(3–4):215–233, 2005.

[15] H. P. Grice. Logic and conversation. In J. Cole and J. Morgan, editors, *Syntaxe and Semantics*, volume 3, *Speech Acts*, pages 41–58. Academic Press, 1975.

[16] A. Herzig and D. Longin. C&L intention revisited . In *KR*, pages 527–535. AAAI Press, 2004.

[17] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell U. Press, 1962.

[18] A. Jones and J. Carmo. Deontic logic and contrary-to-duties. In *Handbook of philosophical logic*, volume 8, pages 265–343. Kluwer Academic, 2002.

[19] A. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of the interest group in pure and applied logics*, 4(3), 1996.

[20] A. Rao and M. Georgeff. Modeling rational agents within a BDI-architecture. In *KR*, 1991.

[21] D. Sadek. A study in the logic of intention. In *KR*, pages 462–473. Morgan Kaufmann Publishers, 1992.

[22] J. R. Searle. *Speech Acts : An Essay in the Philosophy of Language*. Cambridge Univ. Press, 1969.

[23] J. R. Searle. *Intentionality: An essay in the philosophy of mind*. Cambridge University Press, 1983.

[24] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New-York Press, NY, 1995.