CNRS - INPT - UPS - UT1 - UTM

RESEARCH REPORT

IRIT/RR–2008-9–FR

The institutional dimension of speech acts: a logical approach based on the concept of acceptance

Emiliano LORINI, Dominique LONGIN & Benoit GAUDOU

Equipe LILaC
{lorini,longin,gaudou}@irit.fr

February 27, 2008

UPS-IRIT, 118 route de Narbonne, 31062 Toulouse CEDEX 9
+33 (0) 561 55 67 65 info@irit.fr www.irit.fr

Abstract

The aim of this paper is to provide an analysis of speech acts as institutional actions, where an institution is grounded on the attitudes of its members. In the first part of the paper we present a logic of acceptance, goal and action. Then, we specify how agents can create and maintain normative and institutional facts on the basis of their acceptances *qua* members of a certain institution. Finally, we propose a formal characterization of the speech act *promise*.

Keywords

Modal Logics, Institutions, Agent Communication Languages

Contents

1 Introduction

Saussure [9] distinguishes language (*la langue*) from speech (*la parole*). The "meaning" of the former is defined by institutions (social aspect), by a set of conventions outside the linguistic activity (static aspect). The latter corresponds to the use of the language (dynamical aspect) by individuals, and it is the intention of these individuals that assigns a "value" to the act of utterance (individual aspect). Metaphorically, language is a sort of score while speech is the execution of this score by musicians.

The Austin's concept of "performative utterance" [2] is the first step to bring together language and speech because: first, performative utterances are regulated by social conventions that determine not only the "meaning" of the utterance, but also the "value" of its enunciation[1]; second, it is practically impossible to attribute a meaning to a performative utterance independently of the value of its enunciation.

Through the concept of illocutionary act, speech act theory [2, 21] generalizes the concept of performative utterance. In such a theory, the value of an enunciation cannot be no longer considered as a consequence of a preliminary meaning; but contrary to the linguistic nihilism of the wittgensteinian trend and their slogan "Meaning is use" (See [21, Section 6.4]), speech act theory does not deny that the use of language is founded on a prior knowledge: it claims that the meaning of an utterance integrates some shared rules that fixe the effects of its enunciation on the speech situation. In speech act theory language is conceived as an informal institution defined by constitutive rules of the form "X count as Y" (See [21, Section 2.7] for more details). These rules create or define new forms of behavior such as promises, orders, requests, assertions, *etc.* (see [21, Section 2.5]). For example, in the context of *ordinary communication* [17], a promise is defined and created on the basis of a constitutive rule (shared by the speaker and the hearer) of the form: a speaker's enunciation of the utterance: "I am going to perform action a!" *counts as* a promise to the hearer to perform action a, under the condition that the hearer wants the speaker to perform action a. Subsequently, to speak a language, it is to adopt a form of intentional behavior (by performing a speech act) governed by such a kind of constitutive rules.

The focus of this contribution is to define speech acts as institutional actions where modeled institutions are social or informal institutions. With social or informal institution, we mean an institution which is grounded on the acceptances of its members. Differently from formal (legal) institutions, informal institutions are *rule-governed social practices* in which no member with 'special' powers to create and eliminate institutional facts is introduced (see Section 4).[2] After a short introduction about the concept of acceptance (Section 2), we present a modal logic of acceptance, goal and action (Section 3). On the basis of the notion of acceptance, we specify how a group of agents can create normative and institutional facts which hold only in an attitude-dependent way (Section 4). In Section 5 we define the promissive speech act.

[1] By "enunciation", we mean "the performance of an utterance".
[2] For the distinction between formal and informal institutions, see also [20].

2 The Concept of acceptance

Whereas beliefs have been studied for decades, acceptances have only been examined since [24] and [6] while studying the nature of argument premises or reformulating Moore's paradox [6]. If a belief that p is an attitude constitutively aimed at the truth of p, an acceptance is the output of "a decision to treat p as true in one's utterances and actions" [15] without being necessarily connected to the actual truth of the proposition. Another difference between belief and acceptance is that beliefs are context-independent, whilst acceptance depends on context [4]. In fact, one can decide (say for prudential reasons) to reason and act by "accepting" the truth of a proposition in a specific context, and possibly rejecting the very same proposition in a different one. Although, usually, this aspect of acceptance is studied in private contexts, here we continue the work initiated in [13] by exploring the role of acceptance in institutional contexts. Institutional contexts are rule-governed social practices on the background of which the agents reason. For example, take the case of a game like Clue. The institutional context is the rule-governed social practice which the agents conform to in order to be competent players. On the background of such contexts, we are interested in the individual and collective attitudes (individual and collective acceptances) that can be formally captured. In the context of Clue, for instance, agents accept that something has happened (see Example 2) *qua* players of Clue. The state of acceptance *qua* member of an institution is the kind of acceptance one is committed to when one is "functioning as member of the institution" [26].

3 The logical framework

Syntax. The syntactic primitives of our logic of acceptance, actions and goals are the following: – a finite set of $n > 0$ agents $AGT = \{1, 2, ..., n\}$; – a nonempty finite set of *atomic actions* $AT = \{a, b, ...\}$; – a finite set of atomic formulas $ATM = \{p, q, ...\}$; – a finite set of labels denoting institutional contexts $INST = \{inst_1, inst_2, ..., inst_m\}$. Moreover, we note $2^{AGT\star} = 2^{AGT} \setminus \{\emptyset\}$ the set of all non empty subsets of agents, $\Delta = \left\{C{:}x | C \in 2^{AGT\star}, x \in INST\right\}$ the set of all couples of non empty subsets of agents and institutional contexts.

The language $\mathcal{LANG}$ of the logic $\mathcal{L}$ is defined as the smallest superset of ATM such that: if $\varphi, \psi \in \mathcal{LANG}$, $i, j \in AGT$, $\alpha \in ACT$ and $C{:}x \in \Delta$ then $\neg\varphi$, $\varphi \vee \psi$, $[i{:}\alpha{:}j]\,\varphi$, $Goal_i\varphi$ and $Accept_{C:x}\varphi \in \mathcal{LANG}$, where the set ACT is the smallest superset of AT such that: if $i, j \in AGT$, $a \in AT$ then $inf(i{:}a{:}j) \in ACT$ (*informative actions*). The classical boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$, $\top$ (tautology) and $\bot$ (contradiction) are defined from $\vee$ and $\neg$ in the usual manner.

The set ACT includes all informative actions of type "informing that agent i is going to perform (atomic) action a for agent j", where, with informative action, we mean the performance of an utterance, the act of making an enunciation. Since AT and AGT are finite sets, it follows that ACT as well is finite.

Operators of the form $Accept_{C:x}$ have been introduced in [13] where a logic of acceptance has been proposed. These operators allow to express those facts that the agents in a group C accept while identifying themselves with a certain institution x.

Formula $Accept_{C:x}\varphi$ has to be read "the agents in C accept that φ while functioning as members of the institution x".

EXAMPLE 1. $Accept_{C:Greenpeace}protectEarth$ *is read "the agents in C accept that the mission of Greenpeace is to protect the Earth while functioning as activists of Greenpeace".*

For $C{:}x \in \Delta$, $Accept_{C:x}\bot$ has to be read "agents in C are not functioning as members of the institution x" because we assume that functioning as a member of an institution is, at least in this minimal sense, a rational activity; conversely, $\neg Accept_{C:x}\bot$ has to be read "agents in C are functioning as members of the institution x"; $\neg Accept_{C:x}\bot \wedge Accept_{C:x}\varphi$ stands for "agents in C are functioning as members of the institution x and they accept that φ while functioning as members of x" or simply "agents in C accept that φ *qua* members of the institution x" (*i.e.* group acceptance).

The dynamic operators of the form $[i{:}\alpha{:}j]$ are just a generalization of standard operators of dynamic logic [16] where both the author (initiator) and the addresse of a certain action α are specified. Formula $[i{:}\alpha{:}j]\,\varphi$ has to be read "after agent i does action α for agent j, it is the case that φ". Operators of the form $Goal_i$ are standard operators for agents' goals [7]. $Goal_i\varphi$ has to be read "agent i has the goal that φ holds" (or "i wants φ to be true").

The following abbreviations are given for any $i, j \in AGT$, $a \in AT$ and $\alpha \in ACT$:

$$\langle i{:}\alpha{:}j\rangle\,\varphi \stackrel{def}{=} \neg\,[i{:}\alpha{:}j]\,\neg\varphi;\, Int_{i,j}(\alpha) \quad \stackrel{def}{=} Goal_i\,\langle i{:}\alpha{:}j\rangle\,\top;\, Inf_{i,j}(a) \stackrel{def}{=} \langle i{:}inf(i{:}a{:}j){:}j\rangle\,\top.$$

Formula $\langle i{:}\alpha{:}j\rangle\,\varphi$ is meant to stand for "agent i performs action a and φ is true afterward". $Int_{i,j}(\alpha)$: agent i intends to perform action α for j. $Inf_{i,j}(a)$: agent i informs agent j that he is going to perform the (atomic) action a for him.

Semantics. We use a possible worlds semantics. A model of the logic $\mathcal{L}$ is a tuple $\mathcal{M} = \langle W, \mathscr{A}, \mathscr{R}, \mathscr{G}, \mathscr{V}\rangle$ where:

- W is a set of possible worlds;
- $\mathscr{A} : \Delta \longrightarrow (W \longrightarrow 2^W)$ associates each $C{:}x \in \Delta$ and possible world w with the set $\mathscr{A}_{C:x}(w)$ of possible worlds accepted by the group C in w, where the agents in C are functioning as members of the institution x;
- $\mathscr{R} : AGT \times ACT \times AGT \longrightarrow (W \longrightarrow 2^W)$ associates each two agents $i, j \in AGT$, action $\alpha \in ACT$ and possible world w with the set $\mathscr{R}_{i:\alpha:j}(w)$ of possible worlds that are reachable from w through the occurrence of action α performed by i for agent j;
- $\mathscr{G} : AGT \longrightarrow (W \longrightarrow 2^W)$ associates each agent $i \in AGT$ and possible world w with the set $\mathscr{G}_i(w)$ of worlds that are compatible with agent i's goals in w;
- $\mathscr{V} : W \longrightarrow 2^{ATM}$ is a truth assignment which associates each world w with the set $\mathscr{V}(w)$ of atomic propositions true in w.

To the standard truth conditions for atomic formulas, negation and disjunction we add
:

- $\mathcal{M}, w \models Accept_{C:x}\varphi$ iff for all $w' \in W$, if $w' \in \mathscr{A}_{C:x}(w)$ then $\mathcal{M}, w' \models \varphi$;
- $\mathcal{M}, w \models [i{:}\alpha{:}j]\,\varphi$ iff for all $w' \in W$, if $w' \in \mathscr{R}_{i:\alpha:j}(w)$ then $\mathcal{M}, w' \models \varphi$;
- $\mathcal{M}, w \models Goal_i\varphi$ iff for all $w' \in W$, if $w' \in \mathscr{G}_i(w)$ then $\mathcal{M}, w' \models \varphi$.

Axiomatization. Every operator of type $Accept_{C:x}$ and $[i{:}\alpha{:}j]$ is supposed to be a normal modal operator satisfying standard axioms and rules of inference of system K. Every operator $Goal_i$ is supposed to be a normal modal operator satisfying standard axioms and rules of inference of system KD.[3] Axiom D for goals corresponds to the following constraint of seriality over $\mathcal{L}$ models. For every $i \in AGT$ and $w \in W$:

$$\mathscr{G}_i(w) \neq \emptyset \qquad \textbf{S1}$$

The rest of the section contains other axioms for acceptance, action and intention and corresponding semantic constraints over $\mathcal{L}$ models.

We suppose that given a set of agents C, all $B \subseteq C$ have access to all the facts that are (not) accepted by agents in C while functioning as members of the institution x. In particular, we suppose that: if agents in C (do not) accept that φ while functioning as members of the institution x then for every subset B of C and institution y while functioning as members of the institution y, agents in B accept that agents in C (do not) accept that φ while functioning as members of the institution x. Such properties are captured by the following two axiom schemas. For every $C{:}x$, $B{:}y \in \Delta$, if $B \subseteq C$ then:

$$Accept_{C:x}\varphi \rightarrow Accept_{B:y}Accept_{C:x}\varphi \qquad \mathbf{4}_{\boldsymbol{Accept}}$$

$$\neg Accept_{C:x}\varphi \rightarrow Accept_{B:y}\neg Accept_{C:x}\varphi \qquad \mathbf{5}_{\boldsymbol{Accept}}$$

Axioms $\mathbf{4}_{\boldsymbol{Accept}}$ and $\mathbf{5}_{\boldsymbol{Accept}}$ together correspond to the following semantic constraint over $\mathcal{L}$ models. For every $w \in W$ and $C{:}x$, $B{:}y \in \Delta$, if $B \subseteq C$ then:

$$\text{if } w' \in \mathscr{A}_{B:y}(w) \text{ then } \mathscr{A}_{C:x}(w') = \mathscr{A}_{C:x}(w) \qquad \textbf{S2}$$

We also suppose that if agents in C accept that φ *qua* members of the institution x then, for every subset B of C, it holds that agents in B accept φ *qua* members of the institution x. Formally, for every $C{:}x$, $B{:}x \in \Delta$, if $B \subseteq C$ then:

$$(\neg Accept_{C:x}\bot \wedge Accept_{C:x}\varphi) \rightarrow (\neg Accept_{B:x}\bot \wedge Accept_{B:x}\varphi) \qquad \mathbf{Inc}_{\boldsymbol{Accept}}$$

EXAMPLE 2. *Imagine three agents i, j, k that,* qua *Clue players, accept that someone called Mrs. Red, has been killed: $\neg Accept_{\{i,j,k\}:Clue}\bot \wedge Accept_{\{i,j,k\}:Clue}killedMrsRed$. This implies that also the two agents i, j* qua *Clue players accept that someone called Mrs. Red has been killed: $\neg Accept_{\{i,j\}:Clue}\bot \wedge Accept_{\{i,j\}:Clue}killedMrsRed$.*

Axiom $\mathbf{Inc}_{\boldsymbol{Accept}}$ corresponds to the following semantic constraint over $\mathcal{L}$ models. For every $w \in W$, $C{:}x$, $B{:}x \in \Delta$, if $B \subseteq C$:

$$\text{if } \mathscr{A}_{C:x}(w) \neq \emptyset \text{ then } \mathscr{A}_{B:x}(w) \neq \emptyset \text{ and } \mathscr{A}_{B:x}(w) \subseteq \mathscr{A}_{C:x}(w) \qquad \textbf{S3}$$

We suppose the following additional constraint over $\mathcal{L}$ models. For every $w \in W$, $i, j, i', j' \in AGT$ and $\alpha, \beta \in ACT$:

$$\text{if } w' \in \mathscr{R}_{i:\alpha:j}(w) \text{ and } w'' \in \mathscr{R}_{i':\beta:j'}(w) \text{ then } w' = w'' \qquad \textbf{S4}$$

The property **S4** says that all actions occurring in a world w lead to the same world. Thus, all actions occur in parallel and they do not have non-deterministic effects. This explains

[3]Axiom D for $Goal_i$ is: $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$.

why we have phrased $\langle i{:}\alpha{:}j\rangle\,\varphi$ "i does α for j and φ holds afterward" rather than "*it is possible that* i does α for j and φ holds afterward". Constraint **S4** corresponds to the following axiom of our logic. For every $i, j, i', j' \in AGT$ and $\alpha, \beta \in ACT$:

$$\langle i{:}\alpha{:}j\rangle\,\varphi \rightarrow [i'{:}\beta{:}j']\,\varphi \qquad \textbf{\textit{Det}}$$

We also suppose that the world is never static in our framework, that is, for every world w there exists at least two agents i, j and action α such that i performs α for j at w. Formally, given a $\mathcal{L}$ model M, for every $w \in W$ we have that:

$$\exists i, j \in AGT, \exists \alpha \in ACT, \exists w' \in W \text{ s.t. } w' \in \mathscr{R}_{i:\alpha:j}(w) \qquad \textbf{S5}$$

Property **S5** of $\mathcal{L}$ models corresponds to the following axiom of our logic.

$$\bigvee_{i,j \in AGT, \alpha \in ACT} \langle i{:}\alpha{:}j\rangle\,\top \qquad \textbf{\textit{Active}}$$

Axiom ***Active*** ensures that for every world w there is a *next* world of w which is reachable from w by the occurrence of some action performed by some agent for another agent. This is the reason why the operator X for *next* of LTL (linear temporal logic) can be defined as follows:[4]

$$\mathbf{X}\varphi \stackrel{def}{=} \bigvee_{i,j \in AGT, \alpha \in ACT} \langle i{:}\alpha{:}j\rangle\,\varphi$$

where $\mathbf{X}\varphi$ is meant to stand for "φ will be true in the next state".

The following axiom relates intentions with actions. For every $i, j \in AGT$ and $\alpha \in ACT$:

$$\langle i{:}\alpha{:}j\rangle\,\top \rightarrow Int_{i,j}(\alpha) \qquad \textbf{\textit{IntAct}}$$

According to Axiom ***IntAct***, an agent i performs action α for agent j only if he has the intention to do α for j. In this sense we suppose that an agent's *doing* is by definition intentional. ***IntAct*** corresponds to the following semantic constraint over $\mathcal{L}$ models. For every $i, j \in AGT$, $\alpha \in ACT$ and $w \in W$:

$$\text{if } \exists v' \text{ such that } v' \in \mathscr{R}_{i:\alpha:j}(w) \text{ then } \forall w' \in \mathscr{G}_i(w), \exists w'' \text{ such that } w'' \in \mathscr{R}_{i:\alpha:j}(w') \qquad \textbf{S6}$$

We call $\mathcal{L}$ the logic axiomatized by the principles presented above and we write $\vdash_{\mathcal{L}} \varphi$ iff formula φ is a theorem of $\mathcal{L}$. Moreover, we write $\models_{\mathcal{L}} \varphi$ iff formula φ is *valid* in all $\mathcal{L}$ models, *i.e.* $\mathcal{M}, w \models \varphi$ for every $\mathcal{L}$ model $\mathcal{M}$ and world w in $\mathcal{M}$. Finally, we say that a formula φ is *satisfiable* if there exists an $\mathcal{L}$ model $\mathcal{M}$ and a world w in $\mathcal{M}$ such that $\mathcal{M}, w \models \varphi$.

4 Institutional concepts

Normative and institutional facts are a class of facts that are typical of institutional contexts. According to [22], such facts have the peculiar feature of being dependent on the agents' attitudes in a way that we are in the position to specify in detail in the logic $\mathcal{L}$. More precisely it has been noted that these facts are characterized at least by two features [18, 22, 25]: *performativity* (an attitude of certain type shared by a group of agents

[4] Note that X satisfies the standard property $X\varphi \leftrightarrow \neg X \neg\varphi$.

towards a normative or an institutional fact may contribute to the truth of a sentence describing the fact); *reflexivity* (if a sentence describing a normative or an institutional fact is true, the relevant attitude is present).

EXAMPLE 3. *If the agents* qua *group members accept a certain piece of paper as money (an institutional fact), then, in the appropriate context, this piece of paper is money for that group (performativity). At the same time, if it is true that a certain piece of paper is money for a group, then the agents* qua *group members accept the piece of paper as money (reflexivity).*

In order to represent in the logic $\mathcal{L}$ these kind of facts, we need first to define the concept of truth with respect to an institutional context in way that respects these two principles.

Truth in an institutional context. We formalize the notion of truth w.r.t. a certain institutional context with the operator $[\![x]\!]$. A formula $[\![x]\!]\,\varphi$ is read "within the institutional context x, it is the case that φ". Here we suppose that "within the institutional context x it is the case that φ" if and only if "for every set of agents C, the agents in C accept that φ while functioning as members of the institution x". Formally:

$$[\![x]\!]\,\varphi \stackrel{def}{=} \bigwedge_{C \in 2^{AGT\star}} Accept_{C:x}\varphi$$

It is straightforward to prove that every $[\![x]\!]$ is a normal modal operator. Given the previous analysis, a fact is true w.r.t. an institutional context x if and only if such fact is accepted by all the agents while they function as members of x (hence the performativity and the reflexivity principles are maintained).

At this point, it might be objected that there are facts which are true in an institutional context but only "special" members of the institution are aware of them and can change them. For instance, there are laws in every country which are known and can be changed only by the specialists of the domain (lawyers, judges, members of the parliament, *etc.*). In order to resist to this objection recall that, at this stage, our model applies to informal institutions of a society, in particular to language. Relative to this restriction, the proposed assumption is justified because, for informal institutions, there is no special agent who has the power to create and eliminate institutional facts characterizing the institution itself (*i.e.* nobody has the power to change the rules of the speech act *promise* in the context of language).

Finally, the following abbreviation is given:

$$[\![Univ]\!]\,\varphi \stackrel{def}{=} \bigwedge_{x \in INST} [\![x]\!]\,\varphi$$

which stands for "φ is universally accepted as true".

Constitutive rules. From the concept of truth with respect to an institutional context a notion of *constitutive rule* of the form "φ counts as ψ in the institutional context x" can be defined. We conceive a constitutive rule as a material implication of the form $\varphi \rightarrow \psi$ in the scope of an operator $[\![x]\!]$. We suppose that a constitutive rule is intrinsically local, that is, a rule that is not universally valid while it is accepted by the members of a certain institution. More generally, for every $x \in INST$ the following abbreviation (that stands for "φ counts as ψ in the institutional context x") is given:[5]

$$\varphi \overset{x}{\triangleright} \psi \overset{def}{=} [\![x]\!]\,(\varphi \rightarrow \psi) \wedge \neg\,[\![Univ]\!]\,(\varphi \rightarrow \psi)$$

EXAMPLE 4. *In the context of gestural language there exists a constitutive rule according to which, the nodding gesture "counts as" an endorsement of what the speaker is suggesting, noted* $nodding \overset{gesture}{\triangleright} yes$. *But in different contexts the same gesture may express exactly the opposite fact (viz.* $\neg\,[\![Univ]\!]\,(nodding \rightarrow yes)$ *holds).*

Obligations. Informal institutions such as language involve a deontic dimension that up to now we have ignored. In order to capture this core feature we extend the logic $\mathcal{L}$ by introducing a *violation* atom V as in Anderson's reduction of deontic logic to alethic logic [1]. By means to this new formal construct we can formally characterize the obligations which are valid in a certain institution by anchoring them in the acceptances of the members of that institution.

We say that "in the institutional context x, agent i has the obligation to perform action α for agent j under the condition φ" (noted $Obl_x(i, \alpha, j, \varphi)$) if and only if "in the institutional context x, the fact that i does not perform action α for j under the condition φ counts as a violation at the next step of time".[6] Formally:

$$Obl_x(i, \alpha, j, \varphi) \overset{def}{=} (\varphi \wedge [i{:}\alpha{:}j]\bot) \overset{x}{\triangleright} \mathbf{X}V$$

EXAMPLE 5. *Formula* $Obl_{EBay}(i, sendGoods, j, paid(j, i))$ *expresses that, in the context of EBay, it is obligatory for an agent* i *to send certain goods to agent* j *under the condition that* j *has paid the goods to* i.

We here distinguish an obligation which is valid in a certain institution x (viz. institutional obligation) from the instantiations of this obligation in specific groups of agents which are members of this institution. We say that i's obligation to do action α for j is instantiated in the group C of members of institution x (noted $IObl_{C:x}(i, \alpha, j)$) if and only if the agents in C, *qua* members of the institution x, accept that if i does not perform action α for j then i will incur a violation in the next state. Formally:

$$IObl_{C:x}(i, \alpha, j) \overset{def}{=} \neg Accept_{C:x}\bot \wedge Accept_{C:x}([i{:}\alpha{:}j]\bot \rightarrow \mathbf{X}V)$$

The following Theorem highlights the relationships between institutional obligations and instantiations of obligations in groups.

Theorem 1

$$\vdash_{\mathcal{L}} (Obl_x(i, \alpha, j, \varphi) \wedge \neg Accept_{C:x}\bot \wedge Accept_{C:x}\varphi) \rightarrow \qquad IObl_{C:x}(i, \alpha, j)$$

[5]Our notion of constitutive rule of the form $\varphi \overset{x}{\triangleright} \psi$ is similar to the notion of *proper classificatory rule* given in [14].

[6]Formulas of type $Obl_x(i, \alpha, j, \varphi)$ can be conceived as particular instances of so-called *regulative rules* in Searle's sense [22]. On the distinction between *regulative rule* and *constitutive rule* see also [3].

According to Theorem 1, if the agents in C, *qua* members of institution x, accept that φ and, in x, i has the obligation to perform action α for agent j under the condition φ then, i's obligation to do action α for j is instantiated in the group C of members of x.

EXAMPLE 6.*From the fact that in EBay, it is obligatory for an agent i to send certain goods to agent j under the condition that j has paid the goods to i (noted $Obl_{EBay}(i, sendGoods, j, paid(j, i))$), and the fact that the group C of EBay surfers accept that j has paid certain goods to i (noted $\neg Accept_{C:EBay}\bot \wedge Accept_{C:EBay}paid(j, i)$), we infer that i's obligation to send the goods to j is instantiated in the group C of EBay surfers (noted $IObl_{C:EBay}(i, sendGoods, j)$).*

Social commitment Social commitment has a fundamental role in the interaction between agents in an institution. It fixes the relations between agents, bounding an agent toward another. Social commitment is thus a relational notion: it relies at least two agents, the agent who is committed (the *debtor* [12]) and the agent to whom the debtor is committed (the *creditor*). According to Castelfranchi [5], there are two crucial aspects of social commitment: a motivational aspect and a deontic aspect. If the debtor i is committed to the creditor j to perform action α for him then, the creditor and the debitor must mutually know that the creditor is interested in the fact that the debtor performs α (motivational aspect) and, the creditor must have specific rights on the debtor being entitled to ask to the debtor to perform action α (deontic aspect).[7]

In the present analysis we specify these two conditions for social commitment on the basis of the notions of acceptance and instantiated obligation. Differently from Castelfranchi, we make explicit the institutional dimension of social commitment by specifying the institution in which the commitment is established. On the one hand, we characterize the deontic aspect of i's commitment to j to perform action α by the fact that i and j function as members of a certain institution x and, as members of x, i and j accept that i is obliged to perform action α for j. That is, i's obligation to do action α for j is instantiated in the group $\{i, j\}$ of members of x. On the other hand, we characterize the motivational aspect of i's commitment to j by the fact that i and j, *qua* members of the institution x, accept that j wants i to do α for him. Formally:

$$SC_x(i, \alpha, j) \stackrel{def}{=} IObl_{\{i,j\}:x}(i, \alpha, j) \wedge \neg Accept_{\{i,j\}:x}\bot \wedge Accept_{\{i,j\}:x}Goal_j \langle i{:}\alpha{:}j\rangle \top$$

which stands for "i is committed to j to do α w.r.t. institution x."

In opposition to [23] and works on ACLs [12, 19, 11], in this work social commitment is not taken primitive and is anchored in group attitudes. In particular, our notion of social commitment is grounded on the acceptance of the debtor and the creditor *qua* members of a certain institution.

Colombetti et al. [12] have developed an ACL semantics based on a *primitive* notion of commitment and its dynamics (through its states change). For example, when an agent promises to another one to perform an action, he creates a *pending commitment*, that becomes *fulfilled* (resp. *violated*) if the agent performs (resp. do not perform) the action. The commitment can also be *canceled* by the creditor. In Colombetti et al.'s account, nothing

[7]Castelfranchi also requires the condition of i and j's mutual knowledge about i's intention to perform α for j. We do not include this condition here since in our view it is too strong. Indeed, after a promise, the social commitment can persist even if the debtor drops the intention that he manifested.

is said about the consequences of the cancelation of a commitment. In our characterization, after the creditor utters that he does not want anymore the action to be performed by the debtor, the debtor's commitment is dropped. But the the debtor's obligation might remain instantiated. Indeed, formula $\neg SC_x(i, \alpha, j) \wedge IObl_{\{i,j\}:x}(i, \alpha, j)$ is satisfiable in our logic.

5 Application to the promise

Speech acts can be viewed as institutional facts represented by constitutive rules. In this section, we focus on a particular kind of speech act: the promissive. We first define the constitutive rule which creates the relation between an utterance (as a physical action) and the speech act *promise* (as an institutional action). Then, we specify the deontic dimension of promise. Finally, we establish the relationship between promise and social commitment.

According to [21], in the context of ordinary communication (noted OC) the speech act *promise* is defined on the basis of a constitutive rule of the form: a speaker's enunciation of the utterance: "I am going to perform action a" counts as a promise to the hearer to perform action a, under the condition that the hearer wants the speaker to perform action a. We here consider the constitutive rule defining the speech act *promise* with respect to the institutional context of ordinary communication (OC) as a global axiom [10]. For every $i, j \in AGT$ and $a \in AT$ we suppose that:

$$(Inf_{i,j}(a) \wedge Goal_j \langle i{:}a{:}j \rangle \top) \overset{OC}{\rhd} \mathbf{X}\, Promise(i,a,j) \tag{1}$$

where $Promise(i,a,j)$ is an atom denoting the *institutional fact* that i has promised to j to do action a for him. The previous constitutive rule says that: in the context of ordinary communication, for every agents i, j and atomic action a, i's act of informing j that is going to perform a for him, under the condition in which j wants i to do a, counts as a promise of i to j to perform a for him. The reason why we have $\mathbf{X}\, Promise(i,a,j)$ instead of $Promise(i,a,j)$ in the consequent of the previous counts-as assertion is that the atom $Promise(i,a,j)$ represents the institutional effect of i's act of informing j that is going to perform a for him. Thus, $Promise(i,a,j)$ must necessarily hold after the occurrence of i's act of informing.

Note that the previous constitutive rule defining the speech act *promise* in the context of ordinary communication can not be generalized to all institutional contexts. For example, in the context of the card game of poker the enunciation of the utterance: "I am going to perform a certain action" does not necessarily count as a promise of the speaker's to the hearer to perform the action in question. Indeed, in the context of poker players are allowed to bluff. More generally, the following formula should be acceptable for the context of poker: $\neg((Inf_{i,j}(a) \wedge Goal_j \langle i{:}a{:}j \rangle \top) \overset{Poker}{\rhd} \mathbf{X}\, Promise(i,a,j))$.

By definition, institutional facts are connected to a deontic dimension. In particular, an institutional fact is intrinsically connected to certain normative facts expressed in terms of obligations and permissions. For example, "being of age" is an institutional fact in many countries to which a certain number of permissions and obligations are associated (*e.g.* in many countries if you are of age you have the permission to vote and the obligation to fulfill the military duties). Therefore, since $Promise(i,a,j)$ is an institutional fact, it must be connected to certain normative facts. The following global axiom is given in order to establish such a connection with the normative level of *promise*. For every $i, j \in AGT$ and $a \in AT$ we suppose that:

$$Obl_{OC}(i, a, j, Promise(i,a,j)) \tag{2}$$

$Obl_{OC}(i, a, j, Promise(i,a,j))$ means that: in the context of ordinary communication, under the condition that i has promised to j to perform action a for him, i is obliged to perform action a for j.

Theorem 2

$$\vdash_{\mathcal{L}} (\neg Accept_{\{i,j\}:OC}\bot \wedge Accept_{\{i,j\}:OC}Goal_j \langle i{:}a{:}j\rangle \top \wedge$$
$$Accept_{\{i,j\}:OC}Promise(i,a,j)) \rightarrow SC_{OC}(i, a, j)$$

This theorem highlights the relationship between promise and social commitment in the context of ordinary communication: the fact that in the context of ordinary communication i and j accept that i has promised to agent j to do action a for him and that j wants i to do a for him entails the fact that i is committed to j to do a.

6 Conclusion

We have introduced a logic of acceptance, action and goal and provided an institution-based semantics of the commissive act of promise. As far as we know, the only approach which is similar to ours is Colombetti's approach [8] where speech acts are also modeled in terms of constitutive rules of the form "X count as Y" and of the institutional effects brought about their performances. However, in Colombetti's approach, there is no connection between institutional level and agentive level of speech acts and the relationship between constitutive rules defining speech acts and agents' attitudes is not investigated. In our approach this relationship is established by anchoring constitutive rules and obligations in agents' acceptances.

In the future will extend our analysis to directive speech acts (not treated here due to space restrictions). To this end, we will need to model the hierarchy between roles in an institution which entitles an agent playing a certain role (e.g. the employer of a company) to give orders to other agents playing other roles (e.g. the employees).

References

[1] A. Anderson. *A reduction of deontic logic to alethic modal logic*. *Mind*, vol. 22, pp. 100–103, 1958.

[2] J. L. Austin. *How To Do Things With Words*. Oxford University Press, 1962.

[3] G. Boella, L. van der Torre. *Regulative and constitutive norms in normative multiagent systems*. In *Proc. of KR 2004*. AAAI Press, pp. 255–266, 2004.

[4] M. E. Bratman. *Practical reasoning and acceptance in context*. *Mind*, vol. 101(401), pp. 1–15, 1992.

[5] C. Castelfranchi. *Commitment: from individual intentions to groups and organizations*. In *Proc. of ICMAS'95*. MIT press, pp. 528–535, 1995.

[6] L. J. Cohen. *An essay on belief and acceptance*. Oxford University Press, New York, USA, 1992.

[7] P. R. Cohen, H. J. Levesque. *Intention is Choice with Commitment*. *Artificial Intelligence Journal*, vol. 42(2–3), pp. 213–261, 1990.

[8] M. Colombetti, M. Verdicchio. *An analysis of agent speech acts as institutional actions*. In *Proc. of AAMAS'02*. ACM Press, pp. 1157–1164, 2002.

[9] F. de Saussure. *Course in General Linguistics*. Open Court Publishing, La Salle, Illinois, 1986.

[10] M. Fitting. *Proof Methods for Modal and Intuitionistic Logics*. D. Reidel, Dordrecht, 1983.

[11] R. A. Flores, P. Pasquier, B. Chaib-draa. *Conversational Semantics with Social Commitments*. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 14, pp. 165–186, 2007.

[12] N. Fornara, F. Viganò, M. Colombetti. *Agent communication and artificial institutions*. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 14, pp. 121–142, 2007.

[13] B. Gaudou, D. Longin, E. Lorini, L. Tummolini. *Anchoring Institutions in Agents' Attitudes: Towards a Logical Framework for Autonomous MAS*. In *Proc. of AAMAS'08*, to appear.

[14] D. Grossi, J.-J. C. Meyer, F. Dignum. *Classificatory aspects of counts-as: An analysis in modal logic*. *Journal of Logic and Computation*, vol. 16(5), pp. 613–643, 2006.

[15] P. Hakli. *Group beliefs and the distinction between belief and acceptance*. *Cognitive Systems Research*, vol. 7, pp. 286–297, 2006.

[16] D. Harel, D. Kozen, J. Tiuryn. *Dynamic Logic*. MIT Press, Cambridge, 2000.

[17] A. J. I. Jones, X. Parent. *Conventional signalling acts and conversation*. In *Advances in Agent Communication*, ed. F. Dignum, Springer, vol. 2922, pp. 1–17, 2004.

[18] E. Lagerspetz. *The opposite mirrors*. Kluwer, Dordrecht, 1995.

[19] A. U. Mallya, M. P. Singh. *An algebra for commitment protocols*. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 14, pp. 143–163, 2007.

[20] D. North. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press, Cambridge, 1990.

[21] J. R. SEARLE. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, New York, 1969.

[22] J. R. SEARLE. *The Construction of Social Reality*. The Free Press, New York, 1995.

[23] M. P. SINGH. *A social semantics for agent communication languages*. In *Issues in Agent Communication*, eds. F. Dignum, M. Greaves, Springer, vol. 1916, pp. 75–88, 2000.

[24] R. STALNAKER. *Inquiry*. MIT Press, Cambridge, 1984.

[25] R. TUOMELA. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge University Press, Cambridge, 2002.

[26] R. TUOMELA. *The Philosophy of Sociality*. Oxford University Press, Oxford, 2007.

CNRS - INPT - UPS - UT1 - UTM

ABSTRACT

The aim of this paper is to provide an analysis of speech acts as institutional actions, where an institution is grounded on the attitudes of its members. In the first part of the paper we present a logic of acceptance, goal and action. Then, we specify how agents can create and maintain normative and institutional facts on the basis of their acceptances *qua* members of a certain institution. Finally, we propose a formal characterization of the speech act *promise*.

KEYWORDS

Modal Logics, Institutions, Agent Communication Languages

UPS-IRIT, 118 route de Narbonne, 31062 Toulouse CEDEX 9

+33 (0) 561 55 67 65 info@irit.fr www.irit.fr