HAL
open science

An ACL for companion agents that can choose (not) to lie

Carole Adam, Benoit Gaudou

To cite this version:

Carole Adam, Benoit Gaudou. An ACL for companion agents that can choose (not) to lie. [Research Report] RR-LIG-043, LIG. 2013. hal-01471961

HAL Id: hal-01471961
https://hal.science/hal-01471961

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.

An ACL for companion agents that can choose (not) to lie

Carole Adam[1] and **Benoit Gaudou**[2]

Abstract. The principles of Agent Communication Languages (ACL) were rethought in 1998 when Singh [21] criticised mentalist approaches for relying on private mental attitudes and strong hypotheses incompatible with heterogeneous Multi-Agent Systems. We believe that now is the time to rethink them again to cope with hybrid MAS where artificial agents now have to communicate with humans. It is thus becoming important to differentiate between uttering words as an attempt to reach some goal, and actually succeeding to perform a speech act, since such an attempt can fail for various reasons. In this paper we thus propose a novel approach to the semantics of ACL that is more faithful to the philosophy of language, by distinguishing the locutionary and illocutionary levels of speech acts, as first defined by Austin and later studied by Searle and Vanderveken.

1 Introduction

Agent Communication Languages (ACLs) are almost as old as the development of Multi-Agent Systems (MAS). They have mainly been developed to allow cognitive agents to communicate in a common and rich language in an anthropomorphic way, as opposed to ant communication via pheromones. Cognitive agents can reason and plan their communication, so each speech act is defined by its pre-conditions (situations in which it can or cannot be executed) and post-conditions (what is created by its execution), in the same way as physical actions.

These pre- and post-conditions have first been encoded in terms of mental attitudes [7]. These so-called *mentalist approaches* work fine for homogeneous MAS where the internal architecture of all agents is known, and hypotheses of sincerity or cooperation make sense. However in heterogeneous MAS, the other agents cannot be assumed to even have mental attitudes, let alone be sincere and cooperative; the need to rely on private mental attitudes and to make such strong hypotheses thus prevents **interoperability**. These drawbacks of the mentalist approach led Singh to "rethink the principles" of ACL [21] and suggest to encode their semantics in terms of public social concepts instead. As a result of this *social approach*, commitments-based semantics of ACL were then developed [8]. Both approaches actually have their own advantages, in particular mental attitudes allow prediction of and reasoning about the agents' behaviour, while public social concepts bring verifiability. We thus believe that semantics of ACL should merge both types of concepts.

However, both approaches hit the same pitfall by imposing preconditions to the use of speech acts that constrain what the agents can say. This leads to three drawbacks. First it limits the agents' **autonomy**: these preconditions prevent agents from even attempting to perform a speech act that would fail (for example one that conflicts with their beliefs or that would create new conflicting commitments); so the agents cannot make their own decision about what speech acts they want to attempt, but are actually forced to respect dialogue conventions embedded in the ACL semantics and protocol. Second it makes the agents' less **anthropomorphic** since they can only use a limited range of speech acts. This can be prohibitive in itself for agents and robots that have a very humanoid look and might trigger a strong rejection phenomenon if their capabilities do not match the resulting expectations (uncanny valley phenomenon, see [23]), or for simulations where realism is important to immerse the user. Third and more importantly, agents are also limited in their ability to handle speech acts, since they are not equipped to cope with the reception of certain speech acts assumed to be impossible. This is not a problem when communicating with other similarly limited artificial agents, but the current trend is for agents to communicate more and more with humans (*e.g.* Embodied Conversational Agents or ECA [12]) who certainly cannot be constrained in what they say; in particular they might lie, be irrelevant or use irony; so such agents cannot **communicate with humans**.

Actually, as long as human speakers can physically utter a sentence, they might do so, regardless of any other conditions, but this sentence might fail to reach its goal. This had been accounted for in Speech Acts Theory [3] where speech acts are formed of three acts that can be attempted and performed separately, but this decomposition was lost in ACL semantics, making agents less anthropomorphic, less autonomous, and less interoperable. We thus argue that it is essential for the agent community to rethink the principles of ACL again and bring them closer to the original Speech Acts theory. The main contribution of this paper is to provide a new approach to the semantics of ACL where each speech act is decomposed into a locutionary and an illocutionary act, as defined by Austin. As a result, agents can attempt any speech act they will, but may only succeed an utterance (the locutionary act) while failing to perform the intended illocutionary act. Besides our semantics merges public and private concepts from both the mentalist and social approaches.

The article is structured as follows: Section 2 gives some background in the philosophy of language about the Speech Acts theory and introduces the problematic; Section 3.1 presents our idea to solve it and justifies our hypothesis; Section 3 defines our logical framework and uses it to formalise our approach in terms of logical axioms linking the different components of speech acts; Section 4 then applies this paradigm by providing an instantiation of the components of a few different types of speech acts. We conclude in Section 5 about the novelty and benefits of this new paradigm.

[1] University Joseph Fourier - Laboratory of Informatics, Grenoble, France, email: carole.adam@imag.fr
[2] University of Toulouse - Institute of Research in Informatics, Toulouse, France, email: benoit.gaudou@univ-tlse1.fr

2 Background and problematic

In this section we provide some background from the philosophy of language about speech act theory, and discuss how it was applied, in a simplified version, to define mentalist semantics of ACL.

2.1 Locutionary and illocutionary acts

Austin [3] introduced three levels of acts. A **locutionary act** consists in "uttering a certain sentence with a certain sense and reference" (*i.e.* with a certain meaning "in the traditional sense"); for example saying "open the door". Then **illocutionary acts** are "utterances which have a certain (conventional) force"; for example to order someone to open the door. Finally the **perlocutionary act** is "what we bring about or achieve by saying something"; for example convincing someone to open the door. Being actions, these three acts can be attempted without being achieved with success.

2.2 Illocutionary force

Then Searle [18] refined the notion of illocutionary act by decomposing it into the illocutionary force and propositional content, and later proposed a formal illocutionary logic with Vanderveken [19]. In particular they describe the illocutionary force as consisting of the seven following components:

- the illocutionary point, which is its internal purpose, and can only take one of five values: assertive, commissive, directive, declarative or expressive;
- the degree of strength of the illocutionary point that differentiates stronger or weaker acts having the same point;
- the mode of achievement, if this illocutionary act requires to be achieved in a certain way or under certain conditions;
- a set of propositional content conditions constraining the content of the act;
- a set of preparatory conditions that are necessary for the successful and non-defective performance of the speech act;
- a set of sincerity conditions that specify the psychological state that needs to be expressed (even if insincerely) for the speech act to succeed;
- the degree of strength of the sincerity conditions, which is the intensity of the psychological state that should be expressed.

2.3 Success and non-defectiveness

Searle and Vandeveken [20] have defined that for the performance of an illocutionary act to be **successful** in a certain context of utterance: the illocutionary point must be achieved on the content with its required characteristic mode of achievement and degree of strength; the expressed proposition must satisfy the propositional content conditions; the speaker must presuppose the preparatory conditions of F; and the speaker must express the psychological state of the sincerity condition with its characteristic degree of strength.

But this successful performance could still be defective (*e.g.* in case of false presupposition, or insincere expression). So they also defined that this performance is **non-defective** if in addition the presupposed preparatory conditions really hold and the speaker really has the expressed psychological state.

2.4 Agent Communication Languages

Speech Act Theory gave rise to the first semantics of ACL, namely the mentalist semantics, that match speech acts with their conditions and effects in terms of agents' mental attitudes, thus allowing prediction of and reasoning about the agents' behaviour. These semantics are therefore often expressed in BDI logics [16, 17], *i.e.* logics inspired by Bratman's work [4] and providing operators for mental attitudes (beliefs, desires, intentions). 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.

But despite their philosophical basis, these semantics of ACL only retained the illocutionary level of speech acts: a speech act is considered as an action, that can only be executed if the success conditions of the illocutionary act hold (FIPA-ACL feasibility precondition), so there is no locutionary level; then hypotheses allow to always deduce the perlocutionary effect (FIPA-ACL rational effect) so there is no perlocutionary level either. As a result (and contrarily to humans), the agents cannot attempt and fail an illocutionary act; when they attempt a speech act, they actually perform it, since the possibility of failure was ruled out beforehand by the feasibility precondition; so after this execution, the conventional affects of the speech act are always deduced, and hypotheses may even ensure that the perlocutionary effect is deduced too.

3 Our novel paradigm for the semantics of ACL

3.1 Hypotheses

In this paper we thus propose a novel paradigm for the semantics of ACL that is more faithful to Speech Acts theory, and whose main characteristic is to differentiates the locutionary and illocutionary acts in speech acts. Therefore agents can *utter* whatever sentence they wish (locutionary act), while further conditions constrain the successful performance of the illocutionary act. In particular the successful performance of each level does not guarantee the successful performance of the next level, and agents can utter a locutionary act even if the intended illocutionary act then fails. On this point we differ from all other semantics of ACL proposed so far. However we still have to make a number of simplifications that are detailed below.

No perlocutionary acts. We will decompose speech acts into the locutionary and illocutionary acts defined by Austin, but will ignore the perlocutionary act here. Indeed, Austin [3, p.109] said that "clearly any, or almost any, perlocutionary act is liable to be brought off, in sufficiently special circumstances, by the issuing, with or without calculation, of any utterance whatsoever, and in particular by a straightforward constative utterance." Searle and Vanderveken [20, p.11] also insisted that perlocutionary acts concern subsequent effects of speech acts (intentional or not) and thus cannot be conventionalised. As a result we will denote a speech act as an action α_{sa} that is composed of a locutionary act α_{loc} (the utterance) and an illocutionary act α_{illoc} (carrying a certain illocutionary force).

Literal use of language. As noted by Austin, these two acts are independent since a speaker can choose various utterances to perform the same illocutionary act, and the same utterance can express various illocutionary acts. In the following we will however suppose that agents use language literally, or non ambiguously, in the sense of [20, p.131], who define a literal performance of an illocutionary

act as "uttering a sentence which expresses literally that force and content in that context". In particular this literal sentence expresses the illocutionary point with its characteristic mode of achievement and degree of strength. So under this assumption the α_{loc} and α_{illoc} acts are linked, since there is only one locutionary act α_{loc} that literally expresses the illocutionary act α_{illoc}. This assumption makes sense for artificial agents that use a somewhat constrained language anyway; besides it is not too restrictive either to impose that humans have to speak literally to these agents to be understood.

Non-verifiability of the non-defectiveness of acts. As shown by the definitions above, the non-defectiveness of illocutionary acts depends on the actual state of the world and on the speaker's actual beliefs. As a result, it can never be verified[3] by the agents involved in the dialogue. Therefore in this paper we are only interested in, and will only formalise, the successful but possibly defective performance of speech acts. Besides, even if it is defective, a speech act can still be successful and achieve the intended effect.

3.2 Logical operators

We argue that to faithfully represent the various concepts of speech acts theory, we need to be able to manipulate agents' beliefs, choices, intentions, obligations and actions, but also what has been expressed by agents, which is crucial to represent a successful act without assumption about its non-defectiveness. We thus choose to use the BDI logic with Grounding operator defined by [10, 11], augmented with a deontic operator in SDL [15]. Below we list the various operators of our logic, but due to space limitations we refer the readers interested in their underlying standard possible world semantics or exhaustive axiomatics to the original papers cited for each of them.

Let $AGT = \{i, j, ...\}$ be a finite set of agents. We denote by $2^{AGT*} = 2^{AGT} \setminus \{\emptyset\}$ the set of all nonempty 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.

Our logic language is defined by the following BNF grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid Before_\alpha\varphi \mid After_\alpha\varphi \mid G\varphi \mid H\varphi \mid Choice_i\varphi \mid Grd_K\varphi \mid O\varphi$$

where p ranges over ATM, K over 2^{AGT*}, α over ACT and i over AGT. 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 $Bel_i\varphi$, $Intend_i\varphi$, $Done_\alpha\varphi$, $Happens_\alpha\varphi$, $F\varphi$, $Past\varphi$, $Perm\varphi$ and $Forbid\varphi$ will be defined as abbreviations.

3.2.1 Time and action operators

Time operators. $G\varphi$ 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 **S4.3**$_t$ ([5]; see [1] for detailed axiomatics). For convenience, we also introduce the dual operators: $F\varphi \stackrel{def}{=} \neg G\neg\varphi$ 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] Mentalist approaches were heavily criticised for grounding on non-verifiable concepts and we want to avoid this trap.

Action operators. We introduce $After_\alpha\varphi$ (reading "φ is true after every execution of action α") and $Before_\alpha\varphi$ (meaning "φ was true before every execution of action α") as primitive action operators, defined in a K$_t$ temporal logic (see [5] 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.

3.2.2 Mental attitude operators

Individual and public belief. $Grd_K\varphi$ means: "φ is publicly grounded for the group K" [9]. 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 [14]: $Grd_{\{i\}}\varphi$ (also written $Grd_i\varphi$ for the sake of readability) means: "i believes φ". The Grd_K operators are rational and public for every subgroup.

Choice. $Choice_i\varphi$ reads "agent i chooses (prefers) that φ". This operator is defined in a KD45 logic (see [13] for more details).

Intention. Intention is defined from choice [13]: 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): $Intend_i\varphi \stackrel{def}{=} Choice_i F\, Grd_i\varphi \wedge \neg Grd_i\varphi \wedge \neg Grd_i F\, Grd_i\varphi$. The formula $Intend_i\, Done_{i:\alpha}\top$ represents i's intention to perform action α.

3.2.3 Obligation operator

Deontic operators. The operator $O\varphi$ (reading "it is obligatory that φ") represents impersonal *obligation to be*, and is defined in Standard Deontic Logic [15], *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 appear in their content. For instance $O\,F\,Done_{i:\alpha}\top$ 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". In this paper we do not need to refer to any explicit institution imposing these obligations, but if that was the case this work could be extended by using the operators defined in [2].

3.3 Formalisation of speech acts

Under our hypotheses stated above, a speech act (denoted α_{sa} in the sequel) is characterised by a unique pair of a locutionary act ($\alpha_{loc} \stackrel{def}{=} \langle i, j, \mathsf{Utter}, \varphi\rangle$) and the associated illocutionary act literally expressed by it ($\alpha_{illoc} \stackrel{def}{=} \langle i, j, \mathsf{Force}, \varphi\rangle$), where i is the agent performing the act (the *speaker*); j is the *addressee* of the act, with $i \neq j$; Force is the *name of the act* (*i.e.* its *illocutionary force*); and φ is its *propositional content*. Thus **performing** a speech act consists in performing, or uttering, this locutionary act and also performing with success this attempted illocutionary act.

Only $Done_{\alpha_{loc}}\varphi$ (or more precisely $Before_{\alpha_{loc}}\varphi$) is defined as a primitive here, while the action operators of α_{sa} and α_{illoc} are abbreviations[4]. We will thus impose the following relation (expressing

[4] Formally, we should define precisely the accessibility relation for particular actions α_{sa} and α_{illoc} from the accessibility relation for actions α_{loc}, but this is out of the scope of this paper.

the links between locutionary, illocutionary and speech acts):

$$Done_{\alpha_{sa}}\top \stackrel{def}{=} Done_{\alpha_{loc}}\top \wedge Done_{\alpha_{illoc}}\top$$

Hypothesis. Moreover, we consider that locutionnary acts are public for the witnesses (the speaker and the hearer in our case) 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}:

$$Done_{\alpha_{loc}}\top \rightarrow Grd_{\{i,j\}}Done_{\alpha_{loc}}\top$$

Notations. An illocutionary act α_{illoc} is characterised by:

- Its illocutionary point $IlcPt$
- Its mode of achievement, that we represent by effects obtained after successful performance $AcvMdEff$
- Its propositional content conditions, that in our account will be logical constraints on the formulas that can be used as content, so we will not explicitly specify any propositional content conditions
- Its preparatory conditions $PrepCd$
- Its sincerity conditions $SincCd$

Remark. The **expression** of a mental attitude is formalised with the grounding operator Grd_K where K is the group of agents aware of the speech act, minimally containing the speaker i and the hearer j of this speech act. In particular the speaker expresses that it believes the preparatory condition: $Grd_K Bel_i PrepCd$, and expresses the sincerity condition: $Grd_K SincCd$ (being a mental state of i this is equivalent to $Grd_K Bel_i SincCd$ see [9] for more details).

3.4 Axioms describing the use of speech acts

3.4.1 Remark about consistency

The deduction of any public belief must be constrained by a condition ensuring that it is not inconsistent with what was previously established in the context of dialogue. The generic consistency condition for the deduction of a public belief $Grd_K\varphi$ is of the form $\neg Grd_K\neg\varphi$. For the sake of readability of the axioms below, we thus introduce an abbreviation for "consistent deductions", denoted $\rightarrow_c$ and defined as follows:

$$(\varphi \rightarrow_c Grd_K\psi) \stackrel{def}{=} (\varphi \wedge \neg Grd_K\neg\psi \rightarrow Grd_k\psi)$$

3.4.2 Axiom 0: utterance at will

If an agent is willing to perform a speech act α_{sa}, it can always **attempt** to do so, by performing the *corresponding* locutionary act α_{loc}, *i.e.* the utterance that literally expresses it:

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

3.4.3 Axiom 1: expressed mental states

By attempting to perform an illocutionary act α_{illoc} (by way of its literal expression by the corresponding α_{loc}), and even if this illocutionary act fails, an agent always expresses that when uttering α_{loc}, it had the intention to achieve the illocutionary point of α_{illoc} with its characteristic mode of achievement.

$$Done_{\alpha_{loc}}\top \rightarrow_c Grd_{i,j}Done_{\alpha_{loc}}Intend_i(IlcPt \wedge AcvMdEff)$$

The speaker also additionnally expresses (on condition that it is consistent) that it presupposes its preparatory conditions:

$$Done_{\alpha_{loc}}\top \rightarrow_c Grd_{i,j}Bel_i PrepCd$$

And finally it expresses (on condition once again that this is consistent) the mental states of the sincerity condition:

$$Done_{\alpha_{loc}}\top \rightarrow_c Grd_{i,j}SincCd$$

3.4.4 Axiom 2: deduced effects

If the speaker performed the locutionary act, with the appropriate achievement mode, and successfully expressed that he has the mental states of the sincerity condition and that he presupposes the preparatory conditions (*i.e.* if these were not inconsistent with what was previously expressed) then his locutionary act reaches some effects: it achieves its illocutionary point, with the added effects deduced from the specific mode of achievement. These two effects are deduced independently from each other, if they are consistent with the context.

$$Done_{\alpha_{loc}}\top \wedge Grd_{i,j}(SincCd \wedge Bel_i PrepCd) \rightarrow_c IlcPt$$

$$Done_{\alpha_{loc}}\top \wedge Grd_{i,j}(SincCd \wedge Bel_i PrepCd) \rightarrow_c AcvMdEff$$

3.4.5 Definition of successful performance

A speaker **successfully and non-defectively performs** an illocutionary act if: it achieves the illocutionary point with the characteristic mode of achievement and degree of strength; the propositional content condition is true; it presupposes the preparatory condition and this condition holds; and it expresses the sincerity condition and this condition holds. So we could define the successful non-defective performance of an illocutionary act as follows:

$$\begin{aligned} &Done_{\alpha_{illoc}}\top \stackrel{def}{=} \\ &\quad Done_{\alpha_{loc}}(PrepCd \wedge SincCd) \wedge IlcPt \wedge AcvMdEff \\ &\quad \wedge Grd_{i,j}(Bel_i PrepCd \wedge SincCd) \end{aligned}$$

But as we highlighted in our hypotheses above, a speech act can also be successful while still being logically defective (for example if the speaker presupposes false preparatory conditions or expresses insincere mental states), and this **possibly defective success** is the only thing that can actually be verified by the other agents, since what counts is what the speaker expresses and not what he actually believes. Therefore in our paradigm we consider $Done_{\alpha_{illoc}}\top$ to

represent this successful performance, whether the act is defective or not, and the correct definition is thus the following:

$$Done_{\alpha_{illoc}}\top \stackrel{def}{=} Done_{\alpha_{loc}}\top \wedge IlcPt \wedge AcvMdEff \wedge Grd_{i,j}(SincCd \wedge Bel_i PrepCd)$$

3.5 Theorems

From this definitions and our axioms we can prove the following theorems that express the consequences of the successful performance of an illocutionary act (be it defective or not).

Effects of successful performance If an illocutionary act is successfully performed, then as a result its illocutionary point is achieved with the additional effects of its specific mode of achievement:

$$Done_{\alpha_{illoc}}\top \rightarrow IlcPt \wedge AcvMdEff$$

Links between performance of locutionary and illocutionary act If an illocutionary act is successfully performed, then the associated (literal) locutionary act was performed too. (The converse is not true since the performance of an attempted illocutionary act can fail.)

$$Done_{\alpha_{illoc}}\top \rightarrow Done_{\alpha_{loc}}\top$$

4 Semantics of a new ACL in our paradigm

We now apply our paradigm by providing the semantics for a novel ACL based on its principles; *i.e.* we formally characterise the various components of the illocutionary force for a set of speech acts. So far we only have speech acts from three different categories: assertives (assert and inform), directives (direct and order), and commissives (commit and promise).

4.1 Assertives

According to Vanderveken [22, p.125], the force of assertion has:

- the assertive point ("which consists in representing an actual state of affairs" [22, p.105])
- the neutral mode of achievement
- the neutral propositional content condition
- the preparatory condition that the speaker has reasons or evidence for the truth of the propositional content
- the sincerity condition that the speaker believes the propositional content
- the neutral degree of strength

Having reasons or evidence to believe φ is part of the philosophical definition of believing φ[5] so we will not impose a preparatory condition for `Assert` , as it is implied by its sincerity condition. So we obtain the following logical characterisation of the primitive assertive force `Assert` :

- $IlcPt \stackrel{def}{=} Grd_K Bel_i \varphi$
- $AcvMdEff \stackrel{def}{=} \top$
- $PrepCd \stackrel{def}{=} \top$
- $SincCd \stackrel{def}{=} Bel_i \varphi$

The `Inform` illocutionary act [22, p.175]: "to inform is hearer directed in that it is to assert with the preparatory condition that the hearer does not already know P." So the only difference from `Assert` is the preparatory condition:

- $PrepCd \stackrel{def}{=} \neg Bel_j \varphi$

4.2 Commissives

The primitive commissive illocutionary force ([22, p.125]) has:

- the commissive point ("which consists in committing the speaker to a future course of action" [22, p.105])
- the neutral mode of achievement and degree of strength
- the condition that the propositional content represents a future course of action of the speaker
- the preparatory condition that the speaker is capable of carrying out that action
- the sincerity condition that he intends to carry it out

The preparatory condition that the speaker is capable to carry out the action is expressed in our formalism by beliefs. More precisely, the speaker is deemed capable of carrying out the action if both he and the hearer envisage that he will carry out this action at some point in the future. This ensures that the speaker can be mistaken when presupposing this condition (which would not be the case with his own beliefs), and that this condition is consistent with the speaker's intention to carry out this action (which implies that the action is not believed to happen anyway). So we obtain the following logical characterisation of the primitive commissive force `Commit` :

- $IlcPt \stackrel{def}{=} Grd_K Intend_i Done_{i:\alpha}\top$
- $AcvMdEff \stackrel{def}{=} \top$
- $PrepCd \stackrel{def}{=} \neg Bel_i \neg F Happens_{i:\alpha}\top \wedge \neg Bel_j \neg F Happens_{i:\alpha}\top$
- $SincCd \stackrel{def}{=} Intend_i Done_{i:\alpha}\top$

The `Promise` act is different from the primitive commissive force on several points [22, p.182]:

- it is always made to someone, and has the special preparatory condition that the promised action is good for this someone;
- it involves the explicit undertaking of an obligation (that may remain implicit in other types of commissives), and this increases the degree of strength of the sincerity conditions.

So formally the differences between `Promise` and `Commit` are:

- $AcvMdEff \stackrel{def}{=} OF Done_{i:\alpha}\top$
- $PrepCd \stackrel{def}{=} Choice_j Done_{i:\alpha}\top$

4.3 Directives

The primitive directive illocutionary force ([22, p.125]) has:

- the directive point ("which consists in making an attempt to get the hearer to do something", [22, p.105])
- the neutral mode of achievement and degree of strength
- the condition that the propositional content represents a future course of action of the hearer
- the preparatory condition that the hearer can carry out that action

[5] See [6] for example: "beliefs are shaped by evidence for what is believed"

- the sincerity condition that the speaker desires or wants the hearer to carry it out

This preparatory condition is similar to the one for commissive speech acts, except for the fact that it is the hearer carrying out the action. So we logically characterise the `Direct` speech act as:

- $IlcPt \stackrel{def}{=} Grd_K Intend_i Done_{j:\alpha} \top$
- $AcvMdEff \stackrel{def}{=} \top$
- $PrepCd \stackrel{def}{=} \neg Bel_i \neg F Happens_{j:\alpha} \top \wedge \neg Bel_j \neg F Happens_{j:\alpha} \top$

Being the primitive directive, `Direct` does not have a special mode of achievement. In particular it does not make explicit if the hearer has an option to refuse or not. Directives that explicitly give a refusal option (*e.g.* `Request`) have a specific polite achievement mode so that the hearer has the option to accept or refuse:

- $AcvMdEff \stackrel{def}{=} Grd_{i,j} Bel_j \neg OF Done_{j:\alpha} \top$

On the contrary to tell someone to do something is to direct them with a more peremptory mode of achievement, *i.e.* without allowing the option of refusal [22, p.194]; so the hearer cannot grant or refuse, but can only obey or disobey. So logically speaking the difference between `Tell` and `Direct` is the specific mode of achievement, whose result is to express that j publicly believes he has to accept what he was told to do (which is not equivalent to j being actually obliged to obey, but prevents him from refusing):

- $AcvMdEff \stackrel{def}{=} Grd_{i,j} Bel_j OF Done_{j:\alpha} \top$

Then to give an order is to invoke a position of authority to add strength to `Tell` , but still without creating the obligation to obey. Some even stronger directives (such as `Dictate`) can actually create the obligation to obey, but they depend on the speaker actually having some institutional power or authority over the hearer, and we do not have the logical operators to express these concepts here (see [2] for a possible formalisation of these institutional concepts).

5 Conclusion and perspectives

In this paper, we have proposed a novel paradigm for semantics of ACL that allows agents to attempt to perform any speech act in the same way that humans do, thus increasing their autonomy, and that describes in which context the corresponding illocutionary act is successful. We also provided an instantiation of this paradigm by describing a particular ACL containing so far only the three most widely used classes of speech acts, namely directives, promissives and assertives.

Artificial agents using our ACL are thus enabled to lie or be irrelevant, and can thus also handle such speech acts on the part of other agents. This is essential to allow them to communicate with humans, who cannot be assumed to conform to strict interaction protocols. This approach will therefore be useful for the new generation of agents who communicate mainly with humans, such as museum guides, pedagogical agents, or companion agents. Besides, their wider range of allowed speech acts will make them much more unpredictable and thus engaging on the long term. With this application in mind, our 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 ACL will thus allow for much more realistic and engaging agents, able to express emotions they do not have or that are inconsistent with what they say.

REFERENCES

[1] Carole Adam, Andreas Herzig, and Dominique Longin, 'A logical formalization of the OCC theory of emotions', *Synthese*, **168**(2), 201–248, (2009).

[2] Carole Adam, Andreas Herzig, Dominique Longin, and Vincent Louis, 'Unifying the intentional and institutional semantics of speech acts', in *DALT*, (2008).

[3] J.L. Austin, *How to Do Things with Words*, Harvard Uni. Press, 1962.

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

[5] J.P. Burgess, 'Basic tense logic', in *Handbook of philosophical logic*, volume 7, Kluwer Academic, (2002).

[6] Pascal Engel, 'Believing, holding true, and accepting', *Philosophical Explorations*, **I**(2), 140–151, (1998).

[7] FIPA, 'FIPA Communicative Act Library Specification', `http://www.fipa.org/specs/fipa00037/`, (2002).

[8] Nicoletta Fornara, Franscesco Viganò, and Marco Colombetti, 'Agent communication and artificial institutions', *JAAMAS*, **14**, 121–142, (2007).

[9] Benoit Gaudou, *Formalizing social attitudes in modal logic*, Ph.D. dissertation, Universite de Toulouse, juillet 2008.

[10] Benoit Gaudou, Andreas Herzig, and Dominique Longin, 'Grounding and the expression of belief', in *Proc. KR 2006*, eds., Patrick Doherty, John Mylopoulos, and Christopher A. Welty, pp. 211–229, Windermere, UK, (2006). AAAI Press.

[11] Benoit Gaudou, Andreas Herzig, and Dominique Longin, 'Logical formalization of social commitments: Application to ACL (poster)', in *AAMAS*, (2009).

[12] Jonathan Gratch and Stacy Marsella, 'Some lessons for emotion psychology for the design of lifelike characters', *Journal of Applied AI*, **19**(3–4), 215–233, (2005).

[13] Andreas Herzig and Dominique Longin, 'C&L intention revisited ', in *KR*, pp. 527–535. AAAI Press, (2004).

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

[15] A. Jones and J. Carmo, 'Deontic logic and contrary-to-duties', in *Handbook of philosophical logic*, volume 8, 265–343, Kluwer Academic, (2002).

[16] A.S. Rao and M.P. Georgeff, 'Modeling rational agents within a BDI-architecture', in *KR*, (1991).

[17] David Sadek, 'A study in the logic of intention', in *KR*, pp. 462–473. Morgan Kaufmann Publishers, (1992).

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

[19] John Searle and Daniel Vanderveken, *Foundations of illocutionary logic*, Cambridge University Press, Cambridge, England, 1985.

[20] J.R. Searle and D. Vanderveken, 'Speech acts and illocutionary logic', in *Logic, Thought and Action*, ed., D. Vanderveken, volume 2 of *Logic, Epistemology, and the Unity of Science*, chapter 5, 109–132, Springer, (2005).

[21] Munindar P. Singh, 'Agent communication languages: Rethinking the principles', *Computer*, **31**(12), 40–47, (1998).

[22] Daniel Vanderveken, *Meaning and Speech Acts*, volume 1: Principles of language use, Cambridge University Press, 1990.

[23] Michael L. Walters, Dag S. Syrdal, Kerstin Dautenhahn, RenÃľ T. Boekhorst, and Kheng L. Koay, 'Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion', *Autonomous Robots*, **24**(2), 159–178, (February 2008).