The Prague Bulletin of Mathematical Linguistics
NUMBER 94 SEPTEMBER 2010 35–56

Pragmatically-Motivated Utterance Fine-Tuning in Human-Computer Dialogue

Vladimir Popescu[a], Jean Caelen[b], Corneliu Burileanu[c]

[a] Laboratoire Informatique d'Avignon, University of Avignon, France
[b] Laboratoire d'Informatique de Grenoble, CNRS, France
[c] Politehnica University of Bucharest, Romania

Abstract

The naturalness and user-friendliness of the utterances generated by the computer when engaging in dialogue with humans is a key point for the success of spoken language interaction-based computer applications. This article addresses the issue by proposing a mechanism for controlling the strength of the illocutionary force conveyed by the utterances produced by the machine. The degree of strength for a speech act roughly quantifies the pressure that this act puts on its recipient. Few research efforts are reported for controlling this pressure in utterance generation. This article provides the means to adjust this force, relying on the discourse structure of the dialogue and on the public commitments that the speakers make during the dialogue. After a concise statement of the formal framework, the proposed approach is presented in a detailed manner and qualitatively assessed via relevant examples.

1. Introduction

In this article we address a particular issue in answer generation for human-computer dialogues. We are concerned with fine-tuning the degree of strength for the utterances generated by the computer in dialogues, so that a natural dialogue is achieved. Tuning this degree of strength is important not only for social reasons, but also for psychological reasons, related to the level of motivation that an utterance determines in its addressee. Hence, this degree of strength – named, in line with (Vanderveken, 1990-1991), *illocutionary force degree*, reflects, from a pragmatical viewpoint, the illocutionary goal set by a speaker when she produces an utterance[1]. This illocutionary goal has, certainly, perlocutionary effects on the hearer, in that it

[1] We will henceforth designate the speaker by *she*, and the hearer by *he*.

Corresponding author: vladimir.popescu@imag.fr
Cite as: Vladimir Popescu, Jean Caelen, Corneliu Burileanu. Pragmatically-Motivated Utterance Fine-Tuning in Human-Computer Dialogue. The Prague Bulletin of Mathematical Linguistics No. 94, 2010, pp. 35–56.
doi: 10.2478/v10108-010-0019-1.

may determine his further dialogue turns. Although the importance of this aspect in linguistic interaction has been pointed out in several studies (Vanderveken, 1990-1991; Faller, 2006), the actual way the illocutionary force degree might be controlled was, to our knowledge, not specifically addressed. Therefore, the goal of this article is to specify a formal framework allowing us to compute the illocutionary force degree for an utterance that the machine is due to produce.

For computing the illocutionary force degree, we rely on two key ideas: (i) using (Maudet et al., 2004, 2006)'s work for computing public commitments starting from SDRT ("Segmented Discourse Representation Theory") rhetorical relations; (ii) using (Vanderveken, 1990-1991)'s ideas for defining and formalizing the notion of illocutionary force degree, adapting it to human-computer dialogue. Moreover, by the "agnostic" attitude that we adopt with respect to the beliefs and intentions of the agents, we follow a research agenda along the lines of (Kibble, 2006b,a). Indeed, we distance ourselves from the mentalistic approaches derived from the BDI ("Belief-Desire-Intention") framework (Cohen and Levesque, 1990b,a; Cohen and Perrault, 1979). We chose to do so, since computing intentions of the agents is considered to be rather unreliable, with respect to the social commitments, that are readily available (Kibble, 2006b, 2007).

The main contribution of this article consists in combining points (i) and (ii) above. In order to do this, we start from a set of assumptions. First, we assume that each dialogue partner (more specifically, the computer and the human speaker) has a *commitment store* that contains the semantic forms for the utterances each speaker overtly commits to during conversation (Kibble and Piwek, 2007), along with the rhetorical relations between them; the latter are accounted for in a framework inspired from SDRT (Asher and Lascarides, 2003). We posit that, from the viewpoint of the aspects tackled in this article, the goal of the machine is to achieve *dialogue* success, as opposed to *task* success, which is the responsibility of the dialogue and task managers (Caelen and Xuereb, 2007). Concerning speakers' commitments, dialogue success is achieved if the machine's and user's commitments are not logically inconsistent. A sufficient (albeit restrictive) and, in our opinion, easier to check, condition that guarantees the consistency of the commitment stores is their equivalence. Informally stated, the more the commitment stores are logically equivalent, the more the dialogue is successful. Although this hypothesis does not hold for the general case of human dialogues, it is supported, for service-oriented dialogues by empirical evidence brought by experiments performed using an artificial dialogue agent, in the context of a room reservation dialogue system (Nguyen, 2005). Unlike Maudet *et alia*, who consider the viewpoint of an *external* observer of the dialogue (Maudet et al., 2006), we adopt the viewpoint of the machine. This implies that the user's and machine's commitment stores are different if and only if there exists a semantic form that is either in the commitment store of the machine and not in the user's commitment store, or vice versa. In this situation, the machine has to adjust the illocutionary force degree[2]

[2]In line with (Caelen and Xuereb, 2007) and (Vanderveken, 1990-1991), we consider the notions of *speech act type* and *illocutionary force* as synonymous.

for this utterance, so that either the user commits to it, or the machine concedes to retract its commitment to this semantic form.

From the assumptions shown above, a general adjustment policy for the degree of illocutionary force can be derived, by taking into account the way the machine (as a speaker) wants to act on the commitments of its interlocutor: first, if the machine wants the interlocutor to integrate something in her/his commitment store, then it produces an utterance having a strong degree of force (that is, stronger than a "neutral" degree of force, usable for pure assertions, irrespective of the status of the commitments); second, if the machine wants its interlocutor to retract something from her/his commitment store, then it produces an utterance with an even stronger degree of force; this utterance is supposed to determine the interlocutor to give up one of her/his commitments. On the other hand, if the machine wants to signal its interlocutor that it integrated the contents of one of her/his utterances in its commitment store, it produces an utterance with a weak degree of force. Moreover, if the machine wants to signal that it concedes to retract one of its commitments (thus endorsing one of its interlocutor's commitments), then it produces an utterance with an even weaker degree of force.

Concerning previous work related to the research described in this article, we can mention (Traum et al., 2003)'s dialogue model, implemented with conversational agents that interact in a multi-party dialogue context, with virtual and human speakers, in several applications, ranging from military troops training, to interactive story-telling. This approach addresses two different aspects of the interaction: (i) the *discursive* character of the agents' linguistic contributions – Traum *et al.* do not explicitly model rhetorical constraints imposed by dialogue history on the speech turns, this is why their approach only *simulates* rhetorical capabilities for the agents; (ii) the *intentionality* of the dialogue contributions – Traum *et alia*'s approach indeed succeeds in altering the surface form of communicative intentions, but they do this by relying on speakers' *emotional attitudes*, which requires access to their intentional state. This can only be done reliably in very limited contexts (concerning a specified task), the approach does not seem readily transposable to different tasks, in that task-specific knowledge is not separate enough from task-independent knowledge.

Another relevant work is (Gupta et al., 2007)'s POLLY system: the authors propose a system architecture for controlling the surface form of the utterances in dual-party human-system conversations. For this, the authors use (Brown and Levinson, 1987)' Politeness Theory. Gupta *et alia* rely on pragmatic elements for choosing an adequate surface form for an utterance that is specified in a deep form (as a plan, i.e. a set of preconditions and actions to realize if the preconditions are satisfied). The authors use constraints imposed by a speaker's *menaces*, which basically boils to taking into account three contextual factors: (i) the power that the hearer has over the speaker, (ii) the social distance between the speaker and the hearer, and (iii) an imposing level of the speech act realized via an utterance. Starting from Brown and Levinson's Politeness Theory, Gupta *et alia* propose a fine-grained control of the surface form of the utterances, according to the type of the menace that the speaker undergoes. The passage from these politeness specifications (called "strategies" in Politeness Theory) to linguistic forms is realized in a rather *ad-hoc* manner, since predefined lexical items are "collated" to the prototypical, neutral, linguistic form of an utterance. Unlike in POLLY, we propose to take into account other facets of the dialogue context, namely discourse structure and public commitments, in order to handle basically the same phenomenon – choosing a contextually-adequate linguistic form, between several alternatives for expressing the same semantic content.

In view of these previous research efforts however, we do not claim that the usage of the users' commitments as a decision criterion in computing the degree of illocutionary force yields better performances than politeness elements, or than emotion-related elements. We only try to show that the rhetorical structure, a resource that is already computed and used for other purposes (computing speech act types (Xuereb and Caelen, 2005), generating discourse connectors in multi-sentential utterances (Popescu and Caelen, 2009), or elliptic constructions (Popescu et al., 2008)), can also be used for acting on the degree of illocutionary force. The issue of *reusing* already available (computed) reasources is of major importance in dialogue systems, where timely response to users's demands is essential from an ergonomical standpoint. However, the issue whether users' commitments are, in general, an appropriate ressource to be taken into account when computing degrees of illocutionary force falls out of the scope of this article. What we do show though, is that there is a certain empirical motivation behind our approach.

This article is structured as follows: in Section 2 we introduce the setting where our contributions are made, i.e. task-oriented dialogue systems; in Section 3 we describe in detail the mechanisms used for adjusting the illocutionary force degree in human-computer dialogue; these mechanisms are also validated on several typical dialogue situations concerning a book reservation service. Lastly, a conclusive discussion ends up the article.

2. Research Context: Pragmatic Control of Utterance Generation

The work described in this article is developed in the context of a formal framework for controlling certain aspects of utterance production in human-computer dialogue. These aspects are either pragmatic in nature (such as the degree of strength that is assigned to a speech act type that is performed by an utterance), or semantic (such as the semantic ellipsis control in utterances). Both these sides are controlled by leaning on a formal account of the rhetorical structure of dialogues: the interaction has to take place so that the *global coherence* of the dialogue is maximized (Asher and Lascarides, 2003). On the other hand, an account of *local relevance* of the utterances is used in order to foster the global coherence; namely, the duality between utterance-level speech act types and rhetorical relations that connect pairs or larger sets of utterances is taken into account (Caelen and Xuereb, 2007), in order to improve the rhetorical structuring of the dialogue.

Thus, in this article we address the issue of how the commitment stores of the dialogue partners (Asher and Lascarides, 2008) trigger decisions regarding the adjustment of the degree of strength for the illocutionary force performed by the utterance. More specifically, we give an account on how illocutionary force degrees are computed from the interplay between the commitments of the speakers in dialogue.

Before continuing with the presentation of our formal framework for controlling the illocutionary force degree, we introduce a set of notational conventions: (i) U and M denote the dialogue partners, a user and the machine, respectively; (ii) L denotes a generic speaker; in our framework, this means either M, or a human speaker, U; (iii) CS_L denotes the commitment store of speaker L; (iv) π_i are utterance labels, for a positive integer i; (v) $K(\pi)$ represents the semantic form for utterance π (expressed in a first-order logic); (vi) Σ_ρ represents the semantics for the rhetorical relation ρ, expressed in the same first-order logic as the utterances; (vii) $\mathsf{emitter}(\pi)$ is a discourse function that returns the identity of the speaker that had produced utterance π; (viii) $\mathsf{equals}(\alpha,\beta)$ is a binary discourse predicate that is true if variable α is bound to the value β; (ix) the operation $\leftarrow_\pi$ states that the left-hand part is updated with the right-hand part, via utterance π.

3. Pragmatics-based Illocutionary Force Tuning

3.1. Formalizing Vanderveken's Illocutionary Force Degree

In formalizing the notion of degree of illocutionary force, we rely somehow on previous work of (Vanderveken, 1990-1991). Thus, in an utterance generation context, we assume that the input to the generator is a logical expression of the form $F(p)$, where

- F is an *illocutionary force* in Vanderveken's terminology, or a speech act type, in Caelen's terminology: F^{F} ("make-do", a directive act), F^{S} ("make-know", an act of informing the hearer on something), F^{FS} ("make-do-know", a request for information from the hearer), F^{P} ("make-can", an act of providing the hearer with several choices), et F^{D} ("make-must", an act of imposing an obligation on the hearer) (Caelen and Xuereb, 2007);
- p is a *propositional content* in Vanderveken and Caelen's terminologies; in the framework proposed here, this propositional content is identified with the logical formula that expresses the semantics of the utterance that realizes the act F; such a propositional content is denoted by $K(\pi)$, as above, if the label of utterance $F(p)$ is π.

Like in the work of Vanderveken and Caelen, we assume that each speech act type F can be realized (via an utterance) with a certain degree of strength; we call this degree *illocutionary force degree* and we denote it by $\partial\phi$. Thus, the relation between an utterance, its propositional content, its illocutionary force and the degree of this force can be seen as a series of functional compositions:

$$\pi \to^K K(\pi) \to^F F(K(\pi)) \to^{\partial\phi} \partial\phi(F(K(\pi)) \times K(\pi)),$$

which can also be written as $(\partial\phi \circ (F \circ K) \times K)(\pi)$. We explicitly present below each of these functions.

The K function assigns meanings (expressed in a first-order logic) to utterances and to rhetorical relations; thus, it maps labels to logical formulas:

$$K : \Pi \cup P \to \Lambda\ ,$$

where $\Pi = \{\pi_1, ...\}$ and $P = \{\rho_1, ...\}$ are two sets of labels – for utterances and for rhetorical relations, $\Lambda = \{\lambda_1, ...\}$ is a set of logical formulas.

Thus, the K function is computed, when utterance generation is concerned, by the dialogue controller and, when user utterance analysis is concerned, by the semantic analyzer. However, we assume that the representations produced by the semantic analyzer and by the dialogue controller are coherent, i.e. the same set of predicates (in the task ontology) is used. When the K function takes the label of a rhetorical relation as argument, its operation is equivalent to a simple search of the semantics of the rhetorical relation under discussion.

Function F assigns speech act types to (the semantics of) utterances; this function is undefined on the semantics of the rhetorical relations:

$F : \Lambda \setminus \{\Sigma_\rho : \rho \in P\} \cup \to \{\mathrm{F^F}, \mathrm{F^S}, \mathrm{F^{FS}}, \mathrm{F^P}, \mathrm{F^D}\}$;

here, $\{\Sigma_\rho : \rho \in P\}$ represents the set of semantics for all known rhetorical relations; if function F is applied on a formula $K(\pi)$ such that it is the machine that produced utterance π (hence, $\mathsf{equals}(M, \mathsf{emitter}(\pi))$), then it is the dialogue controller that computes the F function.

Function $\partial\phi$ assigns degrees of strength to speech act types, that are in turn applied to the logical forms of utterances; thus, $\partial\phi$ maps speech act types to integers.

In order to endow the utterances with more variability, we stipulate that function $\partial\phi$ can have, for the same speech act type, effects that depend on the propositional content of the utterance, $K(\pi)$. Thus, we define function $\partial\phi$ on the Cartesian product between the set of speech act types and the set of propositional contents:

$\partial\phi : \{F\} \times (\Lambda \setminus \{\Sigma_\rho : \rho \in P\}) \to \{-2, -1, 0, 1, 2\}$,

where $\{F\}$ and Λ denote, respectively, the set of speech act types, and the set (theoretically infinite, but practically limited by the particular task considered) of logical formulas that can be generated from a specified (e.g. in the task ontology) set of predicates. We assign five possible levels to the illocutionary force degree, classifying them in **very strong**, **strong**, **neutral**, **weak** and **very weak** ; -2 and -1 represent very weak and weak forces respectively, 0 represents the neutral force, and 1 and 2 represent strong and very strong forces, respectively.

The goal of this article is to define "analytically" the $\partial\phi$ function. Before doing this, we are first going to motivate the illocutionary force degree via an example. We will thus show that there is a relevant relationship between the illocutionary force degree that is assigned to an utterance, and its linguistic form. Consider for instance a dialogue between a user U and a machine M, where the user says:

U: Sorry, can you tell me, please where I can find book 'B'?

The dialogue manager in M produces, as a response to U's question, a communicative intention of informing the user that the book can be found on the first floor, to the left; this communicative intention is expressed, in logical form, and offered to the generator (π denotes the label of this utterance):

$\mathrm{F^S}(\exists X, Y, Z, T, U, V : \mathsf{agent}(X) \wedge \mathsf{equals}(X, \neg\mathsf{emitter}(\pi)) \wedge \mathsf{object}(Y) \wedge \mathsf{equals}(Y, \text{'book'}) \wedge \mathsf{feature}(Y, Z) \wedge \mathsf{equals}(Z, \text{'title'}) \wedge \mathsf{equals}(Z, \text{'B'}) \wedge \mathsf{feature}(Y, T) \wedge \mathsf{equals}(T, \text{'location'}) \wedge \mathsf{feature}(T, U) \wedge \mathsf{equals}(U, \text{'level'}) \wedge \mathsf{equals}(U, \text{'1'}) \wedge \mathsf{feature}(T, V) \wedge \mathsf{equals}(V, \text{'direction'}) \wedge \mathsf{equals}(V, \text{'left'}))$.

Some possible linguistic forms, for this communicative intention and different degrees for the illocutionary force F^S, are (linguistic markers associated to illocutionary force degrees are written in boldface):

- $\partial\phi(F^S(K(\pi)), K(\pi)) = -2$:
  M: On the first floor, to the left, **I think**.
- $\partial\phi(F^S(K(\pi)), K(\pi)) = -1$:
  M: **Wait a minute please... here it is:** book 'B' is on the first floor, to the left.
- $\partial\phi(F^S(K(\pi)), K(\pi)) = 1$:
  M: You can **certainly** find book 'B' on the first floor, **just** to the left.
- $\partial\phi(F^S(K(\pi)), K(\pi)) = 2$:
  M: **Listen** you can **certainly** find book 'B' on the first floor, **just** to the left, **do you understand?**

Given that the speech act type considered is an F^S (provision of information), all these forms stem from a neutral illocutionary force degree:

$\partial\phi(F^S(K(\pi)), K(\pi)) = 0$:

M: You can find book 'B' on the first floor, to the left.

After having seen this example, we observe that, from an answer generation viewpoint, there are two issues to deal with, concerning the illocutionary force degree: (i) computing the appropriate illocutionary force degree, for a given utterance, specified in logical form, and (ii) mapping a specified illocutionary force degree to an appropriate linguistic form. In this article, only the first issue is explicitly addressed. The second problem, of mapping specified illocutionary force degrees to surface forms, is tackled by manual annotation of canned utterances.

Before introducing the illocutionary force handling mechanisms, we discuss on the adequacy of the five-level scale for the degree of illocutionary force. In our view, this issue has two facets. First, why a discrete and not a continuous scale? This has been argued for at length in several articles by Searle and Vandeveken *inter alia*, e.g. (Searle and Vanderveken, 2006), or (Motsch, 1980). Essentially, a discrete scale seems to be rather a theoretical choice, partly supported by the idea that language is a discrete phenomenon, than a necessary constraint. Second, why a five-level scale? An answer is that the choice of five levels is arbitrary (in theory, one can have a denumerable set of levels if we wish, but for a practical system one has to make a choice for a finite number of degrees). However, we chose five levels that correspond, roughly, to the informal (and intuitive) distiction between very weak, weak, neutral, strong and very strong degrees. Moreover, given that in the following subsection we discuss and illustrate the coupling between the relations between interlocutors' commitments and the very weak/weak/strong/very strong degrees of force, the -2 to $+2$ scale seems a methodologically convenient choice, i.e., canned phrases can easily be annotated in this way, so that an answer generation module can choose a suitable utterance (i.e., parameterized by the required degree of force).

3.2. Illocutionary Force Handling Mechanism

The main idea that guides the computation of the illocutionary force degree stems from the connection between public commitments and speech acts. As we emphasized before, we rely on (Maudet et al., 2006)'s framework for computing commitment stores from discourse structure.

For each user U in a dialogue, there exists a commitment store CS_U that contains the semantics of the utterances that U has produced, along with the semantics of the machine's utterances, that U has agreed with (this is indicated by rhetorical relations between these utterances and utterances of U), and finally, along with the *negated* semantics of the machine's utterances that U did *not* agree with, along with the rhetorical relations that emphasize this fact, e.g. *P-Corr* (Plan Correction) or *Contrast* (Asher and Lascarides, 2003). For example, consider the following dialogue, between a human user U and the machine M, which simulates a librarian:

M: You can still borrow three books, sir!
U: So, I can take this one as well?
M: Yes, you can take it, sir.

This interaction contains a question of U, that is in an $Elab_q$ relation to the first utterance of M; the subsequent answer of M is in an *Elaboration* relation to the first utterance, since, indeed the two turns of M achieve the same effect (from the point of view of the task that the dialogue tries to help resolving) as a unique turn of M:

M: You can still borrow three books, so, for instance, you can take book 'X' that you want.

where book 'X' and 'this one' in the user's question, refer to the same object in the physical world.

The way that commitment stores are defined and used is inspired from (Maudet et al., 2006); for instance, in the previous dialogue example, the commitment store of U, after she had asked the question, is a set:

$CS_U = \{K(\pi_1), K(\pi_2), \Sigma_{Elab_q(\pi_1,\pi_2)}\}$,

where π_1 and π_2 denote the first utterance of M and the first utterance of U (the question) respectively, the function $K(\pi)$ denotes the semantic content of utterance π, and $\Sigma_{Elab_q(\pi_1,\pi_2)}$ denotes a Prolog-style semantics of the rhetorical relation $Elab_q(\pi_1, \pi_2)$, which specifies that utterance π_2 is a question such that any relevant answer elaborates on utterance π_1 (Asher and Lascarides, 2003).

The notion of commitment store and the way such a structure is computed for each speaker in dialogue could have been given a better account, by using the notion of *common ground*. More specifically, we could have relied on the idea that common ground in dialogue can be seen as the joint entailments of all the speakers' public commitments (Lascarides and Asher, 2009). Thus, commitments could have been seen as SDRSs, one such discourse structure for each speaker. Moreover, in an adequate framework, the commitments should be computed by taking into account the *preferences* of the speakers, which are learned from and affected by conversational moves. The dynamic interplay between preferences and conversational actions could be seen as a game-theoretical problem and modeled as such (Asher and Lascarides, 2008). However, we believe that in the context of this article, where the emphasis is put on the way these commitments are *used* for computing illocutionary force degrees, the approximation provided above for the commitment stores is sufficient as a departure point.

Then, for computing a scalar value for the illocutionary force degree, we rely on the intuition that the closer the speakers' public commitments are, the closer to zero (a *neutral* value) the illocutionary force degree is. In other words, in an ideal situation where U and M share exactly the same commitments throughout the dialogue, the machine can express its utterances in a neutral manner, from an illocutionary standpoint. That is, M assigns a zero degree to the illocutionary force.

For most of the speech turns in dialogue we have that $CS_U \not\equiv CS_M$, i.e., U and M's commitment stores are not equivalent[3] after each speech turn. This non-equivalence results in several possible cases, concerning the relationship between the interlocutors' commitments:

1. $CS_M \subset CS_U$, that is, M's commitment store is strictly included in U's commitment store; this is typical for dialogues where the user teaches the machine new tasks;
2. $CS_M \supset CS_U$, that is, conversely, M's commitments strictly include U's commitments; this situation is typical for tutoring dialogues;
3. $CS_M \cap CS_U \equiv CS_\cap \neq \emptyset$, that is, there is a certain overlapping between U and M's commitments; this is typical for service-oriented dialogues, where the machine interacts with previously unknown users;
4. $CS_M \cap CS_U \equiv \emptyset$, that is, M and U have no commitment in common; this is typical for dialogues where speakers have fundamentally different cultural backgrounds, or where the user teases the system; this type of dialogue represents only unsuccessful interactions.

Before furthering the discussion, we have to explain what we mean by intersecting two commitment stores: this notion can be considered either at a purely syntactic level (i.e. the logical forms contained *ad litteram* in the two commitment stores), or at a semantic level (i.e. the logical forms present in the commitment stores and satisfiable in the same set of models). Similarly to the commitments equivalence, we consider the intersection of commitment stores at a semantic level (a syntactic definition would have been too constrained). Thus, in order to automatically determine the intersection of two commitment stores, or to evaluate their equivalence, we go through several steps, stated below:

for two commitment stores CS_i and CS_j:

(a) **for** each logical formula $\lambda_k^{(i)} \in CS_i$ and $\lambda_l^{(j)} \in CS_j$:

i. compute the models where these formulas are satisfiable:

$$\mu_k^{(i)}[m] : \mu_k^{(i)}[m] \models \lambda_k^{(i)}; \mu_l^{(j)}[m] : \mu_l^{(j)}[n] \models \lambda_l^{(j)};$$

ii. compute the conjunctions of these models:

$$M_k^{(i)} = \bigwedge_{m:\mu_k^{(i)}[m]\neq\emptyset} (\mu_k^{(i)})[m]; M_l^{(j)} = \bigwedge_{n:\mu_k^{(i)}[n]\neq\emptyset} (\mu_l^{(j)})[n];$$

[3] This equivalence is considered at a semantic level, stemming from the identity of the models where the logical forms in the commitment stores are satisfiable.

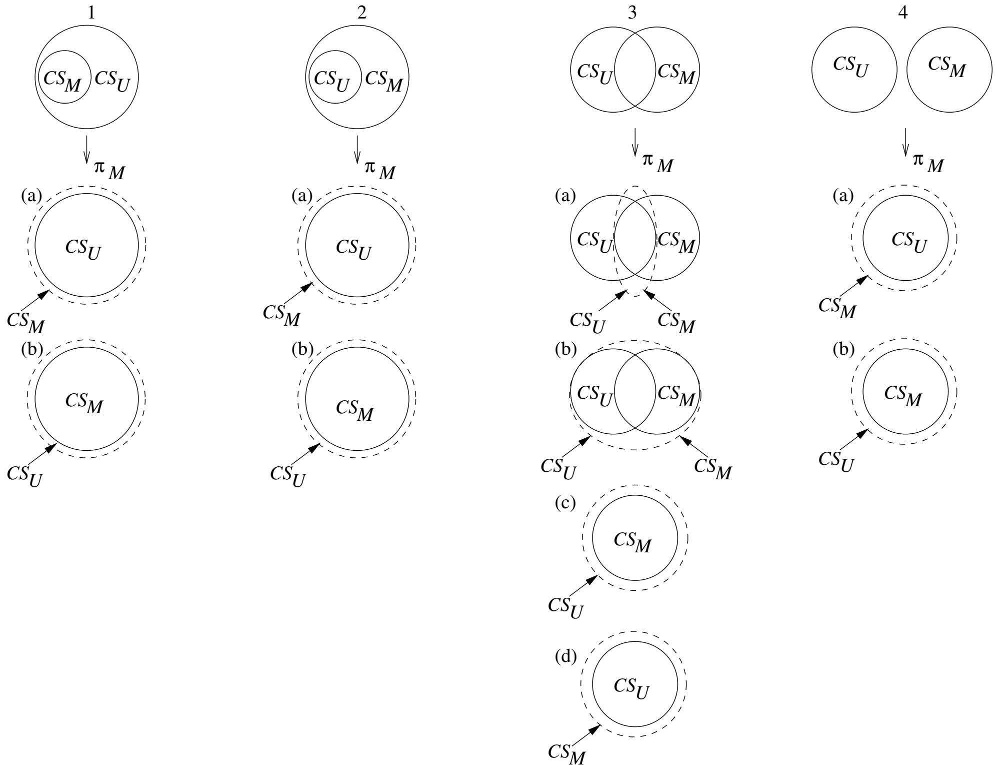

Figure 1. Commitment stores for U and M

(b) compute the unions of the conjunctions of the models:
$$M_i = \bigcup_{k:M_k^{(i)} \neq \emptyset} (M_k^{(i)});\ M_j = \bigcup_{l:M_l^{(j)} \neq \emptyset} (M_l^{(j)}).$$

Then, we compare these sets M_i and M_j for the two commitment stores CS_i and CS_j, in order to evaluate their equivalence or their intersection. The crucial issue with this approach lies in determining the set of models that satisfy the logical formulas. Essentially, this is accomplished by considering the whole task ontology as a model, from which we first eliminate the predicates that make the logical formula unsatisfiable in the model. Then, we can still generate models where the logical formula is satisfiable, by further eliminating predicates in the task ontology. However, we thus obtain a set of models strictly ordered with respect to inclusion. We therefore observe that the sets $M_k^{(i)}$, for each logical formula $\lambda_k^{(i)}$ in the commitment store CS_i, are subsets of predicates in the task ontology, such that adding $\lambda_k^{(i)}$ to this subset does not lead to a contradiction.

The strict inclusion relation ($\subset$) between commitment stores is defined in semantic terms as well, hence of models where the formulas contained in the commitment stores are satisfiable. Thus, an inclusion relation between two commitment stores CS_i and CS_j is equivalent to the *inverse* inclusion relation between the corresponding sets of models M_i and M_j:

$CS_i \subset CS_j \Leftrightarrow M_j \subset M_i$.

This is explained by the fact that the more formulas are contained in a commitment store (modulo the logical equivalence), the less there are models that satisfy *all* these formulas. The difference operation between two commitment stores is defined in model-theoretic terms as well: the result of a difference between two commitment stores represents the set of models that do not satisfy any formula in the second commitment store, but satisfy all the formulas in the first commitment store that are not equivalent to the formulas in the second commitment store:

$CS_i \setminus CS_j = \{\mu : \forall\lambda_j \in CS_j\, \mu \not\models \lambda_j \wedge \forall\lambda_i \in CS_i : \lambda_i \not\equiv \lambda_j\, \mu \models \lambda_i\}$.

Given the fact that throughout the article we adopt the machine's viewpoint, in order to predict its behavior in producing a speech turn, we assume that, in each of the four cases presented above, the next dialogue contribution to be produced belongs to the machine; we thus denote by π_M the *next* speech turn waited from M. In this context, we analyze in a detailed manner below each of the four situations indicated above (concerning the relations between commitment stores); these situations are illustrated in Figure 1:

Case 1: $CS_M \subset CS_U$

M's goal in dialogue is to make the commitment stores of the speakers equivalent, that is, $CS_M \equiv CS_U$. We assume that, when the machine produces turn π_M, it tries to modify as scarcely as possible its own commitment store. Moreover, for simplicity we assume that π_M consists of a single utterance; in fact, this does not reduce the generality of the problem, because we can assume that the machine produces each utterance sequentially, even in a speech turn that contains several utterances. Thus, two possible situations arise:

(a) $CS_M \leftarrow_{\pi_M} CS_M \cup (CS_U \setminus CS_M)$:
In this case, M produces an utterance π_M through which it marks the acceptance of all the preceding utterances in U's commitment store, but absent from his own commitments, i.e., the machine commits to $K(\pi_{U-}) = CS_U \setminus CS_M$, where π_{U-} denotes a (set of) utterance(s) previously produced by U. Thus, the machine chooses a *weak* degree of illocutionary force for π_M[4] : $\partial\phi(F(K(\pi_M))) = \partial\phi(\pi_M) = -1$.

(b) $CS_U \leftarrow_{\pi_M} CS_U \setminus (CS_U \setminus CS_M)$:
In this case, M produces an utterance π_M through which it tries to convince U to withdraw her commitment to $K(\pi_{U-})$ (i.e., to facts the machine is not committed to). Thus, the machine chooses a *very strong* degree of illocutionary force: $\partial\phi(\pi_M) = 2$.

[4]We will henceforth simplify the notation and write $\partial\phi(\pi)$ instead of $\partial\phi(F(K(\pi)), K(\pi))$ for the illocutionary force degree of an utterance labeled π.

The machine will first try the second possibility and, only if this fails (i.e. the user does not cooperate with the machine in adjusting its commitments to *M*'s), *M* will choose the first possibility.

Case 2: $CS_M \supset CS_U$

Assuming that the machine has the same goals as in the preceding case (i.e., to make its commitments equivalent to the user's commitments) two possibilities arise as well:

(a) $CS_M \leftarrow_{\pi_M} CS_M \setminus (CS_M \setminus CS_U)$:
In this case, *M* produces utterance π_M through which it concedes to withdraw previous commitments not shared by *U* (i.e., $K(\pi_{M-}) = CS_M \setminus CS_U$, where π_{M-} denotes an utterance previously produced by *M*). Thus, the machine marks this withdrawal by a *very weak* degree of illocutionary force for its utterance: $\partial\phi(\pi_M) = -2$.

(b) $CS_U \leftarrow_{\pi_M} CS_U \cup (CS_M \setminus CS_U)$:
In this case, *M* produces utterance π_M through which it tries to convince *U* to commit to $K(\pi_{M-}) = CS_M \setminus CS_U$. Thus, the machine chooses a *strong* degree of illocutionary force for π_M: $\partial\phi(\pi_M) = 1$.

As in the previous case, the machine first tries the second possibility and, only if this fails, *M* will go for the first possibility.

Case 3: $CS_M \cap CS_U = CS_\cap$

For the subsequent presentation, we introduce the following notations: (i) $K(\pi_{M-}) = CS_M \setminus CS_U$, (ii) $K(\pi_{U-}) = CS_U \setminus CS_M$, (iii) $K(\pi_{UM-}) = CS_U \cap CS_M$, (iv) $K(\pi_{M\neg U-}) = CS_M \setminus CS_\cap$, (v) $K(\pi_{U\neg M-}) = CS_U \setminus CS_\cap$. We thus have that $CS_U \cup CS_M = \{K(\pi_{M-}), K(\pi_{U-}), K(\pi_{UM-})\}$.

Under the same assumptions concerning *M*'s goal in dialogue and the heuristics on the preferences for updating the commitment stores, four possibilities arise in this case:

(a) $(CS_M \leftarrow_{\pi_M} CS_\cap) \wedge (CS_U \leftarrow_{\pi_M} CS_\cap)$:
In this situation, *M* produces utterance π_M through which it concedes to withdraw its commitment to $K(\pi_{M\neg U-})$ *and* it tries to convince *U* to withdraw its commitment to $K(\pi_{U\neg M-})$. Thus, the machine first aggregates π_M in two utterances, π'_M and π''_M, such that $K(\pi'_M) \wedge K(\pi''_M) = K(\pi_M)$. Secondly, *M* assigns a *very weak* degree of illocutionary force to π'_M: $\partial\phi(\pi'_M) = -2$, and a *very strong* degree of illocutionary force to π''_M: $\partial\phi(\pi''_M) = 2$.

(b) $(CS_M \leftarrow_{\pi_M} CS_M \cup CS_U) \wedge (CS_U \leftarrow_{\pi_M} CS_M \cup CS_U)$:
In this case, *M* produces utterance π_M through which it commits to $K(\pi_{U-})$ *and* it tries to convince *U* to commit to $K(\pi_{M-})$. Thus, the machine first aggregates π_M in two utterances, π'_M and π''_M, such that $K(\pi'_M) \wedge K(\pi''_M) = K(\pi_M)$. Secondly, *M* assigns a *weak* degree of illocutionary force to π'_M: $\partial\phi(\pi'_M) = -1$, and a *strong* degree of illocutionary force to π''_M: $\partial\phi(\pi''_M) = 1$.

(c) $CS_U \leftarrow_{\pi_M} CS_M \equiv CS_U \setminus (CS_U \setminus CS_M) \cup (CS_M \setminus CS_U)$:

In this case, M produces utterance π_M through which it tries to convince U to withdraw its commitment to $K(\pi_{U-})$ *and* to commit to $K(\pi_{M-})$. Thus, the machine first aggregates π_M in two utterances, π'_M and π''_M such that $K(\pi'_M) \wedge K(\pi''_M) = K(\pi_M)$. Secondly, M chooses a *very strong* degree of illocutionary force for the first utterance, and a *strong* degree for the second utterance: $\partial\phi(\pi'_M) = 2$ and, respectively, $\partial\phi(\pi''_M) = 1$.

(d) $CS_M \leftarrow_{\pi_M} CS_U \equiv CS_M \setminus (CS_M \setminus CS_U) \cup (CS_U \setminus CS_M)$:
In this case, M produces utterance π_M through which it withdraws its commitment to $K(\pi_{M\neg U-})$ *and* commits to $K(\pi_{U\neg M-})$. Thus, the machine first aggregates π_M in two utterances, π'_M and π''_M such that $K(\pi'_M) \wedge K(\pi''_M) = K(\pi_M)$. Secondly, M assigns a *very weak* degree of illocutionary force to π'_M and a *weak* degree to π''_M: $\partial\phi(\pi'_M) = -2$, and $\partial\phi(\pi''_M) = -1$, respectively.

We assume that the machine adopts the heuristic of trying to modify as scarcely as possible its commitments, preferring to add utterances, rather that removing them. Thus, the order of preference for the possibilities listed above is: (c) $\rightarrow$ (b) $\rightarrow$ (a) $\rightarrow$ (d).

Case 4: $CS_M \cap CS_U = \emptyset$

Adopting the constraint of not having empty commitment stores following an updating operation, and keeping the same assumption on M's goal in dialogue, as well as on the heuristic concerning the preferences in commitment stores updating, we have two possibilities:

(a) $CS_M \leftarrow_{\pi_M} CS_U$:
In this situation, via π_M M withdraws all its commitments, accepting all the user's commitments instead. This represents an extreme version of the case 3.(d), but the formal treatment is identical to that case.

(b) $CS_U \leftarrow_{\pi_M} CS_M$:
In this case, M produces utterance π_M through which it tries to convince U to give up all his previous commitments, while accepting all of M's commitments. This situation represents an extreme version of the case 3.(c), but the formal treatment remains unchanged.

Given the heuristic whereby M modifies as scarcely as possible its commitment store, preferring additions to it instead of removals, the order of preferences for the possibilities presented above for this case is: (b) $\rightarrow$ (a).

Concerning the assignment of illocutionary force degrees, several comments arise, in order to defend the choice of a five-level scalar representation, and to clarify the distinction between the five possible values for this degree of force. Thus, given a speaker L and an utterance π produced by L, we can depict several situations:

- when via π, L concedes to remove something from its commitment store CS_L, like $CS_L \leftarrow_\pi CS_L \setminus \{\pi_{L-}\}$ for some previous utterances π_{L-} of L, we have that $\partial\phi(\pi) = -2$, hence a **very weak** degree of illocutionary force;

- when via π, L adds something to its commitment store CS_L, like $CS_L \leftarrow_\pi CS_L \cup \{\pi_{\bar{L}-}\}$ for some previous utterances of another speaker, $\bar{L}$, we have that $\partial\phi(\pi) = -1$, hence a **weak** degree of the illocutionary force;
- when via π, L tries to convince another speaker $\bar{L}$ to add something to its commitment store $CS_{\bar{L}}$ (i.e., $CS_{\bar{L}} \leftarrow_\pi CS_{\bar{L}} \cup \{\pi_{L-}\}$), we have that $\partial\phi(\pi) = 1$, hence a **strong** degree of illocutionary force;
- when via π, L tries to convince another speaker $\bar{L}$ to remove something from its commitment store $CS_{\bar{L}}$ (i.e., $CS_{\bar{L}} \leftarrow_\pi CS_{\bar{L}} \setminus \{\pi_{\bar{L}-}\}$ for a previous utterance $\pi_{\bar{L}-}$ of $\bar{L}$), we have that $\partial\phi(\pi) = 2$, hence a **very strong** degree of illocutionary force.

For the **neutral** degree of illocutionary force (i.e., of value 0), we stipulate that an utterance can be produced with a neutral degree only if it asserts un-contestable facts in the world (which are encoded in the task ontology), for instance the fact that 'Normally, a book cannot be read in the total absence of light', it realizes a speech act of the type F^{S} *and* is uttered for the first time in dialogue (i.e. it is not re-uttered in an attempt to make the interlocutor accept the utterance, after having previously rejected it). By consequence, any other type of utterance, realizing any other speech act type, cannot have a neutral degree of illocutionary force.

We thus observe how the set of possible relations between the commitment stores for the speakers in dialogue is mapped into a set of values for the degree of illocutionary force. This provides a first justification *a posteriori* for the five-level scale for this degree of force. Yet, the choice of a discrete scale for the degrees of illocutionary force can still seem methodological in nature: indeed, it makes it possible to annotate easier corpora of pre-defined utterances that thus map degrees of force into linguistic realizations. However, this five-level scale accounts well for the (set-theoretic) relations between the commitment stores of the speakers. For instance, if we had chosen a continuous scale of values in an interval (e.g.,$[-1;+1]$), we would have had difficulties in mapping these values to relationships between commitment stores.

A possible criticism to this approach concerns the dialogue success criterion. Indeed, we could consider the following dialogue example, between the machine M (a virtual librarian) and a user U:

> U_1: Hello, I would like to have some books on the French revolution!
> M_1: Hello, Sir, I have two books: an introductory one, 'X', and a more advanced one, 'Y'.
> U_2: I would like to have book 'Y', it seems more interesting.
> M_2: OK, give me your card, please.

In this dialogue, the utterance 'there is a book 'Y' on the French revolution in the library' should belong to CS_M and to CS_U. Nevertheless, whereas the utterance 'there is a book 'X' on the French revolution in the library' belongs to CS_M; U has not committed to its content, either explicitly or implicitly; U is neutral about the contents of this utterance; hence, the utterance does not belong to CS_U. Thus, after these speech turns, we have that $CS_U \subset CS_M$; yet, the dialogue is successful, since the user manages to make its reservation. The dialogue would not have been more successful if, for instance, U had produced 'OK, I agree' as well, in the beginning of speech turn U_2, thus committing to the entire content of the previous speech turn of M (that is, to the contents of M_1):

U_2': OK, I agree. I would like to have book 'Y', it seems more interesting.

On the one hand, the equivalence between the commitment stores of U and M in dialogue represents a *sufficient* condition to dialogue success; even though this condition is not always *necessary*, its fulfillment ensures that the dialogue is successful. On the other hand, speech turn U_2 may represent an *indirect implicit* confirmation (essentially, if U does not contradict, explicitly or implicitly, any part of the preceding speech turn of M, then we stipulate that *by default* U commits to the entire speech turn of M) to the *whole* speech turn M_1.

We further argue why we considered commitments' equivalence instead of commitments' mutual consistency as a dialogue success criterion. The argument pertains to computational considerations. We consider that it is easier to perform a mere "set-theoretic" commitments equivalence check, than a consistency check, because a consistency check would require us to verify for revisions in the commitment stores, that is, to check the *entire* commitment stores for *each* commitment store updating process. However, if equivalence is checked, only what is being *currently* added to the commitment stores needs to be checked, since their equivalence is aimed at during the entire dialogue. Even though this equivalence is never attained, it is "asymptotically" aimed at, in that the machine (not necessarily like the human interlocutor, and this is important, since, in our view, the machine need neither imitate, nor even approximate human behavior, but only be a useful and "friendly" assistant to the human) tries to adjust its commitments (and, if necessarily, its interlocutors' commitments) so that they are the closest possible, for each dialogue turn pair. Hence, altough restrictive, the condition on the equivalence of the commitment stores is, we believe, operationally more feasible than the check on their consistency.

Lastly, another relevant criticism stems from the fact that the degree of illocutionary force is not determined only from the discourse structure and the public commitments of the speakers: social roles are involved as well, in limiting the set of values that the degree of illocutionary force might take. According to our framework as it was presented until now, in a situation where the user U states α and the machine M states $\neg\alpha$, and the machine tries to convince U to give up his commitment to α, then it should utter $\neg\alpha$ with an illocutionary force degree of $+2$. This would lead the machine to produce an utterance like: 'But Sir, α is certainly not true, do you understand?'. In a service scenario, where U is the customer and M the "server", this utterance is not appropriate, because it would lead U to complain about M's impoliteness. This is a case where M has much less power than U in conversation. M simply does not have the right to produce an utterance with an illocutionary force degree of $+2$. All the machine can do in this case is to produce an utterance with a degree of force of $+1$: 'Sir, I am convinced that α is not true'. We therefore observe how a hierarchical social status with respect to the interlocutor induces a tighter upper bound for the degree of illocutionary force. In the same context, we can ask ourselves whether, conversely, there are relational configurations that induce a tighter lower bound for the degree of illocutionary force as well. We believe that here such a lower bound can only be imposed by the personality of the speaker: for example, if the latter has a proud nature, then she will tend not to produce utterances with a degree of force of -2 (for thus giving up its previous commitments), preferring to produce utterances with a degree of force of -1 instead. Thus, the speaker only concedes to accept the interlocutor's utterances (that could in turn *entail* the withdrawal of certain commitments). Consider for instance the following dialogue, between a user and a "proud" machine, in the sense stated above:

U_1: But book 'X' is not on the first floor, as you told me! Actually, this book is on the second floor!
M_1: Hm... I think you are right; I will send one of my colleagues on the second floor bring you the book.

In M_1 (which has a degree of illocutionary force of -1), the machine does not explicitly withdraw its previous commitment to the fact that book 'X' were on the first floor, preferring to confirm vaguely user's turn U_1. Had the machine been less "proud", it would have explicitly given up its commitment to the fact that book 'X' were on the first floor, as in a speech turn $M_1^{(\prime)}$: 'Yes, indeed, I was wrong, I am sorry; hence, I will send one of my colleagues on the second floor bring you the book'.

To conclude on the social roles, we propose, in a first approximation, that the domain of values for the degree of illocutionary force be limited by an upper bound that stems from a lower position on the social hierarchy, with respect to the interlocutor. In the "virtual librarian" example, the latter could only produce utterances with illocutionary force degrees in the set $\{-2; -1; 0; +1\}$. On the contrary, the human user could produce utterances with any degree of illocutionary force, between -2 and $+2$. The issue of a finer analysis of the effects that social roles induce on the illocutionary force in producing utterances in conversation remains open to further study.

3.3. Assessment of the Framework

In this section we assess the proposed framework via two typical dialogues, illustrating the mechanism for adjusting the degree of illocutionary force. The mapping between degrees of force and linguistic form is realized by using canned phrases, annotated with illocutionary force degrees, based on linguistic intuitions. The two dialogues considered differ only in certain speech turns, the task context being the same: a user U tries to find a book 'B' in a library, whose (virtual) librarian is the machine M.

Dialogue 1:
U_1: Sorry, can you tell me, please, where I can find book 'B'? — $\text{F}^{FS}(\pi_1)$
M_1: Hello, well, you can find the book just at the end of the corridor, to the left. — $\text{F}^{S}(\pi_2); QAP(\pi_1, \pi_2)$
U_2: But I've just looked there, and I couldn't find the book! — $\text{F}^{S}(\pi_3); P-Corr(\pi_2, \pi_3)$
M_2: Oh, I was wrong, I am sorry; indeed, you can find book 'B' on the first floor, to the right. — $\text{F}^{S}(\pi_4); Contrast(\pi_2, \pi_4); P-Elab(\pi_3, \pi_4); QAP(\pi_1, \pi_4)$

Dialogue 2:
U_1: Sorry, can you tell me, please, where I can find book ‘B’? — $F^{FS}(\pi_1)$
M'_1: Hello, wait a minute please... here it is: you can find this book on the first floor, to the left. — $F^{S}(\pi'_2); QAP(\pi_1, \pi'_2)$
U_2: But I've just looked there, and I couldn't find the book! — $F^{S}(\pi_3); P - Corr(\pi'_2, \pi_3)$
M'_2: Listen, you can certainly find book ‘B’ on the first floor, just to the left, I have just checked in my data base! — $F^{S}(\pi'_4); P - Corr(\pi_3, \pi'_4); QAP(\pi_1, \pi'_4)$

Next to each speech turn in these dialogues we have marked the speech act type, the label of the utterance and the rhetorical relations that connect this speech turn to previous turns[5] Moreover, for M's speech turns, we have marked the illocutionary force degree as well. The rhetorical relations are computed in the framework of (Asher and Lascarides, 2003) SDRT, whereas the speech act types are assigned based on linguistic markers (for user utterances), using (Colineau, 1997)'s connectionist approach, or provided directly by the dialogue controller (for machine utterances) (Caelen and Xuereb, 2007). By consequence, we limit our description to the commitment stores updating, and to the manner whereby the commitments allow us to compute the illocutionary force degrees. These processes are presented in detail below, for each of the two dialogues:

Dialogue 1:
1. when U produces π_1, its commitment store remains unchanged, and so does M's commitment store: $CS_U \leftarrow_{\pi_1} CS_U \wedge CS_M \leftarrow_{\pi_1} CS_M$;
2. when M produces π_2, its commitment store is updated, whereas U's commitment store remains, for the moment, unchanged: $CS_M \leftarrow_{\pi_2} CS_M \cup \{K(\pi_2), \Sigma_{QAP(\pi_1,\pi_2)}\} \wedge CS_U \leftarrow_{\pi_2} CS_U$; M's goal is that U update its commitment store with $K(\pi_2)$ as well: $CS_U \leftarrow_{\pi_2} CS_U \cup \{K(\pi_2)\}$, hence M assigns a value of $+1$ to the illocutionary force degree of π_2: $\partial\phi(\pi_2) = +1$;
3. when U produces π_3, its commitment store is updated by adding $\neg K(\pi_2)$, whereas M's commitments do not change for the moment: $CS_U \leftarrow_{\pi_3} CS_U \cup \{\neg K(\pi_2), \Sigma_{P-Corr(\pi_2,\pi_3)}\} \wedge CS_M \leftarrow_{\pi_3} CS_M$;
4. when M produces π_4, its commitment store is updated by removing $K(\pi_2)$ from it, and by adding $K(\pi_4)$ to it, whereas U's commitments do not change for the moment: $CS_M \leftarrow_{\pi_4} CS_M \setminus \{K(\pi_2)\} \cup \{K(\pi_4), \Sigma_{Contrast(\pi_2,\pi_4)}, \Sigma_{P-Elab(\pi_3,\pi_4)}, \Sigma_{QAP(\pi_1,\pi_4)}\}$; M's goal is that U's commitment store be updated by adding $K(\pi_4)$: $CS_U \leftarrow_{\pi_4} CS_U \cup \{K(\pi_4)\}$ such that $K(\pi_4) \Rightarrow \neg K(\pi_2)$, hence M assigns a value of -2 to the illocutionary force degree of π_4: $\partial\phi(\pi_4) = -2$.

Dialogue 2:
1. when U produces π_1, its commitment store remains unchanged, and so does M's commitment store: $CS_U \leftarrow_{\pi_1} CS_U \wedge CS_M \leftarrow_{\pi_1} CS_M$;

[5] For the simplicity of the presentation, we consider that each speech turn contains only one utterance.

2. when M produces π'_2, its commitment store is updated, whereas U's commitment store remains unchanged for the moment: $CS_M \leftarrow_{\pi'_2} CS_M \cup \{K(\pi'_2), \Sigma_{QAP(\pi_1,\pi'_2)}\} \wedge CS_U \leftarrow_{\pi'_2} CS_U$; M's goal is that U's commitment store be updated with $K(\pi'_2)$ as well: $CS_U \leftarrow_{\pi'_2} CS_U \cup \{K(\pi'_2)\}$; but first, M has to update its own commitment store with $K(\pi'_2)$, hence M assigns a value of -1 to the illocutionary force degree of π_2: $\partial\phi(\pi_2) = -1$ (unlike in *Dialogue 1*, M does not have $K(\pi'_2)$ in its commitment store, hence it has to ask the task manager for this information);
3. when U produces π_3, its commitment store is updated by adding $\neg K(\pi'_2)$, whereas M's commitments do not change for the moment: $CS_U \leftarrow_{\pi_3} CS_U \cup \{\neg K(\pi'_2), \Sigma_{P-Corr(\pi'_2,\pi_3)}\} \wedge CS_M \leftarrow_{\pi_3} CS_M$;
4. when M produces π'_4, its commitment store still contains $K(\pi'_2)$, whereas U's commitments do not change for the moment: $CS_M \leftarrow_{\pi'_4} CS_M \cup \{\Sigma_{P-Corr(\pi_3,\pi'_4)}, \Sigma_{QAP(\pi_1,\pi'_4)}\}$; M's goal is that U's commitment store be updated by removing $\neg K(\pi'_2)$, and by adding $K(\pi'_2)$: $CS_U \leftarrow_{\pi'_4} CS_U \setminus \{\neg K(\pi'_2)\} \cup \{K(\pi'_2)\}$, hence M assigns a value of $+2$ to the illocutionary force degree of π'_4: $\partial\phi(\pi'_4) = 2$.

For these two dialogues, we have been in two different cases, concerning the (set theoretic) relation between U and M's commitment stores:

- In dialogue 1 we are in the case $CS_U \cap CS_M = CS_\cap \neq \emptyset$ after speech turn M_2 were produced;
- In dialogue 2 we are in the case $CS_U \cap CS_M = \emptyset$ after speech turn M'_2 were produced.

If the degree of illocutionary force had not been adjusted, then we would have obtained in all cases a single speech turn, $M^{(0)}$, instead of M_2, or M'_1 and M'_2; speech turn $M^{(0)}$ would have contained an utterance like:

$M^{(0)}$: You can find book ‘B’ on the first floor, to the left.

This would have led the machine to produce more annoying and less natural turns.

However, the approach has several limits. First, the machine is sometimes “rude” for a public service, especially when choosing very strong degrees of force for determining its interlocutor to give up some of her/his commitments. However, such as behavior is not totally implausible, if we think at more informal interactions, especially those that are less subject to politeness conventions: consider for example a dialogue situation between a bartender and a (drunk) customer in a bar next to a highway at night, where very strong illocutionary degrees and behaviors such as those shown in the article seem casual. Yet, there is another problem, that stems from the imprecisions in determining the interlocutors' public commitments: indeed, if, for some reason (i.e. sudden attack) the interlocutor stops in the middle of its utterance, or the interlocutor tries to tease the system by feeding incoherent turns into it, the machine might perform wrong calculations, and hence, for instance, inadequately produce utterances with a $+1$ (strong) degree of illocutionary force, in trying to determine the interlocutor to commit to something to which the latter has already committed, or which is irrelevant to the dialogue at hand; these limitations are discussed more thoroughly in (Popescu et al., 2009).

4. Discussion and Conclusions

In this article we have shown that fine-tuning the degree of illocutionary force is a crucial aspect in generating (machine) utterances in dialogue. We have proposed a computational framework for computing the illocutionary force degree. The mechanism is based on the public commitments of the speakers (the machine, which emulates a public service, and a human customer of this service). The commitment stores are computed from the discourse structure in a manner borrowed from (Maudet et al., 2006). Furthermore, we have given a precise formalization of Vanderveken's notion of illocutionary force degree, coupling it with the public commitment, in a condition on dialogue success. Although limited and arguable, this condition (of aiming at identical commitments for the interlocutors) is relevant in service-oriented dialogues, that are restrained to a very specific task, priorly specified.

However, the approach described in this article has certain other limits and arguable aspects. First, out of the six components of the illocutionary force, as defined by (Vanderveken, 1990-1991), we take into account explicitly only three: propositional content, degree, and illocutionary point (i.e., speech act type). Another arguable point, although in agreement with (Vanderveken, 1990-1991): p. 120, is the domain of values for the illocutionary force degree: $\{-2, -1, 0, 1, 2\}$; Vanderveken does not limit the number of possible illocutionary force degree, but we do it, on the one hand because the five-level scale emerges from the relationships between the public commitments of the speakers, and on the other hand, for methodological reasons, related to the possibility to conveniently annotate a set of canned phrases that can be used in surface generation.

Another possible criticism concerns the manner whereby we define the negation of an utterance, especially when this utterance is a question. One could object that an interrogative utterance is of the wrong semantic type for the negation operation. This is why we need to clarify this point. Let us consider for example a question as 'Is this book OK for you', labeled π. At a semantic level, this utterance is logically represented via the $K/1$ function. Since it is a question, the utterance contains a predicate which takes a non-initialized variable as argument:

$\exists Y, Z : \mathsf{object}(\text{'book'}) \wedge \mathsf{feature}(\text{'book'}, Y) \wedge \mathsf{title}(Y) \wedge \mathsf{equals}(Y, \langle\text{book\_title}\rangle) \wedge \mathsf{want}(\neg\mathsf{emitter}(\pi), \text{'book'}, Z) \wedge \mathsf{equals}(Z, \text{'?'})$.

Here, the non-initialized variable is the boolean Z that contains the truth value of the predicate $\mathsf{want}/3$, which is true if the entity designated by its first argument (in our case, the recipient of π) wants the entity designated by the second argument (in our case, the book 'book', whose title is specified by the value of variable Y). The negation of such a question does not boil down to the classical negation of each predicate in the conjunction, followed by the substitution of the conjunctions with disjunctions, but to assigning the value 0 to the boolean Z; hence, in our case, $\neg K(\pi)$ has the same form as $K(\pi)$, excepting the last predicate, which has the form $\mathsf{equals}(Z, 0)$.

Another aspect left untackled in this article concerns the importance of the conversational genre in determining relevant illocutionary force degrees for the utterances. Thus, for instance, in absurd theatrical plays very strong illocutionary force degrees can be appropriate even in social configurations where the speaker has a lower position on the social hierarchy, with respect to the hearer, simply because ethical adequacy is sometimes distorted. Moreover, in such plays (or in informal dialogues between close friends, for instance), a very strong illocutionary force degree might as well be used for withdrawing a commitment, in an utterance like 'Listen, I was certainly wrong, I am really sorry, do you understand me?'. However, in the service oriented customer-institution dialogues concerned by this study, such illocutionary configurations seem rather inadequate, at least in the context of dialogue corpora for task-oriented dialogues, such as the PVE ("Portail Vocal pour l'Entreprise") system (Nguyen, 2005). In our view, genre acts in the form of supplementary constraints in human-system dialogues. These constraints can perform a post-filtering on the set of authorized illocutionary forces, in a manner akin to that discussed for social roles, in the end of Section 3.2. Nevertheless, the issue of formalizing the intricate interaction between social and genre constraints in producing dialogue contributions that are adequate from an illocutionary standpoint remains, in our view, open to further study.

In spite of several limitations of the current framework, it represents a first version of a principled way of computing the adequate illocutionary force degree of machine utterances. A first prospect for this research would be to evaluate in quantitative manner the framework described, by comparing automatically generated utterances to human utterances generated via Wizard-of-Oz techniques, or present in real dialogue corpora.

Acknowledgement

The authors wish to thank Prof. Andrew Kehler from the University of California, San Diego, for his valuable comments and thorough remarks concerning the contents of this article.

Bibliography

Asher, Nicholas and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, United Kingdom, 2003.

Asher, Nicholas and Alex Lascarides. Making the right commitments in dialogue. In *Fall 2008 Workshop in Philosophy and Linguistics*. University of Michigan, 2008.

Brown, Penelope and Stephen Levinson. *Politeness: Some Universals in Language Use*. Cambridge University Press, UK, 1987.

Caelen, Jean and Anne Xuereb. *Interaction et pragmatique – jeux de dialogue et de langage*. Editions Hermès - Lavoisier, Paris, France, 2007.

Cohen, Philip and Hector Levesque. Persistence, intention and commitment. In Cohen, Philip, Jerry Morgan, and Martha Pollack, editors, *Intention in Communication*, pages 33–69. MIT Press, 1990a.

Cohen, Philip and Hector Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42 (2-3):213–261, 1990b.

Cohen, Philip and Raymond Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212, 1979.

Colineau, Nathalie. *Etude des marqueurs discursifs dans le dialogue finalisé*. Ph D Thesis, Joseph Fourier University, Grenoble, http://tel.archives-ouvertes.fr/tel-00004928/en/, 1997.

Faller, Martina. Evidentiality and epistemic modality at the semantics / pragmatics interface. In *Fall 2006 Workshop in Philosophy and Linguistics*. University of Michigan, 2006.

Gupta, Swati, Marilyn Walker, and Daniela Romano. Generating politeness in task based interaction: An evaluation of the effect of linguistic form and culture. In *Proceedings of the 11th European Workshop on Natural Language Generation ENLG'07*, Schloss Dagstuhl, 2007.

Kibble, Rodger. Speech acts, commitment and multi-agent communication. *Computational and Mathematical Organization Theory*, 12:127–145, 2006a.

Kibble, Rodger. Reasoning about propositional commitments in dialogue. *Research on Language and Computation*, 4:179–202, 2006b.

Kibble, Rodger. Generating coherence relations via internal argumentation. *Journal of Logic, Language and Information*, 16:387–402, 2007.

Kibble, Rodger and Paul Piwek. Introducing dialogue games. In *Lecture Notes, ESSLLI 2007*. FoLLI, Dublin, Ireland, 2007.

Lascarides, Alex and Nicholas Asher. Agreement, disputes and commitments in dialogue. *Journal of Semantics*, 26(2):109–158, 2009.

Maudet, Nicolas, Philippe Muller, and Laurent Prévot. Tableaux conversationnels en SDRT. In *Proceedings of the SDRT Workshop, Traitement Automatique des Langues Naturelles*, Fès, Maroc, 2004.

Maudet, Nicolas, Philippe Muller, and Laurent Prévot. Social constraints on rhetorical relations in dialogue. In *Workshop Constraints in Discourse*, Maynooth, Ireland, 2006.

Motsch, Wolfgang. Situational context and illocutionary force. In Searle, John, Ferenc Kiefer, and Manfred Bierwisch, editors, *Speech Act Theory and Pragmatics*, pages 155–168. D. Reidel Publishing Company, 1980.

Nguyen, Hoá. *Dialogue homme-machine : modélisation de multisession*. Ph D Thesis, Joseph Fourier University, Grenoble, http://tel.archives-ouvertes.fr/tel-00008789/en/, 2005.

Popescu, Vladimir and Jean Caelen. Argumentative ordering of utterances for language generation in multi-party human-computer dialogue. *Argumentation*, 23(2):205–237, 2009.

Popescu, Vladimir, Jean Caelen, and Corneliu Burileanu. Contrôle rhétorique de l'ellipse sémantique en génération du langage pour le dialogue homme-machine à plusieurs locuteurs. *Traitement Automatique des Langues*, 49(1):115–139, 2008.

Popescu, Vladimir, Jean Caelen, and Corneliu Burileanu. A constraint satisfaction approach to context-sensitive utterance generation in multi-party dialogue systems. *International Journal of Speech Technology*, 12(2-3):95–112, 2009.

Searle, John and Daniel Vanderveken. Speech acts and illocutionary logic. In Vanderveken, Daniel, editor, *Logic, Thought and Action*, volume 2 of *Logic, Epistemology and the Unity of Science*, pages 109–132. Springer, 2006.

Traum, David, Michael Fleischman, and Eduard Hovy. Nl generation for virtual humans in a complex social environment. In *Proceedings of the AAAI 2003 Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, Palo Alto, 2003.

Vanderveken, Daniel. *Meaning and Speech Acts*. Cambridge University Press, United Kingdom, 1990-1991.

Xuereb, Anne and Jean Caelen. Actes de langage et relations rhétoriques en dialogue homme-machine. *Revue de l'Université de Moncton*, 36(2):5–51, 2005.

Address for correspondence:
Vladimir Popescu
vladimir.popescu@imag.fr
339, chemin des Meinajaries
Agroparc BP 91228
84911 Avignon Cedex 9
France