Imperatives, Commitment and Action: Towards a Constraint-based Model * †

Paul Piwek
ITRI – University of Brighton
Watts Building, Moulsecoomb, Brighton BN2 4GJ, UK
Email: Paul.Piwek@itri.brighton.ac.uk
Web: http://www.itri.brighton.ac.uk/∼Paul.Piwek

July, 2000

Abstract

The aim of this paper is to provide an analysis of the function of imperatives in dialogue. In particular, the focus is on the use of logically complex imperatives (e.g., 'Say hello to John, if you meet him') and temporal reference in imperatives (e.g., 'Open the bottle of champagne at midnight'). Two specific problems involving logically complex and temporally referring imperatives are introduced. These problems are addressed within a framework for communicating agents. It is argued that such a framework needs to allow for partial models of communicating agents and their environment and for the intentionality of agent states.

1 Introduction

This paper reports on a study into the function of imperatives in dialogue. Imperative clauses are traditionally characterized by the lack of a subject, the use of a base form of the verb, and the absence of modals as well as tense and aspect markers. Examples of naturally occurring imperatives are: 'Get off the table', 'Don't forget about the deposit' and 'Hold on, are we late?'. (The definition and the examples are taken from Biber et al., 1999:219). The function of imperatives (i.e., the task which they perform in discourse)[1] is typically to get the addressee to do something. The aim of this paper is to provide a formal analysis of this function. Eventually, this analysis may prove to be useful beyond the study imperatives, since the same function can be expressed by different means (E.g., by a question such as 'Will you be quiet for goodness sake!'; from Biber et al., 1999:220). Let me now introduce two specific problems concerning the function of imperatives which I address in this paper.

* This paper is to appear in a special issue of LDV-Forum (the journal of the GLDV, full title: LDV-Forum, Zeitschrift fr Computerlinguistik und Sprachtechnologie, Journal for Computational Linguistics and Language Technology), 2000.

† I could not have written this paper without first having been introduced to and taught about natural language pragmatics by Robbert-Jan Beun and Harry Bunt. I am particularly indebted to Robbert-Jan Beun who took care of the day to day supervision when I was writing my PhD dissertation. I am grateful that he never ceased to urge me to start my studies into natural language semantics and pragmatics from an overall view on communicating agents. I have tried to write this paper in that spirit.

[1] I use the term *discourse* to refer to both written and spoken discourse. This paper focuses on spoken discourse, i.e., dialogue. Biber et al. (1999:221) found that imperatives are most frequent in spoken discourse.

(**P.1**) Imperative clauses, like declarative ones, can be used in combination with connectives such as 'or' 'and' and 'if'. For instance, 'if' can be used to construct rule-like imperatives such as: 'Say hello to John, if you meet him' (cf. Hamblin, 1987:84). A problem emerges when we consider the translation of the truth conditional content of this imperative into first-order logic (FOL). Suppose we choose the following translation: $meet(addressee,john) \rightarrow say\_hello\_to(addressee,john)$, i.e., if the addressee meets John, then he says hello to John.[2] Let us assume that such a formalization of the content of imperatives is a legitimate move. Then we can use that formalization as a basis for formalizing the function of imperatives. It can, however, be shown quite easily that one very tempting analysis of imperatives cannot be correct. According to this analysis the function of an imperative is to instruct the addressee to make its content true. In FOL a conditional is true if its antecedent is false or its consequent is true. Thus such an analysis predicts that an addressee of a conditional imperative can comply with the imperative by simply making its antecedent false. For the aforementioned example this means that if the addressee were to avoid meeting John, this would count as acting in accordance with the imperative. This counterintuitive prediction casts doubt on the correctness of the aforementioned analysis.

(**P.2**) A further problem concerning logically complex imperatives has been attributed to Ross (1941). Ross points out that the relation of logical consequence between imperative clauses appears to be different from that between declarative clauses: in FOL, $\phi \vee \psi$ is a logical consequence of ϕ, whereas the imperative 'Post the letter or burn it' does not seem to follow from 'Post the letter'.

Generally speaking, I want to provide an alternative to what Hamblin (1987) has termed 'Giddap' – 'Whoa' theories of imperatives. According to such theories, imperatives in dialogue are immediately followed by an action, as in 'Instructor: Change to second gear. Pupil: (Changes to second gear)'. Such theories fail to extend to logically complex imperatives and imperatives which contain explicit temporal references. In order to explicate the function of imperatives I construct a framework of communicating agents in which a formal characterization of this function can be specified. More specifically, I use this framework to provide an outline of a succession of increasingly more elaborate and, hopefully, more realistic theories. I acknowledge that the end result is still only a rough approximation of the use of imperatives in real human–human conversations.

2 Framework: Models for Communicating Agents

In order to study the function of imperatives a model is constructed of the situations in which imperatives are being used. This model includes the agents who issue and receive imperatives and also the environment in which these agents operate. The idea is that the aforementioned agents are approximations of human beings, although only very rough ones.[3]

The aforementioned approach is not very different from that in the other sciences such as biology or physics. However, for a functional analysis of imperatives it will not do to construct a purely physical model of communicating agents and their environment. Such a description would gloss over the fact that communicating agents carry and use *information* to guide their actions. For instance, when a person is told to 'Open a bottle of Champagne at midnight', s/he will have to store this information and act upon it at the right time. The fact that an agent carries some piece of information

[2]This translation avoids several issues, such as the interpretation of pronouns, which are beyond the scope of this paper. Furthermore, the example has been formalized without any explicit reference to time. Temporal aspects of imperatives are, however, dealt with later on in this paper.

[3]By describing the agents in sufficient formal detail, they might also be realized as software agents. This could facilitate the evaluation of the model. However, the currently described model has not yet been implemented.

with it can be analysed as a state of that agent. Such a state, which an imperative can give rise to, is traditionally known as an Intentional state. Intentional states are states which are directed at or about objects or states of affairs in the world.[4] The crucial characteristic of Intentional states is that we can speak about whether they have been satisfied or not (cf. Searle, 1983); something which a strictly physical model of agent states would not allow us to do. The analysis which is proposed here starts from the assumption that the satisfaction of an imperative is grounded in the satisfaction of the Intentional state (of the addressee) which the imperative gave rise to.[5] It will, however, transpire that quite a few technical difficulties arise when we try to formalize this idea. Primarily, the difficulty is that the information conveyed by an imperative has to be integrated with the information which an agent already avails over. For instance, if I know that it rains and I am told that I should bring my umbrella if it rains, then the combined information should incite my to take my umbrella with me. Thus, the relation between an action of an agent and an imperative can be influenced by other information available to the agent.

In summary, I am going to construct a surrogate world, populate it with a couple of agents and then examine which constraints this world should obey if it were to mirror the real world; in particular, situations in the real world where one individual issues an imperative to another individual. The world together with the constraints to which this world is subjected will be called a model (of reality). As external observers of this model, we could choose to view such a model from an 'eternal point of view'. In other words, all of the past and future in the model would lay bare open for us to inspect. This approach is, however, not very fruitful. In practice, we eventually want to compare our constructed model with reality and determine whether it is a satisfactory picture of it. In reality we have, however, no direct access to the future and even our information about the past may be incomplete. Thus if we want our model to be comparable with reality, it should also allow for partial information.

Two desiderata for the framework have emerged: it should enable us to model both partiality (of information) and Intentionality. In order satisfy the first requirement, the framework is based on a variant of partial predicate logic (cf. Muskens, 1995a). Intentionality is dealt with by extending the standard notion of a model for a language. I propose to add a function to my models which maps specific objects of the model (representing information which an agent carries with it) to (sets of) formulae of the language. These formulae can then again be evaluated in the model as usual. Thus, certain objects in the model become Intentional: they are representing information about the model itself (this information can be true or false depending on how the translations into the language evaluate in the model).

The foundations of the framework are build by means of techniques from formal logic (specifically, model theory). In formal logic, models are specified to provide a semantics for a formal language. I also specify a formal language and its models. The models are used as a representation of a collection of communicating agents and their environment. More specifically, the models include a world and a set of constraints on this world. These constraints will be formulated in terms of the aforementioned formal language. In this respect, the formal language is used in a non-standard way.[6]

[4]Following Searle (1983), I assume that such states need not be conscious and that they are not tied to the verb 'intend'. When I intend to do something, I am in an intentional state, but there are many other types of intentional states such as belief, desire, love and hate.

[5]I am referring to the state which ensues if the addressee *accepts* the imperative. For the moment, let us keep out of consideration situations where the addressee refuses to comply with an imperative.

[6]Normally, constraints on models are formulated as axioms which are independent of the models. For our purposes, it is essential that the constraints are part of the models. This allows us to capture in a model the possible extensions of a partial world (roughly speaking, the total extensions of the world in which all the constraints are true). We can then provide an alternative definition of 'possibility', given a model, as as truth in at least one of the possible extensions of the partial world. Possibility is a concept which will be needed in the analysis of imperatives.

So let us start by introducing the formal language (henceforth L). L consists of operators, individual constants (i.e., names), individual variables and predicate constants. We assume that there are various operators of the form $\Diamond_C$, where the C is a meta-variable which ranges over sets of constraints (these are discussed in detail shortly). The operators will be interpreted as various forms of possibility. An operator $\Diamond_C$ takes a formula ϕ as its argument. $\Diamond_C\phi$ is read as 'Relative to the constraints C it is possible that ϕ'.

We use a multi-sorted logic. This means that our models contain several domains. There are (individual) constants to name the individuals in most of these different domains. The constants are grouped according to the domains over which they range. For each of the domains, there is furthermore a sufficiently large stock of (individual) variables ranging over the domain:

(1) (SORTED) INDIVIDUAL CONSTANTS AND VARIABLES

1. AGENTS: We have logical constants for two agents. They are **A** and **B**. Furthermore there are variables $a_1, a_2, \ldots$ ranging over the domain of agents. Note that we use boldface for individual constants and italics for variables.
2. OBJECTS: We assume a stock of constants for objects such as **door1**, **window1**, **window2**, **car1**, etc. Furthermore there are the variables $o_1, o_2, \ldots$ ranging over objects.
3. EVENTUALITIES: There are individual constants and the variables $e_1, e_2, \ldots$ ranging over eventualities. Our treatment of eventualities is along the lines of Parsons (1990); we assume that an eventuality is either a state or an event and some events can be seen as the actions of particular agents.
4. INSTANTS OF TIME: We have both constants and variables for instants of time. Examples of constants are **at_5**, **at_midnight**, etc. Additionally, we have the variables $t_1, t_2, \ldots$ ranging over instants of time.
5. MOODS: In this domain there are two sentence moods: imperative and declarative. They are named by the constants **imp** and **decl**.
6. COMMITMENT PEGS: A commitment peg is basically a commitment of a particular agent which this agent acquired at a particular time. We call it a commitment peg because the content and the time of introduction of the commitment are attached to it as if it were a peg by means of the two-place predicates c_time (the first argument is a commitment peg and the second one an instant of time) and c_content (the first argument is a commitment peg and the second one a singleton set consisting of a formula representing the content of the commitment). For this domain we have only variables.
7. SETS OF (WELL-FORMED) FORMULAE: There are constants and variables available which range of sets of formulae (or more specically, sentences, i.e., formulae with no free variables). The constants are given mnenomic names from which the formula which they represent can be read off directly. In general, a set of formula $\{\phi\}$ is referred to by the constant $\gg \{\phi\} \ll$. The available variables are: $s_1, s_2, \ldots$
8. COMMITMENT CONTEXTS: Agents can be committed to sets of formulae. In a moment, we will explain in more detail what it means for an agent to be committed to some piece of information. For now, we want to discern between different types of commitments. In particular, an agent, say **A**, can be committed to a particular set of formulae $\{\phi\}$, but s/he can also be committed to, for instance, $\{\phi\}$ being a joint or mutual commitment between **A** and **B**. Whereas in the former instance, we will say that $\{\phi\}$ inhabits **A**'s nil context

(represented with **-**), in the latter instance, we speak of **A**'s **AB** context. We have a domain of such different commitment contexts and a name for each of them: **-**, **AB**, **AB-A**, **AB-B**, **AB-G(A)** and **AB-G(B)**. That is, we have, respectively, the nil context, the mutual commitment context, the context in which **A** and **B** are mutually committed that **A** has a particular commitment, a similar context for commitments of **B** and mutual commitments that **A** or **B** has a particular goal.

In addition to this stock of variables and individual constants we have a collection of predicate constants. For each predicate constant we indicate which types of variables and constants it can take as its arguments. We use the notation: predicate name (arg1:type of argument, arg2:type of argument, ...). The set of predicate constants is presented in groups of constants which more or less belong together.

(2) PREDICATE CONSTANTS

1. (TIME) $\leq$ (arg1: time, arg2: time). This predicate is to be interpreted as temporal precedence/identity of instants of time. succ(arg1: time, arg2: time). succ(t, t') means that the instant t' *immediately* succeeds the instant t.
2. (EVENTUALITIES) agent (arg1: agent, arg2: eventuality). addressee (arg1: agent, arg2: eventuality). patient (arg1: agent/object, arg2: eventuality). occur_at (arg1: eventuality, arg2: time). hold (arg1: eventuality, arg2: time). say (arg1: eventuality). open (arg1: eventuality). be_open (arg1: eventuality). u_content (arg1: eventuality, arg2: set of formulae). u_type (arg1: eventuality, arg2: mood). state (arg1:eventuality). event (arg1: eventuality). The names of most constants should betray their intended interpretation. u_content and u_type stand for utterance content and utterance type, respectively.
3. (INTENTIONALITY) true (arg1: set of formulae). false (arg1: set of formulae). With these predicates we can express that a particular object from the domain which contains sets of formulae is true or false (i.e., the conjunction of the formulae is true or false). Earlier on we have called such objects intentional objects. The truth conditions of the predicates true and false make use of an (Intentionality) function from objects in the domain to propositions of the formal language L. The details are spelled out below.
4. (COMMITMENT) commit (arg1: agent, arg2: context, arg3: set of formulae, arg4: time). With this predicate we can express that a certain set of formulae can be derived from a particular subcontext of an agent's commitments. For instance, commit(**A**, **AB** ,$\{\phi\}$,**at_5**) can be paraphrased as: 'from the subcontext which represents the commitments which **A** thinks to be joint commitments with **B** all the formulae in $\{\phi\}$ can be derived at the instant of time denoted by the constant **at_5**'.

   base_commit (arg1: agent, arg2: context, arg3: commitment peg, arg4:time). These are base commitments (as oppposed to derived or inferred commitments) which are tagged with the time at which they were introduced into the agent's base of commitments.

   c_content (arg1: commitment peg, arg2: set of formulae). c_time (arg1: commitment peg, arg2: time). With these predicates we can record the time and the content of a commitment. The latter predicate will not be used in this paper. We have included it to illustrate how we can model that the agent stores information about a commitment which is different from its content. We could introduce further predicates to store, for instance, details of the surface realization of the utterance with which the commitment was introduced.

5. (FORMULAE) $\subseteq$ (arg1:set of formulae,arg2:set of formulae). $\vdash$ (arg1: set of formulae, arg2: set of formulae). This is a derivation relation between sets of formulae. singleton (arg1: set of formulae). agent_action (arg1: agent, arg2: set of formulae, arg3: time). This should be read as the set of formulae (which is stipulated to be a singleton) denotes an action of which the individual denoted by the first argument is the agent at a particular time (the third argument).

We define terms to be either individual constants or variables (we use the meta variables $T_1, \ldots, T_n$ for terms). The set of formulae is defined as follows:

(3) FORMULAE

1. If P is an n-ary predicate constant and $T_1, \ldots, T_n$ are terms of the correct sort, then $P(T_1, \ldots, T_n)$ is an atomic formula;
2. If T_1 and T_2 are terms of the same sort, then $T_1 = T_2$ is an atomic formula;
3. If ϕ and ψ are formulae, then $\neg\phi$, $\phi \wedge \psi$ and $\phi \vee \psi$ and $\phi \rightarrow \psi$ are formulae;
4. If ϕ is a formula and X is an individual variable, then $\forall X.\phi$ is a formula;
5. If S and S' are variables ranging over the sort 'set of formulae' and ϕ is a formula, then $S = \bigcup S' : \phi$ is a formula;
6. If ϕ is a formula and C_i denotes a set of constraints, then $\Diamond_{C_i}\phi$ is a formula.

The truth conditions of these formulae (in particular, those with no free variables, i.e., the sentences or closed formulae of L) are defined with respect to their intended models. Formally, a model is defined as follows:

(4) CONSTRAINT-BASED INTENTIONAL PARTIAL MODEL

A model is a tuple $\langle W, C \rangle$ consisting of a (possibly) partial world W and a list of sets of constraints $C = C_1, \ldots, C_n$ on W (constraints are expressed with formulae of L). A world W is a quadruple $\langle D, I^+, I^-, Int \rangle$. It consists of a list of domains $D = D_1, \ldots, D_n$, a positive and a negative interpretation function I^+ and I^-, and an intentionality function Int.

Standard models for predicate logic are tuples $\langle D, I \rangle$ consisting of a domain and an interpretation function. The interpretation function maps individual constants to members of D and n-ary predicate constants to subsets of D^n. For instance, a constant **john** is mapped to the corresponding individual in the domain and a two-place predicate constant such as $love$ is mapped to a subset of $D^2 = D \times D$. In other words, $love$ is mapped to a set consisting of tuples of individuals, such that the first individual of the tuple stands in the relation $love$ to the second one. In our models, there are two interpretation functions (I^+ and I^-) instead of one. We stipulate that for individual constants I^+ and I^- are identical. In order to bring partiality into our models, the interpretation functions differ for predicate constants. Consider again the predicate constant $love$. In ordinary models, the interpretation function gives us the pairs of individuals which stand in this relation. Any pair of individuals which is not returned by the interpretation function is automatically classified as not standing in the relation. It is impossible to express lack of information with regard to whether a particular pair of individuals stands in the relation or not. By introducing two interpretation functions, partiality of information becomes expressible: a pair of individuals can be standing in the relation (i.e., it is a member of $I^+(love)$), not stand in the relation (i.e., it is a member of $I^-(love)$) or there is no information about whether the relation holds between the individuals or not (i.e., it is neither a member of $I^+(love)$ nor of $I^-(love)$).[7]

Another non-standard ingredient of our models is the (Intentionality) function Int. It is a function from a subdomain D_k of D to sets of formulae of the language L. Note that within this framework we can express an analogue of the Liar sentence, i.e., 'This sentence is false'. We can form a formula which says that the singleton set consisting of that formula (denoted by a constant from D_k) is false: false(**c**) (where **c** denotes $\{$false(**c**)$\}$). In our framework, it is not possible to show that this formula is true or false with respect to a model (it can be shown that both attempts to construct a proof of truth and falsity lead to an infinite recursion). Since we use a partial logic, this does, however, not lead to a paradox: rather the formula comes out as undefined. Intuititively, the introduction of Int means that the members of D_k can carry (true or false) information about the model itself.

We have given a general definition of Constraint-based Intentional Partial models. Let me now define a particular subclass of these models, that is, communicating agents models:

(5) COMMUNICATING AGENTS MODEL

A communicating agents model M is a Constraint-based Intentional Partial Model $\langle\langle D, I^+, I^-, Int\rangle, C\rangle$ such that:

1. $D = D_a, D_o, D_e, D_t, D_m, D_{cp}, D_{sf}, D_{cc}$. We have domains for, respectively, agents, objects, eventualities, instants of time, moods, commitment pegs, sets of formulae and commitment contexts;
2. $Int : D_{sf} \mapsto$ Formulae of L;
3. $C = C_{utt}, C_{comm}, C_{env}, C_{time}, C_{action}$. We have sets of constraints pertaining to, respectively, utterances, the environment, the temporal structure of reality and actions by the agents. (In fact, strictly speaking C_{utt} is a subset of C_{action}).

Before we can finally give our semantics for L, we need to introduce some further notions. Our partial models come with a natural structure, that of *informational subsumption*; we write $\sqsubseteq$. Given two models M_1 and M_2 such that $C_1 = C_2$, we say that $M_1 \sqsubseteq M_2$ (read: M_2 informationally subsumes M_1) if and only if for both $p = +$ and $p = -$: $I_1^p(c) = I_2^p(c)$ for all individual constants c and $I_1^p(c) \subseteq I_2^p(c)$ for all predicate constants c. A model is called *total* if there is no other model which informationally subsumes it. Furthermore, we write $M_{\leq t}$ for models which are undefined for precisely those n-tuples such that the n-th member is an instant of time and this instant is bigger than t. In other words, these are models which are fully defined up till time t.

We use the usual notion of an assignment a given a model M. Such an assignment maps variables of a particular sort z to members of the corresponding domain D_z of D. An assignment $a[d/x]$ is defined as being identical to the assignment a except for assigning d to x. The value of a term T with respect a model M and an assignment a (written $\|T\|^{M,a}$, and abbreviated as $\|T\|$) is $I^+(T)$ if T is a constant and $a(T)$ if T is a variable.

(6) SEMANTICS: TARSKI TRUTH DEFINITION

1. $M \models P(T_1, \ldots, T_n)[a]$ iff $\langle\|T_1\|, \ldots, \|T_n\|\rangle \in I^+(P)$
   $M \mathrel{=\!\!|} P(T_1, \ldots, T_n)[a]$ iff $\langle\|T_1\|, \ldots, \|T_n\|\rangle \in I^-(P)$
   (where P is not equal to true or false);

[7]We exclude overdefinedness, that is, situations where a (sequence) of individuals is both a member of $I^+(R)$ and $I^-(R)$ for some relation R. Our intention is to use our models as representations of reality. Since reality is supposed to be consistent, we require our models to be so as well.

2. $M \models T_1 = T_2[a]$ iff $\|T1\| = \|T2\|$
   $M \mathrel{=\!\!|} T_1 = T_2[a]$ iff $\|T1\| \neq \|T2\|$;
3. $M \models \neg\phi[a]$ iff $M \mathrel{=\!\!|} \phi[a]$
   $M \mathrel{=\!\!|} \neg\phi[a]$ iff $M \models \phi[a]$;

   $M \models \phi \wedge \psi[a]$ iff $M \models \phi[a]$ and $M \models \psi[a]$
   $M \mathrel{=\!\!|} \phi \wedge \psi[a]$ iff $M \mathrel{=\!\!|} \phi[a]$ or $M \mathrel{=\!\!|} \psi[a]$;

   $M \models \phi \vee \psi[a]$ iff $M \models \phi[a]$ or $M \models \psi[a]$
   $M \mathrel{=\!\!|} \phi \vee \psi[a]$ iff $M \mathrel{=\!\!|} \phi[a]$ and $M \mathrel{=\!\!|} \psi[a]$;

   $M \models \phi \rightarrow \psi[a]$ iff $M \mathrel{=\!\!|} \phi[a]$ or $M \models \psi[a]$
   $M \mathrel{=\!\!|} \phi \rightarrow \psi[a]$ iff $M \models \phi[a]$ and $M \mathrel{=\!\!|} \psi[a]$;
4. $M \models \forall X\phi[a]$ iff $M \models \phi[a[d/x]]$ for all $d \in D_{sort(X)}$
   $M \mathrel{=\!\!|} \forall X\phi[a]$ iff $M \mathrel{=\!\!|} \phi[a[d/x]]$ for some $d \in D_{sort(X)}$;
5. $M \models true(S)[a]$ iff $M \models \bigwedge(Int(S))[a]$
   $M \mathrel{=\!\!|} true(S)[a]$ iff $M \mathrel{=\!\!|} \bigwedge(Int(S))[a]$;

   $M \models false(S)[a]$ iff $M \mathrel{=\!\!|} \bigwedge(Int(S))[a]$
   $M \mathrel{=\!\!|} false(S)[a]$ iff $M \models \bigwedge(Int(S))[a]$;
6. $M \models S_1 = \bigcup S_2 : \phi[a]$ iff $\|S1\| = \bigcup \|S2\|$: $M \models \phi[a]$
   $M \mathrel{=\!\!|} S_1 = \bigcup S_2 : \phi[a]$ iff $\|S1\| \neq \bigcup \|S2\|$: $M \models \phi[a]$;
7. $M \models \Diamond_{C_i}\phi[a]$ iff there is an M' such that $M \sqsubset M'$, M' is total,
   $M' \models \bigwedge C_i[a]$ and $M' \models \phi[a]$
   $M \mathrel{=\!\!|} \Diamond_{C_i}\phi[a]$ iff there is no M' such that $M \sqsubset M'$, M' is total,
   $M' \models \bigwedge C_i[a]$ and $M' \models \phi[a]$.

We read '$M \models/\mathrel{=\!\!|}\ \phi[a]$' as ϕ is true/false in model M under assignment a. The clauses 1., 2., 3. and 4. are along the lines of those of Muskens' (1995a) partial predicate logic. Clause 5. makes essential use of our Intentionality function. For instance, the first item of this clause says that the formula $true(S)$ (where S is a term denoting a set of formulae) is true iff the conjunction of the members of $Int(S)$ is true. $Int(S)$ stands for the set of formulae which is denoted by the term S. Clause 6. allows us to construct a set which is the union of a set of formulae which have a particular property ϕ. Finally, according to the first item of clause 7., ϕ is possible given a model M and a set of constraints C_i (and an assignment) if and only if there is model M' which properly subsumes M ($M \sqsubset M'$), is total, makes the conjunction of the constraints which are a member of C_i true and, finally, makes ϕ itself true.

Our semantics for possibility is similar to the semantics which Landman (1986:53) assigns to the word 'may'. Landman's definition is, however, given for propositional logic and does not relativize possibility to a set of constraints. In particular, our proposal to make the constraints part of the model is different from Landman's treatment. Furthermore, Landman argues that the definition requires some refinements to deal with all the correct inferential patterns in which the word 'may' can occur. For our purposes, such refinements would, however, unnecessarily complicate our treatment of possibility.

3 Theory

In this section, I specify the sets of constraints C_{utt}, C_{comm}, C_{env}, C_{time} and C_{action} which feature in our models. For reasons of space it will be impossible to provide a complete formal version of each and every constraint which I use. The ones which are most pertinent to the problems (P.1) and (P.2) are, however, spelled out in detail. I hope that this section provides the reader with an idea of how to use the framework which has been introduced in the previous section to formulate concrete theories of communicating agents.

Time and Eventualities Let me start with C_{time}. I assume that the time line is a discrete linear order which is infinite in both directions. We stipulate that events are seen as *transitions* between instants of time. Given an event e and an instant of time t, we write $occur\_at(e,t)$ to say that the event e took place between t and the instant of time which immediately succeeds t. Furthermore, we assume that each event is associated with at most one transition between two instants of time:

(7) $\forall e.((\exists t.occur\_at(e,t) \wedge event(e)) \rightarrow (\forall t'.occur\_at(e,t') \rightarrow t = t'))$

Whereas we have the predicate *occur_at* for events, we have the predicate $hold$ for states. A state can hold at several instants of time. However, these instants have to be a connected series. In other words, if a state occurs at two times, then there exists no time in between those times at which it does not occur:

(8) $\forall e.\forall t.\forall t'.((hold(e,t) \wedge hold(e,t') \wedge state(e)) \rightarrow \forall t''.(t \leq t'' \leq t' \rightarrow hold(e,t'')))$

Commitment Let us proceed with C_{comm}. At each instant of time an agent has for each of its commitment contexts a (possibly empty) set of base commitments. For each base commitment, the agents also has information about the time at which that commitment was taken on. These base commitments give rise to a set of derived commitments. Basically, a set of formulae s is a derived commitment for an agent a at some given instant of time t iff s can be derived from the union of the base commitments which the agent maintains at that moment of time:[8]

(9) $\forall a.\forall c.\forall s.\forall t.commit(a,c,s,t) \leftrightarrow$
$\exists s'.((s' = \bigcup s'' : (\exists p.base\_commit(a,c,p,t) \wedge c\_content(p,s'')) \wedge \vdash (s',s))$

It is beyond the scope of this paper to provide any constraints for the relation '$\vdash$' which relates direct commitments to indirect (inferrable) ones. It is assumed that it corresponds to some (computable) relation of derivation between sets of formulae. I already pointed out that an agent has several different commitment contexts. There are, however, constraints which relate the content of these contexts to each other. Again, I will only provide an example of such a constraint (cf. Zeevat, 1997):

(10) $\forall a.\forall s.\forall t.(((a = \mathbf{A} \vee a = \mathbf{B}) \wedge commit(a, \mathbf{AB\text{-}A}, s, t) \wedge commit(a, \mathbf{AB\text{-}B}, s, t))$
$\rightarrow commit(a, \mathbf{AB}, s, t))$

[8] For the sake of uniformity it would have been desirable to represent (base) commitment as any of the other states which can hold at a given instant of time. For reasons of conciseness, I have, however, chosen for a more compact notation.

This constraint says that if both the contexts **AB-A** and **AB-B** of an agent a (**A** or **B**) contain some commitment s, then the context **AB** contains this commitment as well. In other words, if an agent thinks that it is a mutual commitment that **A** is committed to s and it is a mutual commitment that **B** is committed to s, then this agent (assuming that the agent is **A** or **B**) thinks that s itself is a mutual commitment.

I have used the notion of commitment without first providing a definition. I have relied on the reader's intuitive understanding of this notion. The definition is given implicitly in the course of this paper by means of the constraints which are imposed on commitments. We have seen some constraints which relate different types of commitments with each other. In a moment, we provide further constraints which explicate the role of commitments in the behaviour of an agent.

Utterances An agent can acquire new commitments through observation and through communication with other agents.[9] In the real world, an agent can also change his or her mind and retract a commitment. For our purposes, this would, however, introduce complications which detract from the main issues which this paper addresses. Therefore, I assume that once an agent has acquired a commitment, that commitment will persist through time.

New commitments can be introduced through communication. The constraints in C_{utt} regulate the relation between utterance events and the introduction of new commitments. We assume that the agents cannot simultaneously produce an utterance. For that purpose we have a constraint (which we will not spell out in formal detail) which says that only one utterance event can occur at an instance of time. This is, of course, a simplification, and at a later time we might want to relax this constraint. The relation between the utterance of an imperative and the commitments of the addressee of an imperative are spelt out by the following constraint:

(11) $\forall e.\forall t.\forall t'.\forall s.(agent(\mathbf{A}, e) \wedge say(e) \wedge addressee(\mathbf{B}, e) \wedge$
$occur\_at(e, t) \wedge u\_content(e, s) \wedge u\_type(e, \mathbf{imp}) \wedge succ(t, t') \rightarrow$
$\exists p.(base\_commit(\mathbf{B}, \mathbf{AB\text{-}A}, p, t') \wedge c\_content(p, s) \wedge c\_time(p, t)))$

Roughly speaking, according to this constraint if agent **A** utters an imperative to **B** with content s between t and $t + 1$, then **B** updates his base commitments in the context **AB-G(A)** with s at time $t + 1$. In other words, an imperative with content ϕ causes the speaker to think that it is now a mutual commitment that ϕ is a goal of the speaker. There is a further constraint which says the same for the **B**'s context **AB-G(A)** and two further constraints which apply when **B** is the speaker instead of **A**. Furthermore, we have a constraint which deals with declaratives. It says that if the content of the declarative is s and the speaker is **A**, then both speaker and addressee (**B**) add s to their **AB-A** context. In other words, they now both think that it is a mutual commitment that **A** is committed to s.

Of course, these constraints are idealizations. We have not taken into account situations which involve miscommunication. Furthermore, the relation which we have posited between sentence mood and the update of the speaker's and addressee's commitments does not take into account indirect speech acts (Searle, 1975). We trust that such more elaborate theories can be formulated within the framework which is presented here. However, such a theory is at the moment not our main concern (but see, e.g., Beun (1994) for a more elaborate account of the relation between utterances and intentional agent states).

Typically, a commitment is taken on through a declarative utterance. For instance, in 'A: Open the door. B: Ok (I will)', B's utterance is taken to be such a declarative. According to an earlier mentioned

[9] Inferred or derived commitments are not taken to be new commitments.

constraint, after the utterance, A and B each think that it is a mutual commitment that B is committed to opening the door. Let us assume that, as described, the addressee of an imperative signals that s/he will comply with the imperative. Thereby the agent agrees that it is a mutual commitment (amongst speaker and addressee) that the agent is committed to content of the imperative. So according to our analysis so far the function of an imperative is to induce such a commitment. The next question then is what such a commitment amounts to. I want to argue that a proper answer to this question requires us to look from two different perspectives at such a commitment. (**I**) On the one hand, we can ask whether the content of the commitment is true, or can still become true in the world. (**II**) On the other hand, we have an agent who's actions are influenced by the commitment, ideally, in such a way that the content of the commitment does become true. In other words, agents execute a certain policy in order to make sure that their commitments satisfied in the truth conditional sense. However, my claim is that satisfaction of a commitment is not judged purely in terms of truth conditional satisfaction but also in terms of further constraints on the policy which led to that satisfaction.

I want to argue that the two problems which are discussed in the introduction of this paper arise from not properly taking into account the second perspective on the satisfaction of imperatives. Consider again the problem (P.1). Suppose an imperative with the content $\phi \to \psi$ is issued to an agent. If the agent subsequently and as a *result* of this imperative goes about making sure that ϕ is false, then that agent does not act in the spirit of the imperative. And yet, according to perspective (I) there is nothing wrong with such behaviour: according to this perspective an imperative is satisfied if its content is true in the world. In order to explain the infelicity of the agent's actions, we need to bring perspective (II) into play. I want to argue that agents are expected to go about satisfying imperatives only in line with policies for action of a certain type. In particular, an adequate policy should exploit the fact that conditionals of the form 'If such and such a state holds or event occurs, then do such and such' can be seen a *rules* for guiding the behaviour of an agent in a given situation. The picture emerges of an agent who moves about in the world, maintaining a clock which indicates the time and checking whether at the current time there are any actions which s/he needs to execute in order to satisfy his or her commitments.

Actions, Commitment and Environment In our framework, the triggering of a rule relative to the other commitments which an agent maintains can be formalized in terms of the (logical) derivability of the consequent of the rule. A rule, i.e., conditional commitment $\phi \to \psi$, is triggered if the agent's commitments, which include the conditional commitment, allow the agent to derive the consequent ψ of the conditional commitment. More generally, we have the following policy which relates an agent's actions to his or her commitments: At every instant of time t, the agent checks which actions (of him or herself) at time t can be derived from the commitments. Precisely these actions, s/he then carries out. We can formulate this as a constraint by saying that for all times if an agent is committed (directly or through a derived commitment) to some action of him- or herself at that specific instance of time, then this action is carried out by her or him:

(12) $\forall a.\forall s.\forall t.((commit(a, -, s, t) \wedge agent\_action(a, s, t)) \to true(s))$

Now compare this with the following constraint which is in the spirit of perspective (I) on imperatives and commitment:

(13) $\forall a.\forall s.\forall t.(commit(a, -, s, t) \to true(s))$

In words, all commitments of an agent should be true in the world. Notice that constraint (13) is stronger than –i.e., entails– (12). The former requires all commitments to come true, not just those pertaining to actions of the agent.[10]

Worlds which satisfy constraint (13) can be seen as ideal worlds, whereas the constraint (12) can be seen as the basis of a policy for an agent's behaviour to end up in such a world. In order for an agent to approximate (13) through such a policy we need to invoke additional assumptions about the behaviour of that agent. Let us start with a very simple model which requires almost no further assumptions about the agent's behaviour. As we progress to more complex models, the number of assumptions will increase. We assume that in this world C_{env} is empty: there are no interactions between states and events of the environment (i.e., no pre- and postconditions on events and no situations in which one event is part of another event). Furthermore, we assume that the agent's commitments are conjunctions of atomic formulae. These formulae contain no variables. All reference to objects, agents, times, eventualities, etc. are achieved by means of individual constants. Such a set of commitments will contain a subset which denote actions by the agent. The remainder will be actions by other agents, events and states. Let us assume that the latter come true by definition.[11] Commitments pertaining to the agent's own actions will come true if s/he simply carries out each and every one of them. Let us add negation to this model. Now, commitments are conjunctions of possibly negated atomic formulae. In this new set up, we have to extend the agent's policy with the following clause: s/he refrains from an action if its negation follows from his or her commitments.

Let us move to a slightly more complicated model. Assume that commitments can also be conditionals (although quantification is still not permitted). Consider a conditional commitment of the form $\phi \to \psi$, where ψ denotes an action by the agent. Assume that this is the only commitment of the agent. It is evident that the policy of simply carrying out all actions which directly follow the commitments will no longer guarantee that is (13) satisfied. For instance, there could be a situation where ϕ is true, without the agent being committed to ϕ. In that situation, the agent ought to carry out ψ in order to make $\phi \to \psi$ true. But since the agent does not avail over the information ϕ, s/he will do no such thing. In this new set up, an agent will have to actively be on the look-out for information which can trigger conditional commitments. A possibility might be that the agent tries to check all the atomic subformulae of his or her commitments (which can not be derived from his or her commitments) for their truth or falsity. The thus obtained information could then be added to the agent's commitments. Given such a set of commitments, the agent's policy would again be complete with respect to constraint (13).

Finally, let us assume that commitments can also be quantified formulae. Now, checking for the truth of subformulae will no longer do, since they can contain unbound variables. A complete policy could be attained by checking for the truth of such formulae under all possible substitions for their variables. For any realistic domain (i.e., with a sufficiently large domain of individuals), this would, however, not be a feasible policy. We might raise the question how human agents solve this problem. Clearly they have no fool proof policy. It would be interesting to investigate whether their actual policy approximates the aforementioned one in any way.

(14) (EXAMPLE) The content of the imperative 'Say hello to John if you see him' (issued to **A**) is formalized as:

[10] The latter is expressed by the subformula $agent\_action(a, s, t)$. We can see this subformula as a partial description of the object s, which itself again stands for a set of formula. Currently, we assume that in our models elements of the domain D_{sf} (of sets of formulae) either satisfy such a predicate or not and that this corresponds with the syntactic structure of the formula. This idea could be worked out more explicitly by using a tree description logic, where formulae are explicitly modelled as trees instead of primitive objects of the domain D_{sf} (see, e.g., Muskens, 1995b).

[11] Roughly speaking, these commitments correspond to the agent's beliefs. The agent does not actively make sure that they come true, although s/he should not perform actions which make them come out false.

$$(\forall t.\forall e.(see(e) \wedge agent(\mathbf{A}, e) \wedge patient(\mathbf{j}, e) \wedge occur\_at(e, t) \wedge \leq (\textbf{utterance\_time}, t))$$
$$\rightarrow (\exists e'.say\_hello(e') \wedge agent(\mathbf{A}, e') \wedge patient(\mathbf{j}, e') \wedge occur\_at(e', t)))$$

I now want to argue that the approach I have sketched also provides a solution to the problem (P.2). (P.2) arises from Ross' (1941) observation that an imperative with the content ϕ does not seem to entail an imperative with the content $\phi \vee \psi$. Hence, imperatives do not appear to conform to the standard laws for logical inference (e.g., in FOL). I agree with Ross' intuitions that an addressee who has been told to 'burn a letter', should not, subsequently, infer that s/he should 'burn or send the letter'. Let me first describe an approach (within my framework) which would lead to such a paradoxical conclusion, and then indicate how the approach I actually advocate solves the problem. According to perspective (I), an agent should make sure that all his commitments come true. How an agent actually goes about making sure that this really happens belongs to the realm of perspective (II). One possible policy is that an agent simply takes a commitment which follows from his commitments and tries to make it true, and then proceeds to make the next commitment which follows true, until all her or his commitments are true.

Now consider an agent which is only committed to ϕ (where ϕ describes an action by the agent and might have been acquired as a result of the result of imperative). The agent might first try to make a commitment true which follows from ϕ, e.g., $\phi \vee \psi$. In particular, the agent might choose to make ψ true. That would, however, be a bad policy, since there might be a logical connection between ϕ and ψ, e.g., it might be the case that ϕ and ψ cannot be true at the same time (one cannot burn a letter and send it at the same time). In that situation, making ψ true prevents the agent from making ϕ true. The policy I propose cannot lead to such an impasse since it requires an agent to only make true those commitments which describe primitive actions by the agent and which can be derived from his or her base commitments. Now, clearly we cannot derive the action description ψ from ϕ. Thus, our solution to the problem is to maintain the logical connection between ϕ and $\phi \vee \psi$, but provide a new account of the role of commitments in the behaviour of an agent.

Up till this point I have assumed that the environment in which the agents operate is not subject to any interactions. For instance, there are no pre- and postconditions on actions. Adding these introduces a further complication. In particular, if we want to evaluate an agent's behaviour at a specific point in time t with respect to the commitment s/he has taken on. We then have to evaluate this agent's behaviour with respect to a world $W_{\leq t}$ which is specified only up till t. Since some of the commitments might be about future actions of the agent, we can no longer simply demand that all his commitment should be true in $W_{\leq t}$. Here, we need to take recourse to the concept of possibility:

(15) $\forall a.\forall s.\forall t.(commit(a, -, s, t) \rightarrow \Diamond_{C_{env}} true(s))$

Whether an agent behaviour satisfies his commitments at a given time t now depends on whether it is still *possible* (with respect to the constraints to which the environment is subjected) to make all his or her commitments come true.

4 Conclusions

A framework for communicating agents has been introduced. This framework combines and extends various techniques from formal logic. In particular, the framework allows for both Intentional states of agents and Partial models of reality. Furthermore, constraints are taken to be explicit ingredients of our models. They regulate possible extensions of partial descriptions of reality. Within the framework, we can model agents who carry commitments around which they took on as a result of imperatives. I provide an outline of several, increasingly more complex, models of how these commitments should influence the behaviour of an agent (so-called *policies*) for him/her to be said to comply/satisfy the imperative which gave rise to the commitment. The policy which I propose provides a solution to the problems (P.1) and (P.2) which were presented in the introduction of this paper.

Although I have attempted to achieve a certain level of formal rigour in my analysis, there are still many loose ends which require further development. For instance, the proposed model does not explicitly deal with the interpretation process; the question of how an agent arrives at a particular commitment given the surface form of an (imperative) utterance. The context dependence of this process will have to be incorporated into the model for phenomena such as Indirect Speech Acts (cf. Power, 1979; Beun, 1994) and anaphora resolution (e.g., Ahn, 2000; Kamp & Reyle, 1993; Krahmer, 1998; Piwek, 1998; Van Deemter, 1991).[12] Furthermore, the proposed model of time and eventualities makes some strong assumptions (e.g., the assumption that events are instantaneous) which cannot be maintained in the long run. A further issue which requires more discussion is the relation of the proposed model to planning theories of discourse (e.g., Lochbaum, 1994). Lochbaum describes how discourse can lead to complex plans (although, not logically complex, as the commitments described in this paper, but rather complex in the sense of involving hierarchical structures of subplans and actions). The focus of this paper is complementary to that work: I have focussed on how an agent's behaviour is influenced given that s/he has adopted a set of (logically) complex commitments (such a set of commitments can be seen as a partial plan). Despite all these shortcomings, I hope that the work provides a formal basis for an agent-based analysis of linguistic phenomena[13], such as, imperatives.

References

Ahn, R. (2000). *Agents, Objects and Information: Modelling the Dynamics of Interaction* (working title), PhD. Dissertation, Tilburg University. (To appear)

Beun, R.J. (1994). 'Mental state recognition and communicative effects', *Journal of Pragmatics*, **21**, 191–214.

Biber, D. S. Johansson, G. Leech, S. Conrad and E. Finegan (1999). *Longman Grammar of Spoken and Written English*, Harlow: Longman.

Hamblin, C. (1987). *Imperatives*, Oxford: Blackwell.

Hausser, R. (1999). *Foundations of Computational Linguistics*, Berlin: Springer Verlag.

Kamp, H. & U. Reyle. (1993). *From Discourse to Logic*, Dordrecht: Kluwer Academic Publishers.

Krahmer, E. (1998), *Presupposition and Anaphora*, Stanford: CSLI Publications and Cambridge University Press, CSLI Lecture Notes Series **89**.

Landman, F. (1986). *Towards a Theory of Information*, Dordrecht: Foris Publications.

Lochbaum, K. (1994) *Using Collaborative Plans to Model the Intentional Structure of Discourse*, PhD. Dissertation, Harvard University.

Muskens, R. (1995a). *Meaning and Partiality*, Stanford: CSLI Publications, Studies in Logic, Language and Information Series.

Muskens, R. (1995b). 'Order-Independence and Underspecification', In: J. Groenendijk (ed)., *Ellipsis, Underspecification, Events and More in Dynamic Semantics*, DYANA Deliverable R.2.2.C.

Parsons, T. (1990). *Events in the Semantics of English*, Cambridge, Massachusetts: MIT Press.

Piwek, P. (1998). *Logic, Information and Conversation*, PhD. Dissertation, Eindhoven University of Technology.

[12]Extending the current framework in that direction could be realized a move from Predicate Logic to Discourse Representation Theory (Kamp & Reyle, 1993). We would need to change the domain D_{sf} to contain objects representing Discourse Representation Structures, instead of sets of Predicate Logical Formulae. The Intentionality function Int would then be defined as a truth preserving mapping of these structures into sets of Predicate Logical formulae of language L of the framework.

[13]Such an approach is argued for at length in, for instance, Hausser (1999).

Power, R. (1979). 'The Organisation of Purposeful Dialogues', *Linguistics*, **17**, 107–152.

Ross, A. (1941). 'Imperatives and Logic', *Theoria*, **7**, 53–71.

Searle, J. (1975), 'Indirect Speech Acts', In: Cole, P. & J. Morgan (eds), *Syntax and Semantics 3 (Speech Acts)*. New York: Academic Press, 59–82.

Searle, J. (1983). *Intentionality*, Cambridge: Cambridge University Press.

Van Deemter, C. (1991). *On the Composition of Meaning*, PhD. Dissertation, University of Amsterdam.

Zeevat, H. (1997). 'The Common Ground as a Dialogue Parameter'. In: Benz, A. & G. Jäger (eds), *Proceedings of MunDial '97*, CIS–Bericht 97-106, Department of Computational Linguistics, University of Munich, 195–214.