Indirect Speech Acts and Politeness: A Computational Approach

Liliana Ardissono and Guido Boella and Leonardo Lesmo
Dipartimento di Informatica - Università di Torino
Corso Svizzera n.185 - 10149 Torino - Italy
Phone: +39-11-7429111; E-mail: {liliana, lesmo}@di.unito.it

This paper appeared on the Proc. of the 17th Cognitive Science Conference, Pittsburgh, 1995

Abstract

This paper describes a framework for the representation and interpretation of indirect speech acts, relating them to the politeness phenomenon, with particular attention to the case of requests. The speech acts are represented as actions of a plan library and are activated on the basis of the presence of syntactic and semantic information in the linguistic form of the input utterance. The speech act analyzer receives in input the semantic representation of the input sentence and uses the politeness indicators to climb up the decomposition and generalization hierarchies of acts encoded in the library. During this process, it eliminates the indicators and collects the negated presuppositions (represented as effects of the indirect speech act) that characterize the politeness forms. Some cyclic paths in the hierarchy allow the system to cope with complex sentences including nested politeness indicators. In the proper places of the hierarchy the semantic representation of the input sentence is converted into a domain action in order to start-up, when needed, the domain-level plan recognition process.

Introduction

Since Austin (1962) and Searle (1969) wrote their papers about speech acts, it was clear that the study of language must take into account the way people use it to move in the world. An utterance is an action, so it is made with some goals in mind. Among these goals, getting cooperation from the audience and maintaining a good relationship with them play a major role. The cooperation can range from simple attention (if you just want to chat), to providing information (in case of questions), to performing some general action (as closing a window if the speaker asks the hearer to do so). In all these cases, speech acts must be planned by taking into account the relation between the speaker and the hearer.

A major step in computational linguistics was made when the study of traditional fields as syntax and semantics was complemented with the computational study of pragmatics. However, this was accomplished by paying attention mainly to the first of the two goals mentioned above. In particular, it was recognized that goals and plans play a basic role in linguistic communication (Allen & Perrault 1980), but their study was centered on *domain plans*. In the last fifteen years various models of recognition of the speaker's plans were developed, some of which gave fundamental formal accounts of the knowledge which it is based on (Cohen & Levesque, 1990; Cohen & Perrault, 1979), while others had a more computational bias (Carberry, 1988). More recently domain plans have been complemented with higher levels plans called discourse plans (Litman & Allen, 1987) and problem solving plans (Lambert, 1993). While Litman and Allen's discourse plans dealt both with communication strategies and problem-solving activities, Lambert separates the discourse level in two parts: in her framework, communicating strategies are represented in the communicative level, while problem-solving plans model the activity of building the speaker's domain plans.

The present work addresses mainly the second goal mentioned in the first paragraph: what linguistic forms enable a speaker to manifest her/his choice to be more or less polite with the hearer? The desire of maintaining some harmony with the hearer is just one of the multiple goals of the conversation, so the problem of modeling this desire can be faced from a general perspective of modeling goals. However, the features that express the choices made are rather special; while the propositional content of a sentence enables the hearer, after some rather complex inferential activity, to understand the speaker's goals, it is the form in which that propositional content is expressed that makes the utterance more or less polite. For example, the following sentences have the same *illocutionary force*, but a different literal interpretation (e.g. **1b** refers to the hearer's capabilities, **1c** projects on a hypothetical perspective the hearer's action, **1d** refers to her/his wants, while **1e** simply mentions an unsatisfied precondition of the desired act):

1a) *Dammi le chiavi della biblioteca!*
[Give me the keys of the library!]
1b) *Potresti darmi le chiavi della biblioteca?*
[Could you give me the keys of the library?]
1c) *Mi daresti le chiavi della biblioteca?*
[Would you give me the keys of the library?]
1d) *Ti dispiace darmi le chiavi della biblioteca?*
[Do you mind giving me the keys of the library?]
1e) *La biblioteca è chiusa*
[The library is closed.]

Our goal is to get rid of these aspects of the literal interpretation, assuming that their role is just to mark the politeness strategies the speaker has adopted in communicating. This approach follows the guidelines drawn by the research of many linguists, that have investigated the notion of politeness and its implications in communication (Brown & Levinson, 1987; Kasper, 1990; Leech,

1983): they have shown that the origin of many indirect forms of expression lies in the necessity of smoothing the interaction for being polite. As far as the notion of politeness is concerned, various more or less precise explanations have been formulated. In our work, we will refer to Brown and Levinson (1987), who motivate the use of indirect forms of expression with the necessity to preserve some wants that every interlocutor has. In order to characterize these wants they introduce the notion of *face* as:[1].

The public self-image that every member [of a society] *wants to claim for himself, consisting in two related aspects:*

a) negative face: the basic claim of territories, personal preserves, rights to non-distraction -i.e. to freedom of action and freedom from imposition

b) positive face: the positive consistent self-image or 'personality' (crucially including the desire that this self-image be appreciated and approved of) claimed by interactants

Brown and Levinson interpret the behavior of speakers on the basis of a taxonomy of linguistic strategies that enable a speaker to satisfy the goal of preserving the negative face of the interlocutor.[2] For example, when a speaker wants the hearer to perform an action, s/he can express her/his request directly, using an imperative form; however, in this way, s/he does not preserve the hearer's negative face: in fact, she does not hide the presupposition that s/he believes that the hearer wants to execute the action. So, a safer strategy is to use an indirect request such as **1d**, which doesn't presuppose any hearer's attitude towards the requested action (in fact, s/he is questioned about that). The conditional mood in sentence **1c** (*mi daresti*: 'would you give me') has a similar role: in this case the presupposition is canceled by projecting the utterance on an hypothetical world.

The various methods for modulating the strength of utterances are chosen according to the degree of familiarity, respect, relative social roles of the interactants, and the impact that the contents of the acts might have on the interlocutors (Brown & Levinson, 1987).

This paper takes into account the suggestions coming from the authors mentioned above to implement a method for processing and evaluating indirect speech acts as politeness forms. This is done within a framework of plan recognition that has already been applied successfully to the recognition of domain plans in an information-seeking environment (Ardissono *etal*, 1993; Ardissono *etal*, 1994; Ardissono & Sestero, 1995). It must be observed that Hinkelman and Allen (1989) challenged the possibility of facing this problem on the sole basis of planning structures. They argue that the variability of politeness forms among different languages calls for the introduction of knowledge about idioms. While we agree on the need of language-specific knowledge, we will show that the required information can be encoded within a plan formalism, so that the homogeneity of the representation is preserved.

The rest of the paper is organized as follows: the first section describes the formalism used for representing the knowledge about speech acts; the second describes how the speech acts library is used in the process of speech act recognition; the third section shows the speech-act recognition process on an example. Finally, some brief conclusions are presented.

The representation of the speech acts

The knowledge about speech acts and the way they relate to each other is stored in the speech acts library, represented in an action hierarchy based on a formalism similar to that by Kautz (1990). In particular, we set apart the decomposition hierarchy (boxed arcs in the figures) and the generalization-specialization hierarchy (thick arrows).[3] When the decomposition includes a single step, the relation between the two actions is a *generation* relation (Pollack, 1990). The leaves of the hierarchy, **surf-imperative, surf-yn-question, surf-wh-question, surf-assertion** correspond to the different syntactic types of sentence, namely imperatives, declaratives and interrogatives (two small portions of the library are reported in Figure 1 and 2. There, the surface types are circled by thick ovals).

The actions of the hierarchy are characterized by the following features:

- **parameters:** the parameters of an action include the speaker, the hearer and a reference to the speech act. The third parameter has different meanings in the various actions of the speech act library: since the interpretation of surface speech acts starts with the analysis of the linguistic aspects in the input utterances (e.g. the detection of politeness features), the actions related to that phase refer to the semantic representation of the input sentences (e.g. consider **ask-if**, ..., **indirect-req** in Figure 1). On the other hand, after considering the linguistic aspects, the analysis goes on taking into account the knowledge about domain actions (in order to relate the speaker's utterance to domain goals). So, the third parameter of actions referring to this phase of the analysis refers to an instance of a domain action involved by the speaker's utterance. The domain action is recognized from the semantic representation by a plan recognition phase (*action identification* (Carberry, 1990), shown in the figures as **act-id**).
- **preconditions:** they represent the presuppositions associated with actions (see Searle's *felicity conditions* (Searle, 1969)). For example, **obtain-info** (the action of asking information) has the precondition that the speaker does not know the requested information.

[1] Although they presented the notion of *face* as a linguistic universal, many linguists think that it is mainly suited for describing the behavior of western societies.

[2] They also explain which forms people use to anoint the positive face of their interlocutors, but we will not deal with this aspect of communication here.

[3] The knowledge about domain actions is represented in a similar way and stored in the domain level of the plan library. Of course, speech acts refer to specific predicates concerning the knowledge and beliefs of the interactants.

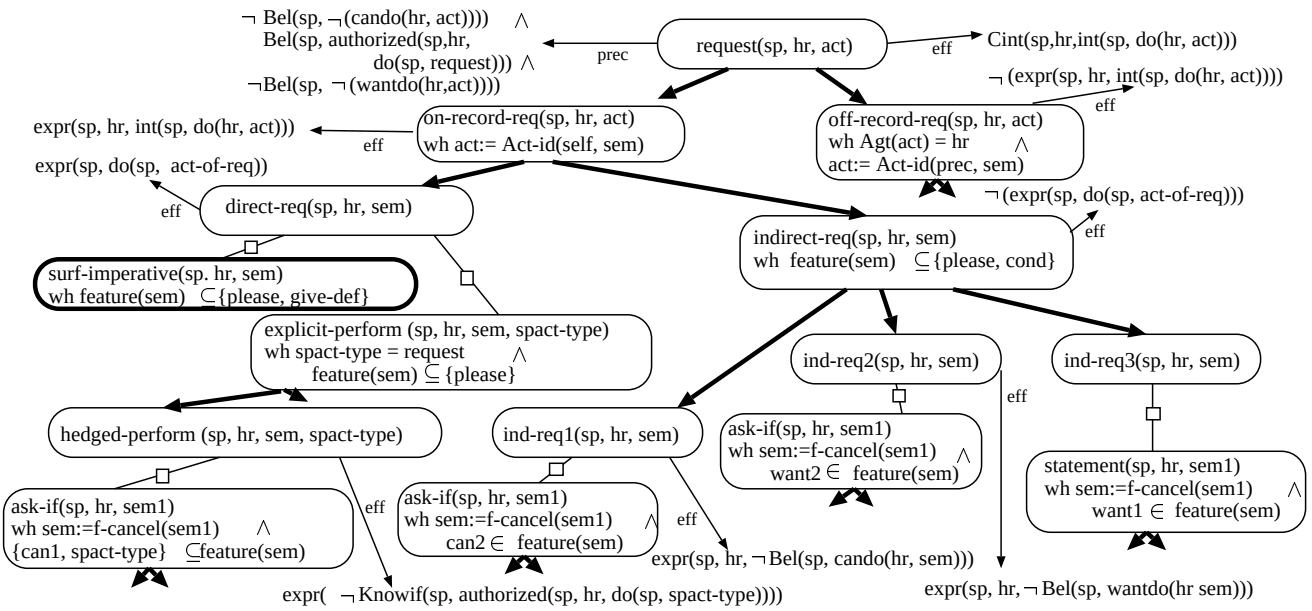

Figure 1: A portion of the Speech-act Library

- **restrictions:** they are included in the **wh** property of actions and, as for parameters, their meaning varies in the different actions of the speech acts library. In the actions related to the analysis of the linguistic aspects of utterances, they concern the linguistic features present in their propositional content. These features are called by Searle (1969) *illocutionary force indicating devices* and allow the hearer to identify the kind of speech act. They are, for example, the form of the sentence (declarative, interrogative, imperative), the tense and mood of verbs, the presence of modal verbs (*can, want, ...*) and performative verbs (*say, ask, order, ...*), or particles like *please, clearly,* etc. An example of this kind of restriction is **can2 ∈ feature(sem)** in **ask-if** in Figure 1, which restricts the main verb of the sentence to be the second person of the modal *potere* ('can'). In some actions referring to the domain actions involved by the input utterances, the restrictions may link the parameters of the actions in the speech acts library with the identified domain actions, or with their parameters. For example, in **off-record-req** in Figure 1, a restriction forces the agent of the identified domain action to be the hearer; this restriction is important in the definition of **off-record-req** because, when the restriction is not respected, a different speech-act is being performed (e.g. if the agent coincides with the speaker, we have an act of stating her/his plan).
- **communicative effects:** the actions of the library produce two types of effects: the first one consists in the communicative intentions of the speaker (e.g. when a request is performed successfully, then the speaker and the hearer share the belief that the speaker intends the hearer to perform an action and intends her/his intention to be a mutual knowledge)[4]. The second type of effects is related with the politeness consequences of the use of direct/indirect expressions in communication:[5] for example, the effect of the indirect request **ind-req1** is to express that the speaker doesn't want to presuppose any hearer's capability in performing the requested action, so that the negative face of the speaker is not threatened.

The recursiveness of natural language implies that illocutionary force indicating devices can be nested inside each other; so, complex utterances including different speech acts can be built and interpreted in a compositional way. For example, the sentence:

2) *Vorrei chiederti se puoi dirmi dove si trova la biblioteca.*

[I would like to ask you whether you can tell me where is the library.]

is composed of an external surface statement with conditional mood (*vorrei*, "I would like"), an explicit performative (*chiedere*, "to ask") and an indirect request expressed by an inner yes/no question (*se ...*, "whether ..."). [6] Because of the freedom in the composition of sentences, the speech acts library contains some cyclic paths (see the **ask-if** action that, in figure 2, occurs in its own definition).

[4] The **Cint** predicate is defined (Airenti *etal*, 1993) as: Cint(sp, hr, p)≡Int(sp, MB(hr, sp, p∧Cint(sp, hr, p)))

[5] Politeness effects are associated with the predicate **expr**. In this way, we model the conventionality of politeness expressions while preserving the base formalism. Basically, the **expr** predicate states which of the various presupposition has been (conventionally) negated to preserve the face of the hearer.

[6] Also the inner indirect request is composed of nested levels: see the use of *puoi*, "you can".

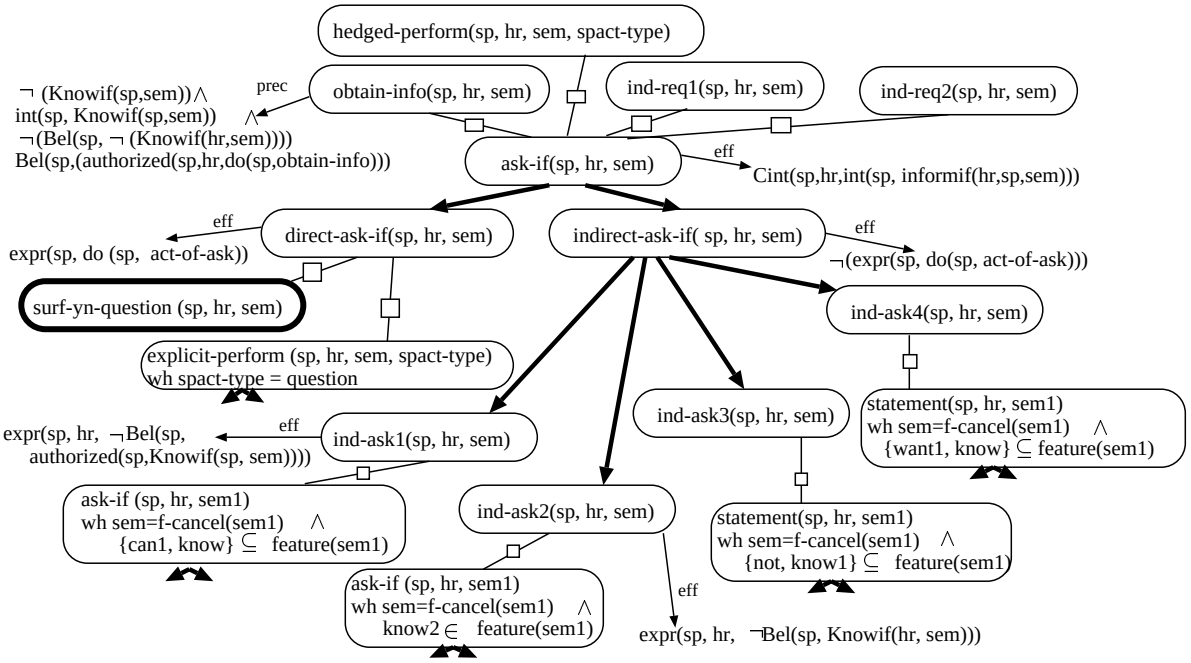

Figure 2: Representation of the **ask-if** speech-act

The speech act recognition process

Communicative actions should be interpreted at three levels: the *phatic* level, referring to the understanding of the single words uttered by the speaker, the *locutionary* level, referring to the comprehension of the meaning of the utterances, and the *illocutionary* level, referring to the interpretation of the sentences as speech acts. While we are not concerned with the phatic level, in our framework the locutionary and illocutionary levels correspond to different phases of analysis of the input sentences. In particular, a NL interpreter (Ardissono *etal*, 1991; DiEugenio & Lesmo, 1987; Lesmo & Terenziani, 1988) carries on the syntactic-semantic analysis and produces the semantic representation (in the formalism of semantic nets); then, the identification of the speech act is performed (this is the main topic of our paper). Finally, the domain-related processing connects the sentence to the previous ones in a single picture of the overall domain plans and goals of the speaker (see Figure 3). These plans are represented by means of hierarchical structures based on the domain level of the plan library and are obtained by applying heuristic rules for action identification and focusing; these rules keep into account contextual information for building coherent hypotheses on the speaker's goals and plans (Ardissono *etal*, 1993; Carberry, 1988).

The input to the second phase (see Figure 3) is a semantic representation of the input (with the contextual - e.g. anaphorical - references already solved) and its output is the recognized speech act, i.e. one of the roots of the hierarchies depicted in the figures. As a side effect of this second step, all "politeness indicators" have been identified, so that just the "pure" propositional content of the input sentence is passed to the third step. Concurrently, a degree of politeness has been evaluated. The goal of this section is to describe how the second step extracts the politeness indicators; nothing will be said about the evaluation of the politeness degree, which is currently obtained via some simple and not yet well developed heuristic rules.

The basic claim is that the whole process is governed by standard plan management procedures: the same procedures used in the third step for the well known domain-dependent analysis of the user's plans and goals.

First, the semantic representation undergoes an **action-identification** phase. Since the interpreter is playing at the *locutionary* level, this phase does not return the main action (as expressed by the main verb) involved in the input, but the surface speech act type (e.g. *surface-yn-question*). This seems reasonable, since, at this level, the term 'act' must refer to *locutionary* acts. The surface type is used as an entry point in the hierarchy, since it must match one of the leaves. Then, starting from the leaf found, an *upward-expansion* procedure is applied. Again, this procedure is the same used within the *focusing* phase of domain-level analysis (Ardissono *etal*, 1993; Carberry, 1990). *Upward-expansion* climbs up the hierarchy along all possible paths (and this can lead to ambiguities).

The key point is the treatment of the **wh** conditions appearing in the nodes of the hierarchy. Most of them refer to standard tests, but there are two types that deserve attention. The first of them is the check of **feature(sem)**; these tests are encoded in a very compact way in the figures; what actually happens is that each of them asks for the inspection of the top-most current

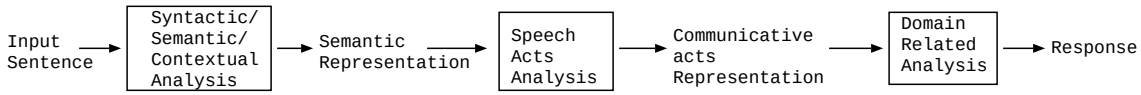

Figure 3: Schema of the interpretation process

node of the semantic representation; if the features mentioned in the test are found, then the node is discarded (**f-cancel**) and the 'main' substructure remains as **sem** (e.g., with modal verbs, the main substructure is the one referring to the 'object' of the proposition; for example, given a sentence like "May I ask you to ..." and its semantic representation "May(User, ask(User, System, ...))"[7] **sem1**, after a **can1** test on the formula, the remaining part is "ask(User, System, ...)", that corresponds to "User asks system to ..." **sem**). So, when the hierarchy is climbed up, the politeness markers disappear and, when one of the roots is reached, what remains is the propositional nucleus of the input sentence. The complete process could require that the root is reached more than once. In fact the process stops only when a root has been reached and no further climbing up is possible. But for nested levels of indirectness, the root can be used as a new entry point in the hierarchy (see the bottom **ask-if** node in the figures). Actually, the process can also fail in case a non-root node has no parent for which the **wh** conditions are met. Hopefully, in this case other alternative paths remain open.

Note however that, given a certain speech act, it is possible to identify more than one primary illocutionary act; so, the upward activation of the actions in the speech acts hierarchy may generate alternative hypotheses. For example, sentence **1b** can be interpreted as a request to have the keys (indirect interpretation) or as an attempt to obtain some information about the capabilities of the hearer. The two interpretations correspond, respectively, to the activation, while moving upward on the speech act hierarchy, of the **request** and **obtain-info** actions.

The second special test concerns the **act-id** predicate (see, for instance, the **on-record-req** node in Fig 1). This prepares the work for the third step (domain-level analysis). As stated above, the output of the speech-act analysis is the recognized speech-act. However, some speech-acts refer to an actual domain action; for instance, a request expresses the intention that the hearer does something, and that something is a domain action that must be encoded within the request (note that this is not the case for **obtain-info**). The speech-act hierarchy specifies this "type coercion" among levels: a **surface imperative** has as argument a semantic representation, while a **request** has, as argument, the corresponding domain action. Procedurally, this means that the usual *action-identification* procedure is executed, so that its role in the overall processing is made explicit in the hierarchy.

[7] For the sake of simplicity, the semantic representation of the sentence has been given in a logical form, instead of as a semantic net.

Example

Given the sentence:

3) *Posso chiederti di darmi le chiavi della biblioteca?*

[May I ask you to give me the keys of the library?][8]

The **surf-yn-question** action is activated on the basis of the interrogative form; the third parameter of the action is instantiated with its propositional content, that refers to the node of the semantic representation (**sem1**), associated to *potere* ('may'). The instantiated surface speech act is: **surf-yn-question (User, System, Sem1)**[9]

After the identification of this speech act, the analysis proceeds with the activation of the speech acts of which it is a substep or a specialization (upward expansion in the speech acts library (Carberry, 1988)): the **direct-ask-if** and then the **ask-if** actions are activated. Note that **surface-yn-question** could be considered as a direct substep of **obtain-info** (in a 'generation' relation). However, the net specifies that a **surface-yn-question** generates an **ask-if**, which in turn generates **obtain-info**. In this way, we are able to factorize an effect (the **Cint** effect of **ask-if**) that is shared by **obtain-info** and the other actions that are generated by **ask-if** (e.g. **ind-req1** or **hedged-perform** as shown in Figure 1 and 2); on the contrary, the peculiarity of the **obtain-info** (i.e. the precondition of not knowing the answer) is kept separate (in fact, in indirect acts performed by means of a question, the speaker almost always knows the answer to the question). Moreover, this effect is inherited both by **indirect-ask-if** and **direct-ask-if** through the specialization hierarchy.

When an action is in the decomposition of more than one speech act, more than one alternative hypothesis can be built (in the example, for the sake of simplicity, we only consider **obtain-info, ind-req1** and **hedged-perform**). However, the domain-level processing rejects the **obtain-info** since here, as usual, it does not make sense that the speaker questions the hearer about her/his own capabilities; **ind-req1** can not be instantiated because the node associated with *potere* ('may') should have the hearer as semantic agent, while in the example the agent is the speaker (compare with sentence **1b**)[10]. So, only **hedged-perform** is activated, because all its restrictions are satisfied. Since only one higher-level action has been instantiated, no ambiguity arises in the

[8] In Italian both verbs 'may' and 'can' correspond to the modal *potere*.

[9] In the actual implementation, the constants **User**, **System** and **Keys** (below) are nodes in the contextual representation standing for the reference to the associated individual.

[10] The analysis of the semantic and syntactic features is performed by the analyzer which, in this example, identifies the verb *potere* ('may') and the performative *chiedere* ('ask').

interpretation of the user's utterance and the upward expansion goes on, extending the unique hypothesis. The **on-record-req** and **request** actions are activated, so interpreting the sentence as a request by the user to perform the domain action: **give (System, Keys, User)**, that is identified by means of an action identification phase. Here, this phase is carried out easily, because the request is posed explicitly and the identified action coincides with the one expressed by the user. The situation is very different for the so called *off-record* requests (Brown & Levinson, 1987), where the speaker doesn't express in an explicit way the requested action, but s/he states one of her/his goals or s/he asks whether some precondition of the action is satisfied: e.g. "I would like to open the library" or "Do you have the keys of the library, please?". In this cases, the requested actions must be inferred from the utterance using the knowledge about domain actions (the task is performed by the domain plan recognition process).

Conclusions

The paper has presented an approach for coping with indirect speech acts in an interpreter of natural language. A plan-based representation of speech acts has been adopted. A major advantage consists in the strict integration of the processes of recognition of speech acts and domain plans. In fact, the same representation underlies both processes; however, the speech act analysis is affected also by the presence of some linguistic (syntactic and semantic) features which have been discussed in the paper. These features are related to the politeness of the request. The next step of our work will be an assessment of the evaluation of the politeness level on the basis of the features detected in the sentences and of the intended impact of this level on the receiver of the message.

The speech-acts recognition algorithm is embedded in a plan-recognition system for information-seeking dialogues in a subset of the University domain. The system is implemented in Common Lisp and runs on workstations. The speech-acts analysis exploits the basic action-identification and upward-expansion procedures written for the recognition of the domain plans of the user of the system.

Acknowledgments

This work was partially supported by MURST 60% and by National Research Council (CNR), project "Pianificazione Automatica".

References

Airenti, G. & Bara, B. & Colombetti, M. (1993). Conversational and behavior games in the pragmatics of discourse. *Cognitive Science*, 17, 197–256.

Ardissono, L. & Lesmo, L. & Pogliano, P. & Terenziani, P. (1991). Representation of determiners in natural language. In *Proc. 12th IJCAI* (pp. 997–1002). Sydney, Australia.

Ardissono, L. & Lesmo, L. & Sestero, D. (1994). Updating the user model on the basis of the recognition of the user's plans. In *Proc. Fourth International Conference on User Modeling* (pp. 5–10). Hyannis, Ma.

Ardissono, L. & Lombardo, A. & Sestero, D. (1993). A flexible approach to cooperative response generation in information-seeking dialogues. In *Proc. 31st Annual Meeting ACL* (pp. 274–276). Columbus, Ohio.

Ardissono, L. & Sestero, D. (1995). Using dynamic user models in the recognition of the plans of the user. *User Modeling and User-Adapted Interaction*, to appear.

Austin, J.L. (1962). *How to Do Things with Words.* Harvard University Press.

Brown, P. & Levinson, S.C. (1987). *Politeness: some universals on language usage.* Cambridge University Press.

Carberry, S. (1988). Modeling the user's plans and goals. *Computational Linguistics*, 2 (pp. 23–37).

Carberry, S. (1990). *Plan Recognition in Natural Language Dialogue.* ACL-MIT Press.

Di Eugenio, B. & Lesmo, L. (1987). Representation and interpretation of determiners in natural language. In *Proc. 10th IJCAI* (pp. 648–653). Milano, Italy.

Hinkelman, E.A. & Allen, J.F. (1989). Two constraints on speech act ambiguity. In *Proc. 27th Annual Meeting of ACL* (pp. 212–219). Vancouver, Canada.

Kasper, G. (1990). Linguistic politeness: Current research issues. *Journal of Pragmatics*, 14 (pp. 187–218).

Kautz, H. (1990). A circumscriptive theory of plan recognition. In P.R. Cohen, J. Morgan, and M.E. Pollack (Eds.), *Intentions in communication* (pp. 105–133). MIT Press.

Lambert, L. (1993). Recognizing Complex Discourse Acts: A Tripartite Plan-Based Model of Dialogue. Doctoral dissertation. University of Delaware.

Leech, G. (1983). *Principles of Pragmatics.* Longman, Cambridge.

Lesmo, L. & Terenziani, P. (1988). Interpretation of noun-phrases in intensional contexts. In *Proc. Coling Budapest* (pp. 378–383), Budapest, Ungheria.

Litman, D. & Allen, J. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11 (pp. 163–200).

Pollack, M.E. (1990). Plans as complex mental attitudes. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication* (pp. 77–103). MIT Press.

Searle, J.R. (1969). *Speech Acts: an Essay in the Philosophy of Language.* Cambridge University Press, Cambridge.