Using Structural Constraints for Speech Act Interpretation

James F. Allen & Elizabeth Hinkelman[1]
Department of Computer Science
University of Rochester
Rochester, NY 14627

ABSTRACT

We present a speech act interpretation system that has the generality of previous plan-based approaches but also can distinguish subtleties of phrasing. As a result, a sentence such as *Can you pass the salt* can be recognized as a request to pass the salt, whereas *Tell me whether you are able to pass the salt* is a question about the hearer's abilities. The system also provides a framework for integrating spoken language information, namely intonation and prosody, into the speech act interpretation process. The resulting system is more accurate than previous approaches and is considerably more efficient in dealing with everyday conventional sentences.

1. INTRODUCTION

A fully functional natural language understanding system will require much more sophisticated interaction between its structural processing (e.g. parsing, semantic interpretation) and its general knowledge and reasoning abilities. In this paper we explore one aspect of this problem, namely the interpretation of intended speech acts. This is an excellent problem to study to explore this problem since the interpretation of the intended act depends strongly on the syntactic and semantic structure of the utterance, but also is highly influenced by context, and hence general reasoning. Neither the structural constraints nor the reasoning about the situation fully determines the interpretation alone - rather each adds partial information to the overall solution. As a result, a system organization where the parser and semantic interpreter produce single specific interpretations that are accepted or rejected by the contextual processing will be highly inefficient.

Systems that perform speech act interpretation to date have attempted to avoid this problem by assuming that the relationship between form and meaning is a simple one. Specifically, almost all work on speech act theory makes what is called the **literal meaning hypothesis** (e.g. Searle 1969, 1975a, Allen & Perrault, 1980, Litman & Allen, 1987). This is the assumption that the syntactic and semantic structure of the sentence uniquely defines some literal meaning independent of the context of the utterance. All indirect meaning can then be derived by reasoning from the literal interpretation. There are many reasons why this assumption is made, but the principle one is to avoid the general issue of how structural processing and general reasoning interact. If there is a context-independent literal meaning, then the structural processing need only produce this interpretation as its output, and it is then used as the starting point for general reasoning. Thus it allows a full separation of the two parts of the interpretation problem.

[1] Elizabeth Hinkelman's current address is Center for Information and Language Studies, 1100 East 57th St., University of Chicago, Chicago IL 60637 (eliz@tira.uchicago.edu). James Allen's email address is james@cs.rochester.edu.

In this paper, I will argue that the literal meaning hypothesis cannot be supported. Primarily, this is because it does not allow for any of the subtleties of phrasing that affect the interpretation of the utterance. In addition, it seems to suggest a model that is contrary to intuition about language learning and processing. These problems result in systems that are insensitive to the syntactic form of the sentence. In addition, systems based on these models could not take advantage of the recognizability of common indirect forms since each sentence was analyzed from first principles. Many people, including the authors, have suggested that certain common forms could be compiled for efficiency reasons. These compiled interpretations, however, still operate from the literal meaning as a starting point and thus are insensitive to syntactic and semantic variation.

If there is no literal meaning, what replaces its role as the connection between structure and inference? We suggest that the result of structural processing is a specification of the space of possible interpretations. This may include specific limits on what the intention could be (e.g. the sentence in question must be a request of some kind), as well as specific interpretations that are favored because of their familiarity and conventionality. This information can then be used to heuristically control the general plan reasoning so that common forms are recognized quickly, and to constrain the possible interpretations that can be suggested from first principles.

The resulting framework has significant practical advantages:

- the common indirect forms are recognized without costly plan reasoning required by earlier models;
- general plan reasoning is constrained in what interpretations it can derive, and thus it more accurate;
- the framework allows for the integration of intonational and prosodic cues in spoken language.

The next section provides some background on speech act theory and the literal meaning hypothesis. Following that, the evidence against the literal meaning hypothesis is discussed in Section 3. We then present our new model in the next two sections: Section 4 outlines the structural interpretation rules that define the interpretation space, and Section 5 discusses the role of plan reasoning in the new framework. Section 6 discusses the implementation briefly, and Section 7 discusses some possible extensions.

2. SPEECH ACT THEORY AND THE LITERAL MEANING HYPOTHESIS

Speech act theory concerns how sentences are used in language. For instance, depending on the context, the sentence *Do you know the time?* may be a genuine yes/no question, a request for the time, an offer to tell someone the time, or even a reminder that it is late. Each one of these different uses can be viewed as a different action performed by saying the utterance in the right context. These actions are called speech acts, and have been the subject of much study in linguistics, philosophy and computer science since they were formulated by Austin (1963).

By formulating speech acts within a theory of action, many difficult problems were solved. In particular, as actions, it makes sense to speak of utterances as being successful or not, depending on whether they achieve the appropriate intended effect. Searle (1975b) produced a fairly comprehensive classification of the different general classes of speech acts based on the intended effects. Since this will be relevant later, let us consider this classification in more detail. Searle classified speech acts into five general classes:

the Representatives - acts that involve a statement about the world, and hence can be judged as true or false (e.g. inform, tell, deny)

the Directives - acts that involve influencing another agent's intentions or behavior (e.g. request, beg, suggest, command)

the Commissives - acts that involve committing the speaker to some intention or behavior (e.g. promise)

the Expressives - acts that involve expression of the speaker's attitude toward some state of affairs (e.g. apologize)

the Declaratives - acts that explicitly involve language as part of their execution (e.g. quit, fire, marry, call out in baseball)

Searle argues convincingly that speech acts can only be defined in terms of the intentions of the speaker, including the intentions that the speaker intends to be recognized from the utterance they make. By adapting work by Grice (1957), he produces quite precise definitions of the general classes of acts, and of particular acts themselves.

The part of the theory that is not clear, but is crucial for building a computational system to recognize speech acts, is how the words and structure of the sentence identify the appropriate act. Searle (1969) suggests that there are certain structures (illocutionary force indicating devices) that serve to identify the intended act. The declarative class of acts gives the best examples: a sentence involving an explicit performative verb indicates the speech act. For example, *I hereby quit* indicates the quitting act quite directly. It is hard to extend this technique to the other classes, however. For instance, what features indicate a request? Clearly, imperative mood sentences seem like a good candidate for requests. But some imperative mood sentences are not requests, while many other forms are requests. For example, in giving instructions the imperative is used extensively (e.g. *First open the lid and take out the hamster*). While instructions may still be directives, they certainly need not be requests. On the other hand, requests can be made by interrogative sentences (*Can you tell me the time?*), representatives (*I want you to open the door*) and sentence fragments (*The door, please*). So, hopes of finding syntactic indicators of requests seem remote.

In response to this problem, Searle (1975) used a version of the literal meaning hypothesis. Utterances are assumed to have a literal speech act interpretation that is derivable from the structural properties of the sentence. If this literal act is inappropriate in the context, the general reasoning processes are used to derive the indirect interpretation. Since speech acts are defined in terms of the recognized intentions of the speaker, if the hearer believes that the speaker intended he or she to perform this reasoning to the indirect act, then the utterance counts as an instance of the indirect act.

This is a promising approach, and was the starting point for the computational model of indirect speech act recognition developed by Allen and Perrault (Perrault & Allen, 1980, Allen, 1983). This system was organized as shown in Figure 1. The parser produced a literal speech act interpretation based on the syntactic mood of the sentence. Imperative mood sentences (*Close the door*) became surface requests, declarative mood sentences (*I want* *you to close the door*) became surface informs, interrogative mood sentences (*Can you close the door?*) became requests to inform. The surface act was the input to the plan inference component which used heuristic plan recognition techniques to derive the speakers plan, identifying the intended indirect act along the way. By carefully keeping track of whether the information used to recognize the plan was based on shared knowledge between the speaker and hearer, or on private knowledge of the hearer, the system was able to distinguish indirect speech acts (where the plan was intended to be recognized) from the simple recognition of the speakers goals (which may or may not have been intended to be recognized).

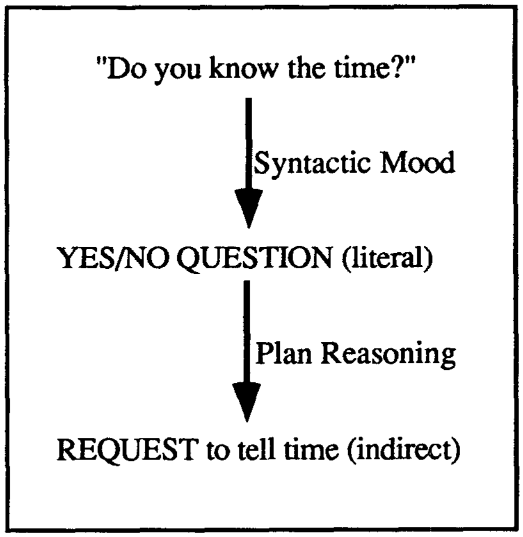

Figure 1: The analysis using the literal meaning hypothesis

This technique has been used almost universally in speech act recognition systems since. A common optimization of this approach that anticipates part of the technique we describe here. Rather than derive all indirect interpretations first principles, common indirect forms are compiled into the act definitions. For instance, the request act might, in addition to its definition from first principles, have a set of decompositions outlining common indirection forms. For example, one of the decompositions of the act REQUEST(Speaker, Hearer, Action) might be:

SURFACE-REQUEST(Speaker, Hearer,
INFORMIF(Hearer, Speaker,
CANDO(Hearer, Action))).

Intuitively, this allows yes/no questions such as *Can you open the door?* to be recognized in one plan recognition step as a request to open the door. This technique has been used extensively (e.g. Sidner & Isreal, 1981, Carberry, 1983, Litman & Allen, 1987, Kautz, 1987).

3. PROBLEMS WITH THE LITERAL MEANING HYPOTHESIS

While the techniques described above have allowed a leap in the sophistication of man-machine dialog systems, the defects arising from the literal meaning hypothesis become more and more pronounced as these methods are used in more general domains. In particular, within a highly-limited domain, there may be only one reasonable interpretation for wide classes of utterances. Subtleties that arise from phrasing do not play a role simply because the limitations of the domain would not allow different interpretations. In a more general system, however, that can handle a domain of useful complexity, these issues become crucial. In this section, we discuss why the problem is so important.

There are two main objections to the literal meaning hypothesis. The principle arguments rest on whether such a literal interpretation can be derived that captures the subtleties of phrasing and information from other sources such as intonation and prosody. The other main objection relates to processing concerns. We will consider both below.

The first problem is that syntactic mood is a very poor indicator of intent. In particular, many other factors are equally good or stronger indicators of the speaker's intention. Intonation, for instance, is an important factor that can override almost anything else. The simple declarative mood sentence *You are going to the party* may be a simple assertion, or with the appropriate intonation a command, or with another intonation a question. The latter case can be indicated in written text using punctuation, viz *You are going to the party?*. Thus it seems that the simple syntactic fact that the subject precedes the verb and its auxiliary verbs indicates little. What is the literal meaning of this sentence independent of context? The declarative mood indicates a surface inform, yet if a question mark completes the sentence, it's a surface question. If the sentence is spoken, the literal interpretation may be indicated solely by intonation. What features should we use to identify the literal form?

Furthermore, certain adverbials might be present that appear to modify (or clarify) the intent. For example, the word *please* seems to indicate the directive class. Thus, *I need the door opened, please* is unambiguously a request, even in contexts where the statement of need would be reasonable. Similarly, the sentence *I quit!* may be a simple inform (as in a reply to the question, *What did you do today?*), or may actually be the quitting act. Which is the literal form? How can we account for the fact that *I hereby quit!* is then unambiguously the quitting act?

These complications are far beyond the abilities of previous analysis techniques. In fact, a system using the literal meaning hypothesis is unable to account for why the sentence *Can you lift that rock?* may very well be a request to lift the rock, whereas a paraphrase of it, *Are you able to lift the rock?* is much less likely to be a request, and *Tell me whether you are able to lift the rock* is definitely not a request! All the systems using the literal meaning hypothesis have not been able to distinguish between these sentences, as they all identify the same surface speech act with the same propositional content. At best, these systems may make it easier for the *Can you* form to be recognized as an indirect act, but they have no way of preventing the other sentences from receiving the same end interpretation.

This problem can be shown in a slightly different way by comparing indirect forms across different languages. If there was a literal meaning, and all indirect forms were computed using general reasoning from that interpretation, then indirect forms should be more or less uniform across languages. But this is not the case. Searle (1975) reports that the literal translation of *Can you open the door?* in Czech is not readily usable as a request. On the other hand, the literal translation of *You want to hang up my coat* in Hebrew is a common indirect form, but a flawed request in English at best. Furthermore, there are differences among dialects within a language in what forms admit indirection.

All this points to a highly conventionalized set of indicators of intention that is at least as important as the general reasoning about the world and plan recognition. Lest we go to far in the other direction, however, note that there remains a wide variety of utterances that can only be analysed in terms of general reasoning. The sentence *Do you know the time?*, for instance, has a wide variety of meanings that differ only by the context. As a more complex example, the phrase *It's cold in here* may be understood in the right context as a request to close the car window even if it has never been heard before in that setting. Language abounds with novel interpretations of utterances that only become relevant in specific circumstances. Thus the problem is how to allow general reasoning to derive an arbitrary number of indirect interpretations in some cases, and how to restrict the interpretations derivable in other cases. The literal meaning hypothesis does not allow us the flexibility to have both.

Finally, anecdotal evidence based on language learning seems contrary to the literal meaning hypothesis. In particular, the first interpretation learned by children of the sentence *Can you pass the salt?* is as a request to pass the salt. The so-called literal reading, in which it is interpreted as a yes/no question, may be learned much later. In this case, it would seem that if there were a literal interpretation, it should be a request, and the yes/no question interpretation should be derived from that!

Given this situation, it seems much more satisfying to view such sentences as ambiguous between the two interpretations, and have that ambiguity resolved in context. As we shall see, the approach outlined below will do exactly that.

4. A NEW APPROACH

While it appears that the literal meaning hypothesis in its pure form must be abandoned, we want to retain the advantages that suggested the approach to begin with. In particular, the literal meaning hypothesis allowed the structural linguistic processing to produce a context independent meaning that could be used as input to the general reasoning processes. This separation had the obvious benefits common to any modular decomposition and should be retained if at all possible.

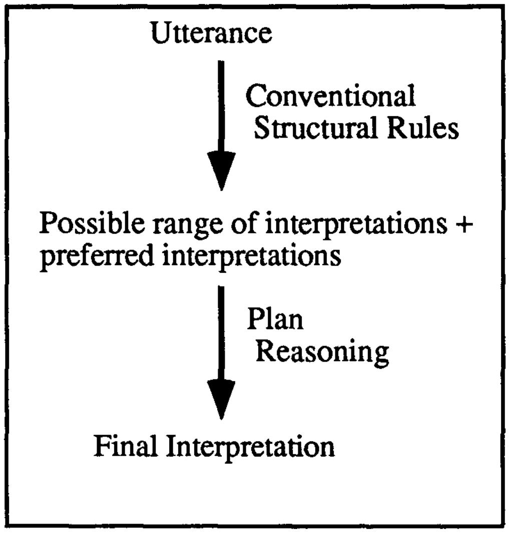
Utterance
Conventional Structural Rules
Possible range of interpretations + preferred interpretations
Plan Reasoning
Final Interpretation

Figure 2: The Analysis using an interpretation space

Rather than produce a literal interpretation, the structural processing produces a description of the space of possible speech act interpretations. This space will be characterized in two ways: the first part specifies the entire range of possible interpretations allowed by the sentence in any context; the second part defines a set of preferred readings suggested by various structural properties of the sentence. In many cases, the space will be quite unconstrained and any speech act interpretation might be possible, but in other cases we will see that there are quite strong structural constraints on the space. The idea is that the general reasoner will first try to select an interpretation from the set of preferred readings, and only if that is not successful will general reasoning be used to derive a speech act interpretation within the space of allowable interpretations. The new architecture, shown in Figure 2, looks much the same as the original architecture except for this change.

In order to specify the space of possible interpretations, we need a language that allows us to concisely specify entire classes of speech acts. Our initial representation to test these ideas is a representation that heavily depends on the notion of action abstraction. Such representations are common in knowledge representations using inheritance hierarchies and the use of such representations in planning systems has been explored in detail by Tenenberg(1988). We will assume his model here. As a start, the initial hierarchy of speech acts we use will simply encode the five major classes of speech acts defined by Searle. It is not essential to this approach that the hierarchy be a tree-structure, it could just as well be a lattice structure. We have not yet found situations that force this and currently the system is restricted to tree structures. The abstraction hierarchy, with a few sample concrete speech acts, is shown in Figure 3. Details on the actual definitions of these acts will be presented in the next section. For now it suffices that the hierarchy provides us with a language to specify entire classes of acts.

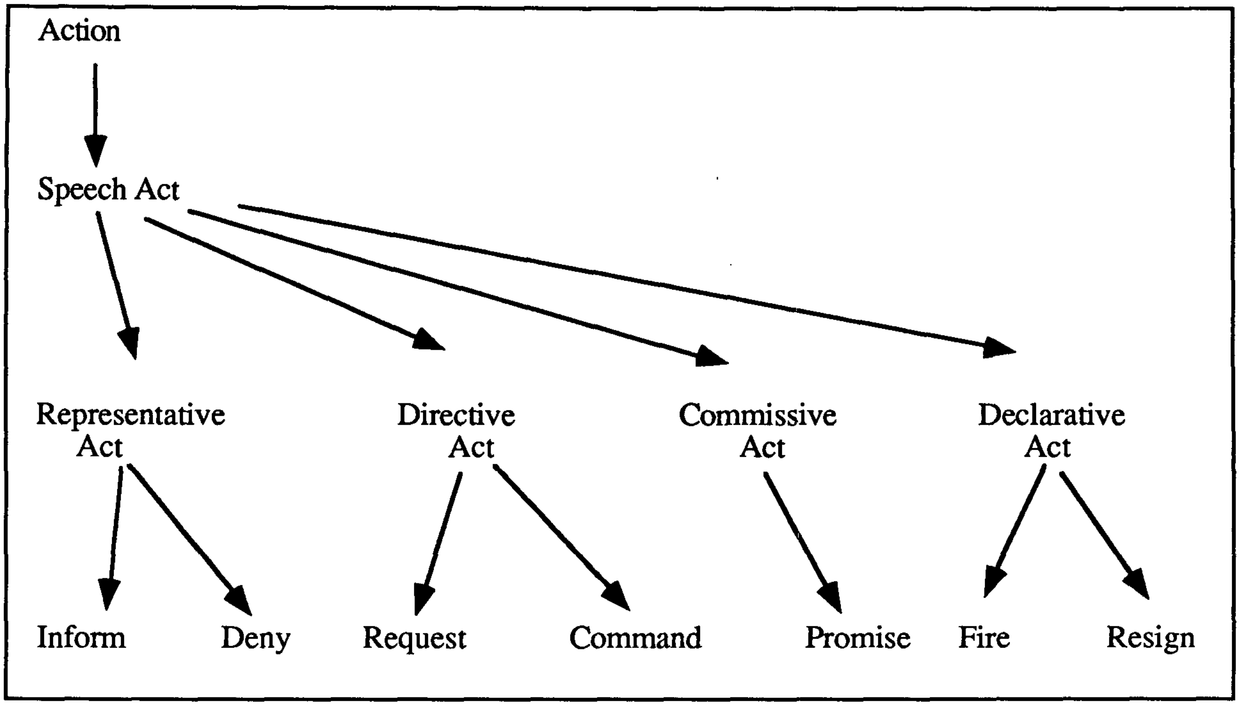

Figure 3: A Speech Act Hierarchy

The language to specify the space of interpretations allows single abstract or concrete speech act descriptions, and complex expressions involving a class of acts with excluded subtypes. For example, the class {REQUEST} would be the class of all requests, whereas the class {DIRECTIVE - COMMAND} would be the entire class of the directive acts with the exception of the commands. Of course, there may be parameter restrictions on the act classes as well - for instance the class of possible interpretations might be restricted so that the speaker and hearer of the speech acts are appropriate to the discourse situation. But let us ignore these complicating factors for the moment and delay the details to the later section on implementation. The other part of the specification of the interpretation space is simply a list of suggested interpretations that are conventionally signalled by the sentence. The importance of these interpretations is that they are considered first before any general reasoning is done to infer new interpretations within the space.

Conventional indicators of the speech acts are captured by pattern-specification rules where the pattern can require any aspect of the syntactic/semantic structure produced by the parser and semantic interpreter, and the specification gives the full range of acceptable interpretations and suggested specific interpretations for any structure matching the pattern. For example, the rule dealing with interrogative mood might be as follows:

If MOOD is Yes-No-Q, then
possible interpretations: any SPEECHACT
suggested interpretations: a yes-no question.

On the other hand, the rule for conventional *can you do X?* sentences would capture the following:

If MOOD is interrogative, SUBJECT is *you*, VOICE is Active, and the AUX is *can*, then
possible interpretations: any SPEECHACT
suggested interpretations: REQUEST to do the act corresponding to the verb phrase.

Both these rules make suggested interpretations without limiting the range of possible interpretations. A rule that restricts the space rather than suggesting a specific act is the rule for the adverbial please:

IF ADV is *please* then
possible interpretations: a DIRECTIVE - COMMAND.

It is very important that these rules can be used incrementally. That is why each rule must specify the complete range of possible interpretations, even if it's essentially unrestrictive. That way, we can gather up the specifications from each of the rules that matched and combine them using an incremental unification algorithm. In particular, two interpretation spaces are combined by computing the cross product of the individual speech acts defining the space, where individual acts are unified using a tree unification algorithm. For example, the sentence *Can you lift that rock?* would have at least two rules defining its space of interpretations. The rule based on simple interrogative mood produces the space:

{ASK, SPEECHACT}

and the rule for *can you* sentences produces the space:

{REQUEST, SPEECHACT}.

The combination of these two spaces is the set consisting of:

ASK * REQUEST -> no interpretation
ASK * SPEECHACT -> ASK
SPEECHACT * REQUEST -> REQUEST
SPEECHACT * SPEECHACT -> SPEECHACT.

The resulting space is:

{ASK, REQUEST, SPEECHACT}.

In other words, as intuition suggests, this sentence is most likely a request to lift the rock, or a question about the hearer's abilities, but could be any other interpretation if these are not reasonable in context. The sentence *Can you lift that rock, please* has three rules defining the space - the two shown for the sentence above, and the other for the *please* rule, which defines the space {DIRECTIVE - COMMAND}. Combining this with the space defined above we get the new space:

ASK * DIRECTIVE-COMMAND -> no interpretation
REQUEST * DIRECTIVE-COMMAND -> REQUEST
SPEECHACT * DIRECTIVE-COMMAND -> DIRECTIVE - COMMAND.

Thus the space of interpretations for this sentence is:

{REQUEST, DIRECTIVE-COMMAND}

which indicates that it is probably a request, but if it is not, it must be some other form of directive (except a command).

Another interesting rule is the one for the word *hereby*. This construct is used with the declarative class of speech acts, as in *I hereby quit* or *I hereby christen this ship the 'Zebra Binge'.* The rule for *hereby* simply indicates that the act described by the performative verb is the act performed:

If ADV is *hereby*
then the space of interpretation = the semantic interpretation of the sentence.

Even without considering any further processing of the interpretation spaces, this framework shows great promise. For example, consider how we now can distinguish between the apparent paraphrases of *Can you lift that rock?*. Above, we saw that this sentence defined the interpretation space:

{REQUEST, ASK, SPEECHACT}.

The sentence *Are you able to lift that rock?*, on the other hand, defines the space {ASK, SPEECHACT}, i.e. it is probably a yes/no question, but if that is not reasonable, it could be any act, including a REQUEST. Thus we see a difference in the ease in which the request interpretation can be found. In one case the request is suggested directly as the meaning of the sentence, whereas in the other case the request interpretation would need to be derived. The sentence *Tell me whether you are able to lift that rock?* is even further restricted. The rule for imperative sentences would indicate a probable interpretation as a request that the hearer inform the speaker of his or her ability to lift the rock, i.e. a yes/no question. The space of interpretations allowed by the imperative mood is approximately the class of directives (although there may be some acts more in the expressive class such as *jump in the lake!*). Finally, the sentence *I hereby request that you tell me whether you are able to lift that rock?* is unambiguously a yes/no question. The hereby rule restricts the act to be the one described in the sentence.

So the new technique allows us to capture the intuitive distinctions between various phrasings quite nicely. It shows how certain forms may be directly identified in their so-called indirect reading, while others require plan reasoning to identify the indirect reading, and others explicitly prohibit the indirect reading altogether.

5. PLAN REASONING

The plan reasoner performs two major roles. Given a set of suggested interpretations, it evaluates each in context to see if it is reasonable. In cases where there are no reasonable interpretations, it uses the defective interpretations as a start to derive a reasonable interpretation. In order to examine the plan reasoner in more detail, the representation of speech acts as actions must be examined further.

We are using a representation of action that is fairly standard across planning systems (e.g Fikes & Nilsson, 1971) and speech act reasoning systems (e.g. Allen, 1983, Litman & Allen, 1987). Each action is represented by a set of formulae in the following classes:

constraints: conditions that must be true for the action to be well-defined;

preconditions: conditions that must be true for the action to be applicable;

effect: a specification of the change in the world caused by the action, typically divided into a set of formulae to add and another set to delete from the initial world description to produce the new world description;

body: a further specification of how the action is performed, either as a sequence of substeps, or as a set of additional goals to achieve.

The complete specification of an act will include those conditions specified directly for the act, plus any conditions defined for abstractions of the act and inherited. For example, it is a common precondition to all speech acts that the speaker and the hearer must be within communication range, and are attending each other. We capture these conditions by simply defining a precondition such as

ATTEND(Speaker,Hearer)

on the abstract action SPEECH-ACT. These conditions are then inherited by all specializations of this act, namely all speech acts. Given this, the full definitions for the request and inform acts, and the act of resigning are given below:

Request(Speaker, Hearer, Action)
constraints: Agent(Action)=Hearer
Want(Speaker, Effects(Action))
precondition: Attending(Speaker,Hearer) <inherited>
effect: Do(Hearer,Action)
body: Achieve Shared(Speaker, Hearer, Want(Speaker, Do(Hearer,Action)))

Inform(Speaker, Hearer, Proposition)
precondition: Attending(Speaker, Hearer)
Believe(Speaker, Proposition)
effect: Believe(Hearer, Proposition)
body: Achieve Shared(Speaker, Hearer, Want(Speaker, Believe(Hearer, Proposition)))

Resign(Speaker, Hearer)
constraints: Employer-of(Hearer, Speaker)
precondition: Attending(Speaker, Hearer) <inherited>
Want(Speaker, ~Employer-of(Hearer,Speaker)
effect: ~Employer-of(Hearer, Speaker)

To make a request, the action request must be an action by the hearer, and the speaker must want the effects of the action. For the request to be successful, the speaker and hearer must be attending each other (and speak the same language, etc). The intended effect is that the hearer wants to do the action. Finally, the way a request is achieved is by any action that makes the speaker's goal to get the hearer to do the action be part of the shared knowledge between the speaker and hearer. More sophisticated definitions are required in some situations (e.g. see Allen 1983, Allen & Perrault, 1980), but this definition is useful for most purposes. More complex definitions could be used in the current system with no additional problem. The definition of the inform act is similar except that it concerns changing the hearer's beliefs rather than his or her goals.

Consider the definition of the design act: You must be employed by the person you are talking to. Furthermore, the standard conditions that the speaker and hearer are attending to each other, etc, must obtain. The speaker must have the goal to resign, and the effect is that the speaker is no longer employed. We could also define a body for the act along the same lines as done with the request act - any action that causes the speaker's goal to resign into shared knowledge would count as a resigning act.

These action definitions are compatible with those used in speech act recognition systems such as Allen (1983) and Litman and Allen (1987), and so the plan reasoning techniques developed there are directly applicable here, if desired. Now that the input is an interpretation space rather than a single literal interpretation, however, new techniques can be used to optimize the plan reasoning, as well as to improve its ability to select the appropriate interpretation. In fact, in most cases, general plan reasoning is not required at all as the interpretation space already contains the correct interpretation. The plan reasoner simply needs to select the appropriate interpretation from the preferred readings.

So the first stage of plan reasoning is evaluation of the specific suggested interpretations. This is done by using the plan selection heuristics used in Allen (1983) to see how plausible each suggested interpretation is in context. In particular, if one of the constraints of an action is false, then the action is eliminated from consideration. If one of the preconditions is false, and not easily achievable in the setting, then the interpretation is marked as defective. Similarly, if the effects of the action are already true, then the interpretation is marked defective. For example, in a setting where both the speaker and hearer know that the hearer knows the time, the so-called literal interpretation of the question *Do you know the time?* is defective because the effect, that the speaker knows whether the hearer knows the time, is already true.

Of course, in many settings, it is unknown whether a certain precondition or effect is true of not. In these cases, the act remains as a valid interpretation. If the interpretation is eventually selected as the final interpretation, then the information determined to be unknown is added as an implicature of the sentence. For instance, the sentence *Open* *the door*, if accepted in context as a request, would then implicate that the speaker had the goal of getting the door open. Thus these conditions serve both as a filter on interpretations, and as the implicatures of the interpretation. As a filter, the conditions must not be false, as implicatures, they must be true. The initial set of conditions is as follows:

Filter Constraint	**Implicature if selected**
The constraints are not false	The constraints are true
The speaker doesn't believe the preconditions are false	The speaker believes the preconditions are true
The speaker doesn't believe the effects are true	The speaker believes the effects are false
The speaker doesn't not want the effects	The speaker wants the effects

Consider the sentence *I quit!* in different contexts. The structural mapping rules would suggest two interpretations for this sentence: an INFORM act (based on the declarative mood sentence), and the RESIGN act (based on the specific lexical rule for *quit*). In the case where the speaker is talking to a friend who is not the speaker's employer, the resign interpretation is eliminated because the constraint that the speaker works for the hearer is false. The inform interpretation remains valid and becomes the preferred interpretation. In the case where the speaker is talking to his or her boss, then the resign act is possible. The inform act is also possible unless it is clear in the situation that the boss knows that the speaker is going to resign. In this last case, the sentence could not be an inform since its effects would already hold before the act was performed. These simple checks can be effective in eliminating a wide range of interpretations that are possible given only the context-independent structure of the sentence. As can be seen, the so-called indirect interpretations can be recognized with the same ease as the so-called literal interpretations within this new framework. As a result, processing everyday conventional forms is highly optimized over the earlier systems that needed to perform plan inference to derive the indirect interpretations.

General plan reasoning is needed only in cases where either there is no specific interpretation that passes the filtering tests, or when more than one interpretation passes and the sentence appears ambiguous. Let us consider the latter case first. Consider the question *Can you lift that rock?* in a setting where it is not obvious whether the hearer is strong enough to perform the act. The two suggested interpretations are a yes/no question (from the interrogative mood) and a request to lift the rock (from the *can you* rule). The implicature checks cannot eliminate either of these readings in the given context. The plan reasoning system must be brought in to explore the further consequences of the two interpretations. A partial plan may be inferred from each and is checked in the context. For instance, if the domain is such that the speaker has a goal that could be furthered by having the rock moved, then the plan inferred from the request interpretation would match this goal well. On the other hand, if the speaker has the goal of evaluating the hearer's strength (say in order to decide whether to take the hearer on a trip), then the plan derived from the yes/no question might best fit. Plan recognition techniques developed in Allen (1983) and Kautz & Allen (1986) can be used directly to suggest such plans and evaluate how well they fit the hearer's known goals.

In the case where no interpretation remains, the defective interpretations are reconsidered as to what plans can be inferred. The techniques for dealing with indirect speech acts (Perrault & Allen, 1980) can be used directly to analyse these cases. There is an added constraint to this process, however. The final indirect interpretation must fall within the class of possible interpretations specified from the structural properties of the sentence. Thus, whereas the sentence *Tell me whether you can lift that rock?* could have been inferred to be a request to lift the rock by Perrault and Allen, this interpretation would not be allowed using our new scheme.

As can be seen above, existing work in speech act interpretation is still used directly in the new approach. The major difference is that the computationally expensive plan recognition techniques are only used as a last resort. Most of the time, the structural constraints and the filtering using the implicatures from the act definitions will serve to identify the appropriate intentions.

6. IMPLEMENTATION

The system is implemented with the RHET knowledge representation system (Allen & Miller, 1989) which provides most of the domain reasoning support. RHET is implemented in COMMONLISP and currently runs on Symbolics machines and TI explorers. It provides a hierarchy of frame-like objects for the action abstraction hierarchy, explicit hierarchical belief spaces for representing the different agent's beliefs as well as their shared beliefs, an explicit temporal reasoning based on interval logic, full equality reasoning between ground terms, and a range of programmer controllable reasoning modes. To give a feel for the representation, consider two example class definitions concerning the top of the speech act hierarchy:

```
(DEFINE-SUBTYPE  SPEECHACT  ACTION
        :ROLES ((R-SPEAKER T-HUMAN) (R-HEARER T-HUMAN))
        :PRECONDITIONS [Attend [f-speaker ?self] [f-hearer ?self]]
                          ([Want [f-speaker ?self] [Do [f-speaker ?self] ?self]]))
```

Paraphrasing, a speech act is a subclass of action with two defined roles, a speaker and hearer, (plus inherited roles such as the time of the act), and two preconditions defined for any instance of this class: the speaker and hearer must be attending each other, and the speaker must intend to perform the speech act.

```
(DEFINE-SUBTYPE  DIRECTIVE  SPEECHACT
        :ROLES ((R-ACT T-ACTION))
        :CONSTRAINTS ([EQ [f-agent [f-act ?self]] [f-hearer ?self]])
        :EFFECTS ([DO [f-hearer ?self] [f-act ?self]]))
```

Paraphrasing, a directive is a subclass of speech acts with an additional role, namely the requested act. Instances of directives must have the agent of the requested act be the hearer, and have the intended effect that the hearer do the requested act. The preconditions and roles from the speech act definition are inherited by the directive acts.

To help distinguish RHET objects from LISP objects, all RHET expressions use square brackets, with regular parentheses used to identify role values. Thus an instance of a directive act with speaker Jack, hearer John, and the action some lifting event L123 would be expressed as

```
[DIRECTIVE     (R-Speaker [JACK1])
               (R-Hearer [JOHN35])
               (R-Act [L123])]
```

In RHET, this same object can be defined incrementally by defining the instance and later specifying the role values by equality assertions, i.e.

```
(DEFINE-INSTANCE d1 DIRECTIVE)
```

```
(EQ [f-speaker d1] [JACK1])
(EQ [f-hearer d1] [JOHN35])
(EQ [f-act d1] [L123])
```

The resulting object would be exactly the same in either case. Note that given the constraints defined for the directive class, this instance only makes sense if the agent of L123 is John, i.e. [f-agent L123]=[JOHN35].

The structural interpretation rules and the plan reasoning system are both implemented in COMMONLISP within the RHET environment. The structural interpretations rules match against a combines syntactic structure and logical form as described in Allen, 1987. As an example, the representation of the sentence *Can you lift the rock?*, simplifying the representation of lexical items, is

```
(S      MOOD YES-NO-Q
        VOICE Active
        SUBJ      (NP    PRO you
                         SEM (PRO p1 PERSON "you")
                         REF Hearer)
        MAINV lift
        AUXS (can)
        TENSE Present
        OBJ       (NP    DET the
                         HEAD rock
                         SEM (DEF/SING r1 ROCK)
                         REF  Rock1235)
        SEM     (PRES c1 CAN
                         (AGENT (PRO p1 PERSON "you"))
                         (THEME (INF l1 LIFT
                                         (AGENT p1]
                                         (THEME (DEF/SING r1 ROCK)))))
        REF [ABLE-TO-DO          (R-AGENT       Hearer)
                                 (R-ACTION    [LIFT       (R-AGENT Hearer)
                                                          (R-THEME Rock1235)])))
```

The interpretation patterns closely resemble the semantic interpretation pattern-action rules described in Allen, 1987. Here is a slightly simplified *can you* rule presented informally earlier in the paper:

```
((S     AUXS can
        MOOD YES-NO-Q
        VOICE Active
        SUBJ (NP PRO you)
        MAINV +action)   ->          {  [REQUEST  (R-ACTION (V R-ACTION REF))],
                                        [SPEECHACT]}
```

This rule would match the above sentence representation and produce the interpretation space consisting of a request to lift the rock and the general class of all speech acts:

```
{[REQUEST  (R-ACT [LIFT          (R-AGENT Hearer)
                                 (R-THEME Rock1235)])],
 [SPEECHACT]}
```

These two RHET objects are the input to the plan reasoning system, which filters them by checking the constraints, preconditions and effects to produce a new interpretation space that is superficially integrated into context. If a single specific interpretation is left, then that is taken as the interpretation. Otherwise, general plan recognition, using a system based on Kautz's algorithm, is used to further eliminate the remaining interpretations, or to suggest new interpretations.The prototype system currently operates on a limited knowledge base of speech acts and interpretation rules. All the examples given in this paper can be run, as well as a few others discussed in Hinkelman (1989). More details on the system and its capabilities can be found in Hinkelman & Allen (1989) as well.

7. DISCUSSION AND EXTENSIONS

While the feasibility of this approach has been demonstrated by the prototype system, the work opens many possibilities for some interesting extensions that would make the system truly useful in general domains. A couple of these possibilities are discussed here.

Perhaps the most interesting result of this work is that we now have a framework in which the intonational and prosodic cues to speech act interpretation can be explored and investigated. Although these cues can be very strong indicators of the intended act, no previous framework was able to integrate information across intonation, prosody, syntax and semantics. Our framework allows this simply: there would be a set of rules that match certain intonation patterns and define interpretation space restrictions just like all the other rules. This information can then be combined in the way described above to affect the interpretation. So one extension is to explore the intonational cues to speech act interpretation and to formalize them as rules within this framework.

The other obvious extension would be a generalization of the current approach to allow likelihoods to be associated with interpretations. Rather than rules simply suggesting all specific interpretations on an equal footing, some interpretations might be preferred. The idea can be motivated by considering the analysis of *Can you open the door?* The interpretation rules would produce at least two specific interpretations: a request, based on the *can you* rule, and a yes-no question, based on the interrogative mood rule. It seems, however, that it is much more likely that this sentence is a request rather than a yes/no question. Currently, each interpretation is treated the same way, and the appropriate one is derived by the plan-based filtering of interpretations. The obvious extension would be to associate a weight with each rule that indicates how common it is. The complication arises in how one might derive the weights, however. Certainly, as many interrogative sentences are used as yes/no questions as are used as requests. So the interrogative mood rule should give a high weight to the yes/no question interpretation. Similarly, *can you* sentences frequently are used as requests, but there are many yes/no readings as well - as in *Can you eat three pizzas in ten minutes?*. Possibly, the contextual effects alone give us these intuitions of preferred readings for forms, and the existing proposal is fine as it stands. This issue should be investigated as it may become crucial as the domain of application becomes more general.

8. SUMMARY

We have described a speech act interpretation system that retains all the advantages of the previous plan-based approaches, yet has the additional characteristics:

- It is much more efficient in dealing with everyday conventional forms;
- It provides a way to integrate in information from intonation and prosody;
- It can identify the implications of different ways of phrasing in paraphrases.

The key idea is to produce a space of possible interpretations, both a range of allowable interpretations and a set of suggested interpretations, rather than a single so-called literal meaning.

Acknowledgements

This work was supported in part by ONR/DARPA grant N00014-82-K-0193, by the Office of Naval research under grant N00014-80-C-0197, and by the National Science Foundation under grant IRI-8351665.

References

Allen, J.F., "Recognizing intentions from natural language utterances," in M. Brady and R.C. Berwick (eds.), *Computational Models of Discourse*. Cambridge, MA: MIT Press, 107-166, 1983.

Allen, J.F., *Natural Language Understanding*, Benjamin Cummings Publishing Co., 1987.

Allen, J.F. and C.R. Perrault, "A plan-based analysis of indirect speech acts," *American Journal of Computational Linguistics 3*, 167-182, 1981.

Allen, J.F. and B.W. Miller, "The Rhetorical knowledge representation system: A user's manual (for Rhet version 14.45)," TR 238, Computer Science Dept., U. Rochester, revised March 1989.

Austin, J.L., "How to Do Things with Words," New York: Oxford U. Press, 1962.

Carberry, S., "Tracking user goals in an information-seeking environment," *Proc., AAAI*, 59-63, 1983.

Fikes, R.E. and N.J. Nilsson, "STRIPS: A new approach to the application of theorem proving to problem solving," *Artificial Intelligence 2*, 3/4, 189-208, 1971.

Grice, H.P., "Meaning," *Philosophical Review 66*, 377-388, 1957; reprinted in D. Steinburg and L. Jakobovits (eds.). *Semantics*. New York: Cambridge U. Press, 1971.

Hinkelman, E.A., "Linguistic and pragmatic constraints on utterance interpretation," Ph.D. Thesis, Computer Science Dept., U. Rochester, 1989.

Hinkelman, E.A. and J.F. Allen, "Two constraints on speech act ambiguity," *Proc., Association for Computational Linguistics*, 212-219, 1989.

Kautz, H.A., "A formal theory of plan recognition," Ph.D. Thesis and TR 215, Computer Science Dept., U. Rochester, 1987.

Kautz, H.A. and J.F. Allen, "Generalized plan recognition," *Proc., AAAI Nat'l. Conf. on Artificial Intelligence*, 1986.

Litman, D.J. and J.F. Allen, "A plan recognition model for subdialogues in conversations," *Cognitive Science 11*, 2, 163-200, 1987.

Perrault, C.R. and J.F. Allen, "A plan-based analysis of indirect speech acts," *AJCL 6*, 3-4, 167-182, 1980.

Searle, J.R., in *Speech Acts*, Cambridge University Press, New York, 1969.

Searle, J.R., "Indirect speech acts," in P. Cole and J. Morgan (eds.). *Syntax and Semantics 3: Speech Acts*. New York: Academic Press, 59-82, 1975a.

Searle, J.R., "A taxonomy of illocutionary acts," in K. (ed), Language, *Mind and Knowledge*, Univ. of Minnesota Press, 1975b.

Sidner, C.L. and Israel, D.J., "Recognizing Intended Meaning and Speakers' Plans," *Proc. IJCAI '81*, 203-208, 1981.

Tenenberg, J.D., "Abstraction in planning," Ph.D. Thesis and TR 250, Computer Science Dept., U. Rochester, 1988.