Predicting and interpreting speech acts in a theatre information and booking system

Toine Andernach
Parlevink Group, Department of Computer Science
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

E-mail: andernac@cs.utwente.nl

Abstract

This paper discusses an approach to speech act analysis in the context of a theatre information and booking system which aims at exploiting as many superficial utterance clues as possible. These clues, combined with contextual information about the dialogue like the dialogue history and predictions derived from the current dialogue and from a test corpus, should enable us to model the course of a dialogue in such a way that a user feels comfortable with it (Dahlbäck & Jönsson 1986).

1 Introduction

In this paper, we present our ongoing work on speech act analysis (communicative functions as we call them) for modelling user intentions in man-machine dialogues. We will give support for our hypothesis that communicative functions of utterances can (at least partially) be determined by using lexical and structural information like words, word order and dialogue history; we will regard communicative functions as partial and multi-dimensional descriptions of what people want in dialogues.

Our empirical basis is a corpus of 64 dialogues collected in a Wizard of Oz environment; the context of the research reported here is Schisma (see section 4), a joint research project of KPN (Royal PTT Nederland) and the University of Twente.

In section 2 we claim that the restrictedness of a dialogue system combined with the awareness of the user of that system enable us to model the course of a dialogue in such a way that a user is satisfied with it.

In section 3 we will discuss some discourse theories, their problems and their attractive aspects and how we profit from the latter.

Section 4 gives an overview of the way we determine the communicative functions of user utterances in our dialogues. We discuss the kinds of communicative functions we consider the ones to be captured in our system and we will discuss some superficial utterance clues; form features of utterances like word order and the presence of a question mark. Other form features we discuss are lexical clues for communicative function.

Section 4.5 gives a quick glance of the rules we will use for mapping form features of utterances with their possible function.

In section 4.6 dialogue structural information like information about the previous utterances is discussed in the light of its use for predicting communciative functions. Special attention is paid to adjacency pairs.

In section 4.7 a method is presented for disclosing relevant information from a corpus and in section 5 we will discuss some future research.

2 Restricted language

In the past decennia, several discourse theories have been proposed. They did not all have the same purpose: some intended to account for the way *people* use (certain phenomena in) language, either in texts, monologues or dialogues. Others aimed at providing computational models for human discourses and yet others aimed at designing discourse models for the development of NLP applications.

The difference between the second and the last is that the latter does not necessarily model man-man linguistic behaviour; users are often aware of the restrictedness of the language by machines; e.g. in a restricted domain the number of (interpretations of) content words, syntactic constructions is limited and there are more semantic and

contextual restrictions. This awareness causes users to adapt their language. As Morel (1989) found in experiments, subjects often adapt their language when they talk to machines; machine-like voice and behaviour do have influence on the linguistic behaviour and the prosody of the users of a dialogue system.

The interplay between man's and machine's behaviour is also observed by Smith, Hipp & Biermann (1992). They assume that the input of the user at a certain point in the dialogue can be predicted in a subdialogue; this subdialogue specifies the focus of the interaction and thereby the number of possible interpretations is limited. More specifically, we follow Waterworth (1987) in his claim that a classification of functions of utterances can be possible and useful as long as we restrict ourselves to identifying and implementing dialogue strategies in a restricted domain.

Therefore, we do not a priori assume that man-machine dialogues proceed the way man-man dialogues do; it is *not* our primary goal to model man-man dialogues in a psychologically and theoretically plausible way although these factors might play a role. People must be able to get information about performances in theatres by using natural language *in such a way that they feel comfortable with it* (Dahlbäck & Jönsson 1986) and that is not necessarily the way they talk to other people.

3 Discourse theories

3.1 Discourse Analysis

In discourse analysis (DA), linguistic techniques are applied to discourse entities larger than the sentence in order to model human linguistic behaviour. An example of a discourse analytic approach is the use of discourse grammars. Basic categories of utterances are identified and concatenation rules are formulated.

According to Levinson (1983) two main problems with discourse analysis are the strict theoretical nature of it and the intuition-based claims of its researchers. These problems could be overcome if we could find generalisations in the structure of *realistic* discourse; discourse grammar rules thus found could be used in a system like ours. The problem however, is that usually, realistic discourses do not obey this kind of rules; the nature of the discourse grammar rules presupposes that the structure of the discourse is fixed. We have seen in our corpus however, that the structure of dialogues is far more flat than often assumed.

3.2 Conversation Analysis

In Conversation Analysis (CA) the emphasis lies on the collection of empirical data while the premature construction of theories is avoided. In naturally occurring conversations, systematic properties of the sequential organisation of conversations are searched for. Conversational analysis is rule-governed and the underlying idea is that shared knowledge of these rules most often enables conversants to have smooth flowing and coherent conversations with one another.

In its main goals CA very well satisfies our needs: first, we base our computational model on empirical data because they simply are more reliable than intuitive data. Second, we think that the dependency relation between an utterance and the utterance immediately preceding it, is more easy to exploit in a dialogue system than the hierarchical structure which is traditionally more emphasized in literature. And third, considering utterances as containing cueing devices used by the speaker seems attractive from an engineering point of view.

3.3 Speech Act theory

The main idea of Speech Act theory (SA) is that utterances do not only have a literal meaning, but perform specific actions (*speech acts*) as well. In SA three aspects of speech acts are distinguished: its *locution*, its *illocution* and its *perlocution* (see (Austin 1962)). The illocution is often considered to be the identifying characteristic of a speech act; it expresses the action executed by the utterance. We will regard the illocution as the most prominent aspect of utterances in dialogues on which the proceeding of a dialogue is based. Therefore, we will concentrate on this aspect here.

Levinson (1983) reports several problems with speech act theory. Among them are:

1. there is no one-to-one mapping between utterances and acts (one utterance can be associated with multiple acts and one act can be performed in multiple utterances)
2. there is no simple form-to-force correlation

We will circumvent these problems by regarding communicative functions as partial and multi-

dimensional descriptions of communicative functions. We assume that every utterance gives at least some clues for these functions.

3.4 Plan theory

Since the late seventies, researchers have tried to apply *Plan Theory* to the generation and interpretation of plans in discourse. A basic assumption among plan theorists is the fact that the linguistic behaviour of agents in information dialogues is *goal-directed*; an agent's goal is to reach a particular state. A term which is often used for this kind of goals is *intention*. A recipe for reaching a particular state is often called *a plan*. It can consist of a number of subgoals each of which can be realised by *a subplan*. Thus, in a plan, goals can be represented in a tree structure in which dominated goals must be reached in order to reach dominating goals (see for instance (Litman & Allen 1990) and (Lambert & Carberry 1991)). Speech acts are considered to be the primitive goals to be met.

A major disadvantage of plans represented as tree structures is that every (non-terminal) node assumes its dominated goals to be fullfilled; all possible plans are fixed for each of their subgoals. That means that we need an extra mechanism to cope with situations in which plans change. Furthermore, Penstein Rosé, Eugenio, Levin & Ess-Dykema (1995) showed that a tree structure is not adequate in cases where dialogues have multiple threads.

Ahrenberg, Jönsson & Dahlbäck (1991) criticised this approach in that they don't consider it to be necessary to model the whole range of plans a user can have; in Grosz & Sidner (1986) intentional structure (i.e. the structure of a user's plans) is isomorphic to dialogue structure and it is as least as difficult to determine the former as it is to determine the latter. Therefore, Ahrenberg et al. (1991) propose a *structural approach* to dialogue modelling; they claim that it is sufficient to use simple discourse plans which consist of two parts, an opening move and a closing move. This idea also stems from CA and fits very well with our idea of dialogue cohesion, which in fact stems from Halliday & Hasan (1976).

4 Communicative functions in Schisma

4.1 Schisma: an introduction

In Schisma we aim at providing a natural language dialogue system which interfaces a database containing information about theatre performances in a certain city or region. The interface should make it possible to ask about performances in general, to tune in to a specific performance and, if desired, make a reservation for this performance. Research until now has concentrated on various aspects of realising such a theatre information and booking system. Among these aspects are the building of a Wizard of Oz environment for the acquisition of a corpus of dialogues for this domain, analysis and tagging of the dialogue corpus, recognition of domain-specific concepts (actors, authors, plays, dates, etc.), syntactic analysis and dialogue modelling.

We are especially interested in the user's goal when he produces an utterance and how he realises that goal in language. We assume that to allow a flexible man-machine dialogue, the *communicative function* of an utterance of a user must be determined. We prefer the term *communicative function* instead of *speech act* because it is a more meaningful term, but we use both terms.

In the Schisma system (see figure 4.1) a special component (the Speech Act Analyser) will be developed. It will get its input from the parser and its output will be transferred to other dialogue managing components.

Like in Conversation Analysis, we assume that there is a strong interdependence between what speakers want and the way they choose their utterances, i.e. between form and function of utterances in a dialogue. We will exploit this interdependence for our system; the more we can rely on superficial information in the utterances for this task, the more computationally attractive this will be. Grosz & Sidner (1986), who were among the first to present a rather integrated computational theory of discourse structure, also stressed the significance of using superficial linguistic clues for identifying structures in discourse. [1]

In an integrated approach to dialogue modelling like (Traum & Hinkelman 1992), traditional speech acts are extended to account for certain

[1] See (Hinkelman 1990) for an illustration of the exploitation of superficial linguistic clues.

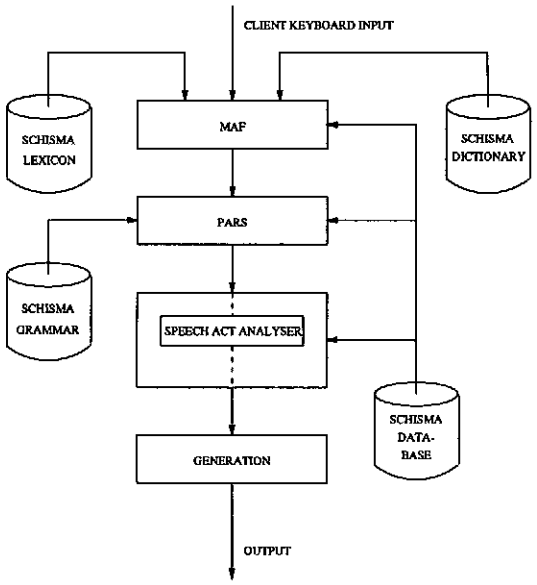

Figure 1: Global architecture of *SCHISMA*

types of coordinated activity that take place between agents in a conversation. Linguistic actions are signalled directly by surface features of the discourse, although usually a combination of surface features and *context* will be necessary to disambiguate acts.

The context in our system at a certain point in the dialogue will consist of a restricted number of possible transitions to communicative functions (determined by a Finite State Automaton), the form and function of previous utterances in the current dialogue, a list of preferred transitions to communicative functions (determined by statistical analysis of a test corpus) and, implicitly, the domain itself, containing certain domain concepts.

4.2 The form-function dichotomy

Since long ago, *the form-function dichotomy* has been recognised as a problem for determining the communicative function of utterances; corresponding utterance forms may have different functions and corresponding functions can be realised by different forms. Examples (1 and (2) taken from the Schisma corpus should make this clearer:

(1) *Kan ik drie kaartjes reserveren voor*
Can I three tickets reserve for
de eerste rij?
the first row?
'Can I reserve three tickets at the first row?'

(2) *Ik wil een kaartje reserveren.*
I want a ticket reserve.
'I would like to reserve a ticket.'

In distinct contexts, (1) can function both as a check for the truth of the proposition and as a request to execute the specified action. Furthermore, (1) and (2) are both requests for reserving (a) ticket(s) while their sentence types differ: (1) is a yes/no question about the speaker's ability to reserve tickets and (2) is a statement.

Crucial in this dichotomy, however, is the definition of the word *form* in this context; usually it is regarded to be the sentence type or mood of an utterance. But would the presumed dichotomy still exist when we take other form features of utterances into account? It is our main hypothesis that the form-function dichotomy can be circumvented by taking into account the form-features and contextual aspects as discussed in the former subsection.

In the next sections we will discuss some form features and our viewpoint of communicative function and how to relate them.

4.3 Communicative functions

After a closer look at the dialogues in our corpus, we found that the main function of all utterances in the corpus are either to *supply* something or to express a *wish* for something (see also (Wachtel 1986)).

We found that the objects of wishing and supplying can be *actions*, *information* and *truth values*. Combining them we get the Cartesian Product of these domain independent dimensions:

1. wish for action
2. supply of action
3. wish for information
4. supply of information
5. wish for truth value
6. supply of truth value

(1) and (2) can have instantiations in a specific domain like *reserve* or dialogue control instantiations like *thank* or *greet*, (3) and (4) concern concepts like *performance* or *actor* and (5) and (6) can for instance be expressed by *yes*, *no* or *ok*.

In an ideal situation, every utterance would give clues for each of these dimensions; in practice however, utterances will give us clues on just a subset of these dimensions. We use the word *dimensions* instead of *levels* because we don't think that a hierarchical classification (i.e. a *taxonomy*) of communicative functions satisfies our needs. Hinkelman (1990)'s taxonomy for instance, expresses that if there is evidence for a certain, say domain-dependent, communicative function, it implies that there is evidence for *all* dominating (more abstract, domain-independent) functions.

In our system, however, we don't want to be forced into a pre-fixed structure of the dialogue. A taxonomy in which domain-independent types of communicative functions dominate more specific domain-dependent functions is not suitable for our purposes because utterances can give clues on each of these dimensions independently:

(3) *En Othello?*
And Othello?
'What about Othello?'

In (3), *Othello* is the name of a performance. The question mark indicates the interrogative force, i.e. in our terminology it is a wish concerning a performance. The word *En* is a clue for the rhetoric relation with the former utterance. However, there is no clue for the kind of question that is meant in (3) (wh or y/n). So, despite of the fact that not every aspect of the domain independent features can be determined, we would still like to be able to account for the domain dependent information in an utterance.

In the following section, we will discuss the superficial clues for the communicative function of an utterance.

4.4 Form features

4.4.1 Sentence type

The first form feature we use for determining the communicative function is *the sentence type* of utterances. Table 4.4.1 is used to determine this sentence type.

The second column labelled with *verb 2nd/1st* indicates whether the finite verb is in second or in first sentence position. The column *subject* indicates the presence of a subject and the column *special* indicates some type-specific features.

The special sentence type *utterance* is introduced for the sentence type of all utterances that cannot be assigned another sentence type. Typical examples are utterances without a finite verb, like noun phrases and other constituents, affirmatives and greetings.

Type	verb 2nd/1st	subject	special
declarative	2nd	+	
imperative	1st	−	imp. verb form
y/n question	1st	+	
wh question	2nd	+	fronted wh-term
utterance			

Table 1: Sentence types

4.4.2 Punctuations

The presence of a question mark is a strong indication for a request. It is a *sufficient* condition in our corpus. Many utterances have a question mark while in the meantime having declarative sentence types:

(4) *2 zei ik toch?*
2 said I, didn't I?
'I said 2, didn't I?'

This observation cannot be generalised; in other contexts for instance, rhetoric questions can occur which in general don't have an interrogative function. See (Beun 1989) for a discussion of so-called *declarative questions* like (4).

On the other side, a question mark does not appear to be a *necessary* condition for an interrogative function: the corpus appears to contain a lot of utterances with interrogative functions, but without a question mark:

(5) *Wanneer is* Silicone Kitty
When is *Silicone Kitty*
'When does *Silicone Kitty* play'

4.4.3 Wh-words

One of the *lexical* clues for the communicative function is the presence of a *wh-word*. In almost all of the cases, the wh-word occupies the first position of the utterance or is part of a preposition phrase in subject position. Of the 62 occurrences of *Wat* for example, 58 are the first word of an utterance. Two of these are used in an exclamatory phrase. Of the other four, two are not interrogative pronouns, one starts with the conjunctive *En* and one is the first word of a subordinate clause.

Examples of utterances containing an interrogative pronoun which is not in first position are (6) and (7):

(6) *Wanneer worden welke voorstellingen*
When are which performances
gegeven van Het nationale toneel
given of *Het nationale toneel*
'When does *Het nationale toneel* perform which plays?'

(7) *Wanneer en hoe laat is*
When and how late is
Under a blue Roof*?*
Under a blue Roof?
'When and at what time will *Under a blue Roof* be played?'

(6) is a special case of a wh-question; two concepts are questioned in one utterance; the same counts for (7) although it is an elliptic utterance, contrary to (6).

4.4.4 Cue phrases

Another kind of lexical clues for the function of utterances in discourses are *cue phrases* (also called *clue words, discourse markers, discourse connectives* or *discourse particles*). Most cue phrases are realised as modal adverbs or adverbial phrases and they are traditionally regarded as explicit indicators of the structure of a discourse. They can e.g. mark a topic introduction, a topic shift (*now*) or a side step (*by the way*).

According to Hirschberg & Litman (1993) structural information conveyed by clue words is crucial to many more tasks:

- anaphora resolution
- inference of speaker intention
- recognition of speaker plans
- generation of explanations and other texts

As, we are mainly interested in the second task and...

> *"...despite the crucial role that cue phrases can play in theories of discourse and their implementation, however, many questions about how cue phrases are identified and defined remain to be examined..."*
> *(Hirschberg & Litman, 1993)*

...we will now have a look at some cue phrases in our corpus, more particularly the words *graag* and *niet*.

Literally, *graag* can be translated as *like to*. In dialogue however, it is often used as a more general politeness marker:

(8) *Ik wil graag naar* Mini en Maxi.
I want very much to *Mini en Maxi.*
'I would like to go to *Mini en Maxi*'

In all 54 cases of *graag* in the corpus it occurs in a declarative utterance. In 36 (67%) of the cases, the word *wil* (want) occurs in the same utterance. 3 other cases the word *zou* has the same function as the word *wil*.

In 10 (18%) of the utterances with *graag*, a verb is lacking and some concept is mentioned as a reply to a question of the Wizard. In 4 cases (7%), *graag* is meant as a confirmation of an immediately preceding yes/no-question of the Wizard. Most of the cases (3) accompanied by the word *ja* (yes). One occurrence of the idiomatic expression *graag gedaan* (it's a pleasure) was found.

A bigram analysis at word level of the utterances of the user yielded the highest frequency for the bigram *Ik wil* while a trigram analyis yielded the highest score for *Ik wil graag*.

To summarise we can say that *graag* supports (strengthens) the wish for information or action; this wish can be implicit (e.g. in the form of a (implicit or explicit) confirmation or choice) or explicit in the form of a wish marker, e.g. the verb *wil*. More specifically:

- (in combination with *ja*) in support of the confirmation of an information or action provision
- in combination with a domain concept in support of the confirmation of a wh or alt question.
- in combination with the explicit wish marker *wil* in support of the request

Another word that can be used as cue phrase is *niet*: in the following examples, *niet* does not function the way it usually to does, as a negation marker:

(9) *Kunt u ze trouwens niet*
Can you them by the way not
opsturen?
send?
'By the way, couldn't you send them to me?'

It seems that if *niet* is omitted, the (logical) meaning remains approximately the same. The

question then is: what does *niet* add in utterances like (9) in the corpus? What they have in common is that they all have interrogative force. This is marked by the y/n question word order.

Let's see what happens if we change (9) in (10):

(10) *U kunt ze trouwens niet*
You can them by the way not
opsturen
send
'You can't send them by the way'

(10) can only have the meaning intended in (9) if it has a rising intonation. With a default declarative intonation, *niet* serves the purpose of negating the proposition expressed in the utterance.

Thus, it seems that the special use of *niet* only occurs in utterances in which the speaker expresses a request. In these directive utterances, the speaker uses *indirectness* techniques to avoid that the speaker will feel forced to obey the speaker. Negating the proposition is one way of doing that. The speaker could also have used the word *misschien* (maybe) which expresses uncertainty by the speaker.

Examples (8) and (9) show that clue words can be very subtle indications for speaker intentions in discourse, very often in combination with other clues in the utterance.

4.5 Formalising the interpretation of communicative functions

Following Hinkelman (1990) we will use rules to determine for a certain input utterance a range of possible partial speech act interpretations. The rule below is an example of the kind of rules given by Hinkelman (1990). It is applicable to (1) above.

```
(S MOOD YES-NO-Q
 VOICE ACT
 SUBJ (NP HEAD ik)
 AUXS {kan}
 MAIN-V +action) => ((REQUEST-ACT ACTION)
                                (SPEECH-ACT))
```

Both structures at the left hand side and the right hand side of the arrow contain features with their values. This rule is applicable if the structure at the left matches (a substructure of) the structure yielded by PARS. The right hand side of the rule is a disjunction of partial descriptions of communicative functions.

4.6 Predicting communicative functions

To optimise the process of assigning communicative functions we could use a Finite State Automaton (FSA) to a priori exclude some communicative functions at a certain point in a dialogue. Such an Automaton is also used in the Verbmobil project (Alexandersson, Maier & Reithinger 1994). In this project, speech acts are both modelled in an FSA which restricts the sequential order in which the speech acts are used and hierarchically modelled in a taxonomy.

The necessary states and transitions in this FSA could be determined by using a test corpus in which the communicative functions are tagged (see section 4.7) or by using common or intuitive knowledge about the sequence of utterances. In the latter case we should be aware of the fact that the word *common* in *common knowledge* does not make the knowledge more reliable.

A rather new way predicting communicative functions is by statistical information (Reithinger & Maier 1995). A finite state model is not sufficient for the prediction task because it is not sufficiently restrictive. Therefore, we will use information about relative frequencies of sequences in a test corpus. This results in information about *adjacency pairs* and *preferred seconds.*

Adjacency pairs consist of two turns each uttered by another speaker. One of the characteristics of the parts of these pairs is their adjacency. Levinson (1983) notices that, instead of occurring strictly adjacent, the parts of an adjacency pair are frequently split up by so-called *insertion sequences* which also consist of adjacency pairs. An example from the corpus is (11):

(11) *S: Hoeveel kaartjes wilt u en met welke reductie?*
C: hoeveel kaartjes zijn er nog?
S: Er zijn nog 400 plaatsen vrij voor deze voorstelling.
C: Doe maar tien

In (11), the second and third turn form an insertion sequence.

Furthermore, Levinson (1983) indicates a problem with the feature of *typedness*, i.e. the fact that a particular first part requires a particular second part. In natural dialogues it is not the case that for instance *an offer* is always followed by an *acceptation*; it can also be followed by a *rejection*. Thus, instead of a strict coupling of

both parts, we could assume a *preference organisation*; an acceptation of an offer is preferred to a rejection of that offer.

This kind of information will be yielded by the statistical analysis of our tagged test corpus.

4.7 Tagging: a systematic way of information disclosure

One way of systematically disclosing the varied amount of information for dialogue management in the corpus is *tagging*; certain characteristics of words, utterances or sequences of utterances are annotated in such a way that common features can be found by statistically processing these annotations.

In order to test our hypothesis that the combination of several superficial clues would enable us to determine (at least some aspects) of its communicative function, we will tag our corpus. To avoid an explosion of feature combinations we will tag three form features: sentence type, presence of a wh-word and presence of a question mark) and four function features (the speaker, the main act (request or provide), the object of the act (information, action or truth value) and the domain-dependent instantiation of the object. The following list gives an overview of the dimensions and their instantiations:

Form:

1. word order
   (a) utterance
   (b) declarative
   (c) y/n
   (d) wh
   (e) imperative
2. presence of a wh word
   (a) yes
   (b) no
3. presence of a question mark
   (a) yes
   (b) no

Function:

1. speaker:
   (a) user
   (b) system
2. domain-independent function classes:
   (a) request
   (b) supply

Domain-independent concept classes with their domain-dependent instantiations:

1. action
   (a) reserve
   (b) annulate
   (c) thank
   (d) greet
2. constant, information
   (a) (ITEM)
   (b) (EMPTY)
3. truth value
   (a) yes
   (b) no
   (c) (NOT KNOWN)

Examples of items are: performance, time, costs, seats, reduction, rank, payment method, address theatre, reservation, information, date, number of people, summary, getting tickets at office

A corpus with utterances characterised this way, can be analysed in several respects: sequences of the tag types, n-grams of clusters of m features of utterances.

Relative frequencies of bigrams of *form* and *function* features of user utterances in the test corpus are used to formulate the rules which map the form with the function features.

Relative frequencies of n-grams of *function* features of both user and system utterances are used in preference rules for predicting communicative functions; after the FSA has restricted the number of potential communicative functions at a certain point in a dialogue, these preference rules will order these functions on a scale ranging from most probable to less probable. This ordening will be used to optimise the process of assigning communicative functions.

5 Future Research

In further research we will test our hypothesis that the communicative function of utterances in man-machine dialogues can be determined using superficial information from the utterances themselves; all utterances will be tagged the way we described in this paper and we will analyse these tagged utterances and improve our current Simulation Environment with a dialogue model based

on the results of this analysis. This environment is used to semi-automatically collect Wizard of OZ dialogues. It will serve as the platform for our eventual prototype; it will be extended with other modules to be developed.

Next step is to collect Wizard of Oz dialogues in a theatre, a more realistic environment, usingthe improved Simulation Environment.

References

Ahrenberg, L., Jönsson, A. & Dahlbäck, N. (1991), Discourse representation and discourse management for a natural language dialogue system, Research report, NLPLAB IDA Linköping University, Linköping, Sweden.

Alexandersson, J., Maier, E. & Reithinger, N. (1994), A robust and efficient three-layered dialogue component for a speech-to-speech translation system, Report 50, DFKI GmbH, Saarbrücken, Germany.

Austin, J. (1962), *How To Do Things With Words?*, Clarendon Press, Oxford.

Beun, R.-J. (1989), The Recognition of Declarative Questions in Information Dialogues, PhD thesis, Instituut voor Perceptie Onderzoek, Eindhoven, The Netherlands.

Dahlbäck, N. & Jönsson, A. (1986), A system for studying human-computer dialogues in natural language, Research report, NLPLAB IDA Linköping University, Linköping, Sweden.

Grosz, B. & Sidner, C. (1986), 'Attention, intentions and the structure of discourse', *Computational Linguistics* **12**(3), 175–204.

Halliday, M. & Hasan, R. (1976), *Cohesion in English*, Longman, London.

Hinkelman, E. (1990), Linguistic and Pragmatic Constraints on Utterance Interpretation, PhD thesis, University of Rochester, Rochester.

Hirschberg, J. & Litman (1993), 'Empirical studies on the disambiguation of cue phrases', *Computational Linguistics* **19**(3), 501–530.

Lambert, L. & Carberry, S. (1991), A tripartite plan-based model of dialogue, *in* '29th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference', ACL, Berkely, California, USA, pp. 47–54.

Levinson, S. (1983), *Pragmatics*, Cambridge University Press, Cambridge.

Litman, D. & Allen, J. (1990), Discourse processing and commonsense plans, *in* P. Cohen, J. Morgan & M. E. Pollack, eds, 'Intentions in Communication', The MIT Press, chapter 17, pp. 365–388.

Morel, M. (1989), Computer-human communication, *in* M. Taylor, E. Néel & D. Bouwhuis, eds, 'The Structure of Multimodal Dialogue', Elsevier Science Publishers B.V., Amsterdam, North-Holland, chapter 24, pp. 323–330.

Penstein Rosé, C., Eugenio, B. D., Levin, L. & Ess-Dykema, C. V. (1995), Discourse processing of dialogues with multiple threads, *in* '33rd Annual Meeting of the Association for Compuational Linguistics', ACL, Cambridge, Massachusetts USA.

Reithinger, N. & Maier, E. (1995), Utilizing statistical dialogue act processing in verbmobil, *in* '33rd Annual Meeting of the Association for Compuational Linguistics', ACL, Cambridge, Massachusetts USA.

Smith, R., Hipp, D. & Biermann, A. (1992), A dialog control algorithm and its performance, *in* 'Proceedings of the Third Conference on Applied Natural Language Processing', ACL, ACL, Trento, Italy, pp. 9–16.

Traum, D. & Hinkelman, E. (1992), Conversation acts in task-oriented spoken dialogue, Technical Report 425, University of Rochester, Rochester.

Wachtel, T. (1986), Pragmatic sensitivity in nl interfaces and the structure of conversation, *in* 'Proceedings of COLING', Bonn, pp. 35–41.

Waterworth, J., ed. (1987), *Speech and Language-Based Interaction with Machines: Towards the Conversational Computer*, Ellis Horwood Limited, Chichester.