Swedish Dialogue Systems
A proposal to the Swedish Language Technology Program (NUTEK/HSFR)

Göteborg University Linköping University Lund University
Royal Institute of Technology Telia Research AB

1st December, 1997

1 The vision

The standard way to interact with a conventional computer is to use a keyboard for input to the computer and a screen for input to the user. Increasingly, however, the need is being felt for interaction using other modes. Currently, the interaction between the user and the machine in these systems is very primitive from a dialogue perspective. The user is asked to enter numeric codes via the telephone or select from a menu of numbers and the machine responds with prerecorded speech or speech assembled from prerecorded fragments. Such systems could be improved by a more adequate treatment of dialogue. A dialogue system for banking would be able to understand the user's spoken request to transfer money rather than present the user with a menu or require that the user have a card in her hand which lists the numeric codes; another step would be a multimodal system which combines different modalities such as speech and vision.

A great deal of research has been conducted both in Sweden and elsewhere on the development of dialogue systems for the travel domain, booking airline or train tickets, planning routes and obtaining relevant tourist information. Some such systems are already in use but there is a great need both of fundamental research into the nature of such dialogues and also the technology needed to allow machines to take part in such dialogues. There is a great deal of activity in the international language engineering research community which concerns the investigation of multimodal dialogue systems. Sweden is unusually well-placed to make a contribution to this area given the current constellation of research groups in the country.

Our aim in this project is to bring together research groups in Sweden who have competence in speech analysis and synthesis, dialogue structure and the interpretation of natural language and move them towards the creation of a common environment which will allow for the creation of generic dialogue systems and the investigation of fundamental issues in dialogue technology. For this to be possible a number of problems, ranging from speech recognition to interaction design, have to be solved. In this project we focus on the higher system levels: dialogue management, robust utterance interpretation, including use of prosodic information, and effective utterance planning. In order to ensure flexibility and ecological validity of dialogue systems we will also create a platform for multimodal spoken language corpora which will contain spoken language audio and video recordings with standardized transcriptions. We believe that the results of a project of this type will be very valuable both as a resource for future research and as a way of raising Swedish competence in the area of multimodal spoken language processing.

2 Goals of the project

Within this two year project we will be able to start this process by investigating areas of fundamental research which we see as logical next steps in improving the state of the art of dialogue technology.

We will work both on the development of dialogue system technology and the empirical investigation of dialogue phenomena. More specifically we aim to work on the following:

1. Research on dialogue systems and development of models for creating spoken natural language interfaces which bring together both speech and dialogue management modules. This will serve as the basis for two implementation tasks: the development of an educational system for instruction concerning dialogue systems and the creation of an environment for the development of robust, interactive and portable spoken dialogue systems. These implementation tasks, in addition to providing valuable input to the empirical investigations, will form the basis for planning a future more generic system which can serve as a testbed for various dialogue modules. We wish to move towards a General Architecture for Dialogue Engineering with a similar philosophy to that of the GATE (General Architecture for Text Engineering) platform developed at Sheffield University, although it would be too ambitious to achieve this during the two year life of this particular project.

2. Empirical investigation of the interaction of focus phenomena with intonation on the one hand and information state update on the other and the consequences of the analysis of these phenomena for the design of dialogue systems. This will involve an investigation of the relationship between prosody, grammar and discourse structure and their implications for speech recognition and synthesis. It will also involve an investigation of the empirical manifestation of communicative functions in dialogue moves and such phenomena as hesitations, self correction and feedback. For these studies a platform for multimodal dialogue corpora and a dialogue move engine for Swedish will be developed. This investigation will be oriented towards fundamental research which is relevant to the kind of dialogue system we are investigating and the future development of such systems.

It is important to note that the two themes are intertwined: theories and models for dialogue technology are developed from both empirical investigations and systems development.

3 Background

Spoken dialogue systems have been an active area of research for quite some time. Internationally we see new products and research projects being developed which involve speech, either solely or as the main communicative modality e.g. TRAINS (Allen *et al.*, 1995), VODIS (Krahmer *et al.*, 1997), VERBMOBIL (Alexandersson *et al.*, 1995). A prototype spoken language interface to a commercial vehicle route planning system has been under development during a number of years at SRI Cambridge, in collaboration with the UK Defence and Evaluation Research Agency (Lewin *et al.* 1993, Lewin and Pulman 1995).

As speech recognition technology has improved, the higher level modules of such systems have become more critical and consideration of these has brought a greater need for empirical and theoretical studies of dialogue phenomena in human-human communication as well as human-machine communication. Moreover, many systems that have speech as an option at the computer screen provide in addition a face or a character for interaction with the user, thus bringing to the forefront the coordination between speech on the one hand, and facial as well as other body gestures, on the other.

3.1 Dialogue system technology

Previous Swedish research systems There are a number of research groups in Sweden which are currently working either on dialogue systems or on research issues which are relevant to dialogue technology. The groups involved in this proposal are Göteborg University, KTH, Linköping University, Lund University and Telia Research.

The Waxholm project was supported by the Language Technology Program. A specific effort in this project was to build a generic system in which speech synthesis and speech recognition can be studied in a man-machine dialogue framework. In addition, the system should facilitate the collection of speech and text data that were required for the development of the system. The demonstrator application, called WAXHOLM, gives information on boat traffic in the Stockholm archipelago. Besides the speech recognition and synthesis components, the system contains modules that handle graphic information such as pictures, maps, charts, and time-tables. This information can be presented to the user at his/her request. The application has similarities to the ATIS domain within the ARPA community and other similar tasks in Europe, for example SUNDIAL. The dialogue system was implemented as a number of independent and specialised modules. The project resulted in a number of research reports and also in public demonstrations. Part of the new proposed project builds on knowledge and software developed in the Waxholm project.

The Department of Linguistics, Göteborg University was a participant of the LRE project PLUS, Pragmatic Language Understanding System. The purpose of the project was to develop a robust, useful and cooperative dialogue system for answering questions about the Yellow Pages. The project built on Wizard of Oz studies and resulted in a dialogue manager and dialogue system which met the project criteria of robustness, usefulness and cooperativeness, cf. Black *et al.* 1991.

The dialogue model to be used as a starting point in this project was developed at the Laboratory of Natural Language Processing at Department of Computer and Information Science, Linköping University (Ahrenberg *et al.*, 1990, Jönsson, 1997). It is based on a number of empirical investigations into human computer natural language interaction carried out in interdisciplinary cooperations involving computer scientists, linguists and cognitive psychologists. The dialogue model does not aim to mimic human interaction for any situation or application. Instead it is restricted to human-computer interaction in natural language, and natural language interfaces for different applications which belong to the domain called simple service systems. Such systems require in essence only that the user identify certain entities, parameters of the service, to the system providing the service, and once they are identified the service can be provided. The research has also resulted in a prototype natural language interface that can be customized to meet the demands from the application based on empirical studies of the user's information needs. Recent investigations show that the dialogue manager developed as part of the LINLIN project can be useful also for spoken interaction to simple service systems.

Robustness, efficiency and portability Spoken dialogue systems require computational models that allow for robust and efficient interaction. User satisfaction with an SDS correlates positively with task success and negatively with delays, misunderstandings and information overload (Walker *et al.* 1997).

A weak point of most SDS:s is the speech recognition. Speech interpretation can be enhanced by the use of syntactic and semantic constraints that are specific to the application, by methods for partial interpretation, by robust dialogue management, and by careful design generally. For information content that is being extracted from an utterance, it is better to go for high precision and low recall than vice versa. Of special interest to this project are prosodic cues, which can be exploited to improve the accuracy of interpretations in many different respects.

It is known that prosody can be exploited in several ways to improve the speech understanding process. Previous studies of English, for example, have shown that prosodic information can be used in discourse analysis in order to improve processes such as lexical access (cf. Munteanu *et al.* 1997, Lieske *et al.* 1997 as well as interpretation/recognition of illocutionary acts (Warnke et al 1997). Within the general area of discourse segmentation (involving the grouping and phrasing of utterances), it has been demonstrated that prosodic cues constitute salient markers of discourse boundaries (Cutler *et al.*, 1997 and references therein). As regards the specific area of man-machine dialogue, it has been claimed that prosody (specifically accentuation) can play a particularly important role in situations where the system requires clarification on referent and/or attribute specification due to defective speech recognition (Terken *et al.* 1997, Swerts & Ostendorf 1997). Furthermore, the whole area involving the role of accentuation in signalling focal vs nonfocal/ground information in spoken discourse has been the subject of intensive research, particularly as regards text-to-speech applications (Hirschberg 1993 and references therein, Cutler *et al.* 1997 and references therein). The important role that prosody will play in the development of spoken dialogue systems in processes such as referent resolution and utterance purpose determination is now becoming recognized (Swerts & Ostendorf 1997).

The capabilities of an SDS are quite limited when compared with a human being. This holds across the board: speech recognition, lexicon, syntax, semantics, world knowledge etc. This is another factor that motivates a careful design. On the other hand, a system that is too customized for a given application is not of general interest. What is needed are portable systems, i.e. systems that are based on components that can be used and adapted for as large a class of applications as possible. In this project we are primarily interested in developing knowledge, methods and systems that pertain to dialogue management and the interpretation and generation of utterances, and are of general use for simple information systems, e.g. systems that handle traffic information. However, to cater for long-term needs we also need general theories based on valid empirical data.

Studies of human dialogue have revealed a large number of features that ensure robust and efficient interaction, e.g. constant feedback (often displayed visually), reliance on shared knowledge and inference capabilities, marked changes of topic, smooth repair mechanisms, and so on. Such features are obviously important for human-computer spoken dialogues as well, but it is an open question how they can best be implemented in an SDS. For instance, Krahmer *et al.* (1997), report that their system verbally confirms all user utterances, before it takes action on them, a feed-back mechanism that humans use only under extreme circumstances. Similarly, while plan recognition enables a system to make more useful responses, it usually also delays response time, thus creating a design trade-off.

Educational systems Traditionally, spoken dialogue systems have required high-level expertise to design and develop. The resulting applications have been large-scaled and complicated and not always easy to modify and extend or even comprehend. Thus, teaching in the subject of spoken dialogue systems and related technologies has mostly been done in lecture format with video taped demonstrations of actual systems. Live demonstrations are typically conducted by a well-behaved researcher who is well acquainted with the given system and knows which questions to ask. Students have mostly been kept at a safe distance.

The aim of the work at KTH has been to put a very simple but fully functioning spoken dialogue system into the hands of our students as an instructional aid. They have been given the opportunity to test it themselves and to examine the system in detail. They have also been given guidance on how to extend and interactively develop the system's functionality.

The system is based on the software technology in our existing spoken dialogue system WAXHOLM (Bertenstam *et al.*, 1995). We have extracted and redesigned components such as speech recognition (Ström, 1996), speech synthesis (Carlson *et al.*, 1990), visual speech synthesis (Beskow, 1995), and parser (Carlson, 1996). This has enabled us to create a toolkit for spoken language technology in the spirit of CSLUsh created at OGI (Sutton *et al.*, 1996) and Sapphire created at MIT (Hetherington and McCandless, 1996).

The dialogue component is so far very limited. In the current grant from the Language Technology Program "Verktygslåda för talbaserade dialogsystem" work has started which will lead to the integration of the more sophisticated dialogue model developed in Linköping. The goal has been twofold. First to include an educational dialogue component based on current science. Second and very important to form a platform for cooperation, integration and sharing of knowledge.

This work has mainly involved adaptation of the various components in Linköping and Stockholm into a common, modular system. We have also begun designing the dialogue manager to allow for input from the speech components and specified a plan for further work which involves the research issues discussed below.

For the current project, there are certain areas where empirical studies of human dialogue, as well as human-computer dialogue, are necessary. These are, primarily, the relationship between prosodic features and semantic/pragmatic features such as focus structure, cue phrases and discourse relations, and similar correlations between visual signals and conversational role (speaking, listening; comprehending vs. non-comprehending listener, etc.). These studies require a platform, where prosody, visual and facial gestures can be studied and analysed alongside the verbal flow of information. Currently there are very few platforms of this type.

3.2 Empirical studies of dialogue

Dialogue corpora There is an increasing interest in studies of human dialogue. For English, the British National Corpus, which is probably the biggest effort so far, includes computer readable transcriptions of 2 million words of dialogue. As far as multimodal corpora of dialogue are concerned, most publicly available corpora only feature speech sounds and transcription. Two examples of this are the Rochester Train corpus and the Edinburgh Map Task corpus which are both focussed on tasks in a single domain.

All three mentioned corpora have been used to investigate communicative functions like speech acts, but also to investigate various types of prosodic features.

The Department of Linguistics, Göteborg University, has been recording and transcribing spoken dialogue since the 1980s. The corpus currently consists of about 1,000,000 machine readable words. A first report based on 300,000 words, giving frequencies of words and collocations, is available in Allwood (1996). A second version of the report is in preparation. The corpus is organized according to activity types, e.g. lecture, negotiation, sermon, patient-doctor consultation etc. It includes some structured tasks and man-machine dialog. There are altogether 25 activity types. Part of the corpus is based on video recordings and part of it is based on audio recordings. The video recordings form the background for the platform for multimodal spoken language corpora proposed below.

The empirical work on spoken dialogue has been combined with work on dialogue theory. Much of this work is summarized in Allwood (1995).

Other work on dialogue has been done in Linköping. The Linköping Dialogue Corpus consists of around 150 dialogues collected in Wizard of Oz experients. Sixty of the dialogues are available over the web. For these dialogues two different scenarios are used, a car purchase scenario and a travel scenario. The analysis of these dialogues was fundamental for the development of the LINLIN dialogue manager.

Studies on prosody Both Horne and Bruce have previously led projects on discourse prosody within the HSFR/NUTEK Speech Technology Program. Horne's work has been mainly geared towards speech synthesis and the development of a number of algorithms in the linguistic component of a text-to-speech system that extract linguistic information from texts which can be used in the prediction of accentuation patterns (Horne *et al.*, 1993, Horne & Filipsson, 1996). Particular attention has been paid to the design of algorithms for tracking coreference/cospecification relations as well as for generating prosodic structure. Bruce's research has concentrated on the prosodic signalling of phrasing in elicited speech as well as the prosodic correlates of topic structure in spontaneous dialogue (Bruce *et al.*, 1993, 1996). He has also led a project on the recognition of prosodic categories in Swedish (House et al., 1987). Whereas most of this previous work has been carried out with the goal of developing models for speech synthesis, the present project application aims at using prosody, together with other information sources, both in the recognition and understanding of spoken dialogue and in the generation of spoken output.

Information states Participants come to a dialogue with a certain "information state", that is to say, a collection of beliefs, knowledge, desires, goals etc. which are collectively known as "attitudes" that the dialogue participant has. A dialogue participant can express one or more of these attitudes in a natural language utterance and thereby succeed in modifying the information state of another dialogue participant. This process of modifying information states during the course of a conversation may be referred to as *dialogue dynamics*.

The notion of information update has become standard in a number of modern semantic theories including work by Irene Heim (Heim, 1982), discourse representation theory (DRT) (Kamp and Reyle, 1993), and dynamic semantics as it has been developed in Amsterdam and elsewhere (Groenendijk and Stokhof, 1991, Dekker, 1993). For a recent paper bringing these approaches together see van Eijck and Kamp (1997). More recently, some researchers have begun to take more seriously the fact that natural language dialogues involve more than one agent and that there are specific phenomena that become apparent when you consider semantics from the perspective of dialogue. Some examples of work in this area are: Ginzburg(1996a, 1996b), Dekker and van Rooy (1997). The formalization of information states has been investigated in Cooper (1996), Cooper and Ginzburg (1996) and Cooper (forthcoming a, forthcoming b).

4 Research Issues

There are two main research issues which this project will address:

4.1 To what extent is it possible to build a general architecture for dialogue engineering?

The project will only be able to begin to explore this question during its two-year span and will attempt to do this in a concrete way by creating one experimental and one pedagogical system which weld together components that have been developed at separate Swedish sites. We regard this issue as a central practical issue which needs to be resolved in the next five years if we are to have a suitable testbed for advanced dialogue technology.

4.1.1 An environment for a robust, interactive and portable multimodal dialogue system.

One specific issue for the project involves to what extent system robustness can be achieved by reducing parsing and interpretation processing and instead rely on simpler processing mechanisms and dialogue management. This means developing methods such as keyword spotting and template matching. Preliminary results will be available from a project carried out within the current planning phase which tentatively identified a number of types of expressions where each type can have its own processing mechanism, e.g. keywords and related synonyms corresponding to entities with information in the database, other concepts utilized in the domain and relational expressions.

Another issue is the role of prosody, both in interpretation and generation. Issues here involve questions on how to utilize such information in interpretation and dialogue management. It also includes questions on how to utilize focus structure to determine prosodic structure when generating a response to the user.

Generation also involves utterance planning. A point at issue here relates to the question of how context and modality influence information aggregation.

A further issue is that of system portability and the development of tools for achieving portability. For example, it is important to design the dialogue manager and the other modules in the system so as to allow for easy adaptation to different applications. Within the project, three different, although related applications will be included: the GULAN database, a database on traffic information and extended ATIS-dialogues. This part of the project requires development of tools for customization of the various components of the system. To what extent can the same dialogue objects and lexical data be used? Can updates and changes be done by automatic, or semi-automatic means?

4.1.2 A toolbox for multimodal dialogue systems

Research issues here include many basic aspects of multimodal dialogue systems that need to be specified in order to develop an optimally flexible system that can be used as an experimental prototype and as a pedagogical tool in education on dialogue structure. The specification and integration of a reasonable dialogue manager is an important area for research in the current application. The manager should be flexible, domain independent and still usable in the specific demonstration/tutorial domain. It should also be robust and open for inspection. This calls for challenging research issues such as how to structure information flow between dialogue system components, sharing of domain knowledge and how to use dialogue knowledge in speech recognition and speech synthesis. It also involves issues on what information is needed for a dialogue manager to cooperate productively with a speech system. Currently work is being carried out on extending lexical information to include semantic information, but it might also be possible to exploit other kinds of information, e.g. prosodic cues in the processing of speech. The project also requires tools for specifying the behaviour of a dialogue system. One important part of this is the design of tools to allow for easy modification of the system in order to illustrate effects of various choices made in the design. The knowledge used by the various modules in the system must also be specified to allow for such customization.

4.2 What is the relationship between focus phenomena, information state update, prosody and gesture?

We see this as one of the central issues given the current state of the art in dialogue research, which combines aspects of speech, gesture and interpretation. Specific issues here involve questions on how to utilize such information in utterance interpretation and generation. Other questions involve the nature of features of dialogue interaction such as interruptions, turntaking, feedback and the nature of communicative functions involved in dialogue moves.

4.2.1 Focus and Information State Update

An important issue that must be addressed is how one should represent focus information in information states. Can, for example, the HPSG analysis of focus phenomena presented by Engdahl and Vallduví (1996) be incorporated in a computational representation of information states which takes account of prosodic phenomena? It looks at the moment as if the kind of representation of information states proposed by Cooper (1996) and Cooper and Ginzburg (1996) can be recast with advantage in terms of the variant of Martin-Löf's type theory including dependent record types recently developed at Göteborg by Betarte and Tasistro (forthcoming) and that this use of records yields an HPSG-like representation of information which could bridge between Cooper and Ginzburg's proposals on the one hand and Engdahl and Vallduvi's proposals on the other. A related research issue concerns what kind of information needs to be included in information states in order to facilitate the definition of updates that are sensitive to focus information. Here are some possibilities:

- questions under discussion (as proposed by Ginzburg, 1996, Roberts, 1996)
- indications of referents (in a database) associated with "dialogue roles" (corresponding to discourse markers in the sense of discourse representation theory) in an agent's information state
- a representation of what the agent views as having been established in the current dialogue (also providing hooks for anaphoric expressions) and possibly in addition a model of the interlocutor's information state
- linguistic information about the previous dialogue contribution (or more than one dialogue contribution)
- micro-conversational events, that is speech events below the level of a complete dialogue contribution, including hesitations and repairs (Poesio and Traum, forthcoming). A high degree of sensitivity to micro-conversational events would involve each word and hesitation being associated with a transition on information states.

4.2.2 Prosody and Dialogue Structure

The project aims at examining a number of issues regarding prosody's role in information processing in Swedish dialogues. Within the general area of discourse segmentation (involving the grouping and phrasing of utterances), the specific area pertaining to the prosody of cue words/discourse markers will constitute one area of investigation. These words are important signals of discourse structure, not only as regards their turn-taking function in dialogues but also as regards their function of signalling an utterance's speech act classification with respect to preceding and following utterances. Cue words, however, are sometimes ambiguous as regards their discourse vs sentential use (Hirschberg & Litman 1993) and it is assumed that accentuation, as well as durational and pausal structure can provide important information as to their interpretation (Hirschberg & Litman 1993, Carletta *et al.* 1997). Since there have been no systematic studies undertaken on Swedish cue word prosody, it is important that our knowledge of this area of discourse structure be increased. How cue word prosody interacts with their semantics and syntax is an equally unexplored area and thus the whole issue of how a comprehensive 'cue word grammar' can be included in/interfaced with current modules in a dialogue management system such as LINLIN must be recognized as a fundamental research area for the proposed project.

Another area of investigation where prosody plays a central role in the exchange of information is in the signalling of focal vs nonfocal/ground information. In this respect, the parameter of accentuation (pitch-range variation) gives important cues as to information structure. Since focal accents correlate to a very significant degree with information that is new to the discourse (Horne *et al.* 1993), a dialogue management system that has access to this prosodic information can use it in order to enhance the updating of information on dialogue referents/objects and attributes/properties. A central issue is how to integrate prosodic information on focus/ground structure in the framework of the LINLIN system which was originally designed for processing written language.

Although prosody is one major cue to focus/ground structure, it is also known from studies on a number of languages that prosody interacts with other kinds of grammatical and discourse information (NP-type and word order) in expressing focus/ground relations (Ariel 1988, Givón 1983, Vallduví and Engdahl, 1996). The question of how the interaction between these different information sources is to be systematized is thus another important issue for the project.

4.2.3 Multimodal Dialogue Corpora

In order to provide an ecologically sound base for the project and an efficient toolbox for empirically examining research issues involved both in the dialogue systems and the dialogue tutorial system discussed above, as well as for continued empirical work on prosody and other dialogue features, we plan to create a platform and toolbox for multimodal spoken language corpora.

This work is also motivated by the need for better theories of dialogue which can serve as input to the construction of dialogue managers and multimodal communication. For example, we need to know what communicative functions (dialogue acts) go with different roles in a dialogue system, e.g. user, system or student, teacher etc. We also need to know more about other features of dialogue interaction such as interruptions, turntaking, feedback, sequencing, hesitation and self-corrections and, not least, we need to know more about the relationship between gesture and speech. In this project, we will provide a platform for examining such issues. We will also provide some initial results on "own communication management" (hesitation and self-corrections - OCM) and "interactive communication management" (turntaking, feedback and sequencing - ICM).

5 Workplan and expected results

In this section we summarize the goals presented in the previous sections and indicate how we plan to achieve them and the kind of results we expect.

5.1 A toolbox for spoken language dialogue systems

The planning project "toolbox for speech based dialogue systems" is well underway. This project has been devoted to three major tasks: to start expanding an existing environment with a simplified dialogue handler, to redesign the software for this new expansion and to plan the main phase of the research and development activity.

During the first year we will have a number of goals. Several of them are continuations of the planning project.

The toolbox should be developed further towards a robust and distributed environment. This means that we need to continue work on a web based architecture, making it possible to use the same educational software in parallel on several computers. This software will be used both in education and demonstrations. The integration of the simple dialogue handler will be completed. It should be possible to include new semantic categories and to relate them to new actions in the system. The work on specifying primitives which describe the behaviour of the dialogue manager for various dialogue situations carried out in 5.2 is an important part of this part, but here with focus on visualization to understand human-computer interaction and dialogue system development.

In the second year the dialogue manager developed in 5.2 will be integrated in the dialogue system. This dialogue handler uses a more elaborate linguistic analysis than is currently used in the toolbox. Thus, interpretation methods developed in 5.2 will also be included. An example based shallow parser will also be added to the dialogue system for comparison and evaluation. The speech based interface will be improved for robustness and domain specific processing. Methods to make a close interaction between the dialogue manager and the recognizer will be explored to illustrate the important tight coupling between modules needed in a complete dialogue system. Multimodal speech synthesis will be integrated and its implications for the man-machine interaction will be included in the educational part.

The last year will also be devoted to improve the flexibility of the system. The customization tool developed in 5.2 will be expanded with facilities to change domains. This allows students to customize the dialogue manager and to further visualise the dialogue and to analyse the dialogue tree. This allows studies of the dialogue system from a generic point of view.

5.2 An environment for robust, interactive and portable spoken dialogue systems

The aim of the Linköping part of the project is to develop and refine methods that support portable robust and efficient interaction in a SDS. The basis for the work is the LINLIN dialogue manager as it has been re-designed and implemented in the planning phase of the project.

The work is organized around three subtasks: (i) tools for customization, (ii) robust interpretation methods, and (iii) design of system utterances.

Tools for customization The first year's work will be focused on further detailed specification of the primitives needed by the dialogue manager and ensuring that the dialogue manager works properly with various combinations of such primitives, for instance for initiating a clarification request and integrating the answer in the current dialogue object or creating a new dialogue unit when applicable.

The development of a programming environment for specifying the behaviour of the dialogue manager will be a major effort the second year. The goal is to produce an environment where system developers can modify the behaviour of the dialogue manager for different domains and applications. This work will also have bearings on the development of the educational toolbox described in 5.1.

Robust interpretation This task is focused on robust methods for analyzing users' utterances into the specific information that are needed by the dialogue manager. The basic idea is to use as simple analysis methods as possible, such as keyword spotting or partial parsing, and combine this with information that can be gained from a prosodic analysis. Information from the dialogue manager and prosody will be used to guide the analysis, as a help to decide what kind of analysis that should be used, and also which parts of the utterance that should be analyzed. This work will be done in close cooperation with the work in Lund described in 5.4.

Currently we are doing a preliminary study on dialogues with written interaction to investigate how the information needed can be obtained by different analysis methods. We are also building a tool that makes it easy to combine different methods of interpretation for a given utterance or text. During the first year of the project this research will continue by analyzing more data, in particular spoken data, and refining the tools.

We expect that this kind of approach also would give higher portability and efficiency since it requires that only small parts of an utterance are given a thorough analysis and thus also requires smaller amount of grammatical and lexical information.

During the second year of the project we will use the results from the first year to build prototype systems for one or two domains. One aim here is to also investigate the limitations of the methods, and explain them in terms of complexity of the utterances and dialogue tasks.

Design of system utterances In this work we will study the role of utterance generation in promoting the efficiency of spoken interaction. The study has two parts, the first dealing with the content to be communicated in relation to dialogue state, and the second with the form of the utterance.

The spoken channel is narrow so the system must not overload the user (caller) with information. This is equally true whether the system responds to a request, takes an initiative or gives feed-back. In the first case, there can be problems if the database returns a number of valid alternatives. In the second and third cases, it is probably important that the user can discover the relevance of the system's utterance from the very first words. This work will be based on analysis of human-human dialogues and existing guide-lines for dialogue design (Bernsen *et al.* 1996, Krahmer *et al.*, 1997) and be performed in Linköping in close cooperation with the work on dialogue management. For efficiency it is also important to structure and pronounce the information in a maximally intelligible way. We are particularly interested in working out models for placement of focal accents. This work will be done in close cooperation with the groups in Lund and Göteborg and with Telia Research.

5.3 A dialogue move engine for Swedish

The aim is build a dialogue move engine which simulates changes of information states in dialogue participants engaging in Swedish dialogues. The targeted users of the dialogue move engine are developers who will be able to use it to simulate dialogic exchanges in Swedish under varying assumptions about dialogue moves and information states. In principle the dialogue move engine will be a module that can be incorporated in a system as well as an experimental aid to the specification of a system. We would try to design it in a way so that it could be incorporated in either the toolbox described in 5.1 or the environment described in 5.2. However, while the attempt will be instructive, we are not sure that this degree of integration can be achieved during the lifetime of the project. The work will consist of the following phases:

1. Examination of a corpus of Swedish dialogues for dialogue move phenomena. The spoken language corpus of dialogues concerning the task of model airplane building may provide the material. Alternatively, or in addition, we will use dialogues relating to travel made available to the project as a whole.
2. Construction of a computational model of dialogue moves, information states and information state transitions suitable for the data found in the corpus. The computational model would allow the use of varying assumptions about the nature of information states, e.g. whether they contain information about the agent's goals in addition to beliefs, whether they contain syntactic information about the utterances. Particular attention will be paid to focus phenomena and related prosodic phenomena. (It is expected that this work will be informed by theoretical results from the Riksbanken project "Information exchange in dialogue".)
3. Coding of a (small) part of the dialogue corpus for dialogue moves and information states according to the computational model in order to illustrate the applicability of the computational model.
4. Implementation of a dialogue move engine which characterizes information states and information state transitions. (It is expected that this work will be carried out in conjunction with the EC project TRINDI.)
5. Creation of a library of dialogue moves for the dialogue move engine which relates to a Swedish grammar (possibly the one incorporated in the SVENSK toolbox, coordinated by SICS under the aegis of the Swedish Language Technology Program) and which accounts for the dialogue moves analyzed in the target dialogues. As a preliminary to this we will investigate the possibility of implementing a prolog system which has a similar behaviour to the LINLIN system developed at Linköping which is to be incorporated in the systems described in 5.1 and 5.2. This will give us a more concrete view of what will be involved in attempting to integrate the dialogue move engine into these environments.

5.4 An investigation of the relationship between prosody, grammar and discourse structure

The structuring of information in spoken discourse is related to the interaction of a number of linguistic and discourse parameters. Prosodic cues such as pitch accentuation, pause-structure and duration interact with grammatical and discourse cues in signalling prominence (focus/ground) relations between discourse referents as well as segmentation into major and minor discourse units. The aim of this work will be to provide results from empirical investigations on these interrelationships that can be implemented in a spoken dialogue system in order to increase its robustness, both as regards speech understanding speech generation.

During the first year of the project, the collection, transcription and prosodic labelling of spontaneous dialogues will be undertaken: The database will include dialogues from two different domains: 1. travel information (recordings to be made at Lund Student Travel Service), and 2. student guidance counselling (recordings to be made at Lund University Student Service). On the basis of this data as well as dialogue data available from Telia Research and the other project groups, a number of empirical studies will be carried out. One investigation will involve an extension of the current study on cue word prosody in monologues to include their interactive function in dialogues, the main goal being to investigate prosody's potential to disambiguate cue words' sentential vs discourse functions. A comparison will be made between the monologue and dialogue data and the number of cue words studied will be extended. A further investigation directed towards the study of the correlation between prosodic structure and the focus/ground status of dialogue referents and attributes will be started. Finally, the more embracing relationship between prosody, syntax and semantics in the expression of prominence will also be examined together with the groups in 5.2 and 5.3.

During the second year of the project, further interaction with the groups in Linköping, Göteborg and the Telia Research group system designers will be undertaken in order to conceptualize how the dialogue manager can optimally incorporate the results of the empirical studies undertaken during the first year of the project. The design of the interfaces relating the prosodic structures with syntactic and semantic structures assumed in other components of the dialogue manager will also be sketched. For example, the question of how cue word prosody interacts with semantics/pragmatics and syntax and the whole issue of how a comprehensive 'cue word grammar' can be included in/interfaced with current modules in a dialogue management system will be addressed. The mechanisms needed for integrating prosodic information on focus/ground structure in the dialogue system will further be specified and some preliminary evaluation of the system's robustness as regards the updating of information on referents and attributes should be available. Some indication as to the portability of the system should also emerge on the basis of the analysis results comparing data from different domains.

5.5 A platform for multimodal dialogue corpora

The aim is to build a platform for a multimodal spoken language corpus, and to create such a corpus by adapting the existing Göteborg Spoken Language Corpus to this platform. The platform will consist of a digitized corpus and the following set of tools. More specifically, the following tasks will be undertaken.

1. A tool for digitizing and compressing audiovisual corpus data
2. A tool for synchronous alignment of audio/video and transcribed text
3. A multimodal corpus presentation tool, preferably accessible from the internet. What we have in mind is an audio/video presentation with simultaneous scrolling in a partiture-formatted transcription with highlighting of the phrase being spoken.
4. A CD containing a sample multimodal corpus with synchronized transcriptions.
5. A generic annotation and coding tool that will include simple corpus presentation using standard and/or partiture format, with optional use of an audio/video presentation extension. The goal is to provide support for as wide a range of annotations and codings as possible.

As a way to validate tool prototypes, actual coding will be carried out during the project period.

6. Tools for analyzing corpora and codings. We will use and compare existing generic statistic tools with some special purpose tools.

7. Tools for presentation of the corpus and analysis results. We have so far in cooperation with IVEE inc. tested visualization tools which allow overview in combination with interactive search. The goal is to produce a tool which combines perspicuous presentation with access to actual transcription, speech and gesture.

8. Coding and theory development (primarily OCM and ICM) using the tools which have been developed in the project.

5.6 Related research activities

Göteborg will be coordinating a European project TRINDI (Task Oriented and Instructional Dialogue) beginning January 1998. The focus of this project will be dialogues between humans and machines that enable the human to make choices in the performance of a certain task (task oriented instructional dialogues). It focuses on changes of information state during the course of a dialogue and models of information revision in such dialogues.

A project called Information Exchange in Dialogue funded by Riksbankens Jubileumsfond will also begin in Göteborg in 1998.

Stockholm and Linköping are part of the DISC project, a European project whose purpose is to develop a Dialogue Engineering Best Practice Model. Other sites include Centre National de la Recherche Scientifique (CNRS-LIMSI) (France), The Maersk Mc-Kinney Moller Institute for Production Technology (MIP) (Denmark), Institut für Maschinelle Sprachverarbeitung (IMS) (Germany), Vocalis Ltd (U.K.), Daimler-Benz AG (Germany) and ELSNET (The Netherlands).

Linköping and Göteborg are, together with major sites in Europe and USA, involved in a project on the development of general schemes for the analysis of discourse, the Discourse Resource Initiative (DRI). Other sites involved include among others Harvard, AT&T, Rochester, University of Pittsburgh, DFKI, Edinburgh, MITRE, University of Pennsylvania, MIT. In 1996 the DRI group met in Philadelphia and developed a first version of a general dialogue coding model for computational dialogue models. This proposal was evaluated and further refined at a second meeting in Dagstuhl 1997.

The Centre for Speech Technology (CTT) at KTH, Stockholm, has been created as a centre of excellence for co-operation between industry and academic research within the area of speech technology. It is planned as a long term activity, concentrating on fundamental research topics of industrial relevance. The research at CTT will initially include the following areas: Methods for automatic speech understanding. Language models for spoken language, Principles for speaker characterisation, Speech production for multimodal speech synthesis, Speech technology in interactive systems.

In Linköping a project has started the purpose of which is is to develop models for human computer interaction combing natural language dialogue with other modalities such as pointing, graphics and hypertext. The project is financed by KFB and uses the same domain as this project. The main issue addressed is how communicative modalities can be selected and combined to provide efficient interaction for different information needs? One aim is to develop a taxonomy of different kinds of situations involving information exchange and to characterize the different kinds of interaction which are suitable for different types of situations.

6 Management

The project represents an equal collaboration between all the institutions involved and will be run by a scientific board consisting of the principal investigators at all of the institutions. Arne Jönsson will oversee the coordination. In order to ensure adequate flow of information between the institutions and the various subprojects all researchers working on the project will meet twice a year in order to present results obtained and plan future work. It is planned that these meetings will lead to further collaboration and joint publications reporting the results of the project. For example, papers or workshops could be presented at the following international events:

- ICSLP (International Conference on Spoken Language Processing), Sydney, Australia, Nov. 30–Dec. 4, 1998
- Eurospeech 99, Budapest, Hungary
- ESCA Dialogue Workshop, spring 1999
- COLING-ACL '98, August 10-14, 1998, Université de Montréal, Montréal, Québec, Canada
- ESSLLI summer school

There will be a kickoff meeting at the beginning of the project where detailed workplans and objectives for the first phase of the project will be discussed to ensure maximum coherence and integration. Quality control within the project will be achieved by evaluating the achievements of each phase at one of the biannual meetings and planning the next phase of the project in terms of which objectives have or have not been achieved.

References

Ahrenberg, L., Dahlbäck, N. and Jönsson, A. (1990) Discourse Representation and Discourse Management for a Natural Language Dialogue System, in *Proceedings of the Second Nordic Conference on Text Comprehension in Man and Machine*.

Aist, G. & Mostow, J. A. (1997) A Time to Be Silent and a Time to Speak: Time-Sensitive Communication Actions in a Reading Tutor that Listens, in *Communicative Actions in Humans and Machines*, ed. by D. Traum, MIT, pp. 1–5.

Alexandersson, J., Maier, E. and Reithinger, N. (1995) A Robust and Efficient Three-Layered Dialogue Component for a Speech-to-Speech Translation System, in *Proceedings of the 7th Conference of the EACL*, pp. 188–193.

Allen, J., Schubert, L. K., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., and Traum, D. R. (1995) The TRAINS Project: a case study in building a conversational planning agent, in *Journal of Experimental and Theoretical Artificial Intelligence*, 7:7–48.

Allwood, J. (1995) An Activity Based Approach to Pragmatics. Gothenburg Papers in Theoretical Linguistics 76, Göteborg University, Department of Linguistics.

Allwood, J. (ed.) (1996) Talspråksfrekvenser. Gothenburg Papers in Theoretical Linguistics S 20. Göteborg University, Department of Linguistics.

Ariel, M. (1988) Referring and accessibility, *Journal of Linguistics*, Vol. 24, pp. 65–87.

van Benthem, Johan and Alice ter Meulen, eds. (1997) *Handbook of Logic and Language*, North Holland, Amsterdam and MIT Press, Cambridge, Mass.

Bernsen, N.O., Dybkjaer, H. and Dybkjaer, L. (1996) Cooperativity in human-machine and human-human spoken dialogue. *Discourse Processes* **21**:2, 1996:213-236.

Bertenstam, J. Blomberg, M., Carlson, R., Elenius, K, Granström, B., Gustafson, J., Hunnicutt, S., Högberg, J., Lindell, R., Neovius, L., de Serpa-Leitao, A., Nord, L. and Ström, N. (1995) The Waxholm system – a progress report, in *Proc. Spoken Dialogue Systems*, Vigsoe, Denmark.

Beskow, J. (1995) Rule-based Visual Speech Synthesis, in *Proceedings of EUROSPEECH95*, Madrid.

Betarte, Gustavo and Alvaro Tasistro (forthcoming) Extension of Martin-Löf's type theory with record types and subtyping. *Proceedings of the conference "25 Years of Constructive Type Theory"*, Oxford University Press.

Black, W., Bunt, H., Dols, F. Donzella, C., Ferrari,G., Gallagher, J., Haidan, R., Imlah, B., Jokinen, K., Lancel, J-M., Nivre, J., Sabah, G., and Wachtel, T. (1991) A Pragmatics Based Language Understanding System, in *Proceedings of the ESPRIT Conference 1991.*

Bruce, G., Filipsson, M., Frid, J., Granström, B., Gustafson, K., Horne, M., House, D., Lastow, B., & Touati, P. (1996) Developing the modelling of Swedish prosody in spontaneous dialogue, *Proceedings ICSLP 96* (Philadelphia), pp. 370–373.

Bruce, G., Granström, B., Gustafson, K. & House, D. (1993) Interaction of F0 and duration in the perception of prosodic phrasing in *Swedish. Nordic Prosody VI*, ed. B. Granström & L. Nord, 7–22. Stockholm: Almqvist & Wiksell.

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. & Anderson, A. (1997) The reliability of a dialogue structure coding scheme. *Computational Linguistics*, Vol. 23, pp. 13–31.

Carlson, R. (1996) The Dialog Component in the Waxholm System, in *Procceedings of the Twente Workshop on Language Technology (TWLT11) Dialogue Management in Natural Language Systems*, University of Twente, the Netherlands.

Carlson R and Hunnicutt S. (1996) Generic and domain-specific aspects of the Waxholm NLP and Dialog modules, in *Proceedings of ICSLP 96*, Philadelphia, USA.

Carlson R, Granström B. and Hunnicutt S (1990) Multilingual text-to-speech development and applications, in *Advances in speech, hearing and language processing*, ed. by Ainsworth W, London, JAI Press, pp. 269–296.

Colton, R. Cole, Novick D., Sutton S. (1996) A laboratory course for designing and testing spoken dialogue systems, in *Proceedings of ICASSP 96*, Atlanta, USA.

Cooper, Robin (1996) The Attitudes in Discourse Representation Theory and Situation Semantics, in *Logic, Language and Computation: Vol. 1*, ed. by Jerry Seligman and Dag Westerståhl, CSLI Publications, Stanford, pp. 137–150

Cooper, Robin and Jonathan Ginzburg (1996) A Compositional Situation Semantics for Attitude Reports, in *Logic, Language and Computation: Vol. 1*, ed. by Jerry Seligman and Dag Westerståhl, CSLI Publications, Stanford, pp. 151–166.

Cooper, Robin (forthcoming a) Using Situations to Reason about the Interpretation of Speech Events, *Current Issues in Computational Semantics*, ed. by Harry Bunt and Reinhard Muskens, Kluwer Academic Publishers, Dordrecht.

Cooper, Robin (forthcoming b) Information States, Attitudes and Dialogue, *Proceedings of the Second Tbilisi Symposium on Language, Logic and Computation, Tbilisi, September 15–20, 1997.*

Cutler, A., Dahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, Vol. 40, pp. 141–201.

Dekker, Paul (1993) *Transsentential Meditations: Ups and Downs in Dynamic Semantics*, PhD thesis, ILLC/Department of Philosophy, University of Amsterdam

Dekker, Paul and Robert van Rooy (1997) Intentional Identity and Information Exchange, paper presented at the Second Tbilisi Symposium on Language, Logic and Computation.

van Eijck, Jan and Hans Kamp (1997) Representing Discourse in Context, in van Benthem and ter Meulen (1997).

Engdahl, Elisabet & Enric Vallduví (1996) Information Packaging in HPSG. Working Papers in Cognitive Science 12, University of Edinburgh.

Ginzburg, Jonathan (1996) Dynamics and the Semantics of Dialogue, in Seligman and Westerståhl (1996).

Ginzburg, Jonathan (1996) Interrogatives: Questions, Facts and Dialogue, in *The Handbook of Contemporary Semantic Theory*, ed. by Shalom Lappin, Blackwell.

Givón, T. (ed.) (1983) *Topic continuity in discourse: a quantitative cross-language study*, Amsterdam: J. Benjamins.

Green, N., Kerpedjiev S., Carenini, G., Moore, J. & Roth, S. (1997) Media-Independent Communicative Actions in Integrated Text and Graphics Generation, in *Communicative Actions in Humans and Machines*, ed. by D. Traum, MIT.

Groenendijk, Jeroen and Martin Stokhof (1991) Dynamic Predicate Logic, *Linguistics and Philosophy*, Vol. 14, pp. 39–100.

Heim, Irene (1982) *The semantics of definite and indefinite noun phrases*, PhD thesis, University of Massachusetts, Amherst.

Hetherington L. and McCandless M. (1996) SAPPHIRE: An extensible speech analysis and recognition tool based on Tcl/Tk, in *Proc. ICASSP 96*, Atlanta, USA.

Hirschberg, J. (1993) Studies of intonation and discourse. *Proceedings ESCA Workshop on Prosody*, Working Papers (Dept. Ling. U. of Lund) No. 41, pp. 90–95.

Horne, M., Filipsson, M., Ljungqvist, M. & Lindström, A. (1993) Referent tracking in restricted texts using a lemmatized lexicon: implications for generation of prosody. *Proceedings Eurospeech '93 (Berlin)*, Vol. 3, pp. 2011–2014.

Horne, M. & Filipsson, M. (1996) Computational extraction of lexico-grammatical information for generation of Swedish intonation. *Progress in speech synthesis*, ed. J. P. H. van Santen, R. Sproat, J. Olive & J. Hirschberg, pp. 443–57. New York: Springer.

House, D., Bruce, G., Lacerda, F. & Lindblom, B. (1987) Automatic prosodic analysis for Swedish speech recognition, Working Papers (Dept. of Linguistics, U. of Lund), No. 31, 87–101.

Jönsson, Arne (1993) A Method for Development of Dialogue Managers for Natural Language Interfaces, in *Proceedings of the Eleventh National Conference of Artificial Intelligence*, Washington DC, pp. 190–195.

Jönsson, Arne (1995) Dialogue Actions for Natural Language Interfaces, in *Proceedings of IJCAI-95*, Montral, Canada.

Jönsson, Arne (1996) A Model for Dialogue Management for Human Computer Interaction, in *Proceedings of ISSD'96*, Philadelphia, pp. 69–72.

Jönsson, Arne (1997) A Model for Habitable and Efficient Dialogue Management for Natural Language Interaction. *Natural Language Engineering*, 3(1).

Kamp, Hans and Uwe Reyle (1993) *From Discourse to Logic*, Kluwer, Dordrecht.

Krahmer, E., Landsbergen, J., Pouteau, X. (1997) How to Obey the 7 Commandments for Spoken Dialogue ? In (eds.) Hirschberg, Kamm and Walker: *Interactive Spoken Dialog Systems: Bringing Speech and NLP Together in Real Applications*. Proceedings of a Workshop, 11–12 July 1997, Madrid, Spain.

Lappin, Shalom, ed., (1996) *The Handbook of Contemporary Semantic Theory*, Blackwell, Oxford.

Lieske, C., Bos, J., Emele, M., Gambäck, B. & Rupp, C. (1997) Giving prosody a meaning. *Proceedings Eurospeech97 (Rhodes)*, pp. 1431–4.

Munteanu, P., Caillaud, B., Serignat, J.-F. & Caelen-Haumont, G. (1997) Automatic word boundary demarcation based on prosody.*Proceedings Eurospeech97 (Rhodes)*, pp. 1487–90.

Poesio, Massimo and David Traum (forthcoming) Conversational Actions and Discourse Situations, *Computational Intelligence.*

Roberts, Craige (1996) Information Structure in Discourse, *Papers in Semantics*, Ohio State Working Papers, No. 49.

Sjölander K., and Gustafson J. (1997) An Integrated System for Teaching Spoken Dialogue Systems Technology, in *Proc. Eurospeech 97*, Rhodes, Greece.

Ström N. (1996) Continuous Speech Recognition in the WAXHOLM Dialogue System, in *STL QPSR* 4/1996, pp. 67–96.

Sutton S., de Veilliers J., Schalkwyk J., Fanty M., Novick D., and Cole, R. (1996) Technical specification of the CSLU toolkit. Tech. Report No. CSLU-013-96, Center for Spoken Language Understanding, Dept. of Computer Science and Engineering, Oregon Graduate Institue of Science and Technology, Portland, Oregon, USA.

Swerts, M. & Ostendorf, M. (1997) Prosodic and lexical indications of discourse structure in human-machine interactions. *Speech Communication* , Vol. 22, pp. 25–41.

Terken, J., Lathouwers, E., Theune, M., Veldhuijzen van Zanten, G. (1997) Prosodic structuring of system utterances in man-machine dialogues. *ProceedingsSPECOM (International Workshop "Speech and Computer") (Romania)*, pp. 71–76.

Vallduví, Enric & Elisabet Engdahl (1996) The linguistic realisation of information packaging. *Linguistics*, 459–519.

Walker, M. A., Litman, D. J., Kamm, C. A., and Abella, A. (1997) Evaluating Interactive Dialogue Systems: Extending Component Evaluation to Integrated System Evaluation. In (eds.) Hirschberg, Kamm and Walker: *Interactive Spoken Dialog Systems: Bringing Speech and NLP Together in Real Applications*. Proceedings of a Workshop, 11–12 July 1997, Madrid, Spain.

Warnke, V., Kompe, R., Niemann, H. & Nöth, E. (1997) Integrated dialog act segmentation and classification using prosodic features and language models, *Proceedings Eurospeech97*, pp. 207–210.