A Dialogue Manager for Natural Language Interfaces

Arne Jönsson*
Computer Science Department
Monash University, Clayton, VICTORIA 3168, AUSTRALIA
email: arnjo@bruce.cs.monash.edu.au.

Abstract

Users of natural language interfaces should conveniently be able to express the commands and queries that the background system can deal with, and the system should react quickly and accurately to all input. Among other things this means that the interface must be able to cope with connected dialogue. Furthermore, the design of natural language interfaces must take into consideration the characteristics of human-computer interaction instead of trying to mimic human interaction. This paper presents a dialogue manager for such natural language interfaces. To control the interaction it uses a dialogue grammar with information on the functional role of an utterance as conveyed in the linguistic structure. Focus structure is handled using dialogue objects recorded in a dialogue tree which can be accessed by the various modules for interpretation, generation and background system access. The Dialogue Manager is designed to facilitate customization to the sublanguage utilized in various applications. The paper also discusses the possibilities of generalizing the Dialogue Manager to other modalities.

Keywords: Natural Language Interface, Dialogue

1 Introduction

This paper presents an engineering approach to the design of dialogue managers for natural language interfaces. It is argued that taking human interaction as the norm when designing natural language interfaces will provide models that are inaccurate and computationally ineffective, based on the erroneous assumption that humans would like to communicate with computers in the same way as they communicate with other people. On the contrary, the language that humans utilize when they are interacting with a computer differs significantly from the language used between humans (cf. Dahlbäck & Jönsson, 1989; Dahlbäck, 1991b; Ogden, 1988; Guindon, Shuldberg, & Conner, 1987; Guindon, 1991).

Thus, it is not sufficient, nor advisable, to design interfaces that resemble human interaction behaviour (cf. Dahlbäck, 1991b, 1991a; Dahlbäck & Jönsson, 1992; Jönsson, 1993a). A dialogue system should be able to react sensibly to all input and handle correctly and efficiently those phenomena that occur, so that the user does not feel constrained or restricted when using the interface. On the other hand the interface should not waste effort on complex computations in order to handle irrelevant or rare phenomena. Instead, the properties of human-computer natural language interaction must be considered and the development of user-friendly natural language interfaces should focus on revealing the minimal model that can accommodate the sublanguages utilized in such interactions (Grishman & Kittredge, 1986). Furthermore, the system's capabilities and limitations should be evident to the user from experience.

The work presented in this paper is restricted to studying written human-computer interaction in natural language, and natural language interfaces for different applications which belong to the domain that Hayes and Reddy (1983) called simple service systems. Simple service systems "require in essence only that the customer or client identify certain entities to the person providing the service; these entities are parameters of the service, and once they are identified the service can be provided" (Hayes & Reddy, 1983, p. 252).

To illustrate some features of such human-

*During the academic year 1994/95 the author is on leave from the Department of Computer and Information Science, Linköping University, S-581 83 LINKÖPING, SWEDEN, arnjo@ida.liu.se.

U11: What is the shape of Ford Fiesta costing 26 800 crowns?
S12: Wait...
Cars cannot answer questions concerning the shape of car models.
U13: Is it rusty?
S14: Wait...
Checking...

Manufacturer	Model	Year	Rust
Ford	Fiesta	1982	2

U15: Does the Mercedes from 1982 have any rust damage?
S16: Wait...
Checking...

Manufacturer	Model	Year	Rust
Mercedes	200	1982	5

U17: How fast is a Mercedes 200?
S18: Wait...
Checking...

Manufacturer	Model	Year	Rust	Top Speed
Mercedes	200	1982	5	160

U19: How fast is the Fiesta?
S20: Wait...
Checking...

Manufacturer	Model	Year	Rust	Top Speed
Ford	Fiesta	1982	2	145

Figure 1: Example of human-computer dialogue. From a corpus of dialogues collected in Wizard-of-Oz-experiments.

computer interaction consider figure 1. In information retrieval systems a common user initiative is, of course, a request for information. Utterance U11: illustrates this. Unfortunately the system could not answer the question, instead, in utterance S12:, metaknowledge, i.e. knowledge about the capabilities of the system, is provided from the system. In U13: a new request for information is presented. This time the pronoun *it* replaces the rephrasing of the specification of the object, i.e. the Ford Fiesta costing 26 800 crowns. Using the flexibility inherent in language, especially various ways to be less verbose, is often utilized in human computer natural language interaction (cf. Dahlbäck & Jönsson, 1989; Guindon et al., 1987). In utterance U19: the user uses a definite description *the Fiesta* referring to the object discussed previously, and originally specified in utterance U11:. This occurs after a series of utterances discussing other objects, U15:-S18:.

Another important issue, also illustrated in the example, is the appearance of the system to the user. The system must provide the user with a correct model of its structure and capabilities. Thus, it is essential that the interface does not mislead the user to expect unrealistic "intellectual" competence by the background system, due to a sophisticated and varied linguistic output. Standardized simple "canned text" answers, such as S12:, can in this respect be preferred to a more flexible text generation system for a natural language interface that is connected to a simple database. The use of tables for output, illustrates another such feature which helps the user stay within the boundaries of the system's capabilities. Users find it convenient to have a table on the screen, containing a set of objects (such as cars), because they can then see not only the object considered most promising at the moment but also see that object in the context of others with similar properties (cf. Jönsson, 1993a; Ahrenberg, Jönsson, & Thurée, 1993).

2 Dialogue Management

The dialogue fragment in figure 1 shows some requirements of dialogue in natural language interfaces such as proper sequencing, meta-communication and focus tracking. A more detailed list is provided by for instance Carbonell and Hayes (1987, pp. 672-673). The requirements can be summarized under two headings:

- Dialogue structure. Handling the relationships between segments in the dialogue, including interaction style, e.g. mixed-initiative, managing the connection between segments, e.g. subdialogues and their relations, multi-sentential utterances, indirect speech acts, meta-communication and cooperative responses.
- Focus structure. The recording of entities mentioned in the discourse to allow a user to refer to them in the course of the interaction, including topical coherence and anaphora resolution where anaphora is taken to include certain sentence fragments and metalinguistic utterances, ellipsis, pronouns and definite descriptions.

Grosz and Sidner (1986) presented a general computational theory of discourse where they divided the problem of managing discourse into three parts: linguistic structure, attentional state and intentional state.

The need for a component which records the objects, properties and relations that are in the focus of attention, the attentional state, is not much debated, although the details of focusing need careful examination. However, the role that is given to the intentional state, i.e. the structure of the discourse purposes, and to the linguistic structure, i.e. the structure of the sequences of utterances in the discourse, provides two competing approaches to dialogue management. One approach uses the linguistic structure to identify the intentional state in terms of the user's goals and intentions. These are then modeled in plans describing the actions which may possibly be carried out in different situations (Cohen & Perrault, 1979; Allen & Perrault, 1980; Litman, 1985; Carberry, 1990). The other approach is to use only the information in the linguistic structure to model the dialogue expectations, i.e. utterances are interpreted on the basis of their functional relation to the previous interaction. The idea is that these constraints on what can be uttered allow us to write a grammar to manage the dialogue (Polanyi & Scha, 1984; Bilange, 1991; Jönsson, 1991). However, the identification of the users' goals is still an important issue, but "Though it is true that conversational moves frequently reflect speakers' goals, it is important to stress that these moves can be identified and interpreted without reference to a speaker's underlying intent for an utterance." (Reichman, 1985, p. 21).

If the goal is to mimic human language capabilities, identifying the users' intentions and utilizing the plan recognition approach might be necessary. However, for the task of managing the dialogue in a natural language interface, less sophisticated approaches, such as using a dialogue grammar, will do just as well.

3 The Dialogue Manager

This section presents a dialogue manager for a natural language interface for different simple service applications (cf. Ahrenberg, Jönsson, & Dahlbäck, 1990; Jönsson, 1991, 1993a). The Dialogue Manager can be viewed as a controller of resources for interpretation, background system access and generation. It receives input from the interpretation modules, inspects the result and accesses the background system with information conveyed in the user input. Eventually an answer is returned from the database access module and the Dialogue Manager then calls the generation modules to generate an answer to the user. If clarification is needed from any of the resources, it is dealt with by the Dialogue Manager.

The Dialogue Manager was initially designed from an analysis of a corpus of 21 dialogues, collected in Wizard-of-Oz-experiments (Dahlbäck, Jönsson, & Ahrenberg, 1993) using five different background systems. It has then been customized for three other applications, one of which is implemented, using a set of 30 new dialogues, and also verified for these applications (Jönsson, 1993b). The verification shows that the proposed principles for focus structure management work well and that the dialogue structure can be accurately handled using context free grammars and information on objects and properties (Jönsson, 1993a), as presented below.

3.1 The dialogue objects

The Dialogue Manager uses information from dialogue objects which model the dialogue segments and moves and information associated with them. The dialogue objects represent the constituents of the dialogue and the Dialogue Manager records instances of dialogue objects in a dialogue tree as the interaction proceeds.

The dialogue object descriptions are domain dependent and can be modified for each new application by manually specifying which dialogue objects parameters to use and the values they can take. The parameters reflect the information needed by the Dialogue Manager and the various processes accessing information stored in the dialogue tree. There is one set of parameters for specifying the initiator, responder, context etc. and another set of parameters specifying focal and non-focal content, as presented below.

3.2 Dialogue structure

The model for dialogue management utilized by the Dialogue Manager assumes that a dialogue is divided into three main classes on the basis of structural complexity. There is one class corresponding to the size of a dialogue (D), another class corresponding to the size of a discourse segment and a third class corresponding to the size of a single speech act, or dialogue move. Thus, a dialogue is structured in terms of discourse segments, and a discourse segment in terms of moves and embedded segments. Utterances are not analyzed as dialogue objects, but as linguistic objects which function as vehicles of one or more moves. An initiative-response (IR) structure is assumed (cf. adjacency-pairs Schegloff & Sacks, 1973) where an initiative opens a segment by introducing a new goal and the response closes the segment (Dahlbäck, 1991b).

There are various other proposals as to the number of categories needed. They differ mainly on the modeling of complex units that consist of sequences of discourse segments, but do not comprise the whole dialogue. For instance, LOKI (Wachtel, 1986) and SUNDIAL (Bilange, 1991) use four. In LOKI the levels are: conversation, dialogue, exchange (corresponding to the notion of IR-unit used in this paper) and move. Bilange calls the levels Transaction level, Exchange level, Intervention level and Dialogue Acts. The feature characterizing the intermediate level (i.e. the Dialogue and Exchange levels respectively in Wachtel's and Bilange's models) is that of having a common topic, i.e. an object whose properties are discussed over a sequence of exchanges. However, a sequence of segments may hang together in a number of different ways; e.g. by being about one object for which different properties are at issue. But it may also be the other way around, so that the same property is topical, while different objects are talked about (cf. Ahrenberg et al., 1990).

To specify the functional role of a move we use the parameters Type and Topic. Type corresponds to the illocutionary type of the move. In simple service systems two sub-goals can be identified (Hayes & Reddy, 1983, p. 266): 1) "specify a parameter to the system" and 2) "obtain the specification of a parameter". Initiatives are categorized accordingly as being of two different types: 1) update, U, where users provide information to the system and 2) question, Q, where users obtain information from the system. Responses are categorized as answer, A, for database answers from the system or answers to clarification requests. Other Type categories are Greeting, Farewell and Discourse Continuation (DC) (Dahlbäck, 1991b) the latter being used for utterances from the system whose purpose is to keep the conversation going.

Topic describes which knowledge source to consult. In information retrieval applications three different topics are used: the database for solving a task (T), acquiring information about the database, system-related, (S) or, finally, the ongoing dialogue (D). If the background system allows update, e.g. ordering of a specified item, a fourth category (O) is needed to account for such utterances.

3.3 Focus structure

In information retrieval dialogues the most common user initiative is a request for information from the database. Users specify a database object, or a set of objects, and ask for concept information, e.g. the value of a property of that object or set of objects. Two[1] focal content parameters, termed Objects and Properties, account for the information structure of a move (query), where Objects denote a set of primary referents, and Properties a complex predicate ascribed to this set (Ahrenberg, 1987). These are focal parameters in the sense that they can be in focus over a sequence of IR-units.

The values to these parameters depend on the background system, and the natural language interface needs to be customized to account for the demands from each application (Jönsson, 1993b). For the application illustrated in figure 1 a relational database is used and the primary referents are cars modeled in the Objects parameter by the sub-parameters (Manufacturer, Model, Year), as seen in the tables in the figure. The Properties parameter models the domain concept in a sub-

[1] For some applications a third parameter, Secondary Objects, is utilized to constrain the database search (Jönsson, 1993a).

parameter termed Aspect which can be specified in another sub-parameter termed Value. For instance, utterance U17 in figure 1, provides Aspect information on the domain concept, *speed* which is specified by the database manager to *160*, i.e. the Value of the Aspect.

The focus parameters contains a description of the items and properties currently in discourse focus and resembles the backward-looking center of centering theories (Grosz, Joshi, & Weinstein, 1983; Joshi & Weinstein, 1981). However, it differs in important respects. First, there is no center in the sense that a center typically is pronominalized. There is also a focus parameter pertaining to properties, in addition to those dealing with objects. Furthermore, the focus parameters are properties of IR units, not moves (cf. Zancanaro, Stock, & Strapparava, 1993). Entities discussed in a segment replace previous values of Objects or Properties. However, such entities can be found in the dialogue tree which accounts for global focus. The focus parameters do not record every possible entity mentioned in the discourse, but only those that can be accessed in the background system. Consequently it is not possible to use anaphoric expressions to refer to entities not found in the background system. This is not a severe limitation. First, users do not utilize this possibility (Jönsson, 1993a), and second, the system cannot provide more information on such entities anyhow. This does not exclude the use of synonyms. It is important for a natural language interface to allow for a generous number of synonyms (Good, Whiteside, Wixon, & Jones, 1984).

The focus structure also needs principles for how the values of the focal parameters Objects and Properties are specified from information in the user initiative and the answer provided from the database. A move can fully specify both Objects and Properties. However, many utterances provide only partial specification of the focal parameters; context information is needed to fully specify them. Two principles account for focus maintenance. A general heuristic principle is that everything not changed in an utterance is copied from one IR-node in the dialogue tree to the newly created IR-node (cf. Seneff, 1992). Another principle is that the value for Objects will be updated with the value from the module accessing the database, if provided.

The details of the copying principles need to be customized for each application to meet the demands of the background system and the focal content parameters (Jönsson, 1993b). An example of such customization for the application on second-hand cars is illustrated in figure 1. Consider utterance U17. The user utterance only specify the Aspect *speed*, however, the system also provides information on the previous Aspect *rust* as seen from the response S18. The modified principle states that if the value of the Objects parameter remains the same (or is a subset of the previous value), the value of the Properties parameter will be the conjunction of the previous value and the new values provided in the new move. This approach is appropriate when information is presented in tables allowing additional information to be presented conveniently (Ahrenberg et al., 1993).

3.4 An Example

During the course of interaction the dialogue tree is built up from instances of dialogue objects with each dialogue object responsible for its own correctness (Jönsson, 1991). Initially, the root-node, the D-node, creates an instance of an IR-node and inserts it into the tree, creating links between the IR-node and the D-node. The IR-node creates an instance of a user move which interprets the first move. Upon receiving an instantiated structure from the interpretation modules, the Dialogue Manager determines the Type and Topic of the utterance based on information in the current active node, the IR-unit which initiated the move, and the information given in the current move.

Consider the utterance *U15: Does the Mercedes from 1982 have any rust damage?* in the dialogue fragment in figure 1. This will be interpreted as a Q_T initiative with the value Mercedes 1982 assigned to the Objects parameter and the value rust to the Properties parameter. The Dialogue Manager sends these values to the database manager. When the database manager has returned an answer to the request the Dialogue Manager will update Objects and Properties with the extension of the request intensionally specified in the move. This allows the user to refer to the objects and properties presented on the screen (cf. Cohen, Perrault, & Allen, 1982).

Next a new IR-unit is created which receives information on local focus to the parameters Objects and Properties from the previous IR-unit. New information from the user initiative then updates the parameters of the current IR-unit as a result of instantiating the output from the parser with information on salient objects and properties. This is illustrated in the utterance *U13: Is it rusty?*. When processing this utterance, the interpreting modules inspecting the dialogue tree looking for a referent to the pronoun *it*, will find the value Ford Fiesta and deliver an instantiated structure to the Dialogue Manager.

Thus, the system behaves as a user-directed interface, i.e. the user initiates the IR-units and the system only initiates an IR-unit for clarification requests. The action to be carried out in regular task related questions depends on how the information in the user initiative and the answer from the database system specify the values to the focal parameters Objects and Properties.

There are also user initiatives which do not depend on the values of Objects and Properties, such as system-related questions, i.e. users' requests for information about the system. These are recognized on the grounds of linguistic information provided by the syntactic/semantic analyzer (Ahrenberg, 1988).

4 Multi-modal systems

The Dialogue Manager presented in this paper is restricted to written human-computer interaction in natural language. However, when communicating with a natural language interface, a user should not be limited to typed keyboard input and screen output. The possibilities of using different modalities must be addressed to ensure efficient system-user interaction. Examples of systems which use a variety of modalities for both interpretation and generation include AlFresco (Stock, 1991), XTRA (Wahlster, 1991), Voyager (Zue, 1994) and CUBRICON (Neal & Shapiro, 1991).

The main difference between these multi-modal systems and conventional natural language interfaces is their ability to use various modalities for interpretation and generation. They all use a combination of input and output modalities such as speech, graphics, pointing and video output. Thus, more advanced interpretation and generation modules are required and principles for determining which media to utilize are needed (Arens, Hovy, & Vossers, 1993). However, the dialogue and focus structures need not necessarily be more complicated. For instance, Voyager successfully utilizes the approach presented here of copying the focus parameters from one segment to the other (Seneff, 1992). Bilange (1991) presents a dialogue grammar for dialogue management for telephone interaction with an information retrieval application. The Vehicle Navigation System (Novick & Sutton, 1994) is a speech system where users can receive driving directions a step at a time by cellular telephone. This type of interaction requires a system capable of recognizing various acknowledgment acts and determining their applicability. A grammar based model for this is defined where exchanges are interpreted in terms of speech acts.

Sitter and Stein (1992) present a theory for dialogue management for information-seeking dialogues. The model assumes that conversation is based on possible sequences of dialogue acts which are modeled in a transition network. In Stein and Thiel (1993) the model is extended to handle multi modal interaction as utilized in the MERIT system (Stein, Thiel, & Tißen, 1992).

IDAS (Reiter & Mellish, 1993) is a multi-modal generation system which mixes proper knowledge base structures with canned text fragments. IDAS does not utilize plan based reasoning as the cost is too high. The cost could be reduced using control heuristics, however this also removes the system's abilities to respond appropriately in unusual situations. This motivates the use of a fixed set of rules to complement the canned text generations. This only works well if the number of tasks to perform is small and fairly predictable. As pointed out in Reiter and Mellish (1993) different applications demand different techniques.

Interfaces to task oriented applications, on the other hand, might require more sophisticated reasoning involving the user's task and goals in order to be helpful (cf. Burger & Marshall, 1993). However, do they need a sophisticated model of the user's intentions or will a hierarchical structure of plans based on the various tasks possible to carry out in the domain do just as well (cf. Wahlster, André, Finkler, Profitlich, & Th, 1993)? Combining telephone interactions with a complex assembly task (Oviatt & Cohen, 1991) involves even more complicated communication. We are faced with a different communicative situation where it is easier to be engaged in a more human-like interaction.

5 Summary

Research on dialogue management for natural language interfaces should focus on models that correctly and efficiently handle those phenomena that actually occur in typed human-computer interaction without having the user feel constrained or restricted when using the interface. For a large number of natural language interface applications this can be achieved using a straightforward solution.

The interaction can be interpreted from the information conveyed in the speech act directly; no reasoning about users' intentions or goals is necessary. Speech act information is assembled into Initiative-Response units which form the basis for interpreting the segment structure. A simple context free grammar can model the interaction and the rules are selected based on information about properties of objects describing the information provided by the system. Referring expressions are handled by copying information from the previous segment to the current segment which in turn is updated with information from the background system.

These principles for dialogue management are applicable in written natural language interaction to simple service systems, but there are indications that they also apply to multi-modal communication for simple service systems.

Acknowledgments

This work results from a project on Dynamic Natural-Language Understanding supported by The Swedish Council of Research in the Humanities and Social Sciences (HSFR) and The Swedish National Board for Industrial and Technical Development (NUTEK) in the joint Research Program for Language Technology. The work has been carried out with the members of the Natural Language Processing Laboratory at Linköping University, Sweden, and I am especially indebted to Lars Ahrenberg, Nils Dahlbäck and Åke Thurée.

References

Ahrenberg, L. (1987). *Interrogative Structures of Swedish. Aspects of the Relation between grammar and speech acts.* Ph.D. thesis, Uppsala University.

Ahrenberg, L. (1988). Functional constraints in knowledge-based natural language understanding. In *Proceedings of the 12th International Conference on Computational Linguistics, Budapest*, pp. 13–18.

Ahrenberg, L., Jönsson, A., & Dahlbäck, N. (1990). Discourse representation and discourse management for natural language interfaces. In *Proceedings of the Second Nordic Conference on Text Comprehension in Man and Machine, Täby, Sweden.*

Ahrenberg, L., Jönsson, A., & Thurée, Å. (1993). Customizing interaction for natural language interfaces. In *Workshop on Pragmatics in Dialogue, The XIV:th Scandinavian Conference of Linguistics and the VIII:th Conference of Nordic and General Linguistics, Göteborg, Sweden.*

Allen, J. F. & Perrault, C. R. (1980). Analysing intention in utterances. *Artificial Intelligence, 15*, 143–178.

Arens, Y., Hovy, E., & Vossers, M. (1993). On the knowledge underlying multimedia presentations. In Maybury, M. T. (Ed.), *Intelligent Multimedia Interfaces*, pp. 280–306. MITPress.

Bilange, E. (1991). A task independent oral dialogue model. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics, Berlin.*

Burger, J. D. & Marshall, R. J. . (1993). The application of natural language models to intelligent multimedia. In Maybury, M. T. (Ed.), *Intelligent Multimedia Interfaces*, pp. 174 – 196. MITPress.

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

Carbonell, J. G. & Hayes, P. J. (1987). Natural language understanding. In Shapiro, S. C. (Ed.), *Encyclopedia of Artificial Intelligence*, pp. 660–677. Wiley-Interscience.

Cohen, P. R. & Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science, 3*, 177–212.

Cohen, P. R., Perrault, C. R., & Allen, J. F. (1982). Beyond question answering. In Lehnert, W. G. & Ringle, M. H. (Eds.), *Strategies for Natural Language Processing.* Lawrence Erlbaum Associates.

Dahlbäck, N. & Jönsson, A. (1989). Empirical studies of discourse representations for natural language interfaces. In *Proceedings from the Fourth Conference of the European Chapter of the association for Computational Linguistics, Manchester.*

Dahlbäck, N. & Jönsson, A. (1992). An empirically based computationally tractable dialogue model. In *Proceedings of the Fourteenth Annual Meeting of The Cognitive Science Society, Bloomington, Indiana.*

Dahlbäck, N. (1991a). Empirical analysis of a discourse model for natural language interfaces. In *Proceedings of the Thirteenth Annual Meeting of The Cognitive Science Society, Chicago, Illinois*, pp. 1–6.

Dahlbäck, N. (1991b). *Representations of Discourse, Cognitive and Computational Aspects.* Ph.D. thesis, Linköping University.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies – why and how. *Knowledge-Based Systems*, *6*(4), 258–266.

Good, M. D., Whiteside, J. A., Wixon, D. R., & Jones, S. J. (1984). Building a user-derived interface. *Communications of the ACM*, *27*(10), 1032–1043.

Grishman, R. & Kittredge, R. I. (1986). *Analysing language in restricted domains.* Lawrence Erlbaum.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the ACL, Cambridge, MA*, pp. 44–50.

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

Guindon, R. (1991). Users request help from advisory systems with simple and restricted language: effects of real-time constraints and limited shared context. *Human-Computer Interaction*, *6*, 47–75.

Guindon, R., Shuldberg, K., & Conner, J. (1987). Grammatical and ungrammatical structures in user-adviser dialogues: evidence for sufficiency of restricted languages in natural language interfaces to advisory systems. In *Proceedings of the 25th Annual Meeting of the ACL, Stanford*, pp. 41–44.

Hayes, P. J. & Reddy, D. R. (1983). Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies*, *19*, 231–284.

Jönsson, A. (1991). A dialogue manager using initiative-response units and distributed control. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics, Berlin.*

Jönsson, A. (1993a). *Dialogue Management for Natural Language Interfaces – An Empirical Approach.* Ph.D. thesis, Linköping University.

Jönsson, A. (1993b). 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.

Joshi, A. K. & Weinstein, S. (1981). Control of inference: role of some aspects of discourse structure - centering. In *Proceedings of the International Joint Conference of Artificial Intelligence, Vancouver, BC.*, pp. 385–387.

Litman, D. J. (1985). *Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues.* Ph.D. thesis, University of Rochester.

Neal, J. G. & Shapiro, S. C. (1991). Intelligent multi-media interface technology. In Sullivan, J. W. & Tyler, S. W. (Eds.), *Intelligent User Interfaces.* ACM Press, Addison-Wesley.

Novick, D. G. & Sutton, S. (1994). An empirical model of acknowledgement for spoken-language systems. In *Proceedings of the 32nd Conference of the Association for Computational Linguistics, New Mexico.*

Ogden, W. C. (1988). Using natural language interfaces. In Helander, M. (Ed.), *Handbook of Human-Computer Interaction.* Elsevier Science Publishers B. V. (North Holland).

Oviatt, S. L. & Cohen, P. R. (1991). The contributing influence of speech and interaction on human discourse patterns. In Sullivan, J. W. & Tyler, S. W. (Eds.), *Intelligent User Interfaces.* ACM Press, Addison-Wesley.

Polanyi, L. & Scha, R. (1984). A syntactic approach to discourse semantics. In *Proceedings of the 10th International Conference on Computational Linguistics, Stanford.*

Reichman, R. (1985). *Getting Computers to Talk Like You and Me.* MIT Press, Cambridge, MA.

Reiter, E. & Mellish, C. (1993). Optimizing the costs and benefits of natural language generation. In *Proceedings of the International Joint Conference of Artificial Intelligence, Chambery, France*, pp. 1164– 1169.

Schegloff, E. A. & Sacks, H. (1973). Opening up closings. *Semiotica*, *7*, 289–327.

Seneff, S. (1992). A relaxation method for understanding spontaneous speech utterances. In *Paper presented at the Fifth DARPA Workshop on Speech and Natural Language*.

Sitter, S. & Stein, A. (1992). Modeling the illocutionary aspects of information-seeking dialogues. *Information Processing & Management*, *28*(2), 165–180.

Stein, A. & Thiel, U. (1993). A conversational model of multimodal interaction in information systems. In *Proceedings of the Eleventh National Conference of Artificial Intelligence, Washington DC*, pp. 283 – 288.

Stein, A., Thiel, U., & Tißen, A. (1992). Knowledge based control of visual dialogues in information systems. In *Proceedings of the 1st International Workshop on Advanced Visual Interfaces, Rome, Italy*.

Stock, O. (1991). Natural language exploration of an information space: the alfresco interactive system. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence, Sydney, Australia*, pp. 972–978.

Wachtel, T. (1986). Pragmatic sensitivity in nl interfaces and the structure of conversation. In *Proceedings of the 11th International Conference of Computational Linguistics, University of Bonn*, pp. 35–42.

Wahlster, W. (1991). User and discourse models for multimodal communication. In Sullivan, J. W. & Tyler, S. W. (Eds.), *Intelligent User Interfaces*. ACM Press, Addison-Wesley.

Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., & Rist, T. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, *63*, 387 – 427.

Zancanaro, M., Stock, O., & Strapparava, C. (1993). Dialogue cohesion sharing and adjusting in an enhanced multimodal environment. In *Proceedings of the International Joint Conference of Artificial Intelligence, Chambery, France*, pp. 1230– 1236.

Zue, V. W. (1994). Toward systems that understand spoken language. *IEEE Expert*, *9*, 51–59.