Acquiring user models for natural language dialogue systems through Wizard-of-Oz techniques[1]

Paul Mc Kevitt

Computing Research Laboratory
Dept. 3CRL, Box 30001
New Mexico State University
Las Cruces, NM 88003-0001, USA.

E-mail: INTERNET: paul@nmsu.edu
Fax: (+1 505)-646-6218
Phone: (+1 505)-646-5109/5466

ABSTRACT

One of the most difficult problems in natural language interface development is that of dialogue processing. Recent empirical research has shown that over one quarter of unknown word errors in the evaluation of a natural language interface were due to dialogue phenomena. A Wizard-of-Oz study has been conducted for written/interactive natural language dialogues in the UNIX help domain showing the intricate relationships between the various types of queries asked by subjects. A speech act categorization and sequencing graph have been developed. Previous research in computational linguistics has argued for speech act analyses. However, such research has concentrated on the use of speech acts for analyzing user's goals, plans and intentions in their own end. From the analysis it is shown that speech act sequencing bears heavily on user modeling issues and may also have implications for dialogue structure. Speech act sequencing can help in the further development of dialogue and user models for natural language interfaces.

1

1 This research is funded by U S *WEST* Advanced Technologies, Denver, Colorado, under their Sponsored Research Program.

1. Introduction

One of the most difficult problems in developing natural language interfaces is to build an interface which will have a good dialogue interaction with its user. A dialogue component is very important in any natural language interface. Another problem in developing such interfaces is that they must have a good model of the user (see Kobsa and Wahlster 1988, Schuster et al. 1988). It has been shown by Whittaker and Stenton (1989) that over one quarter of unknown word errors in the evaluation of a natural language interface were due to dialogue phenomena.

It is our goal to build a dialogue modeling component for a computer program which answers queries about computer operating systems. The program is called OSCON and is described further in Guthrie et al. (1989), Mc Kevitt (1988), Mc Kevitt and Wilks (1987), and Mc Kevitt and Pan (1989). However, in obtaining this goal it was considered necessary to conduct an experiment to find out what sort of queries subjects really asked and what sort of relationships existed between these queries. A Wizard-of-Oz[2] study was conducted for 14 subjects and the results analyzed.

2. Background research

The implications of structural constraints for the analysis of intention and speech acts (see Austin 1975, Bach and Harnish 1979, Searle 1969, and Stampe 1975) has been well studied in the field of natural language processing (see Allen and Hinkelman 1989, Cohen et al. 1982, Hinkelman and Allen 1989, and Reichman 1986). Also, previous natural language dialogue research has concentrated on the use of linguistic items to identify dialogue structure. Examples are the use of conversational cues and phrases to mark changes in topic or attentional state (see Grosz and Sidner 1986), the use of conversational cues or phrases to mark relations between structures (see Reichman 1986), the use of intonation to help in segmenting dialogues (see Hirschberg and Pierrehumbert 1986), and the use of referring expressions and grammatical structure to indirectly identify dialogue structure (see Brennan et al. 1987, Grosz and Sidner 1986, Guindon et al. 1986, Reichman 1978, Reichman-Adar 1984, and Reichman 1986). In fact, recent empirical results by Whittaker and Stenton (1988) not only show that the recognition of intention is useful for indicating shifts in control in dialogue, but that structural conversational cues are not reliable, or good enough on their own, for predicting shifts in control.

It is argued here that the implications of intention, or speech acts, for other dialogue phenomena has been, on the whole, left uncovered. For example speech acts have implications for user modeling and dialogue structure. Recent research has been moving in the direction of the implications of intention for other dialogue phenomena (see Carberry 1989).

3. Wizard-of-Oz

The Wizard-of-Oz technique is useful for obtaining data on human-computer interactions and for providing data for designing user models. The technique is useful as (1) it models human-computer typed interaction, (2) there is no failure of the program answering typed utterances because there is no program, and (3) the setting is controlled as the subject has a set number of tasks to accomplish. The Wizard-of-Oz setting has been strongly argued for in the design and evaluation of natural language interfaces (see Guindon 1988).

There have been a number of attempts at building computer programs which help users in the domain of computer operating systems. These programs vary from menu-based systems (see Shneiderman 1987) to natural language systems (see Guthrie, Mc Kevitt and Wilks 1989, Hecking et al. 1988, Hegner and Douglass 1984, Mc Kevitt 1986, Mc Kevitt and Wilks 1987, Wilensky et al. 1984, 1986, 1988). Although there have been many papers discussing problems and solutions to building natural language consultant systems for operating systems there have been

[2] The Wizard-of-Oz technique is one where subjects interact with a computer through typed dialogue at a monitor and are told that they are conversing with the computer. Subject's utterances are sent to another monitor where a "Wizard" sends back a reply to the subject monitor.

few empirical studies to find out exactly what sorts of queries users ask. Some protocol studies were done at Los Alamos but these studies were done with a user actually using a natural language consultant system called UCC (Hegner and Douglass 1984, Douglass and Hegner 1982) rather than with the Wizard-of-Oz technique. As the UCC system did not always answer user queries these protocols were not useful for some of the analyses we wanted to carry out. A Wizard-of-Oz experiment was conducted for the UNIX operating system and some interesting results are described in Chin (1984). We argue for more elaborate analyses than Chin's.

There have been many studies conducted on user-adviser dialogues using the Wizard-of-Oz experiment for domains such as statistical packages (Guindon 1988, Guindon et al. 1986, Slator et al. 1986), travel agent advice (see Brunner et al. 1989a, 1989b) and information retrieval (Walker and Whittaker 1989, Whittaker and Stenton 1989). The objective of many of these studies is to investigate written/interactive dialogue and use those investigations to help in building better and more robust natural language interfaces.

4. A Wizard-of-Oz experiment

In order to evaluate the use of Wizard-of-Oz studies an experiment was conducted. The experiment took about 1 hour per subject to complete and each subject had 17 tasks to do. Instructions were given to each subject before the experiment commenced. A brief description of the experiment is given here. A more detailed description is found in Mc Kevitt and Ogden (1989).

The selected subjects were read information about the experiment explaining the tasks and conduction of the experiment. The subjects were told that the wizard would be monitoring their input and they could call for help at any time. The subject sat at a Sun-3 monitor in a large laboratory. The Wizard was situated at another Sun-3 monitor with his back to the subject in the same laboratory about 8 Metres away. The subject could not see what was on the Wizard's screen. An effort was made to help the subjects feel at home with the experiment by telling them that it was not them that we were testing.

There were three windows displayed on the subject's monitor. One window in the upper left of the monitor contained a reduced version of the UNIX[3] operating system. This reduced version did not contain all operating system commands as only a limited set of commands were necessary for the experiment. The subject typed all commands to UNIX in this window.

Another window to the right contained three frames: (1) task frame, (2) subject question frame, and (3) wizard answer frame. The task frame was on top, the subject query frame in the middle, and the wizard answer frame below that. The task frame displayed 17 tasks in turn, each one randomly selected from a computer file of 17. The subject selected tasks by clicking in a box marked TASK. A small window in the task frame showed the number of tasks to be completed. The question frame allowed the subject to type queries and then send them to the computer by clicking a box marked QUESTION. There was also a box marked CLEAR which the subject used to clear the question buffer before asking a new question. The answer frame marked ANSWER displayed the Wizard's answer. This frame was not cleared but scrolled up as the answers filled up the frame.

Below the answer frame there were three boxes. One box marked RESET was used for resetting the wizard program for a new user. Another box marked QUIT was used for quitting the program. The third box marked Printscreen allowed one to print the screen on the Laserwriter at any given moment during the experiment.

The wizard's screen layout was virtually the same as the subject's. A small window showed the number of tasks to be completed but the Wizard did not see the tasks themselves.

At the commencement of the experiment the subject was instructed to try a sample question by clicking on TASK. The sample question was the same for each user and involved printing a

[3] UNIX is a trademark of AT&T Bell Laboratories.

file on the printer. The computer mapped a canned answer about printing files on the printer to the ANSWER window. The subject was also instructed to try the answer given by the computer, in the UNIX window. >From then on the subject was on his/her own for the rest of the experiment unless of course he/she called the wizard for help. The last task the subject had to do was log out of the system. Each time the subject or the wizard sent a message there was a beep at the wizard's or user's monitors respectively. Also, when the subject typed a command to the reduced UNIX shell or went on to a new task there was a beep on the wizard's monitor. Information from the screens was stored in log files. The subject and wizard did not see keystrokes or mistyping as only messages actually sent by the subject or wizard were shown on the screen.

The tasks that subjects had to execute were simple operating system tasks. Some of the tasks were specific to UNIX, others possible in any operating system. The tasks involved simple operating basics such as creating, displaying, printing, copying, moving, and removing files and directories. Some tasks involved displaying information about the system too. During each subject session data was collected and placed in a log file. A number of codes were used to mark each item of data. Only sent messages were recorded in the log files and keystrokes were not recorded.

Each subject was asked to fill out a questionnaire which recorded data about the subject. The name, age, sex, and major of users were recorded. Also, the computer experience, operating systems experience, and UNIX experience of subjects was recorded. Levels of expertise of users are shown in Appendix C.

5. Analysis of results

The results of the experiment can be analyzed for many different phenomena. >From the questionnaire that was given to the subjects the following information was obtained. Of the 14 subjects 11 were female and 3 were male. Their majors at the University included the following: Business Computer Systems, Community Health/ Nursing, Accounting, Dental Hygiene, Business Management, Psychology, Wildlife Science, Computer Science, and Marketing. Of the 14 subjects 3 suspected that the "Wizard" could be answering their questions though none of them were sure of this. The other 11 did not suspect this at all. One subject said that it was the time lag in answering that caused her suspicions. The other 2 contemplated the possibility that the answers were being given by the Wizard.

Three of the subjects reported that they had no computer experience the rest reporting experience from less than 3 months to 1-2 years. The same two subjects reported having not used any operating system while the other subjects had used some operating system(s), mainly MS-DOS[4] and PC-DOS[5]. One subject had used a number of different operating systems. Most of the subjects had not used UNIX although three subjects had 1-2 years of UNIX. Therefore, the subjects were on the whole novice UNIX users.

6. Speech act analysis

In doing a speech act analysis of the 14 subject dialogues the set of speech acts shown in Appendix A was obtained. A sequence analysis of these results was conducted and the sequencing graph in Appendix B was developed.

6.1. Speech act categorization

The categorization is my own and may be incorrect and need some modification, yet it gives an initial guide. However, it is important to note that the categorization is derived from that data a postiori. There are seven speech act types: (1) request-for-direction, (2) request-for-guidance, (3) request-for-explanation, (4) request-for-information, (5) request-for-conformation,

[4] MS-DOS is a trademark of Microsoft Corporation.

[5] PC-DOS is a trademark of International Business Machines Corporation.

(6) request-for-form, and (7) request-for-elaboration. The previous set are listed in order of frequency of occurrence in the dialogues.

request-for-direction describes those requests where subjects inquire about means for doing some operation. For example in this domain subjects often asked about operations such as "printing" or "deleting".

request-for-guidance indicates requests where subjects ask about some guidance on what to do next, or what to do when they are stuck. Examples are queries asking for help.

request-for-explanation indicates requests where subjects ask for explanations of some information such as a command or some phenomena that has happened.

request-for-information is a request for information about some topic. Examples are requests for information about the operating system and its status.

request-for-confirmation is a request by the subject for confirmation of some belief that he/she holds. Examples are confirmations of some command form or some goal or plan.

request-for-form includes requests about the form of items. Examples here include the form of commands to be typed to the system.

request-for-elaboration involves requests for elaboration of some item that has been introduced in the dialogue. Examples shown are for more information on commands and goals and plans.

6.2. Speech act sequencing

After a speech act typography was conducted, a speech act sequencing graph was produced (see Appendix B). This graph shown the various speech act types in boxes and the sequences these speech acts occurred in. Each box link has a set of numbers attached to it in square brackets. These numbers indicate the subject, i.e. whether it was subject 1 or subject 14. Each subject had a particular level of expertise with respect to computers, operating systems, and UNIX although most of the subjects were novice UNIX users.

Looking at the sequencing graph it is noted that request-for-direction seems to be the central node, i.e. this node has the most arrows leading out from it, six in all. Also, the other boxes have only one or no arrows leading from them. It is understandable that the request-for-direction should be the most central node as subjects will tend to ask queries about how to do something before they ask about extensions of such information. It is noted that the request-for-direction can precede every other type of request.

It is also interesting to look at the arrows leading from other boxes. In one case a request-for-guidance follows a request-for-elaboration. This might happen if the subject has asked for an elaboration and has not understood it. Therefore, the subject asks for guidance. Such relationships have important implications for user modeling. For example, any dialogue modeling program which recognizes speech acts would note that the subject has got a problem with this particular item of information and should help with its solution. It is also noted in Appendix C that this subject had some experience with UNIX and still had problems with some command. This should help in the design of a user modeling system where even experienced subjects could be helped over problems like this.

In another case a request-for-explanation is followed by a request-for-confirmation. This is understandable as a subject may ask for confirmation after being given an explanation. Interestingly, only one subject asked for confirmation although many subjects asked for explanations.

A request-for-information is followed by a request-for-guidance where again the subject is lost even after being given some information. It is noted from Appendix C that this subject has only 3 months computer experience. Also, request-for-guidance is mapped into itself a number of times. This would indicate that these subjects are really lost.

Correlating Appendix B and C all of the subjects, except one, that asked for elaboration, information or guidance were subjects who had no UNIX experience. Also, request-for-guidance was the most common request and all of the subjects making this request had no UNIX experience. Therefore, it is inferred from this that any user model finding many requests-for-guidance in a natural language dialogue should be aware that the user has little or no knowledge about the topic. Also, a speech act processor in the dialogue model should pass this information to the user model and in turn the user model should act accordingly to affect the answering of queries.

The only cases where the more experienced subjects asked follow up queries or sequenced speech acts were for request-for-explanation, request-for-elaboration, request-for-confirmation, and request-for-form and these only happened a total of five times, with request for explanation being the most densely occupied. Therefore, more experienced users will tend to ask requests about explanations and elaborations rather than requests for guidance or information.

On another plane it is possible to look at the most common types of requests following the more central request-for-direction. The most busy node is the request-for-guidance, and the request-for-confirmation and request-for-explanation tie for second place. Then there is request-for-confirmation not falling to far behind. However, request-for-form and request-for-elaboration fall far behind tying for last place. This all makes sense as many of the subjects had little UNIX experience and therefore they tended to ask for lots of guidance and then just as many explanations as information with less emphasis on confirmation. Of course form and elaboration fell last as the form was usually made quite clear by the Wizard and from previous answers and inexperienced subjects seldom ask for elaboration but wish to increase breadth rather than depth. Such constrains should be taken into account in user modelling for natural language interfaces.

Finally, there are some local observations about subjects. It is noted that certain subjects tended to ask certain types of sequences. For example subjects 8, 11, and 12 made requests-for-guidance a number of times. It is noted that subjects 11 and 12 had no computer experience at all and subject 8 had only 6 months to 1 year. Also, subject 6 and 8 asked for explanations twice each. Subjects 5 and 8 asked for confirmations straight from directions twice each indicating they were unsure of what was going on. Also, subject 7 asked requests for information four times. These request differences are probably more dependent on the personal characteristics of these subjects. Such information should also be taken into account for user modeling.

7. Conclusion and future work

The central conclusion here is that speech acts are not only useful for analyzing the intention of subjects in dialogue and hence their goals and plans as has been shown extensively in dialogue processing research (see Allen and Hinkelman, Cohen at al. 1982, Hinkelman and Allen 1989). Speech acts are also very useful for showing up user modelling data and have implications for user models.

We are also interested in the possibility of using automatic techniques to derive user models from Wizard-of-Oz data. Such techniques would be used to derive user syntactic and semantic usage for queries. The word references for operating system objects and commands could be analyzed with a parser and used in the construction of user-specific models. Users using terminology from one operating system and bringing that over into the UNIX domain could also be recognized. Also, more ambitious tasks such as the recognition of user-specific intention through speech act recognition and follow-on queries needs to be tackled automatically. This information could then be loaded into a natural language system and then the system itself could be tested in an augmented Wizard-of-Oz technique, with a wizard in the loop, where the system responded to queries and if it failed the wizard intervened without the user knowing this, and then using the marked interventions to further the system and make it less brittle. Therefore, what is argued for here are automatic and iterative techniques for the development of user models in natural language dialogue systems, rather than the subjective construction of such models. We would hope that such data would help in the development of user model components in systems, such as KNOME in UC (see Chin, 1988), or the development of belief models (see Ballim and Wilks, 1990). We are also interested in how speech acts sequences provide links between dialogue models and user models in general.

Further work on this analysis involves looking at how speech acts affect dialogue structure. It is believed that intention in general, and speech acts in particular, have implications for dialogue structure and that a computer program can be implemented to demonstrate and use this fact for typed human-computer natural language dialogue. It seems intuitive that not only does SAYING give a clue to INTENTION but that INTENTION gives a clue to SAYING. However, the implications of intention, or speech acts, for dialogue structure have been, on the whole, left uncovered.

If the above hypothesis is true then speech acts can be used to predict shifts in dialogue structure and help tackle one of the largest problems in dialogue processing, i.e. the problem of resolving references in dialogue. If speech acts have implications on dialogue structure then they can be used to segment dialogue into spaces and hence reduce the search space for references. It has been shown by Grosz and Sidner (1986) that by reducing dialogue into segmented spaces the search for references can be reduced to a search in one of the spaces. We report preliminary research on the use of speech acts for segmentation in Mc Kevitt (1990).

8. Acknowledgements

I would like to thank Zhaoxin Pan for implementing the Wizard program and Bill Ogden for help in gathering the data and helping with the analysis of that data. Cari Soderlund is to be thanked for the use of her MAND inter-process communication software (see Soderlund 1988). Ted Dunning, David Farwell, Louise Guthrie, Steve Helmreich and Sylvia Candelaria De Ram and other members of the Natural Language Group at the CRL provided numerous comments on task specifications and program design for the wizard studies as did Hans Brunner at U S *WEST* Advanced Technologies, Denver, Colorado and Philip Stenton at Hewlett-Packard Laboratories, Bristol, United Kingdom.

9. References

Allen, James F. and Elizabeth Hinkelman (1989) *Using structural constraints for speech act interpretation.* In DARPA Speech and Natural Language Workshop, Cape Cod, MA., October 1989.

Austin, J.L. (1975) *How to do things with words.* Second Edition, J.O. Urmson and Marina Sbisa, (Eds.), Harvard University Press: Harvard, Mass.

Bach, Kent and Hamish, Robert M. (1979) *Linguistic communication and speech acts.* The MIT Press: Cambridge, Massachusetts; London, England.

Ballim, Afzal and Yorick Wilks (1990) *Stereotypical Belief and Dynamic Agent Modeling.* In Proceedings of the Second International Workshop on User Modeling, University of Hawaii at Manoa, Honolulu, Hawaii, USA, March.

Brennan, S.E., M.W. Friedman, and C. Pollard (1987) *A centering approach to pronouns.* In Proc. of the 25th Annual Meeting of the Association of Computational Linguistics.

Brunner, Hans, Kathleen Ferrara, and Greg Whittemore (1989a) *On the implications of frequency and distribution for anaphora resolution in written/interactive dialogue.* MCC Technical Report Number ACT-HI-243-89, (MCC Nonconfidential), Microelectronics and Computer Technology Corporation (MCC), Advanced Computing Technology, Human Interface Laboratory, 3500 West Balcones Center Drive, Austin, Texas.

Brunner, Hans, Kathleen Ferrara, and Greg Whittemore (1989b) *An assessment of written/interactive dialogue for information retrieval applications.* MCC Technical Report Number ACT-HI-245-89, (MCC Nonconfidential), Microelectronics and Computer Technology Corporation (MCC), Advanced Computing Technology, Human Interface Laboratory, 3500 West Balcones Center Drive, Austin, Texas.

Carberry, Sandra (1989) *A pragmatics-based approach to ellipsis resolution.* In Computational Linguistics, Vol 15 (2), June, pp. 75-96.

Chin, David (1984) *An analysis of scripts generated in writing between users and computer consultants.* In AFIPS Proceedings of the National Computer Conference 53, Las Vagas, NV, July.

Chin, David (1988) *Exploiting user expertise in answer expression.* In Proceedings of the Seventh National American Conference on Artificial Intelligence (AAAI-88), St. Paul, Minnesota, Vol. 2, 756-760, August.

Cohen, Philip R., C. Raymond Perrault and James F. Allen (1982) "Beyond question answering" In "Strategies for natural language processing", pp. 245-274, Wendy G. Lehnert and Martin H. Ringle (Eds.), Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Douglass, Robert J. and Stephen J. Hegner (1982) *An expert consultant for the UNIX operating system: Bridging the gap between the user and command language semantics.* Proc. Fourth National Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI)/SCIEO Conference, Saskatoon, Saskatchewan, pp. 119-127, May.

Grosz, B.J. and C.L. Sidner (1986) *Attention, intentions and the structure of discourse,* Computational Linguistics, Vol. 12, 175-204.

Guindon, Raymonde (1988) *A multidisciplinary perspective on dialogue structure in user-advisor dialogues.* In Cognitive Science and its Applications for Human-Computer Interaction, R. Guindon (Ed.), Hillsdale, N.J.: Lawrence Earlbaum and Associates.

Guindon, Raymonde, Paul Sladky, Hans Brunner and Joyce Conner (1986) *The structure of user-advisor dialogues: is there method in their madness?.* In Proc. of the 24th Annual Conference of the Association for Computational Linguistics, 224-230, Columbia University, New York.

Guthrie, Louise, Paul Mc Kevitt and Yorick Wilks (1989) *OSCON: An operating system consultant.* In Proc. of the Fourth Annual Rocky Mountain Conference on Artificial Intelligence (RMCAI-89), Subtitled "Augmenting Human Intellect By Computer", pp. 103-113, Registry Hotel, Denver, Colorado, June.

Hecking, M. C. Kemke, E. Nessen, D. Dengler, M. Gutmann and G. Hector (1988) "The SINIX consultant —a progress report" Memo Nr. 28, Universitat des Saarlandes, FB 10 Informatik IV, Im Stanwald 15, D — 6600 Saarbrucken 11, Fed. Rep. of Germany, August.

Hegner, Stephen J. and Robert J. Douglass (1984) *Knowledge base design for an operating system expert consultant.* Proc. of the Fifth National Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI), London, Ontario, pp. 159-161, December.

Hinkelman, Elizabeth and James Allen (1989) *Two constraints on speech act ambiguity.* In Proc. of the 27th Annual Meeting of the Association for Computational Linguistics, pp. 212-219, University of British Columbia, Vancouver, British Columbia, Canada, June.

Hirschberg, J. and J. B. Pierrehumbert (1986) *The intonational structuring of discourse.* In Proc. of the of the 24th Annual Meeting of the Association for Computational Linguistic.

Kobsa, Alfred and Wolfgang Wahlster (1988) "User models in dialog systems" Springer Verlag, Berlin, NY

Mc Kevitt, Paul (1986) *Formalization in an English interface to a UNIX database.* Memoranda in Computer and Cognitive Science, MCCS-86-73, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001.

Mc Kevitt, Paul (1990) *Using speech acts for segmentation in user-adviser dialogue.* In Proceedings of the International Pragmatics Conference, Barcelona, Spain, EC, June.

Mc Kevitt, Paul (1988) *Rules of inference in an operating system consultant.* In Proc. of the First Irish National Conference on Artificial Intelligence and Cognitive Science (AI/CS-88), Vol. 1, University Industry Center, University College Dublin, Dublin, Ireland, European Community (EC), September.

Mc Kevitt, Paul and Yorick Wilks (1987) *Transfer Semantics in an Operating System Consultant: the formalization of actions involving object transfer..* In Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87), Vol. 1, 569-575, Milan, Italy, August.

Mc Kevitt, Paul and Zhaoxin Pan (1989) *A general effect representation for Operating System Commands.* In Proc. of the Second Irish National Conference on Artificial Intelligence and Cognitive Science (AI/CS-89), School of Computer Applications, Dublin City University (DCU), Dublin, Ireland, European Community (EC), September.

Mc Kevitt, Paul and William C. Ogden (1989) *Wizard-of-Oz dialogues in the computer operating systems domain.* Memoranda in Computer and Cognitive Science, MCCS-89-167, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001.

Reichman, R. (1978) *Conversational coherency.* Cognitive Science 2:283-327.

Reichman-Adar, R. (1984) *Extended person-machine interface.* Artificial Intelligence 22(2), 157-218.

Reichman, R. (1986) *Getting computers to talk like you and me.* MIT press: Cambridge, Mass.

Schuster, Ethel; David Chin, Robin Cohen, Alfred Kobsa, Kathrina Morik, Karen Sparck Jones, Wolfgang Wahlster (1988) *Discussion section on the relationship between user models and discourse models.* In Computational Linguistics, Vol. 14, No. 3, Special Issue on User Modeling, Alfred Kobsa and Wolfgang Wahlster (Guest Eds.), 79-103, September.

Searle, John R. (1969) *Speech acts: An essay in the philosophy of language.* Cambridge University Press: London.

Shneiderman, Ben (1987) *Designing the user interface: Strategies for effective human-computer interaction.* Reading, Massachusetts, Addison Wesley.

Slator, Brian M., Matthew P. Anderson and Walt Conley (1986) *Pygmalion at the Interface.* Communications of the ACM, Special Section on the Human Aspects of Computing, Henry Ledgard (Ed.), Vol. 29, No. 7, July, pp. 599-605. Also, as *Pygmalion at the interface: impatient users and foreign-speak.* MCCS-85-26, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001.

Soderlund, Cari (1988) *A TCP remote message passing service.* Memoranda in Computer and Cognitive Science, MCCS-88-130, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001.

Stampe, Dennis W. (1975) *Meaning and truth in the theory of speech acts.* In Syntax and Semantics, Speech Acts, Vol. 3, Cole, Peter and Morgan, Jerry L. (Eds.), pp. 1-39, Academic Press: New York.

Walker, Marilyn and Steve Whittaker (1989) *When natural language is better than menu: a field study.* Technical Report, Hewlett Packard Laboratories, Advanced Information Management Dept., Filton Road, Stoke Gifford, Bristol, United Kingdom, European Community (EC).

Whittaker, Steve and Phil Stenton (1988) *Cues and control in expert-client dialogues.* In Proc. of the 26th Annual Meeting of the Association for Computational Linguistics, pp. 123-130, State University of New York at Buffalo, Buffalo, New York, USA.

Whittaker, Steve and Phil Stenton (1989) *User studies and the design of natural language systems.* In Proc. of the 00th Annual Meeting of the European Association of Computational Linguistics, Manchester, United Kingdom, European Community (EC).

Wilensky, Robert; Arens, Yigal and Chin, David (1984) *Talking to UNIX in English: An overview of UC.* Communications of the ACM, Vol. 27, No. 6, pp. 574-593, June.

Wilensky, Robert; Mayfield, Jim; Albert, Anthony; Chin, David; Cox, Charles; Luria, Marc; Martin, James and Wu, Dekai (1986) *UC — a progress report.* Report No. UCB/CSD 87/303, Computer Science Division (EECS), University of California, Berkeley, California 94720, July.

Wilensky, Robert; David N. Chin, Marc Luria, James Martin, James Mayfield and Dekai Wu (1988) *The Berkeley UNIX Consultant project.* Computational Linguistics, Vol. 14, No. 4, 35-84, December.

Appendix A: Speech act categories

The following list shows the Speech Act query types found in a set of Wizard-of-Oz dialogues for 14 subjects. Except for type (1) all cases of each type are listed. request-for-information occurred most commonly in the dialogues. The ordering of the speech act types here is in terms of frequency the most frequent shown first.

(1) **request-for-direction**

e.g. how do i copy a file
how do i remove a directory
how do i get the date
how do I know if it has produced a printed copy
what do i type to see who is using the computer?

(2) **request-for-guidance** (frequency: 21)

e.g. what now, I can't seem to get it to go to the previous directory? [5]
how? [6]
[how] will you help? [6]
now what do i do next? [6]
what do i do after i find out [6]
if the file is not in that directory, how can I find out
what directory it is in? [7]
after I use cd how will I know if file1 is there? [7]
if the file1 is not there where do I go? [7]
what if there is no such file in diectory? [7]
how do I get to the previous task [8]
I don't understand what I am suppose to do [8]
the rmdir rubbish doesn't do the task, so what now [8]
What do I do if it is not empty? [11]
What do I do if there is no such file or directory? [11]
what if movemeto cannot access? [11]
What do you do you do if you use the command rmdir and the file
name but the directory doesn't empty? [12]
What do you do when there is no file or directory to remove with
the name rubbish [12]
What do you do if it still tells you there is no such file or
directory? [12]
Now what do I do? [13, 13, 13]

(3) **request-for-explanation** (frequency: 10)

e.g. what does directory not empty mean [1, 4]
what does cp -r mean? [2]
explain more [6]
where have all the files gone [6]
why is a directory not copied? [7]
what do you mean by "trash" [8]
when I command it to remove directory it says that
there is no such file or directory [8]
what is oscon [9]
why is my file movemeto not found? [10]

(4) **request-for-information** (frequency: 9)

is the computer hooked up to a printer [4]
what other users are on the system today? [5]
has oscon been printed [6]
am i in the help directory [6]
am i in oscon [6]
is task complete [6]
what other directory can I go to? [7]
Does uswest now exist in your directory? [11]
what is today's date [14]

(5) **request-for-confirmation** (frequency: 6)

e.g. to remove a directory do i need to remove all of the files [1]
can i make a copy of a subdirectory? [5]
can i remove a directory with files in it? [5]
do I use the same method as the last task [8]
DO I TYPE LS "MOVEMETO""TOHERE" [8]
is the directory named "rubbish" now gone? [10]

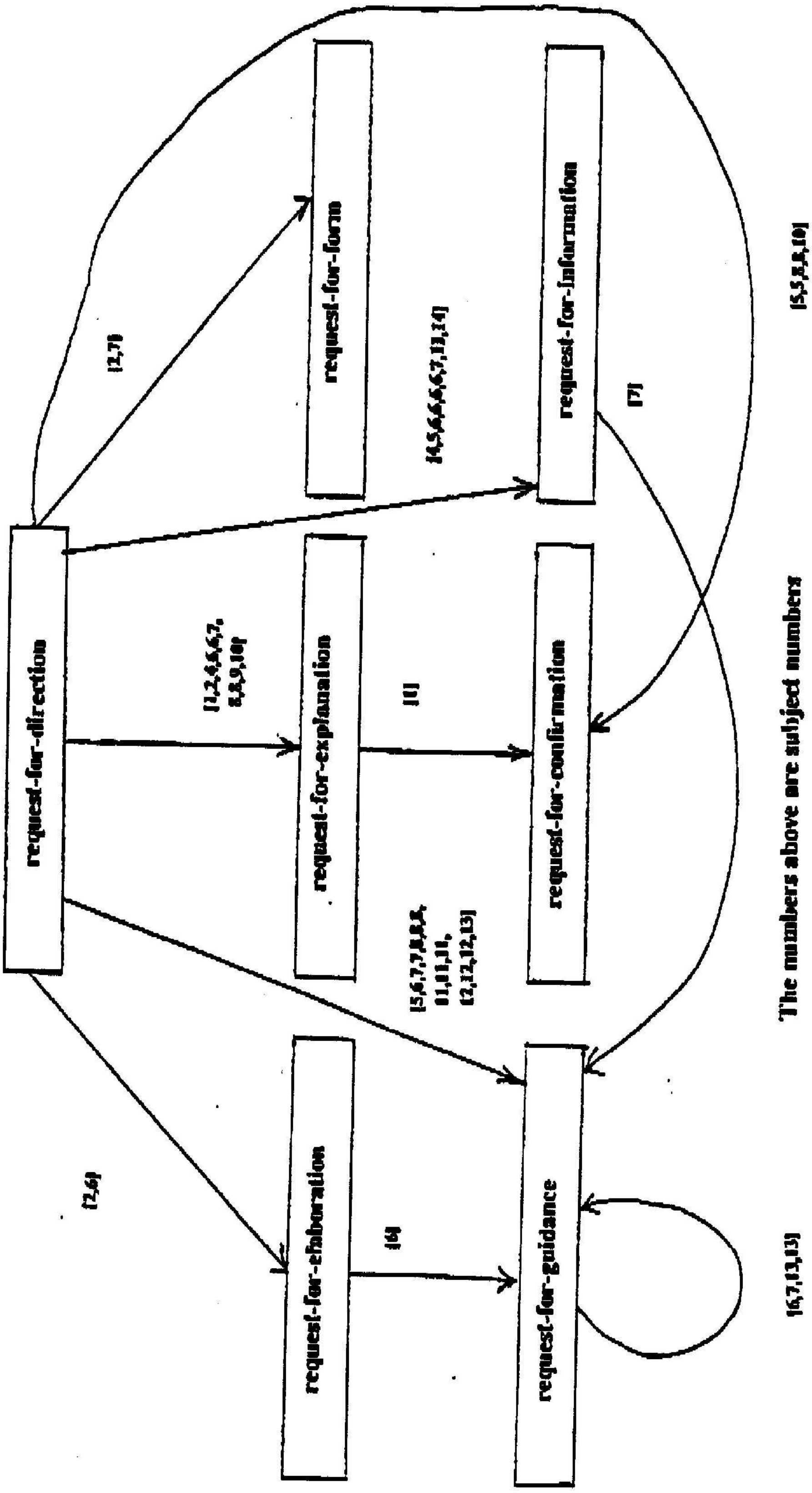
(6) **request-for-form** (frequency: 2)

e.g. how do i use logout? [2]
how will "is" be entered? [7]

(7) **request-for-elaboration** (frequency: 2)

e.g. how do i use more? [2]
can you give me more information on how to find dir and change to dirl [6]

Appendix B: Sequence graph



The numbers above are subject numbers

Appendix C: User expertise levels

The following list shown the subject expertise in (1) computers, (2) operating systems, and (3) UNIX. Most of the subjects had little or no experience with UNIX.

([1] computers: Little; OS: MS-DOS, UNIX; UNIX: Little)
([2] computers: 1-2 Yrs; OS: MS-DOS, UNIX; UNIX: 6 Mth-1 Yr)
([3] computers: 1-2 Yrs; OS: MS-DOS, PC-DOS, CMS, VMS, UNIX; UNIX: 1-2 Yrs)
([4] computers: 1-2 Yrs; OS: MS-DOS; UNIX: None)
([5] computers: 1-2 Yrs; OS: PC-DOS; UNIX: None)
([6] computers: < 3 Mth; OS: PC-DOS, MAC; UNIX: None)
([7] computers: < 3 Mth; OS: PC-DOS; UNIX: None)
([8] computers: 6 Mth-1 Yr; OS: PC-DOS; UNIX: None)
([9] computers: 6 Mth-1 Yr; OS: PC-DOS; UNIX: None)
([10] computers: < 3 Mth; OS: MS-DOS; UNIX: None)
([11] computers: None; OS: ; UNIX:)
([12] computers: None; OS: ; UNIX:)
([13] computers: 6 Mth-1 Yr; OS: PC-DOS, MAC; UNIX: None)
([14] computers: None; OS: ; UNIX: None)