ON THE NATURE AND FUNCTION OF EXPLANATION IN INTELLIGENT INFORMATION RETRIEVAL

N.J. Belkin

School of Communication, Information and Library Studies
Rutgers University
New Brunswick, NJ 08903
U.S.A.

ABSTRACT
We discuss the complexity of explanation activity in human-human goal-directed dialogue, and suggest that this complexity ought to be taken account of in the design of explanation in human-computer interaction. We propose a general model of clatiry in human-computer systems, of which explanation is one component. On the bases of: this model; of a model of human-intermediary interaction in the document retrieval istuation as one of cooperative model-building for the prupose of developing an appropriate search formulation; and, on the results of empirical observation of human user-human intermediary interaction in information systems, we propose a model for explanation by the computer intermediary in information retrieval.

1.INTRODUCTION

Explanation ('the act or process of making plain or comprehensible' (American Heritage Dictionary)) can serve a number of purposes in ordinary goal-directed dialogue between human beings. For Instance, it can be used to describe how one party has arrived at a certain conclusion, or to resolve misunderstandings that one party has of the other, or to be certain that both parties have the same, or equivalent, conceptions of the situation within which they find themselves, or to justify a particular course of action or dialogue pattern, or to convince the other to behave in some particular way, or to encourage the other to believe or have confidence in oneself, or to modify the other's state of knowledge. This brief list is of course not exhaustive, nor are the differenct purposes named exclusive as actually performed, which means that the uses of explanation in dialogue are both very extensive and highly complex.

Many factors influence the actual use and acceptance of explanation in ordinary dialogue. For instance, the nature of the speech situation, the social roles of the participants, the expectations and stereotypes the hold of one another, the cognitive authority they acscribe to one another, will all condition whether explanations will be forthcoming, on the one hand, or accepted, on the other. And, of course, the states of knowledge the participants bring to the situation determine the nature, necessity and possibility of explanation, as does the task or goal on which they are cooperating.

Many people have suggested that at least the functional pattern of human human interaction ought to be the model for human-computer interaction in equivalent circumstances, e.g. [BELKIN82] [HAYES79] [REICHMAN85]. In accepting this point of view, we accept necessarily that the circumstances leading to explanation, its functions, and its uses in human-human dialogues ought to be the bases for explanation in human-computer dialogues, at the functional level. This is the major issue with which this paper is concerned.

Although there has been some significant systematic investigation of the relationships among the various factors mentioned above and the activity of explanation, e.g. [LALLJEE83], little of this work appears to have informed the design and implementation of explanation in knowledge-based or 'intelligent' systems which are meant to operate in a human-computer interactive environment (e.g. 'expert systems'). The primary form of explanation in such systems has been justification of a conclusion or piece of advice, the usual means being a straightforward escription of the logic which led to the conclusion, the mode a response by the computer to a human request for explanation, with the underlying communicative goal rarely being made explicit, although the implied goal appears usually to be instantiation of the computer's expertise, for the purpose either of demonstrating 'intelligence', or of convincing the partner to accept the conclusion. The major exception to this paradigm for explanation in expert systems is the work reported in [CHANDRASEKARAN88], which makes explicit different roles for and types of explanation, in a manner similar to that proposed here. However, in that work, they continue to concentrate on developing explanation in the first instance as the general expert system type.

Certainly there has been much other important work done on explanation in intelligent systems, particularly in the generation of explanation, e.g. [MCKEOWN85]. We note, however, that this kind of explanation is only one of the many possibilities that could exist, and may be relevant in only a small fraction of the kinds of dialogue situations that exist, or could exist between two partners, as discussed above. The intention of this paper is to specify, on the bases of a theoretcial model, and of empirical study of human-human and human-computer interaction, a model for the functions, forms and modes of explanation in the interaction between human user and computer intermediary in the document retrieval situation.

2. 'EXPERT' INTERMEDIARIES FOR DOCUMENT RETRIEVAL, AND EXPLANATION IN EXPERT SYSTEMS

There is by now a reasonably extensive literature concerned with proposals for, and, more rarely, implementations of computer systems which are intended to act as 'intelligent' or 'expert' intermediaries between end users and document data bases, e.g. [BELKIN83] [BRAJNIK87] [BROOKS85] [CROFT87] [FIDEL 86] [FOX87] [MARCUS83] [SHOVAL85]. In most such systems, explanation as a system function is ignored. In some, such as CONIT [MARCUS83], whose goals are instruction as much as mediation, explanation is inherent as teaching, but without reference to the interaction between user and computer in search, or search formulation. In a few, such as [BROOKS85], explanation is mentioned as a function that should be performed, but without further specification. [BRAJNIK87] suggest some ways in which a user model might be used for explanation purposes, but do not develop this idea any further. This appears to us to be a fairly serious lacuna, not for the purely formal reason of conforming with definitions of expert systems, but because, on the basis of empirical investigations by us, e.g. [BELKIN84] [DANIELS85] and others, e.g. [COOMBS85], explanation appears to play an important role in information interaction of the type exemplified by document retrieval or advisory situations.

Unfortunately, the model of explanation which appears to be most prevalent in standard expert systems seems to us insufficient, or inadequate

not only within its own context, but especially for the document retrieval situation. There are several reasons for this, including the fact that the standard model only refers to one type and function of explanation, and also that the expert intermediary interacts not with another expert, as is the case with most expert systems, but rather with a novice (at least in the area of document retrieval). The latter reason suggests tht the issues such as the social roles of the participants, and the goals of explanation, will be important in defining how explanation is to be implemented in a truely intelligent intermediary for document retrieval.

3. A MODEL FOR CLARITY AND EXPLANATION

We propose a general model of clarity[1] as a framework within which appropriate explanation strategies could be developed. Clarity, as used here, is the characteristic that the user in the system understand the workings and/or other components of the system. By system, in this context, we explicitly mean the 'joint cognitive system' as proposed by [HOLLNAGEL83]. Obviously, one means to achieving clarity is overt explanation; in particular, explanation of:

what has occurred;
what is occurring;
what will be done, or is intended;
the process itself;
the data base.

Clarity will also be achieved through description, as for instance, of each party's model of the:

world;
task;
system;
other;
self.

These appear to be the two major forms of clarity. Both of these forms can be in any of several modes; that is,

implicit (pre-agreement, or general conversational postulates);
on demand by user (prompted);
during the process (unprompted);
process as clarification;
control by other.

In any system, at any particular time, there will be some optimum form and mode of clarity. The task of the computer half of the dialogue (or of the system designer), is to choose the appropriate form and mode. This will be done on the basis of the following factors:

relationship beteen the parties;
familiarity with other;
familiarity with system;
complexity of system;
understanding of problem;
complexity of problem;
immediacy/importance of problem.

Thus, we consider that clarity is guided by, or directed towards affective, cognitive and functional requirements of the interaction, and will be based upon the models that the participants hold of one another, the system, and themselves in the system.

This general specification of clarity is used as the basis for determining clarification activity (including explanation) in the human-computer intermediary document retrieval interaction.

4. MODEL BUILDING AND EXPLANATION IN AN EXPERT INTERMEDIARY FOR DOCUMENT RETRIEVAL

In a series of earlier papers [BELKIN83] [BELKIN84] [BELKIN87]

[BROOKS85] [DANIELS85] we and others have suggested that the interaction between the user and intermediary in document retrieval consists, to a great extent, of mutual cooperation in the tasks of the intermediary's building models of the user, and the user's problem, situation and requirements, which will be useful in jointly producing an appropriate search formulation. This conclusion has been based on the detailed observation ofa number of human-human information interactions, and has resulted in a functional specification of user-intermediary interaction in the document retrieval situation. These functions are identified and briefly explained in figure 1.

Problem Mode	Determine if system's capabilities are appropriate to user's situation
Problem State	Determine position of user in problem management process
User Model	Develop model of user type, status, knowledge, experience, goals, etc.
Problem Description	Develop model of problem type, topic, structure, context
Dialogue Mode	Determine appropriate dialogue type and mode for situation
Retrieval Strategy	Choose and apply appropriate retrieval strategies and techniques
Response Generator	Determine appropriate response to user
Explanation	Explain and describe system capabilities, etc., and activities
Input Analyst	Convert input from user to form appropriate for system
Output Generator	Convert response specification to appropriate output form

Figure 1. The functions of an intelligent intermediary for document retrieval (after [BELKIN83]).

One of the functions identified in this work was Explanation, which was primarily unprompted justification of intermediary activities, or unprompted wxplanation of system capabilites and characteristics. In addition, there was a great deal of dexription and matching of models of one another. Although this type of explanation/clarification activity was based on several of the different models that the intermediaries held, such as the User Model, Problem Description, and the Retrieval Strategy, it appeared that most of the unprompted explanation was based upon the intermediary's model of the user's model of the system, as conditioned by such other models as User Model and Problem State. The goal of this kind of explanation usually appeared to be the attempt to modify the user's model of the system in such a way as to facilitate cooperative interaction on the search formulation problem.

5. METHODS

In order to investigate the nature of clarification activity in the information system more thoroughly, we continued with our original data and methods, concentrating espceially upon clarification activity, and adding to those data with new methods of collection and analysis. Our original data are audio recordings of about 100 interactions between human users and human intermediaries, primarily in document retrieval settings, but also in some advisory interactions. Our methods of collection and analysis of these data are reported in detail in [BELKIN87]. Briefly, we collected the data in real settings, transcribed the audiotapes, segmented them into utterances, and groups of utterances, called foci, and then assigned task and functions to the utterances and foci according to the coding scheme of figure 1, and a more detailed scheme of lower-level tasks. We are now supplementing these data, and categories, by videotapeing such interactions. These new data give us access to non-verbal interaction, which is especially important in establishing the social relations between the parties, and in interpreting the locutionary force of ambiguous utterances.

The methods we are now using are basically our scheme of functional discourse analysis, in this instance to discover when clarification occurs, and why, and how it interacts with other functions. In addition, we analyze the transcripts accoring to the categories specified in our model of clarity, and interview the parties to the new interactions we record. We then relate the performance of the explanation functions in each interaction to the determining factors for clarity form and mode.

6. RESULTS

The research is still in progress, so that our results are preliminary, but nevertheless highly suggestive. Figures 2, 3 and 4 are examples from different interactions which indicate some of the types of clarification activity that are engaged in in our data.

```
I  now i gather you're (cough) excuse me you're a visitor/3
U

I  i'm        yes are you part of the university or/5
U  yes i am/4                                          well

I                                      ya(,) um i i jus (,)
U  i teach at a canadian university/6

I  we ask you this because i- it's awful to bring up
U

I  charges straight away (laugh) but just so that you know(,)
U

I  you know that it's a ten pound basic an it's (inaud)/7
U                                                        yes/8

I = Intermediary;  U = User;  /n = utterance number
```

Figure 2. Portion of focus 1, interaction 190684.

I there's Social Sciences Citation Index (...) it sounds as
U

I though it's (,) very much Social Sciences but its- its a
U

I very broad based um (,) indexing source/95 and i
U mm hmm/96

I think that would be worth trying/97 what we've got to
U ok/98

I do there is concentrate on just on title words (.) so it's
U

I not so easy(,) you know so if you say forestry and they talk
U

I about (.) a (inaud) plantation or something you wouldn't
U

I actually get it/99
U mm/100

Figure 3. Portion of interaction 120684hba

I so (.) tell me first something about the research you're
U

I doing(,) and then the topic of your search(,) and then
U

I we'll choose some terms for the search/4
U ok (..) um

Figure 4. Extract from interaction 260684ksa

The three examples of figures 2, 3 and 4 indicate several different types of clarification patterns which we have discovered in the data. In the interaction of figure 2, for instance, utterances 3 and 5 are used to establish a particular model of the user's conceptual model by the intermediary, who on the basis of that model offers the information (clarification) about the cost of using the system, having judged that the user is unlikely to know this. In figure 3, utterance 95 offers a justification for using SSCI, and utterance 99 an explanation and description/justification of how to use that data base. In figure 4, the intermediary begins the interaction by dexribing the plan or structure of the interaction in general.

On the basis of data of this type, we draw the following tentative reulst. First, it appears that the basic goals of clarification in the information interaction enviornment which we studied are:

bringing the user's model of the system to a level which the intermediary feels sufficient for effective interaction;

making the intermediary's plan of action evident or plain;

making the model of user and prblem explicit; and,

making sure that the system is appropriate to the user's problem and situation.

The typical forms of clarity which we have observed are, in approximate order of frequency:

direct explanation, usually with model elicitation;

prosp ective description of process or procedure;

description and matching of models;

direct explanation and matching, together with model elicitation.

The modes of clarity which we have observed, again in approximate order of frequency of use, are:

Unprompted explanation during the process;

user instigation (rare);

mutual agreement.

The most usual mode of clarification is unprompted; that is, offered by the intermediary. It seems, indeed, to be the case that prompted clarification (usually explanation) by the user is most often a sign of some discourse disfunction. That is, the users typically expect that the intermediary inform them sufficiently in advance of any problem. Many aspects of this result appear to reflect the general role character of the interaction, in which the intermediary is usually 'on top', as expert in interaction with the data base. In situations in which the subject and topic negotiation become more important, the user becomes the instigator and controller, and is thus allowed to ask for clarification. The major exceptions to the finding that this aspect of role is important occur in dierect interaction with external sources such as thesauri, where user's subject knowledge may conflict with the thesaural structure, which conflict may require clarification, in problem description model elicitation on the part of the user, and in explanation of features of the interface which the intermediary considers basically unimportant.

This last point leads to what appears to be a fairly general result. That is, most clarification activity on the part of the intermediary appears to be based on the comparison, by the intermediary,of a model of what a user ought to know about the system and her/his role in it (including the interaction), with the intermediary's model of what the user does know about this (the intemediary's model of the user's conceptual model). Thus, clarification often serves an explicit teaching function, with the intermediary attempting to modify what s/he preceives the user's conceptual model to be to what it should be, for effective interaction. The intermediary's model of the conceptual model is built very early in the interaction, and depends to a great extent upon inferences derived from a small amount of directly elicited data, primarily in construction of the User Model. That is, the intermediary has very stron stereotypes of what users of particular types, and particular levels of experience are likely to know about the system, and proceeds with unprompted clarification activity based almost exclusively on these stereotypes. The reason for this choice of mode appears to be two-fold. First, it seems that effective information ineraction in this situation depends upon the user's already having an appropriate conceptual model of the relevant aspect of the system when it is invoked. Second, it also appears that both intermediaries and users believe that, especially in the case of naive user, the user typically does not know what needs to be explained.

Another aspect of the relationship between normative conceptual model an user's actual conceptual model is that the normative model (that held by the intermediary) changes from user to user, and even sometimes within one interaction. That is, intermediaries take account of various factors associated with users in order to decide how much any one particular user needs to know about the system in order to interact effectively on the specific problem of interest. Thus, for highly complex topics, it may be very important for the user to understand thesaural structure and relationships very well, as well as detailed aspects of search logic, whereas for users with less demanding requirements, a simpler model of the system might be sufficient. These normative models can change in the course of interactions not only when new hypotheses about the complexity of the problem are developed by the intermediary, but also as previous models or stereotypes are changed in the light of new evidence from the user.

The status of the intermediary's model of the conceptual model aslo has a strong effect on the temporal order of clarification activities. The general ordering of clarification, in terms of what is clarified and how, is conditioned on the performance of the other interaction functions. That is, explanation occurs piecemeal and opportunistically, according to what the intermediary believes the user believes about aspects of the specific functional focus of the dialogue at any one time. When the conceptual model is believed to be quite inadequate, then detailed explanation is the preferred form; when the conceptual model is thought to be rather close to that required, then description or model elicitation is more likely to be invoked. Thus, although the intermediary's model of the user's conceptual model is established early, in broad form, it is modified incrementally, rather than wholescale, and only as required. With a user whose conceptual model is deemed to be adequate at the start of the interaction, as, for instance, with long-time users, the clarification is hidden almost completely. That is, it takes on an implicit mode, and only becomes explicit when, from the intermediary's point of view, a new situation arises, or, from the user's point of view, a contradiction between expectations and interaction occurs. These situations lead to unprompted description and prompted explanation, respectively.

Other clarification goals occur generally as part of the discourse function to which they refer; that is, clarification of the intermediary's model of the problem occurs during problem description foci, generally as unprompted description. This is not, however, a strict rule, for contradictions of aspects of the intermediary's model of the user can occur during performance of other functions which are based on knowledge of that aspect of the model. For instance, during a problem description focus, the user might make a requirement on the search topic which would cause the intermediary to check on her/his model of the user's goals. And clarification of the plan of interaction, although usually occuring at the beginning of the dialogue, often recurs when the plan needs to be changed, or when it appears that it has been changed. In both cases, the most appropriate mode is unprompted, prospective explanation.

7. CONCLUSION

The results outlined in the previous section, although still incomplete, show that clarification in this particular kind of dialogue follows some fairly strong rules, in terms of goals, forms and modes of clarification, and in terms of its temporal sequencing. The data seem to indicate that the preferred mode of clarification is prospective and unprompted, that whether it is descriptive or explanatory depends strongly on the intermediary's view of the adequacy of the conceptual model, and that the particular clarification goal is strongly dependent both upon that view and on the specific discourse focus at any one time. It seems also the case that the social roles of the two parties are important, especially in mode of clarification, and that these roles in turn depend to some extent on

whose expertise is being tapped at the moment. These results, particularly the last, are still tentative, since more data are being analyzed, and very little information from non-verbal data has yet been incorporated. Nevertheless, they are suggestive not lnly as a description of what goes on in human-human information interaction, but also in terms of how clarification might proceed in human computer-interaction.

Althoughat this point such suggestions are necessarily speculative, it seems clear that the usual mode of prompted explanation in retrospect is in general inappropriate for this form of human-computer interaction. Furthermore, it appears that effective clarification requires detailed models of various aspects of the user, and in particular a strong model of the condeptual model, as well as an internal view of what constitutes an adequate conceptual model, for a specific user and situation. Therefore, we can prob ably say that clarification activity, on the part of the intermediary, can only proceed effectively if this 'ideal' conceptual model is built into the system in some way, but especially in a way that allows its adaptation to specific circumstances. This will be no mean feat, since the status of such a normative model is very uncertain.

Furthermore, most clarification will depend upon early construction of the user model, which can be used to build the intermediary's model of the user's conceptual model. Since clarification seems to be most effective prospectively, a potentially productive strategy would be to display a model of the interaction process as a whole at the beginning of the dialogue, and to maintain this model throughout the interaction. But since clarification also seems most effective when invoked according th the general discourse function being performed, a clarification window (say) could be invoked with each shift of focus, which would offer information about aspects of that function which the intermediary judges liekley to be required by the user (based on the model of the conceptual model). A further form of clarification could take place by invoking another window in which the intermediary's model of the particular aspect of the user being investigated then were made explicit, perhaps for direct manipulation. And for the case of role shifts, a facility fro prompting for explanation must be included throughout.

The sort of interface described above is certainly well within technical possibilites, but it does require an intelligent intermediary of a rather particular sort; that is, one which does quitee a lot of model buildigg about the user during the course of the interaction. Such systems have been proposed, and in cases built, by a number of people, e.g. [BELKIN83] [BRAJNIK87] [CROFT87] [FOX87], on several grounds. It appears from the work described here, that effective clarification, within the context of effective human-computer interaction, is yet another reason for suggesting that this type of system is required for truly intelligent information retrieval.

NOTE

[1] This model was first presented at the Danish Artificial Intelligence Society Seminar on architecturte, complexity and transparency in artificial intelligence, Copenhagen, 1985. It has subsequently benefitted by comments from preeentations at Bellcore, Morristown, NJH OCLC, and the Potomac Valley Chapter of ASIS.

REFERENCES

[BELKIN82]. Belkin, N.J. Models of dialogue for information retrieval. In Proceedings of the 4th International Research Forum in Information Science, Boras, Sweden, 1981. Boras: Hogskolan i Boras; 1982: 15-36.

[BELKIN83]. Belkin, N.J., Seeger, T. & Wersig, G. Distributed expert problem treatment as a model for information system analysis and design. Journal of Information Science, 5(1983): 153-167.

[BELKIN84] Belkin, N.J. Cognitive models and information transfer. Social Science Information Studies, 4 (1984): 111-129.

[BELKIN87] Belkin, N.J, Brooks, H.M. & Daniels, P.J. Knowledge elicitation using discourse analysis. International Journal of Man-Machine Studies, 27 (1987): 127-144.

[BRAJNIK87] Brajnik, G., Guida, G. & Tasso, C. User modelling in intelligent information retrieval. Information Processing and Management, 23 (1987) 305-320.

[BROOKS84] Brooks, H.M. Information retrieval and expert systems - approaches and methods of development. In: Informatics 7: Intelligent Information Retrieval. Londo: Aslib; 1984.

[BROOKS85] Brooks, H.M., Daniels, P.J. & Belkin, N.J. Problem description and user models: developing an intelligent interface for document retrieval systems. In: Informatiocs 8: Advances in intelligent retreival. London: Aslib; 1985: 191-214.

[CHANDRASEKARAN88] Chandrasekaran, B., Tanner, M.D. & Josephson, J. R. Explanation: the role of control strategies and deep models. In: Expert systems: the user interface, J. A. Hendler, ed. Norwood, NJ: Ablex; 1988: 219-247.

[COOMBS85] Coombs, M.J. & Alty, J.L. An application of the Birmingham discourse analysis system to the study of computer guidance interactions. Human-Computer Interaction, 1 (1985): 243-282.

[CROFT87] Croft, W.B. & Thompson, R.H. I^3R: A new approach to the design of document retrieval systems. Journal of the American Society for Information Science, 38 (1987): 398-404.

[DANIELS85] Daniels, P.J., Brooks, H.M. & Belkin, N.J. Using problem structures for driving human-computer dialogues. In: RIAO '85. Grenoble: I.M.A.G.; 1985: 131-149.

[FOX87] Fox, E.A. Development of the CODER system: a testbed for artificial intelligence methods in information retrieval. Information Processing and Management, 23 (1987): 341-366.

[FIDEL86] Fidel, R. Towards expert systems for the selection of search keys. Journal of the American Society for Information Science, 37 (1986): 37-44.

[HAYES79] Hayes, P.J. & Reddy, R. An anatomy of graceful interaction in man-machine communication. Technical Report, Computer Science Department, Carnegie-Mellon University, 1979.

[HOLLNAGEL83] Hollnagel, E. & Woods, D.D. Cognitive systems engineering: new wine in new bottles. International Journal of Man-Machine Studies, v. 18 (1983): 583-600.

[LALLJEE83] Lalljee, M. & Abelson, R.P. The organization of explanations. In: Attribution theory: Social and functional extensions. Oxford. Basil Blackwell, 1983: 65-80.

[MARCUS83] Marcus, R. S. An experimental comparison of the effectiveness of computers and humans as search intermediaries. Journal of the American Society for Information Science, v. 34 (1983): 381-404.

[McKEOWN85] McKeown, K., Wish, M. & Matthews, K. Tailoring explanations for the user. In: Proceedings of the International Joint Conference on Artificial Intelligence, Los Angeles, August 1985. Los Altos, Calif., William Kaufman, 1985: 794-798.

[CHANDRASEKARAN88] Chandrasekaran, B., Tanner, M.C. & Josephson, J.R. Explanation: the role of control strategies and deep models. In: Expert systems: the user interface, J.A. Hendler, ed. Norwood, NJ: Ablex, 1988: 219-247.

[REICHMAN85] Reichman, R. Getting computers to talk like you and me:... Cambridge, Mass., MIT Press, 1985.

[SHOVAL85] Shoval, P. Principles, procedures and rules in an expert system for information retrieval. Information Processing and Management, v. 21 (1985): 475-487.

[VICKERY86] Vickery, A., Brooks, H.M. & Vickery, B.C. An expert system for referral: The PLEXUS project. In: Intelligent information systems: progress and prospects, R. Davis, ed. Chichester, England, Ellis Horwood, 1986: 154-183.