AN INTELLIGENT INTERFACE FOR DOCUMENT RETRIEVAL SYSTEMS: DEVELOPING THE PROBLEM DESCRIPTION AND RETRIEVAL STRATEGY COMPONENTS.

Helen M. Brooks

Submitted for the degree of Doctor of Philosophy.

Department of Information Science,
The City University,
London.

June 1986.

Marco Polo describes a bridge, stone by stone.

"But which is the stone that supports the bridge ?" Kublai Khan asks.

"The bridge is not supported by one stone or another," Marco answers, "but by the line of the arch that they form."

Kublai Khan remains silent, reflecting. Then he adds: "Why do you speak of stones? It is only the arch that matters to me."

Polo answers: "Without stones there is no arch."

Italo Calvino
(Invisible Cities)

AN INTELLIGENT INTERFACE FOR DOCUMENT RETRIEVAL SYSTEMS: DEVELOPING THE PROBLEM DESCRIPTION AND RETRIEVAL STRATEGY COMPONENTS.

CONTENTS.

ACKNOWLEDGEMENTS.

I would like to thank all those who have contributed to this work and helped make it possible. In particular I would like to thank:

Steve Robertson, whose patient supervision and help saw me through this thesis.

Nick Belkin, whose ideas initiated this research and whose encouragement and support enabled it to reach completion.

Alina Vickery, who has over many years helped, advised, discussed and encouraged.

Penny Daniels, co-team member, transcriber and coder.

Angela and Jane, who were never to busy to discuss "how intermediaries do this.." and whose professional expertise informed much of this research.

The intermediaries at the University of London - at CIS and at the Institute of Education library - who agreed to be recorded.

Katy Stinton, who did some of the recording and transcription.

My parents, whose help (and child-minding) enabled this research to be carried out and written-up.

All the friends who variously encouraged, commented, discussed, baby-sat and supplied the coffee!

And my son, who has had to put up with a preoccupied parent and her "thesius" for quite long enough !!

ABSTRACT.

This research is aimed towards the eventual design and implementation of an intelligent interface for document retrieval systems. A number of functions have been identified as being necessary for intelligent interaction with the user. The research presented here is concerned with two of these functions in particular: Problem Description and Retrieval Strategy. Problem Description involves the development of a model of the user's problem. Retrieval Strategy is the selection and application of appropriate retrieval strategies to the knowledge resources (e.g. bibliographic databases). It is assumed that an intelligent interface should simulate the functional behaviour of a good human intermediary. Therefore the way in which human intermediaries carry out these two functions, and the knowledge resources required to enable them to do this, have been investigated. Audio-recordings were made of human user-human intermediary interactions in online search service settings. The transcripts of these interviews have been subjected to detailed functional discourse analysis. This analysis has been used to develop a specification for both functions and to identify the categories and specific knowledge resources each function would need to employ in order to carry out its tasks successfully. Interactions between these functions and other functions, such as User Modelling, were also examined.

Results show that both Problem Description and Retrieval Strategy comprise a number of subfunctions. Each function requires extensive knowledge resources, encompassing many different types of knowledge. Analysis of the interactions between the Problem Description and Retrieval Strategy function indicates that retrieval strategies are formulated on the basis of the problem description and conversely the developing retrieval strategy may also affect the way in which the problem description evolves. Formalisms for representing problem descriptions and the knowledge resources employed by the Problem Description are suggested. The implications of these findings for the design and implementation of the Problem Description and Retrieval Strategy functions in an intelligent interface, and for the design and implementation of the interface as a whole, are outlined.

I. INTRODUCTION

In ancient Greece, individuals or cities with problems sent a deputation to the oracle at Delphi. There they explained their problem to the Pythia (priestess) who communicated with the god Apollo and delivered the often highly ambiguous response back to the supplicant in written form. Today, researchers with problems visit online search services and explain their problem to an intermediary. The search intermediary develops and structures the user's description of his/her problem and uses it to formulate a request to an online document retrieval system. This request is then passed by the intermediary to the system and the system's response is delivered to the user.

These two information systems may be thousands of years apart and yet their basic structure is similar. There is a user, a wealthy Greek or researcher, a knowledge resource, Apollo or the document retrieval system, and an intermediary mechanism, the Pythia or the search intermediary. Together the knowledge resource and the intermediary mechanism comprise the information provision mechanism; the oracle at Delphi or the online search service.

As far as we know, the ancient Greeks accepted the necessity for the services of the Pythia. The debate about the necessity of the human search intermediary, however, has been raging ever since the introduction of online services. Many users would prefer to carry out their own searches and in some circumstances this would produce the most satisfactory results. Human intermediaries are needed however because effective accessing of the knowledge resource requires knowledge of the contents, structure and access mechanisms of the knowledge resource. Users normally do not possess this knowledge and since their use of the information provision mechanism is infrequent and irregular, are unlikely to acquire it.

Unfortunately it has not been in the interests of either producers of online databases or online host systems to carry out radical changes which would create more "user-friendly" systems. Recent moves towards making end-user searching feasible have concentrated on the development of front-ends to interface with the existing online systems. This process is most noticable in the development of "gateway" software (Janke, 1984). Gateway packages assist users with some of the mechanics of searching e.g. connection, logging-on, use of query language, but not with the conceptual aspects of query formulation. Despite the availability of such search aids, there has been no dramatic increase in the amount of end-user searching (Ojala, 1985). This suggests that the problems search aids address are not the main difficulties users experience with online searching.

It is the central assumption of this thesis that many of the difficulties experienced when searching online document retrieval systems are rooted in the inability of users to specify precisely what information they require and the demand by the system that they do so. Users may recognise that they have some problem and that their state of knowledge with respect to this problem is anomalous (Belkin, 1980), but are usually unable to say what would help to resolve that anomaly. If users do not know what they need to know, then they will be unable to specify what it is that they need. The access mechanisms of the retrieval systems, on the other hand, require that the user makes a request, expressed in a formal system language, that is an accurate description, in the systems' terms, of the documents that will help the user manage his/her problems. Bridging the gap between users with their problems and the system with its formal requirements, necessitates the services of an intermediary mechanism.

In order to mediate effectively between user and knowledge resource, the intermediary must have some understanding of both the user, his/her problems, goals, intentions and beliefs, and the knowledge resource and its contents, structure and access mechanisms. That is, the intermediary needs to construct models or representations of the user and of the knowledge resources, which can be used to guide the interaction with the user and direct the search of the system. The construction of such models is in part derived from the intermediary's own knowledge and experience and in part through the process of interaction and negotiation between user and intermediary (Belkin, 1984; 1985).

On the one hand therefore, an intermediary mechanism between knowledge resource and user is necessary and a human intermediary does more than simply translate a user request into a valid system request, but on the other hand, users would often prefer to do their own searches and sometimes this would be the most expedient approach. One solution would be to automate at least some of the tasks at present carried out by the human intermediary. That is, create a computer system which would mediate between the user and the knowledge resource. Such an interface would act as an expert assistant: helping users of document retrieval systems to carry out effective searches on online document retrieval systems and reducing the need to consult a human search analyst.

An intelligent interface in this context can be defined as a computer system which stands or mediates between the user and the online bibliographic systems. The interface must be able to act intelligently in the sense that it should accomplish the same functions as a good human intermediary (Daniels, Brooks & Belkin, 1985). This last point is important because in order to show the same functional behaviour as a human intermediary, the intelligent interface must employ, in some sense, at least part of the expertise, knowledge and problem-solving skills a good human intermediary would use in the same situation.

In order to achieve an intelligent interface for document retrieval systems it is necessary to identify and specify a minimal set of functions for such an interface, the knowledge required to carry out these functions and the structure of the dialogue between interface and user within which the functions are performed. To do this, it is necessary to find out what <u>human</u> intermediaries do, why and how they do it. That is, to specify the functions intermediaries carry out during the information interaction, the knowledge and expertise they use and how they use it (Belkin, 1985).

Since research (e.g. Belkin & Windel, 1984) indicates that interaction between the user and the intermediary involves model building by both participants within a co-operative dialog, it is also necessary to discover how intermediaries model the user and his/her problem, both in terms of the structure and content of the models themselves and the methods used to develop them.

For several years, research groups at the City University and at the Free University, Berlin, have been attempting to characterise the intermediary mechanism of an information system (Belkin, Seeger & Wersig, 1983; Belkin, Hennings & Seeger, 1984; Brooks & Belkin, 1983; Daniels, Brooks & Belkin, 1985). This research has resulted in the characterisation of the intermediary mechanism as a set of independent "experts". Each expert is responsible for carrying out a specific function, yet draws on the other functions for information necessary for accomplishing its tasks. These individual tasks all have the ultimate goal of providing an appropriate response to the user.

The functions defined as necessary for intelligent interaction with the user are: Problem State, Problem Mode, Problem Description, User Model, Retrieval Strategy, Response Generator, Dialogue Mode, Input Analyst, Output Generator (for a more detailed description see chapter II, section 5 and chapter III, section 1). For each of these functions, there will exist in the interface a corresponding "expert" with its own tasks, problem solving skills and knowledge.

The research presented in this thesis, is directed towards the development of an intelligent interface for document retrieval systems designed along the lines indicated above. That is, an interface which simulates the behaviour of a good human intermediary and which is based on a functional model in which each function is carried out by an independent expert which co-operates and interacts with the other experts. In particular, this thesis will concentrate on the development of two of the interface's functions:

(i) PROBLEM DESCRIPTION: developing a description of the user's problem.

(ii) RETRIEVAL STRATEGY: choosing and applying appropriate strategies to the knowledge resource.

These functions will only be considered within the context of the presearch information interaction between the user and the intermediary (human or machine). The other interface functions which will not be discussed or will be dealt with at a superficial level only, include: the natural language and dialogue handling components, user modelling, generating appropriate and informative explanations, and adjusting the search output along various user specified parameters.

Within these constraints, an attempt will be made to specify the Problem Description and Retrieval Strategy functions of an intelligent interface by investigating in detail:

(a) how human intermediaries carry out these functions;

(b) the sub-functions, goals and sub-goals of each function;

(c) the structure of the problem underlying the function;

(d) modelling activities carried out by the functions;

(e) the knowledge resources required by each function;

(f) a formalism in which the functions might be implemented in an operational interface.

The intention of this research is not to design an operational intelligent interface but rather to determine what such a design should encompass and what the constraints will be. This preliminary "knowledge acquisition" phase is a necessary pre-requisite to the development of a truely "intelligent" (and therefore useful) interface to document retrieval systems. Suggestions will be made as to how the two functions concerned might be implemented within the overall interface.

It will be assumed throughout that the information system's knowledge resources are bibliographic databases hosted by one of the commercially available online systems. This is the environment within which the research has been carried out. However it is envisaged that much of this research can be generalised and extended to other types of human-computer interface and in particular, to interfaces in other types of information system e.g. student advisory systems.

Chapter II reviews the state-of-the-art in interface design for document retrieval systems. Chapter III presents empirical research directed at the specification of the Problem Description and Retrieval Strategy functions and their knowledge resources. Formalisms for the implementation of these functions are discussed in Chapter IV. Conclusions and suggestions for further research are presented in Chapter V.

II. INTERFACES FOR DOCUMENT RETRIEVAL SYSTEMS.

1. INTRODUCTION

Online information retrieval is now such an accepted part of information and library practice that is is being referred to as a "traditional library tool". Yet, in the last decade, these systems have not become easier to search. On the contrary, ever increasing demands are made on the skills and knowledge of searchers. Areas of difficulty include:

- Complex logging-on procedures: the increase in telecommunication networks and database hosts has resulted in an equivalent proliferation of network addresses, system passwords etc.

- Query languages: the query languages of online information retrieval systems tend to be complex and effective searching usually requires a trained searcher. Moreover the increasing number of hosts, (at least 300 (Williams, Lannom and Robins, 1984)), and the lack of standardization means that few regular searchers can manage with a knowledge of one query language. In some subject areas, searchers report the need for familiarity with 6-8 query languages.

- Database selection: there are now upwards of 2,400 databases available online (Williams, Lannom & Robins, 1984). Several hosts have responded to this upsurge by providing a facility which matches user-entered terms with descriptions of database content and ranks the databases that seem most appropriate for searching. Such facilities are limited in scope and do not provide the depth of information on database content that experienced searchers possess. However, the growth in the number of databases means that one searcher cannot know the content of every databases and the more databases there are, the more difficult it becomes to establish what is held where.

- Database structure: lack of standardization in database design has resulted in enormous variations in the record structures of different databases and sometimes even within a database. Effective searching requires a high degree of familiarity with at least two or three frequently searched databases and the ability to use printed guides and information sheets to seek out information on less well known databases.

- Indexing languages and thesauri: Some databases index records using a controlled vocabulary of terms. These may be listed in a printed thesaurus. Effective searching on such a database requires a knowledge of how the indexers apply the controlled vocabulary and how to extract such information from the thesaurus. The intermediary must also be aware of changes in indexing policy within a particular database.

- Boolean logic: despite many years of research into alternative retrieval techniques, most commercial vendors have DBMS which use boolean logic as the primary retrieval technique. The concept of boolean operators is not easy to grasp and confusion arises between the operators and the english language terms "and" "or" and "not".

- Search strategy formulation: the cost of online means that search time must be kept to a minimum. This in turn requires efficient search strategies. The formulation of an efficient search strategy is a skill that is only built up through long experience with searching.

Other more general problems, common to many present-day computer systems, are the need for accurate and rapid keyboarding, the need to invest considerable amounts of time in learning how to use the system, the overall user unfriendliness, and the requirement to search frequently in order to maintain search skills.

The particular difficulties presented by online bibliographic retrieval systems have, in a sense, been by-passed by having trained, experienced searchers - human intermediaries - to interface between users and online systems. This has worked reasonably well but has always been recognised not to be the ideal solution. Many of the more routine searches could be carried out satisfactorily by end-users if the online systems were more amenable to use by inexperienced, irregular searchers. Further there are certain categories of search where it would be more desirable per se for end-users to do their own searching e.g. where the subject is specialised and searching requires the use of this specialised knowledge or where the database to be searched is more data/news oriented (Ojala, 1985).

In many institutions there are now insufficient intermediaries to meet the demand by users for searches. At the same time, many end-users are expressing interest in the idea of accessing online systems directly. Recent advances in information technology and related areas, have led to the ready availability of microcomputers, modems and access to international telecommunication networks. These advances have placed the <u>possibility</u> of searching within the reach of many end-users. The fact that this opportunity is infrequently taken-up and even more rarely pursued successfully, says much about the ease-of-use of online systems.

The growth of computer literacy amongst end-users means that many feel that they <u>ought</u> to be able to carry out their own searches and experience a real sense of frustration when this proves to be more difficult than anticipated. Many potential users do not have access to a search service, others are located too far away for the search service to be used conveniently and at short notice. For such users, self-searching seems an obvious solution. Another category of users dislike libraries, distrust librarians and would not use a library-based service (Ojala, 1985).

On the whole, end-users still choose to rely on the services of a human intermediary. There is evidence, for instance, that the level of user satisfaction with the search is higher if it was carried out by an intermediary (Hurt, 1983). Often end-users are disuaded from carrying out their own searches by the complexities of the online access mechanisms, the abundance of databases and the search process itself. End-users resist the use of thesauri, find it difficult to remember commands and dislike typing (Leipzig, Kozak & Schwartz, 1983). Another factor which mitigates against end-user searching is the increased cost of the search. Not only are inexperienced searchers slower, but end-users often prefer to "think" online, thereby lengthening the search time. For those end-users who have "taken the plunge", searching has often proved a costly and unsatisfactory experience (Borgman, Case & Meadow, 1985; Kirby & Miller, 1985; Peart, 1985; Tatalias, 1985).

Researchers have become increasingly concerned with issues such as:

> Given the increase in demands, both intellectual and institutional, being made on the intermediary, what can be done to provide more support and assistance?
>
> Given the technical feasibility of end-user searching, and the general desirability of end-user access, what can be done to assist users to carry out effective, satisfactory searches themselves?

The pressure to increase access to online document retrieval systems, and broaden the user base to include a larger proportion of end-users, has always been offset by the cost to the vendors of developing more "user-friendly" access software. Consequently the database management systems hosting the bibliographic databases are often perceived as being "fixed" and unalterable. Change is seen as lying in the development of interfaces and front-ends to these systems which would in some way assist searchers through the automation of some of the search tasks.

Most of the interfaces proposed or developed to date are "non-expert" in that they do not attempt to incorporate and use the expert knowledge and search skills of a human intermediary. Instead they aim to help users with tasks such as connecting with the remote computer, request entry, query language commands, search strategy formulation based on non-boolean retrieval algorithms. Recent developments in Artificial Intelligence have inspired researchers into contemplating the development of intelligent interfaces which embody some of the human intermediary's expertise. Several "expert" intelligent interfaces are now being developed. However, neither category of interface addresses certain crucial problems concerned with the problem domain itself. The domain, i.e. the whole search strategy formulation process from pre-search interview through to reference retrieval, is not well understood. Consequently, there is considerable debate as to what a interface <u>needs</u> to do in order to be of assistance, and how it should be organised to do it. Some research has therefore been directed to studying the problem domain "in situ" and attempting to draw from this some general conclusions about the tasks, functions and knowledge resources an intelligent interface should possess.

2. "NON-EXPERT" INTERFACES FOR ONLINE IR SYSTEMS

The design and development of interfaces for computerised document retrieval systems has almost as old a history as the online systems themselves - a clear reflection of the difficulties experienced in effectively accessing these bibliographic systems by both experienced and inexperienced searchers alike. All the early interface designs, and many of the current, are "non-expert". They aim to automate some search tasks without explicitly representing and using the knowledge and expertise of a human intermediary. The majority can also be said to be non-intelligent because the tasks that are automated are clerical and/or routine and because the way in which the tasks are carried out is by the application of an algorithm rather than by use of reasoning or problem-solving skills. This is not to say that such interfaces are not worthwhile or useful, many are, but to indicate that they are limited in what they achieve and usually concentrate on automating or assisting with one small aspect of searching.

"Non-expert" interfaces to document retrieval systems can be categorised according to the search processes they aim to automate into the following groups:

(a) Communications and request entry: i.e. gateway software. Typically packages include automatic dialling and logon facilities, off-line search formulation and downloading.

(b) Natural language access: enabling users to enter their queries in natural language rather than using a formal query language.

(c) Search strategy formulation: i.e. systems which wholly or partly construct a search strategy given some initial information by the user.

(d) Searching online: systems which monitor the user's actions online, provide explanatory error messages and detect "faulty" search step sequences.

These categories are not exclusive but indicate the main emphasis in the development of the interface.

2.1. Gateway software.

Gateway software is essentially dedicated communications software which allows microcomputer users to access remote computer systems via a telecommunications network. The earliest gateway packages were communications software and nothing else. These were soon expanded to included other facilities such as auto-dial, automatic logging-on, assistance with request data-entry and the downloading of retrieved records. More recent packages have included additional options such as editing/word processing and database management software (Levy, 1984; Kehoe, 1985). Some gateway software is produced by, or with, the bibliographic database producers. Understandably such software tends to be specialised for searching on the databases of the producer and adapted to make use of particular database features. Gateway software is now also being produced by independent software houses and these products tend to be more general in scope, able to search a number of vendors, but therefore not able to take into account the special features of any host or database. Descriptions of some of the better known gateway packages are presented below.

Sci-Mate

Sci-Mate (also known as the Universal Online Searcher), is a gateway package developed by the Institute for Scientific Information (ISI). It has two major components: a communications and request data entry system; and a database management system. The communications subsystem gives the user the choice of entering his/her request either in menu-driven mode through a series of menus, or in direct query mode using the host's own query language. Users can access databases held by five vendors: ISI, BRS, DIALOG, SDC and NLM. Retrieved references can be downloaded and then automatically reformulated into the database management system's database format. The database management subsystem gives users the facility to create personal databases from the retrieved references, which can be searched and sorted (Stout & Marcinko, 1983).

The menu-driven mode allows the user to select options such as the database to be searched or the type of search to be carried out. Users must specify keywords but the system adds the appropriate system commands. Whilst menu-mode has the advantage of freeing the user from having to memorise host commands, the Sci-Mate menu system has been criticised as being cumbersome in nature with a substantial number of menus to which the user must respond. For those who have already mastered the program's procedures this becomes tedious (Levy, 1984). The other main criticism is that during the menu-selection process, the user remains online. This makes the search costly, particularly for non-US users who have to pay additional international telecommunication charges. Experienced users can avoid these problems by switching to direct command mode but then, of course, the user-support is lost. Other difficulties are that users must select the appropriate Boolean operators themselves and must remember the file numbers and file name acronyms for databases (Lamb, Auster and Westel, 1985). Although Sci-Mate allows access to a wide variety of hosts, it does not support all the search facilities available on these systems, limiting the sophistication of the search that can be attempted.

Search Helper.

Search Helper, produced by Information Access Company (IAC), is designed to serve as a gateway to several of the IAC databases on Dialog e.g. Magazine Index, National Newspaper Index, Management Contents, Newsearch. The software is menu-driven and users are required to enter the numbers of selected menu options or to respond to Yes/No questions. The user's selections are then "translated" into a Dialog database search. The search formulation can be stored on disk for future use. Retrieved items may be downloaded and stored for review and printing (Evans & Pisciotta,

1985).

Search Helper assumes that users' searches will be simple and that the user will take more complex ones to an intermediary. The user must outline the search topic through selections from the sequence of menus. Brief instructions are given on how to enter keywords and the user must then proceed to enter his/her query. The search strategy can be entered offline which reduces search costs. Search Helper does not exploit the full power of the Dialog query language. The boolean operator "NOT", limiting, sorting and display format choices are not available, nor is the downloading of retrieved items (Levy, 1984).

An evaluation of Search Helper was carried out by the library at Carnegie Mellon University (Evans & Pisciotta, 1985). End-users approaching the library's search service were asked to carry out their own searches using the package. The findings showed that end-users liked using Search Helper and found no significant difficulties in using it. Since the searches were offered at no charge, it is not clear to what extent this influenced the user's responses. The survey found infact that few users expressed willingness to pay for the service (despite the relatively low cost - $2.50 a search on average).

<u>Insearch</u>

Insearch has been developed by a software company (Menlo Corporation). It employs graphics and windowing techniques to create visual effects. A list of commands is displayed at the bottom of the screen and the user can select one by placing the cursor over it. Insearch is able to access all Dialog databases and makes use of all the capabilities of Dialog's query language. The package includes a tutorial disk and disks containing the Dialog "bluesheets" describing the databases. Requests may be entered offline and help is given with selecting an appropriate database. Criticisms of this package have included the fact that many find it difficult to use (despite the comprehensive tutorial program included with the package) and the narrowness of the range of microcomputers for which the package is available. There are also complaints that the graphic displays change too quickly (Levy, 1984).

The latest version of this software has been renamed Pro-Search and is intended for intermediaries rather than end-users. The system has been expanded to include the option of accessing BRS. Users can now choose to search using the host system's query language as well as Pro-Search's own query system.

CAN/OLE Interface

The CAN/OLE gateway software is designed to search databases on CAN/OLE, one of the most used host systems in Canada, which has a difficult query language. The software is written in PASCAL for an Apple IIe microcomputer. It has a menu-driven data entry system supplemented with an easy-to-use query language. Not all CAN/OLE facilities are accessible through the interface. The user is allowed access to seven basic operations: selecting a database, examining database vocabulary, searching, displaying retrieved records, ordering offline prints, displaying search histories, asking for assistance and logging-off. The user can enter his/her request either in menu-mode or in command language mode. The idea is that inexperience users will make use of the menu-driven mode and experienced users will make use of the command language mode. The user is given assistance with logging on and step-by-step guidance and explanations of commands, are provided (Lamb, Auster & Westel, 1985).

MicroDISCLOSURE

MicroDISCLOSURE is specifically designed to assist with searching the Disclosure databases produced by IAC and which are held on Dialog. Queries are specified using menus. and there is no alternative command mode (cf Sci-Mate). Not only does this prove a cumbersome mode of searching for experienced users, but the user is required to interact with the menus online. Since Disclosure databases are expensive, this adds considerably to the search costs. Several of the host query language facilities are not supported e.g. Boolean "NOT" or index browsing. Retrieved references can be downloaded, and reformatted into databases supported by DBMS software included as part of the package. A spreadsheet analysis program is available as an additional package and can be used to analyse downloaded data (Levy, 1984).

Connect

Connect is an intelligent terminal package developed by Learned Information Ltd. specifically for use with online information retrieval and electronic mail services. It is able to access databases provided by BRS, the Source, Pergammon Infoline and Dialog. It can make use of auto-dial modems, which simplifies logging-on procedures. Search strategies can be entered and stored offline and then released and executed once the appropriate database is logged-on. A help facility provides on request brief explanations to the user about how to use the system.

EasyNet

EasyNet facilitates access to a wide range of hosts. At present, ten database processors can be used, including Dialog, BRS, NLM, Pergammon Infoline, SDC and VU/TEXT. Queries are specified via a series of menus. The user is required to answer menu questions about the topic or author of interest. A suitable database is selected by the system. The user then must enter the search terms and these are then reformulated by EasyNet into queries couched in the database host's query language. Once the designated database has been logged into, the query is automatically executed.

Gateway software packages have proved popular if for no other reason than that they allow users to carry out online searches using a microcomputer. Thus libraries, institutions and home users can use one piece of equipment for a searching, word processing and variety of other tasks. In addition, most gateway packages offer facilities which support the searcher, albeit at a non-conceptual level e.g. auto-dialling, automatic logging on, the entry of search formulations offline. Auto-dialling and automatic logging-on save the searcher much frustration and the need to memorise long passwords and network addresses. The entry of search formulations offline, allows the user to enter the request in his/her own time and at no cost. Many gateway systems also enable the retrieved items to be downloaded and stored as a personal file for subsequent use.

Useful as these facilities are, it is interesting that the gateway producers seem more concerned to promote the request-entry support facilities - usually some form of data entry by menu selection. Menu-driven modes can assist inexperienced searchers by suggesting possible options and by generating the appropriate system commands. This is some remove though from being able to handle requests entered in natural language, despite claims such as:

> " (it is a) menu-driven system that permits searches in plain English of the major commercially available databases " (Stout & Marcinko, 1983).

> " these interfaces effectively translate a search statement formulated in plain English into the common command language of the system being searched " (Levy, 1984).

Not only do these systems <u>not</u> possess natural language understanding capabilities but the menu driven modes can often not support the full range of query language facilities. Experienced users complain that menu-driven modes are cumbersome with too many menus to work through. Most unpopular is software which requires the user to enter the request online in menu-mode. Further, users of gateway software are often required to select the appropriate boolean operator themselves and must decide which terms to enter in their search statement. The intellectual onus very definitely remains with the user.

So, whilst gateway software is a useful tool for experienced users with respect to facilities such as downloading and offline request entry, the help offered with search formulation is often mis-directed and inappropriate. Too much is left to the user for real assistance to be afforded to inexperienced users while experienced searchers find menu-driven modes tedious and limited.

2.2. Natural language interfaces.

Natural language understanding is one of the "pillars" of artificial intelligence research and a great deal of research has been carried out in this area (Waltz, 1982; Grishman, 1984). Part of that research has been concerned specifically with developing natural language interfaces for database systems (Rich, 1982). The function of such an interface is to reformulate the user's natural language questions into a formal search specification which can be applied to the database (Sparck Jones, 1982). Natural language understanding is seen as making possible the construction of user-friendly interfaces which will allow users to make requests to the database without needing to be familiar with the DBMS's formal query language. A study carried out in the early eighties found over one hundred natural language understanding systems had been developed, of which a substantial proportion were natural language interfaces (Kaplan, 1982).

Many natural language interfaces to database systems, (sometimes known as query systems because they are restricted to processing natural language search requests), are now operational e.g. LUNAR (Woods, Kaplan & Nash-Webber, 1972), REQUEST (Plath, 1976), INTELLECT (Harris, 1984), ROBOT (Harris, 1979), TQA (Damereau, 1980). They are designed for use with hierarchical or relational databases and to retrieve factual information given queries such as "how many blue wiggles were made in 1985" or "who is the head of the sales department and what is her salary". Such database systems have highly formal, complex, database semantics and correspondingly formal query languages. The natural language interface is often tailored to the semantics of a particular database, with consequent portability problems. An exception is the system developed by Boguraev and Sparck Jones which combines a powerful general natural language analyser with a relatively restricted database specific translator (Boguraev & Sparck Jones, 1982).

Very little work has been done to develop natural language interfaces for online bibliographic retrieval systems, which represent quite a different application domain from factual databases. Most interfaces to document retrieval systems, intelligent or otherwise, use menu or query language driven request modes. (For exceptions see IR-NLI (3.2) and PLEXUS (3.5)) What work has been carried out in this area has been mainly concerned with automatic vocabulary mapping, employing little or no linguistic analysis or knowledge. An example of this approach is the CITE system developed at the National Library of Medicine, NLM, (Doszcos, 1983). CITE is intended to facilitate end-user searching of CATLINE, the online catalog file of NLM, by mapping the user's request into controlled vocabulary terms. The user's natural language request is first stemmed and once the word roots have been identified, a "best-match" search algorithm retrieves close matching MesH terms. Additional look-up procedures are used to identify subheadings that might be included in the term list. Text words, MeSH headings and subheadings are then displayed, ranked in order of degree of similarity to the original request terms. This static vocabulary mapping is complemented by dynamic associative mapping based on user relevance feedback about displayed records. At present, CITE has no syntactic analysing capabilities.

One of the only natural language interfaces for document retrieval systems that employs linguistic analysis is that developed by Tait (Tait, 1983). It uses as a basis the natural language analyser developed by Boguraev and Sparck Jones. (Boguraev & Sparck Jones, 1982). The user's natural language request is translated into an internal representation (a case-labelled dependency structure) by the natural language analyser. An attempt is made to identify and isolate the content part of the user's request. The result is then passed to a term extraction processor which makes use of linguistic information to do the extraction and is sophisticated enough to extract "oscillators using transistors" as a concept from the phrase "high frequency oscillators using slow switching germanium transistors". Fragments are passed onto the inference processor which can infer if prepositions can be inserted and can generate variants of the term. The outcome of the term extraction and inference processing is passed to the english language generator which converts the internal representation back into english. The set of english strings is then processed into a search formulation.

Natural language interfaces have not been seen as a serious option for online bibliographic retrieval systems and very few have been developed, either as part of some larger system or in their own right. None of the gateway packages, for instance, has anything approaching natural language understanding capabilities. There are a number of possible reasons for this. The most significant is that the natural language interfaces developed for factual databases are limited to very narrow subject domains (and hence vocabulary subsets) and usually to specific database semantics. Most bibliographic databases, even when they cover quite narrow domains e.g. concrete or aluminium, still involve larger vocabulary profiles than most natural language interfaces can manage. The lack of explicit database semantics such as are found in relational databases creates additional problems, particularly with respect to disambiguation. On an implementation level, natural language interfaces have not been more extensively investigated because they involve a substantial effort to develop. However, the real limitation with natural language interfaces per se, is that while they may prove useful for translating a natural language request into a system query, they do not assist in the process of getting from the user's problem to that request. Nor do they take into account factors such as the desired form of the outcome, number of items retrieved, appropriate database to search. To be effective, an interface to a document retrieval system must assist the user and that requires more than natural language understanding. The Unix Consultant (UC) help system, for instance, employs planning and goal analysis as well as semantic analysis (Wilensky, Arens & Chin, 1984).

2.3. Interfaces which assist with search strategy formulation.

The design and development of improved retrieval techniques has been a topic of research for some years. In general the researchers are concerned to develop a total information retrieval system which uses one or more "advanced" technique e.g term weighting, document ranking, relevance feedback. Now some of this research is being incorporated into the design of interfaces to conventional online bibliographic retrieval systems, thus enabling the "advanced" techniques to be employed without adjustment to the host systems.

The Cirt system is just such an interface. It was developed at the City University as part of a British Library funded project and is at present undergoing testing and evaluation (Robertson et al.,1986). Cirt is intended as a tool for research and development, showing the feasibility of using particular techniques in "real" searching environments, rather than as a commercial product. Cirt applies probabilistic retrieval methods to searches carried out on a conventional database hosted by a system using traditional boolean based retrieval methods. The user logs onto the online host, through the interface, and enters a list of terms (natural language or descriptors). Cirt determines the frequency of each of these terms (by carrying out a single term search on the database) and uses this information to calculate a weighting for each one. The weighting and ranking algorithm embodied in Cirt constructs boolean search statements which are then applied to the database. The retrieved documents are shown to the user one-by-one, in ranked order, and the user is asked to evaluate each item for relevance. Documents judged relevant are analysed and the information used to re-weight the search terms. The user can also add or delete terms if required. The search can then be re-processed.

On completion of the current project, if the interface proves effective, it is likely that future interfaces, expert or otherwise, will employ probabilistic methods of retrieval, perhaps as part of a repertoire of search techniques (Croft & Thompson, 1985).

An interface which also assists with search strategy formulation but which has rather different aims and objectives is CONIT. CONIT (Connector for Networked Information Transfer) was designed and implemented at M.I.T. (Marcus & Reintjes, 1981a; 1981b). The original intention was to provide a uniform virtual interface to online systems so that, as far as the user is concerned, a single system is being searched. Searches are carried out using a common command language which CONIT translates (using a rule-table) into the appropriate query language command. The common command language is based on english language words and a command-argument structure. Only the basic search functions of each online host are supported. To ensure cross-file compatability, the user must select free-text search terms. CONIT also provides instruction and other search aids for the user. Early trials showed that untrained users could manage to perform their own searches on CONIT, although these searches were less effective than those carried by an experience intermediary.

Further enhancements to the system were then made. These included the automatic saving of all searches (so that they could be re-run at a later date), and some individualised database searching (converting authors' names into database specific formats). In addition, an automatic keyword/stemming search function was developed. Each term in the user's phrase is stemmed and then searched for individually. The resulting sets are then combined using AND. If any of the subsets has no postings, CONIT suggests scanning alphabetically similar terms in the index. If too many postings are retrieved, CONIT replaces the stem with the original word and searches again. This simple, automatic searching mode seems to be of assistance to inexperienced users, although the search times were rather lengthy - an average of 81.7 minutes

(Marcus, 1983).

2.4. Interfaces which provide online assistance.

The interfaces so far reviewed are concerned with mapping the user's input into appropriate search terms, or constructing a search strategy or carrying out the search automatically. The task of assisting with the online search itself, where the role of the interface is to monitor and advise the user rather than do the search has been neglected. The exception is IIDA (Individualise Instruction Data Access) which was developed within a training/teaching context. IIDA is designed both to teach inexperienced searchers how to search and to provide assistance with the carrying out the online search. (Meadow, Hewett & Aversa, 1982a; 1982b).

The IIDA software is adapted from CONIT (see 2.3). Unlike CONIT, IIDA only accesses DIALOG and makes use of DIALOG's own query language (cf CONIT's common command language). When carrying out a search, the user enters commands by way of IIDA but uses the DIALOG language and, in normal circumstances, receives the same response as if s/he was logged in directly to DIALOG IIDA monitors the user's interaction with DIALOG, examining both commands and responses looking for errors or indications of problems. If any of these occurs, the user is offered help facilities which are designed to offer assistance with the specific problem in hand. For instance, the standard DIALOG error messages have been supplemented so that they are more meaningful to the user and more indicative of where the error lies and of what should be done next. Since IIDA does not "know" anything about the user, as an individual, or his/her problem, IIDA's assistance is limited to what can be detected by monitoring the user-host dialogue.

One of the most interesting features of IIDA is that it is able to detect "faulty" search patterns (Toliver, 1981). The diagnostic support programs embody knowledge about which search patterns indicate problems or failures. A "healthy" cycle is defined as a progression of command types from BEGIN to SELECT-EXPAND to COMBINE to DISPLAY-TYPE. A new cycle begins when the search reverts to a prior command type. The diagnostic support system contains two command components: the threshold analysis program (TAP) and the warning control program (WCP). TAP monitors the user's search looking for features such as:

- consecutive commands of the same type
- creation and use of zero-postings sets
- use and non-use of sets created in the current cycle
- repetition of commands
- delays by users in entering commands

TAP incorporates about 50 rules specifying particular search behaviour faults. A given search pattern may trigger one or more of these rules. An algorithm determines whether there is sufficient evidence that the behaviour is faulty and if the threshold is reached, WCP is called. The WCP program processes the error code sent by TAP. Error codes are translated into error messages and if more than one message is to be sent, these are joined by a connecting phrase. When the WCP recognises severely recurring problems, it can direct control to offer the help facility, issue an EXPAND command or even automatically log-off.

3. INTELLIGENT INTERFACES FOR DOCUMENT RETRIEVAL SYSTEMS

Since the early eighties, several projects have been initiated to develop intelligent interfaces for online information retrieval systems, based on expert system techniques (Clarke & Cronin, 1983; Yaghmai & Maxin, 1984; Belkin & Vickery, 1985; Kehoe, 1985). Expert systems are computer-based systems which attempt to embody and use some of the knowledge and skills of a human expert. The formal definition of an expert system proposed by the BCS Expert System specialist group states:

> "an expert system is regarded as the embodiment within a computer of a knowledge-based component from an expert skill in such a form that the system can offer intelligent advice or take an intelligent decision about a processing function.
>
> A desirable additional characteristic, which many would consider fundamental, is the capacity of the system, on demand, to justify its own line of reasoning in a manner directly intelligible to the enquirer. The style adopted to attain these characteristics is rule-based programming."

Other definitions are to be found in Feigenbaum, (1979), Addis, (1981); Hayes-Roth, Waterman & Lenat (1983); Sparck Jones, (1983a); Brachman et al., (1983); Johnson, (1984); Muller, (1985). Unfortunately the phrase "expert system" has been used very widely to refer a wide range of computer systems, from the very sophisticated to the rather simple with seemingly little in common. The confusion over terminology, and the lack of any very precise definition, means that given a knowledge-based system, there may well be no consensus as to whether that system should properly be termed an expert system. Moreover there are certainly systems called "expert" which are no more than knowledge-based, if that!

The attraction of expert systems lies in the fact that they provide a means to incorporate and use human expert knowledge in a computer system, thereby making the the development of "intelligent" systems a possibility. The domains to which expert system techniques have been successfully applied, though, are areas such as medical diagnosis, geological prospecting and computer configuration, (Feigenbaum, 1979; Hayes-Roth, Waterman & Lenat, 1983; Buchanan & Shortliffe, 1984) which are very different from that of an interface to an information retrieval system. For example, conventional expert systems act directly to help solve a user's problem. They recommend a course of treatment, or suggest candidate molecular structures for unknown compounds, or indicate where ore deposits might be found. The role of an intelligent interface to a document retrieval system, however, would be to help the user find documents in which there might be the information which will help the user better manage his/her problem.

Other critical differences include the very wide subject domain covered by most document retrieval systems. Conventional wisdom is that expert systems should be built for domains which are well-defined, narrow and discrete. Little work has been done on expert systems which can cope with domains which are broad but relatively shallow, and which involve large numbers of concepts and relations (Welbank, 1983). Another difficulty is the language-dependence of document retrieval systems. Knowledge in traditional expert systems is expressed "incidentally" in linguistic form. The same is not true for documents or document descriptions. "Information conveyed by a document is conveyed by the specific linguistic expressions of that document" (Sparck Jones, 1983b).

The successful use of expert system techniques in this domain may not be possible until such techniques have been further advanced. A further constraint is lack of knowledge about the domain itself which makes it unclear how the problem solving process might be simplified (and therefore incorporated into a system using existing techniques). Consequently it is surprising how little of the current expert system-interface work is explicitly based on a preliminary study of either the problem domain or the human intermediary and his/her knowledge. To quote Fidel (1986):

" one wonders why the experience of intermediaries has been neglected by system designers when it is an important source of knowledge."

Most of the work on intelligent interfaces for document retrieval systems is still at the design stage or the systems are operational on an experimental basis only. An outline of some current work in this area is presented below:

3.1. Expert

EXPERT was developed by Marcus′ team at M.I.T. and implemented in MacLisp on the M.I.T. Multics system (Marcus, 1981; Yip, 1981; Marcus, 1985). It is a rule-based system, concerned with automating query formulation and online searching. It embodies a number of rules for deciding which database to search , how to select synonyms and how to combine terms e.g.

```
IF: the field of the search topic is Cancer research
    and therapy
THEN:
    it is definite (1.0) that CANCERLIT is the
      appropriate database, and
    there is strongly suggestive evidence (0.8)
      that CANCERPROJ is appropriate, and
    there is suggestive evidence (0.7)
      that MEDLINE is appropriate, and
    activate RULE d1000 for next consideration
```

It is intended that EXPERT should:

(i) enable users to represent their search topic;

(ii) assist users with selecting a database to search;

(iii) automatically construct a search strategy from the topic representation and execute it;

(iv) use relevance feedback to reformulate the search strategy;

Representing the user's search topic is achieved by asking the user to enter concepts associate with the topic and then to suggest terms to express each concept (see fig.2.1)

The immediate task is to set up an initial search strategy for the search problem.

....

Let's find a major concept of your search topic:
Terminate your input with a space followed by the dollar sigh symbol ($) and a carriage return.
------------> linguistic theory $
Let's denote <linguistic theory> by CONCEPT-1
Give another major concept of your search topic (if there is none, answer nil):
------------> language learning $

....

Please give another keyword for this concept (if no more keyword, answer nil):
-----------> transformational grammar $

Figure 2.1: a screen from EXPERT (1)

The topic concepts and other terms entered by the user are automatically processed and transformed into a keyword/stem Boolean search strategy. EXPERT displays the outcome to the user (see fig. 2.2).

```
----------------------------------------------------------------
  CONCEPT-1                  AND              CONCEPT-2

| linguistic theory          | |      language learning        |
|     OR                     | |            OR                 |
|                            | |                               |
| transformational grammar   | |      language acquisition     |
|                            | |                               |
|    OR                      | |                               |
|                            | |                               |
| generative grammar         | |                               |

Does the above table correctly represent the major concepts and
their respective search terms that you have supplied?
Please answer yes or no:
----------->
----------------------------------------------------------------
```

Figure 2.2: a screen from EXPERT (2)

Database selection is carried out by asking the user to indicate the topic area, one of eight broad areas displayed on the menu e.g. social sciences, education, humanities. On the basis of this choice, the user is then presented with a display of subjects within the topic area e.g. education for exceptional children, education media and materials, and asked to select one or more subject areas that are related to the search topic. EXPERT then ranks its list of databases according to their prospective relevance to the user's search topic (as indicated by the user's choice of subject areas). The user is then shown the four databases the system thinks most relevant and is able to select one database from this list.

Once a database has been selected, EXPERT automatically connects to the online host and then the search strategy formulated by EXPERT is translated into a sequence of search statements. Terms within a concept are connected with "OR" and the query prefixed with the appropriate system command. Search statements are transmitted to the online host without user intervention. The number of postings obtained is compared with the number postings desired by the user and if there is a discrepancy, EXPERT suggests to the user how the search might be refined. In addition, brief citation details of the first ten documents are displayed to the user for evaluation. Items evaluated positively are redisplayed and the user asked to select any terms from the title or abstract which seem useful. The user must then modify the concept table, using the system's suggestions and/or adding the selected terms. The modified concept table is used to form a new search strategy which is then re-executed. This process can be re-iterated until the user is satisfied with the results.

Testing of EXPERT was carried out on a very limited basis and it does not appear to have been evaluated formally. Search times were in the order of 30-40 minutes (very long by European standards). Problems with lack of robustness of the system are reported, particularly when there are less-than-perfect conditions on the telecommunications network or with the hosts. Yip (1981) lists the limitations of EXPERT as: able to do subject searches only; unable to formulate database dependent strategies; lacking sufficient user involvement in search strategy reformulation; unable to suggest search terms.

The EXPERT project was not taken further but certain design features were later incorporated into CONIT (Marcus, 1985).

The rules in EXPERT concerning database selection were evolved after some discussion with information specialists (Yip, 1981). These are the only rules that appear to embody search heuristics. The remaining rules concerned with constructing the concept table, displaying messages, logging-on, formulating the strategy etc. could have been implemented using non-rule based methods. Apart from its use of production rule techniques, EXPERT seems to be little more than a gateway plus relevance feedback retrieval system. With the exceptions of the construction of search statements and assistance with database selection, the intellectual effort is still largely on the user's side.

3.2. IR-NLI

IR-NLI (Guida & Tasso, 1983) is intended to be an intelligent interface for commercially available online bibliographic retrieval systems, which will allow users to enter their request in natural language. The natural language processing facilities are supplemented by using search tactics such as those described by Bates (1979a; 1979b).

The user's request is parsed and the relevant concepts extracted. Additional terms are then selected for each concept, which the user can accept or reject. A search strategy is automatically generated. Terms within a concept combined using OR and the concept groups themselves combined with AND. Appropriate system commands are also assigned.

It is not clear to what extent this interface has been implemented (see Sparck Jones, 1983). However the design suggests an advance on that of EXPERT since the IR-NLI user is not required to break his/her request down into the component concepts. An extract from a sample session is presented in fig. 2.3 (Taken from Sparck Jones, 1983).

3.3. An expert system for IR [Shoval]

Many online bibliographic databases use controlled vocabularies to index the documents. A search strategy may require the use of these terms but expertise is needed to select terms appropriately from the thesaurus. The system developed by Shoval (1985) is designed to assist users with the selection of terms for a query. It employs a knowledge base which "contains the knowledge available to the human expert" - a thesaurus and additional knowledge of terms, meanings and associations. This knowledge is represented as a semantic network. The usual hierarchical and synonymity relations found in thesauri are enhanced by the addition of two other types of relation, generator and modal. Generator relations link the components of multi-word phrases e.g. the terms "information" and "retrieval". Modal links

---

> PLEASE SPECIFY YOUR REQUEST
* I would like to have information about intelligent terminals for online information retrieval
IR : concept-1 : INTELLIGENT TERMINALS
concept-2 : ONLINE INFORMATION RETRIEVAL

> DO YOU PREFER AN HIGH RECALL OR AN HIGH PRECISION SEARCH ?

....

> BUILDING BLOCK APPROACH SELECTED
> TRYING TACTICS :
SELECT
PARALLEL
RELATE
RESPELL
RESPACE
> THE CONCEPT "ONLINE INFORMATION RETRIEVAL" COULD BE SPLITTED INTO:
- ONLINE
- INFORMATION RETRIEVAL
PLEASE VALIDATE
* yes
> THE FOLLOWING ADDITIONAL TERMINOLOGY IS PROPOSED FOR THE CONCEPT "ONLINE" :
INTERACTIVE, REALTIME
PLEASE VALIDATE
* ok

Figure 2.3: a screen from IR-NLI

are used to express more general relations e.g. the relation between "business" and "organisational structure".

The user decides which terms represent his/her query and enters them one-by-one. The system then conducts a search of its knowledge base and additional terms of probable relevance are selected. An evaluation process is applied to distinguish terms which are most representative of the user's request. Using a variety of algorithms, the selected terms are ranked. A further procedure is then activated to determine the order in which terms should be presented to the user, taking into account various term features. The user is able to select the terms which s/he regards as relevant. The list of suggested and accepted terms is printed out and can be employed by the user when formulating his/her query.

Shoval intends to extend the system's capabilities so that it will handle user requests entered in natural language and will organise the final selection of terms into a query which would be accepted by the online host. The present system employs more powerful term selection heuristics than either EXPERT or IR-NLI and does so as a co-operative effort with the user. Whilst the system "knows" about the vocabulary of particular subject domains and some of the relationships between the concepts, it does not appear to embody knowledge about how these terms are applied by the indexers to documents in the database. Thus however well the final set of terms may express the user's request it is no indication as to whether useful documents will be retrieved using these terms. Moreover it is doubtful whether the semantic knowledge in the knowledge base can be described as "expert" since most of it appears to be knowledge common to english language speakers. A more detailed critique of this approach is given in Fidel (1986).

3.4. CANSEARCH

CANSEARCH, (Pollitt, 1981; 1984; 1985), is designed to construct search strategies for end-users wishing to search the CANCERLIT database on subjects related to cancer therapy. The rationale behind the project was to develop an interface which would enable medical practitioners to access directly the online databases which cover the cancer literature.

The major database in this area is CANCERLIT, produced by NLM, which contains records indexed with MeSH. In order to search CANCERLIT effectively it is necessary to use MeSH headings. However, it is difficult for first-time or infrequent users to select appropriate MeSH terms to express their query. The thesaurus is not easy to use and they are unlikely to be sufficiently aware of indexing policies to select suitable terms.

Another difficulty for users is the mechanics of entering their queries. The spread of microcomputers has resulted in increased familiarity with keyboarding, but entering lists of technical terms, which must be spelt correctly, and may extend to more than 30 characters in length, is no simple matter. Misspellings and slow query entry can considerably lengthen a search and therefore the search costs. To overcome these difficulties, CANSEARCH uses a touch-screen interface. By "touching" a term on the screen, the user automatically selects that term and incorporates it into his/her query.

Pollitt's system assumes that a cancer therapy search topic involves certain facets e.g. site of tumour, therapy, type of tumour. Using these facets as a framework, the system guides the user through relevant sections of the MeSh thesaurus and user touches the screen to select one or more of the displayed terms. The facet structure also guides the combination of terms with Boolean operators to form the search strategy. The outcome is a more sophisticated strategy than those produced by either EXPERT or IR-NLI. For example:

Test query:
Ovarian granulosa cell cancer treatment

Search formulation

"SUBS APPLY TH,DT,RT,DH,NU,PC,RH,SU
1: OVARIAN NEOPLASMS
2: GRANULOSA CELL TUMOR
3: 1 AND 2
"SUBS CANCEL
4: CRYOTHERAPY OR HYPERTHERMIA INDUCED OR PALLATIVE THERAPY OR TERMINAL CARE
5: OVARIAN NEOPLASMS
6: GRANULOSA CELL TUMOR
7: 5 AND 6
8: 4 AND 7
9: 3 OR 8
10: 9 AND HUMAN

Pollitt considers CANSEARCH to incorporate the following kinds of knowledge:

(i) Subject knowledge: the system knows what categories of concept are likely to be present in a request concerned with cancer therapy. This knowledge is employed to provide the facet meta-structure for assisting users enter their requests.

(ii) Indexing knowledge: the system knows about MeSH and about how the indexers apply these terms.

(iii) Knowledge about the online query language: the system knows about the query language commands and how these should be used.

(iv) Knowledge about searching: the system knows how constraints such as recall, precision, specificity and exhaustivity affect search strategy formulation.

Future plans include the incorporation of knowledge about users, their journal preferences and previous searches, and about the environment - which periodicals are available locally etc.

CANSEARCH is implemented in PROLOG on a PRIME 750 minicomputer. Some 40 rules are required to handle the different display frames. There are 20 frame selection rules and a further 35 rules which deal with search statement generation. CANSEARCH is at present completing testing and evaluation. It represents a useful tool for end-users and intermediaries alike, since the construction of effective search strategies using MeSH is a difficult and complex task.

3.5. PLEXUS - an expert system for referral.

The expert system for referral project, funded by the British Library Research and Development Department, is being carried out at Central Information Service (CIS) at the University of London. It has produced a working prototype called PLEXUS (Vickery et al., 1986). The overall objective of the project was to construct an expert system for a library application which could be used to demonstrate to librarians the nature and potentialities of expert systems. The referral task was chosen as a suitable application because it is a task common to all libraries and should therefore have widespread appeal (Vickery & Brooks, 1983). Since it would not be feasible to develop an intelligent referral system which covered the whole range of possible topics, the prototype is restricted to one subject area - gardening/horticulture.

Referral is the process by which a librarian guides a user with a problem to information sources, usually outside the library itself, which potentially might provide the user with the means of managing or even resolving his/her problem. The information sources, (or referral resources), may include other libraries, institutions, societies or even human experts. In most libraries, information about referral resources is found on a card index or in printed guides and handbooks. To facilitate access, it was decided to compile a list of referral resources in gardening and to store it as a database on a standard microcomputer DBMS.

Thus PLEXUS differs from most of the other interfaces described in this section in that it does not interface with a knowledge resource which is a database of bibliographic references as such, but rather with a database of referral resources i.e. of information sources themselves. However since some of these referral resources are documents and since some of the ideas could be readily transferred to an interface to a document retrieval system, it is included here.

The task of the expert system for referral, PLEXUS, therefore, is to assist users who has some problem to identify information sources which potentially might be of use in helping to resolve that problem. The design of the PLEXUS is based on a functional model comprising five functions:

(i) constructing a model of the user;

(ii) obtaining a description of the user's problem;

(iii) formulating, and if necessary reformulating, a search strategy;

(iv) presenting the results of the search to the user in an informative manner and obtaining the user's evaluation of them;

(v) providing the user with explanations of the system's capabilities, the system's activities and the search outcome.

Only the first four of these functions have been developed to date. Each function has been implemented as a separate sytem module with its own set of knowledge resources. The GETUM module is concerned with user modelling, the GETSTAT module with obtaining a statement of the user's problem, SEARCH with search strategy construction and reformulation, and EVALUAT with presenting the outcome of the search and obtaining the user's evaluation of it. The system also has at its disposal a number of resources:

- a dictionary of terms in gardening, horticulture and related areas;

- a stoplist of terms;

- the database of referral resources in gardening;

- a hierarchical classification of the subject domain (the BSO classification system).

At the start of a session the user is presented with a brief explanation of the scope and capabilities of PLEXUS. The GETUM module then acts to construct a model of the user. The user characteristics represented in this model are:

(i) the extent of the user's knowledge of the PLEXUS system;

(ii) the relationship between the topic of the query and the user's job;

(iii) the extent of the user's practical experience in the field;

(iv) the user's familiarity with information resources in the domain;

(v) what advice seeking activity the user has undertaken already;

(vi) the user's geographic location (in broad terms);

Once the model has been constructed, and the user is satisfied that it is accurate, the system proceeds to try to develop a model of the user's problem. The user is not asked to indicate the referral resources s/he would like to consult, nor to present the system with a formal query, but rather to describe what the problem is about. The user is presented with an open question of the form:

"Please tell me a little about your problem or interest"

The user's response may be couched in more or less any form. A single term, list of terms, phrase, sentence or sentences are all equally acceptable. Procedures in the OPEN subfunction take the initial input, remove any stopwords, stem the remaining terms and then search for these terms in the dictionary. If the terms are present in the dictionary, OPEN creates a data structure for that term and places in it information found in the dictionary about the semantic category to which the term belongs, synonyms and/or other equivalent terms and any pointers into the BSO classification. Terms which are not recognised are placed on a separate stack for possible processing later.

At this stage, one of three subfunctions - STRAT1, STRAT2 or STRAT3 - may be activated, depending on the proportion of recognised terms in the input. The flexibility afforded by this multi-strategy approach means that only in a very few cases will the system be unable to obtain an adequate problem statement, no matter what the user's initial input.

If PLEXUS is unable to recognise any of the terms in the input, (or there are no terms left after the input has been processed through the stoplist), then STRAT3 is called. STRAT3 causes a display of concepts in the domain to be presented to the user (i.e. a display of sections of BSO). The user is able to browse through this display and select any terms that seem appropriate. This is intended to assist users who have difficulty with standard data entry and also to serve as an outline of the system's scope. If the user is unable to find anything suitable, PLEXUS will inform the user that it is unable to help and terminate the session. If terms have been selected, these will be treated in the same way as recognised terms entered by the user in the initial statement.

If PLEXUS recognises only a small proportion of terms, STRAT2 is called which takes each of the unknown terms in turn and tries to identify it. PLEXUS "knows" the types of concept to expect in statements about the subject domain. The user is asked a series of question about the term, to try to identify the semantic category to which that term belongs. For example the user may be asked "Is it a plant?" or "Is it a part of a plant?" or "Is it the name of a tool used in gardening?". Once the category of the unknown term has been ascertained, it can be incorporated into the system's dictionary and the term can be treated as a recognised word.

If PLEXUS recognised most of the terms in the input it calls STRAT1. This does some preliminary disambiguation of multi-meanings and then checks that the range of concept categories present is "reasonable". For each of the terms, a frame is instantiated. The structure of the frame depends on the semantic category of that term. Rules attempt to fill as many of the slots as possible from existing information. The set of instantiated frames constitutes the system's model of the user's problem.

Procedures check the problem model for sufficient completeness. Sufficient completeness is where there is sufficient information in the problem model to construct an effective search strategy. If the model is incomplete, rules direct what questions should be posed in order to elicit the missing information. If the model is judged sufficiently complete, a set of rules uses the model to generate a confirmatory statement i.e. a statement of what the system understands about the user's problem. If the user indicates that this statement is correct, the system can proceed to the next function, if not the whole GETSTAT module is reiterated.

Once a sufficiently complete model of the user's problem has been obtained, this is then transformed into a search strategy. The SEARCH function maps the elements of the problem model into a boolean query which the DBMS search procedures will accept. The DBMS and database of referral resources are in no way modified for PLEXUS and PLEXUS has no effect on the retrieval process itself.

The number of items retrieved by the initial application of the query to the referral database is assessed. The SEARCH function employs knowledge of search strategies and tactics (in this case for small databases), to broaden or narrow the search if necessary. For instance if there are too few postings the system may:

ENTER A BROADER DESCRIPTOR: The BSO classification is used to find broader or narrower descriptors for terms

EXCLUDE CONCEPTS: PLEXUS employs knowledge about which are the most and least important categories of concept for searching in this domain to determine which concept to drop first.

ADD SYNONYMS: Information is available in the dictionary about synonyms or equivalent terms. These can be added to the query to broaden it.

The re-formulated query is presented to the DBMS for processing. The outcome is again assessed and further modifications made if necessary.

When an acceptable number of items has been retrieved, the EVALUAT function organises the display of these items to the user. Information in the user model is used to assist with this process e.g. institutions located near to the user are displayed before those located some distance away.

PLEXUS has been implemented on a SIRIUS I microcomputer with a 20Mb external hard disk and 850Kb of internal memory. The software is written in PASCAL and currently runs to some 10,000 line of code. The prototype system took approximately 40 person months to develop (Vickery et al., 1986).

3.6. Other intelligent interface projects

Smith and Chignall (1984) intend to develop an expert system which will concentrate in the first instance on helping users of Chemical Abstracts to select appropriate search terms. The knowledge base will embody the knowledge of an expert Chemical Abstracts searcher concerning the choice of suitable search terms (and, implicitly, knowledge about how the indexers apply these terms to the database records). To construct such a knowledge base, they plan to develop software which can re-organise the Chemical Abstracts database into abstract scripts, where the scripts reflect the deep structure underlying the document descriptions. Using the knowledge base, it would be possible to ascertain all the values of a particular slot/primitive and Smith and Chignall propose that on the basis of such information it would be possible to construct an effective query formulation. Inference rules and heuristics would be employed to assist with this process.

Crystal & Jakobson (1982) are constructing a FRont End for Databases (FRED) which will connect users to hosts and provide them with a common interface to all the online hosts. The user would enter his.her request in natural language and this would be translated into the syntax required for each of the databases and hosts to be accesses. Fred will comprise three major subsystems:

(i) a user interface: which processes user input and generates system responses.

(ii) a database interface: which interacts with the target databases

(iii) a knowledge base: which contains facts and heuristics about users and available databases.

The knowledge base will contain information about previous user queries, the structure of databases, command languages, which vendors host which databases and protocols for connecting to the online systems. One objective of this project is to provide the front-end with enough intelligence to understand the end-user's problem so that it can help locate the data required.

Some initial work has been reported on an expert system information counsellor (Obermeier & Cooper, 1984) and on an expert system search intermediary (Walker & Janes, 1984).

4. INTELLIGENT IR SYSTEMS : RELATED RESEARCH

Work relevant to the development of intelligent interfaces for document retrieval systems has been carried out in the design of systems which are complete information systems in their own right. In particular, THOMAS, the ASK project and GRUNDY have been influential in the design of more recent interfaces (e.g. see Croft & Thompson, 1985).

4.1. THOMAS

THOMAS (Oddy, 1977a; 1977b) was designed to assist users with query formulation. Oddy was concerned with developing a means for retrieving bibliographic information which did not require the user to enter an explicit query. Underlying this objective was the assumption that users have difficulty putting requests to information retrieval systems because doing this involved specifying what the user does not know.

THOMAS was designed to simulate the behaviour of a human librarian during the pre-search interview. It has a model of the way in which the literature of the subject domain is organised. This model is implemented as a network of associations between documents, authors and subject terms. THOMAS interacts with the user and during this process, forms an image of the searcher's interests, derived from its world model. Refences are selected for display according to the state of the user model. The user's reactions to these retrieved items are sought and the user model modified accordingly. If the user rejects several references in turn, this is taken to mean that the user model must be altered substantially. In order to determine how to do this, THOMAS searches for an earlier reference that the user evaluated positively and which is as different as possible from the current user model. This reference is then shown to the user for reconsideration. If there is no suitable reference, the user is shown the topic he/she has specified and any associated search terms. If no explicit topic has been entered, the user is asked to suggest a new term.

THOMAS is very flexible and allows searchers to take as little or as much initiative as they want. There are no commands to learn and no formal requests to make. In effect the user browses through the document collection, but this browsing is directed and aided by the system.

4.2. GRUNDY

GRUNDY is not an interface as such in its present form. However it has been influential in the design of intelligent interfaces and since it is document oriented it seems appropriate to discuss it here. GRUNDY was developed by Rich (1979; 1983) to test out ideas about user modelling in interactive systems. In particular, she was concerned to explore the idea of building models of individual users with the aid of stereotypes and then exploiting the resulting model to adapt the system's responses to an individual user.

The application domain chosen was that of a library system which recommends fiction books to readers. This domain involves a heterogeneous set of users and a highly flexible, non-trivial task; anyone of a large number of books could be selected. The intention was to come as close as possible to simulating the performance of a librarian, at a public library dealing with a user who has walked in and asked for something to read,(given the limitations of the computer system which cannot make judgements about the user based on visual characteristics or speech patterns).

In order to achieve its objectives, GRUNDY makes use of two data collections. One collection contains descriptions of individual books . Each description comprises a set of facets with attached values. The other collection is the set of user stereotypes. Each stereotype comprises a number of facets and their associated values, which refer to characteristics which might have a bearing on people's choice of books.

GRUNDY constructs models of individual users on the basis of information provided by the user. This process involves making inferences from the user's statement e.g. the user's name is used to infer whether the user is male or female. The user is asked to type in a few single words about his/herself. These words trigger other stereotypes e.g. sports-person, radical. The stereotypes selected are compiled into a single user synopsis, which contains all the facets in all the activated stereotypes. Against each facet is a value and a confidence rating. The values indicate whether the characteristic represented by the facet is positive or negative e.g. the facet tolerate-sex can take values ranging from -5 (no tolerance) to +5 (indicating total acceptance). The confidence values (0-1000) indicate how certain GRUNDY is that the values assigned to the facets are true for a particular user. The more GRUNDY knows about a user, the more specific its knowledge, the higher the confidence ratings. A sample stereotype is given in fig. 2.4.

Sorts-person stereotype.

Activated-by : Atheletic-word-trigger
Generalization : ANY-PERSON

FACET	VALUE	RATING
Motivations		
Excite	800	600
Interests		
Sport	900	800
Thrill	5	700
Tolerate-violence	4	600
Romance	-5	500
Education	-2	500
Tolerate-suffering	4	600
Strengths		
Physical strength	900	900
Perseverance	800	600

Fig. 2.4: a GRUNDY stereotype.

Once there is a sufficiently developed user model, GRUNDY uses the model to select books. A facet is selected from the user synopsis which has a very high or a very low value, and associated with this value a high confidence rating. This facet is then searched for in the inverted index to the database of documents.

All the books suggested by the facet are selected and the description of each book is then matched against the user model. The book whose description most closely matches the user model is then selected. GRUNDY informs the user of the title and author of the book and asks the user if s/he has read it before. If the user answers in the affirmative, GRUNDY then asks if s/he liked it. If the user has read the book and likes it, GRUNDY reinforces the aspects of the user synopsis which match the book description. If the user has read the book and did not like it, GRUNDY tries to find out the reason by asking the user which aspects (i.e. facets) of the book the user disliked. This information is used to adjust the values and confidence ratings in the user synopsis. If the user has not read the book at all, GRUNDY suggests it as a possible choice. Again, if the user indicates s/he doesn't like the sound of the book, GRUNDY tries to refine the user synopsis.

GRUNDY was not originally conceived as an expert system, though in present terms it almost certainly would be counted as one. Clancey (1985) has commented, however, that GRUNDY does not use expert heuristics in its problem solving, since the document descriptions are couched in the same terms as the user model, e.g. "fast plots" is a book characteristic but in the implementation "likes fast plots" is also a user characteristic, associated with a person stereotype. Clancey considers that the relation between a user characteristic, represented in a person stereotype, and a document attribute, represented in the document description, is heuristic and should be distinguished from abstractions of people and books (Clancey, 1985). Precisely such a separation would be required if GRUNDY were to be developed as a true interface, since no commercially available document retrieval system represents its documents in the same way as GRUNDY. GRUNDY would need to incorporate additional knowledge to be able to infer how a user model might be matched against existing document descriptions. The real importance of GRUNDY is that it demonstrated the feasibility and the importance of considering user characteristics in the retrieval process. User modelling is now seen as a vitally important function of any intelligent interface (e.g. Sparck Jones,

1984).

4.3. The ASK project

The overall objective of the ASK project is to design and implement a retrieval system which incorporates several theoretical ideas about information retrieval systems, including the ASK hypothesis (Belkin, Oddy & Brooks, 1982a; 1982b). The ASk hypothesis states that users engage in information seeking behaviour because they recognise that they have some gap or anomaly in their state of knowledge (ASK) about some topic or situation. Since the user does not know, or is uncertain, about something s/he will be unable to specify what is required to resolve the anomaly. However, most conventional information retrieval systems require users do precisely that - specify the information they require (Belkin, 1977; 1980).

A major concern of the ASK project is to develop means for representing ASKs such that these representations can be used as the basis for retrieval. A method for generating graphical representations of users' ASKs has been devised based on a textual analysis of users' problem statements. The problem statements are collected by asking users to talk about their problem. A statistical text analysis program is then applied to these problem statements to identify relevant concepts, their relations and the strength of association between concepts. The resulting association matrices can then be transformed into a network representation, with concepts signified by nodes and associations by arcs. Strength of association is indicated by arc length and width. These representations were found to have structural features which seemed to indicate the presence of particular types of anomaly. For instance, a node with a large number of weakly associated concepts or a group of strongly associated concepts unconnected with the remainder of the network (Belkin, Oddy & Brooks, 1982a; 1982b; Hapeshi & Belkin, 1985). Further research is at present underway to categorise ASK representations in terms of structural features which can be used to define search strategies (Belkin & Kwasnik,

1986).

The importance of the ASK project in this context is that it explores methods for representing user's <u>problems</u> and the environment surrounding such problems. Interface research has tended to concentrate on automating the processing of user's requests leaving the onus of formulating such a request to the user. In ASK terms, it is not reasonable to expect the user to do this. It is interesting that a similar conclusion was reached by the designers of the UC (Unix Consultant) help system. Conventional operating system help facilities (manual or online) require users to specify the command they are interested in, and then display information pertaining to that command. But, the command <u>name</u> may be precisely what the user wishes to find out. UC allows users to describe what they wish to accomplish and tries to put forward a plan to achieve these goals (Wilensky, Arens & Chin, 1984).

5. DISTRIBUTED EXPERT PROBLEM TREATMENT

Most intelligent interface design, in the context of interfaces for information systems, is based on two assumptions:

- that users can specify the information they require;

- that the tasks involved in transforming the user's request into a search strategy can be implemented as a linear sequence, with some simple back-tracking if necessary.

The effectiveness of these interfaces depends, crucially, on the extent to which the user is able to formulate a request i.e. a statement which describes the information the user requires to solve his/her problem. The inability of most users to do this successfully, because the gaps or uncertainties in their knowledge preclude it, (see section 4.3), has serious consequences therefore. At best, the interface allows the user to enter a descriptive request statement in natural language, although minimal natural language processing, in any linguistic sense, will be carried out on it. Often the user is restricted to entering single terms or must indicate the main concepts of his/her request. The system's ability to assist the user even with this task is limited because the system has no understanding of the individual user or of his/her problem (Meadow, Hewett & Avesa, 1982a).

The elicitation and use of information about users, their status, goals, intentions, beliefs and so on, during the reference/presearch interview is something that reference librarians and human intermediaries do as a matter of course (Taylor, 1968; Auster, 1983; Daniels, 1985; White, 1985). Taylor (1968), for instance, found that librarians felt it was important to ascertain the motivations of the inquirer. Inquirers may not be able to define what they want but they are usually able to explain why they need it. The user's personal background was seen as helpful in determining the urgency of the search, the level of any dialogue, how critical of the outcome the user is likely to be etc, thereby providing a context for the search. Librarians also tried to find out the kind of answer that would be acceptable to the user. With the exception of Grundy (section 4.2) and PLEXUS (section 3.5), these aspects of the intermediary's task in the presearch/reference interview and their effect on the outcome, have not been taken into account in intelligent interface designs.

Similarly, the linearity of event sequencing during human-human presearch interviews has been refuted by numerous researchers (e.g. Hitchingham 1979; Smith 1979; Cochrane, 1981). Certain tasks are more likely to be carried out at particular stages in the process but it is difficult to predict, given a task, which other task has preceded or will succeed it. There are large amounts of backtracking, reiteration and side-sequencing. This has been ignored by system designers who have produced highly linear, sequential systems. Clearly the development of an intelligent interface which approaches human functional behaviour requires an understanding not only of <u>what</u> tasks are carried out by human intermediaries but also a model of functional interaction which will account for the complex patterns observed and which can be then incorporated into the design of the system.

<u>5.1. The MONSTRAT model.</u>

The MONSTRAT project sought to address the problem of what an intelligent interface must do, in order to understand the user and the user's situation and thus provide an appropriate response to the user. A functional analysis of information interactions was carried out. These interactions were taken from a wide range of situations, including student-advisory interactions, rent-advice situations, online intermediary-user presearch interviews. (Belkin, Seeger & Wersig, 1983; Belkin, 1984). The basic assumption behind this analysis was that any functional model should reflect the functional behaviour of a "good" human intermediary or advisor or counsellor.

The analysis identified a minimal set of functions that take place in information interactions (see fig 3.1). The MONSTRAT model specifies that an intelligent interface for document retrieval systems needs to:

(a) understand the state the user has reached in the problem solving process;

(b) determine the kind of resources or system capabilities appropriate for this user and problem;

(c) construct a model of the user, his/her goals, intentions and experience;

(d) develop a description of the user's problem and what the user knows about it;

(e) choose and apply retrieval strategies to knowledge resources.

This is a far more extensive view of the tasks of an intelligent interface that that of most other interfaces for document retrieval which have tended to concentrate on the retrieval strategy aspects.

The MONSTRAT model is too complex to conceive of as a unitary design. Using this model, a number of information interactions were analysed and it was found that there was no simple sequencing of functions and that the functions themselves interacted in complex ways. It was proposed, therefore, that in an intelligent interface, each function would be carried out by a seperate "expert". These experts would be independent but able to interact with each other; co-operating to achieve a single goal. This idea of distributed expert problem treatment builds on distributed problem solving approaches developed in Artificial Intelligence research to handle complex problems.

To investigate the general validity of the proposed model and to resolve the issue of architecture and control mechanisms, a number of simulation experiments were carried out. In the simulation, each of the functions was assigned to a human. A well-specified problem was presented to this human-interfaced information system, and the human experts co-operated with each other to try to resolve it. The functions interacted with each other, sending messages as appropriate, and communicated with the user until the user was helped or gave up.

The simulation experiments confirmed the validity of the overall functional model. The most effective control structure was the blackboard model in which messages from experts are posted on a "blackboard" where they can be read by other experts monitoring the blackboard. This type of control structure was employed in the Hearsay II speech understanding system (Erman & Lesser, 1979). The pattern of interactions observed in the simulation experiments closely followed the type of temporal patterning observed in human-human information interactions. The sequencing of the functions was circumstance driven and the activation of a function depended on output from several functions being available (Belkin, Hennings & Seegar, 1984).

Thus the MONSTRAT model of an intelligent interface specifies a distributed system of functional experts. Each of these experts is a particular function, one of a minimal set of functions, organised to communicate with one another through a blackboard structure until their goal of helping the user with his/her problem has been achieved. This model represents the most complete functional specification of an intelligent interface to date. Its power lies in its derivation from analysis of human intermediary/advisor behaviour in information interactions and from its stated aim of simulating that functional behaviour.

An interface based on the MONSTRAT model has yet to be implemented. Many issues remain to be resolved e.g. the detailed specification of the individual functions; specification of the knowledge resources required by each function. However, some of the central ideas i.e. distributed expert problem treatment and blackboard architecture, have been employed by Croft in his expert assistant which is at present under construction (see below).

5.2. An expert system assistant for IR [Croft]

This project, still under development, is being carried out at the University of Massachusetts by Croft and his team (Croft, 1985; Croft & Thompson, 1985). They intend to build into a document retrieval systems an expert assistant which knows about different types of users, methods of formulating queries, different retrieval strategies and so on. The assistant will help the user to formulate his/her query and will provide a number of search techniques for retrieving documents. The functions of the system are seen to be:

(i) Constructing and maintaining user models. These models will represent the characteristics of individual users.

(ii) Constructing and maintaining a model of current information need. This request model is built during query formulation and may be modified by subsequent user feedback.

(iii) Selecting search strategies based on information in the request and user models.

(iv) Presenting information to the user and obtaining feedback. The type of information displayed and how it is displayed will depend on the request and user models and on the capabilities of the display mechanisms.

(v) Providing explanations of the system's actions.

The system will comprise three major components:

An INTERFACE MANAGER: whose task it is to collect information from the user and to display to the user information from the system.

SYSTEMS EXPERTS: which will independently carry out specific functions but will interact and share information.

The KNOWLEDGE BASE: which is divided into Long term memory (LTM) and short term memory (STM).

These three components will interact within a blackboard style architecture similar to that employed by HEARSAY-II (Erman & Lesser, 1979). The blackboard is equivalent to the system's shprt term memory. It store information pertaining to a particular session only. The system's long term memory will consist of the document collection and statistical information about it, user models for particular individuals and user stereotypes. The documents in the collection are represented in a network formalism.

Croft envisages that seven system experts will be required:

(a) BROWSING EXPERT (BE): this expert will provide an informal way for the user to search for documents and to look at other parts of the long term memory. The evaluations given by the user to documents seen when browsing will be posted to short term memory and may be used by the other experts.

(b) EXPLAINER (EXP) : which will allow the user to interrogate the system about its actions, the models it has constructed and about its use, capabilities and scope.

(c) THESAURUS EXPERT (ThE): the goal of this expert is to find more specific descriptors for any higher level concepts found in the user's request. These narrower terms can be used to make the request more specific.

(d) REQUEST MODEL BUILDER (RMB): this expert must construct and maintain a model of the user's information need. Its sources of information are evaluations by the user of terms, documents, authors, concepts, (posted by the other experts), and some statistical data about the document collection. The Request Model Builder will also compute weights for the terms in the user's request. Three weights will be generated for each term: the frequency of the term in the user's statement; a user supplied emphasis weight (where the user has indicated that the term is of particular interest or importance); and a weight computed using the probabilistic retrieval model.

(e) USER MODEL BUILDER (UMB): whose function is to build and maintain models which describe aspects of the user e.g. journal preference, and in particular, the user's knowledge of the subject area.

(d) SEARCH CONTROLLER (SC): the search controller must select and apply formal search techniques e.g. cluster searching, extended boolean searching, citation searching. Which retrieval techniques are applied will be decided on the basis of the user and request models.

(e) NATURAL LANGUAGE EXPERT (NLE): this expert transforms the natural language input from the user into an internal form - basically a list of stemmed terms and their frequency in the input.

The expert assistant is being implemented on a VAX workstation in Common LISP. Once the infrastructural software supporting the blackboard architecture is complete work will commence on the design of the individual experts. Issues such as the user characteristics which will be modelled or the nature and levels of explanations to be given, have yet to be determined.

III. SPECIFYING THE PROBLEM DESCRIPTION AND RETRIEVAL STRATEGY COMPONENTS OF AN INTELLIGENT INTERFACE FOR DOCUMENT RETRIEVAL SYSTEMS.

1. INTRODUCTION.

The ultimate goal towards which this research is directed, is the development of an intelligent interface for document retrieval systems. The design of this interface will be based on the assumption that the goals or functions which are present in information interactions between human intermediary and human end-user should be simulated in human-computer interactions. An intelligent interface should act "intelligently" in that it should be able to carry out the same set of tasks and should accomplish the same functions that a good human intermediary would in the same situation.

Given this basic assumption, it follows that designing and building an interface depends on identifying and specifying what a human intermediary in an information provision mechanism actually does and needs to do, for the search to be effective and the outcome satisfactory for the user. What are the goals and tasks undertaken by an intermediary during the pre-search interview with the user? What information does the intermediary need to obtain from the user? What knowledge does the intermediary bring to the situation herself? How are the two elements combined to form the intermediary's cognitive model of the situation? These are the issues that must be investigated.

The theoretical basis for this study derives from two earlier research projects, ASK and INSTRAT. The ASK project (Belkin, Oddy & Brooks, 1982a; 1982b) was concerned with the issues of non-specifiability of information requirements (see chapter II section 4.3). Users are able to recognise that their own state of knowledge with respect to some problem is anomalous and will be able to describe the problem itself but are usually unable to state precisely what is required to resolve that anomaly. The ASK hypothesis has important consequences for interface design. In particular, it indicates that users should not be required by the interface to present a request but rather they should be given the opportunity to describe their problem.

The functional model on which the interface will be based is that developed by project INSTRAT. The INSTRAT project derived a functional model of information interactions from an abstract analysis of what an ideal intermediary mechanism must know and do, in order to respond in the best possible manner to the user's request for assistance in managing his/her problematic situation. Later work involved simulating, testing and evaluating this model (Belkin, Seeger & Wersig, 1983; Belkin, Hennings and Seeger, 1984; Brooks and Belkin, 1983). The outcome of the project was the specification of a minimal set of functions that must be performed by the intermediary mechanism of an information provision mechanism (see fig.3.1), a suggested design for the implementation of an information provision mechanism based on this model and a schema for the analysis of information systems.

Subsequently, the MONSTRAT model has been used to analyse human-human interactions such as student advisory sessions, radio phone-in advice sessions, (Belkin & Windel, 1984) and online presearch interviews (Brooks & Belkin, 1983; Price, 1983). It has also been used to identify and define a problem-structure which might drive human-computer dialogues and assist with the interpretation of the human-user input (Daniels, Brooks & Belkin, 1985).

In order to build an operational interface based on the models and assumptions outlined, it is necessary to investigate a number of problem areas at a detailed level. In particular, the following issues need to be resolved (Brooks, Daniels & Belkin, 1985).

(i) **Function Specification:** Identification and detailed specification of the individual functions.

FUNCTION NAME	DESCRIPTION
1. Problem State (PS)	Determine the position of the user in problem treatment process, e.g. formulating the problem, problem well specified.
2. Problem Mode (PM)	Determine appropriate mechanism capability e.g. document retrieval
3. User Model (UM)	Generate description of user type, goals, beliefs, knowledge e.g. graduate student, thesis.
4. Problem Description (PD)	Generate a description of the problem - the type, topic, structure, environment etc.
5. Dialogue Mode (DM)	Determine the appropriate dialog type and level for the situation e.g. menu, natural language.
6. Retrieval Strategy (RS)	Choose and apply appropriate retrieval strategies to knowledge resource.
7. Response Generator (RG)	Determine propositional structure of response to user which is appropriate to the situation.
8. Explanation (EX)	Describe mechanism of operation, restrictions etc. to the user as appropriate.
9. Input Analyst (IA)	Convert input from the user into structures usable by other functions.
10. Output Generator (OG)	Convert propositional response into a form appropriate to the user, situation and dialog mode.

Figure 3.1 : (after Belkin, Hennings & Seeger, 1984).
The functions of an intelligent interface for document retrieval systems.

The functions which are necessary for successful interaction need to be discovered, and each of them needs to be specified in terms of the tasks it is supposed to carry out and how these tasks should be accomplished. This needs to be done to a level of detail sufficient to allow implementation of the function.

(ii) **Knowledge Resources:** Identification and specification of the knowledge resources required for each function.

Each function will require some knowledge to enable it to achieve its goals. It is necessary to discover what kinds of knowledge are necessary for each function and what specific knowledge is required, at a level of detail which will allow machine representation and use of the knowledge.

(iii) **Function Interaction:** Specification of the interactions among functions.

The functions contribute information to one another. It is necessary to identify which functions interact with one another and how. This needs to be done at a level of detail which will allow the construction of a communication and control architecture.

(iv) **Model Building:** Identifying the types of model that are necessary to successful information system performance.

Interaction between user and intermediary in successful information system performance seems to involve model building by both participants, with special emphasis on the intermediary's building of appropriate models of various aspects of the user and his/her problem. The method of representation for each type of model must also be specified.

(v) **Dialogue Structure:** Specification of the dialogue structure.

A means for driving and interpreting a human-computer dialogue in the information interaction context needs to be specified.

Accomplishing all these goals would provide a framework within which the intelligent interface could be implemented.

The design of a complete intelligent interface is beyond the scope of this study. Instead the emphasis will be on the accomplishment of the first four goals for two of the functions:

Problem Description and Retrieval Strategy.

That is, for both the Problem Description and Retrieval Strategy functions an attempt will be made to specify these functions, identify their knowledge resources, analyse their interactions with other functions, and define their model building activities.

2. METHODS

2.1. Background.

The aim of this research is to specify two of the functions of an intelligent interface for document retrieval systems and the knowledge these functions require to achieve their goals. The successful development of an intelligent knowledge-based system, in whatever application area, depends on the ability of the knowledge engineer to elicit the human expertise, problem solving skills and knowledge required to tackle problems in that domain. What appears on the surface to be a relatively straightforward task turns out in practice to be extremely difficult, time consuming and complex (Buchanan, 1982, Welbank, 1983). This is because:

- human knowledge is complex, messy and ill-formulated.
- experts find it difficult, if not impossible, to articulate their knowledge
- the more expert the human expert, the more "unconcious" their problem solving
- knowledge elicitation techniques are poorly understood and of limited applicability

Within the field of Artificial Intelligence itself there is little in the way of methods or techniques that could be used to facilitate knowledge elicitation. Attention has been turned to other disciplines where knowledge transfer is of interest e.g. cognitive psychology.

The main techniques for knowledge elicitation can be summarised as follows:

(i) Interviewing the expert: either informally or making use of structured interviewing techniques.

(ii) Verbal protocol analysis: that is, analysing recordings of experts thinking aloud as they carry out a task.

(iii) Observational studies: observing and recording the behaviour of the expert as s/he works on a real problem, in their normal working environment,in as unobtrusive way as possible.

(iv) Machine induction: that is, the generation of rules by a machine on the basis of a set of case studies presented to it.

In addition, some use has been made of "psychological" techniques based on multi-dimensional scaling (Gammack & Young, 1985; Shaw, 1985). A review of knowledge acquisition methods is presented by Welbank (1983).

Interviewing is a technique frequently used to elicit knowledge from human experts. It provides a relatively simple method of acquiring a substantial amount of basic knowledge about a problem domain, particularly concerning the concepts and terminology of the domain.

Interviews can be "structured" to various degrees and in different ways. The expert, for instance, can be asked initially to prepare and deliver an introductory lecture. More structured interviews can then be used to probe relevant areas in greater depth (Gammack & Young, 1985). The asking of specific questions, though, requires that the knowledge engineer possesses a certain amount of knowledge about the problem domain already. Where this is not the case, structured interviewing may not be possible.

Particular questioning techniques that have been made use of in acquiring expert knowledge include generalized checklists, critical incident reporting and autobiographical memory procedures (Gammack & Young, 1985); goal decomposition and reclassification (Grover, 1983).

Verbal protocol analysis is the analysis of recordings of experts thinking aloud as they carry out a task. In true protocol analysis, the expert describes his/her thoughts and impressions but does not explain them. Frequently the task given to the expert is a standard case or even an artificial problem. The use of verbal protocol analysis to acquire expert knowledge is common and has been used in the construction of several expert systems (e.g. Clancey, 1981; Kuipers & Kassirer, 1983).

A comprehensive survey of the use of verbal protocol analysis was carried out by Ericsson and Simon (1980). This work is primarily concerned with research methods in cognitive psychology but its findings, and the authors' reflections on appropriate methodology are relevant to knowledge elicitation in other contexts. They note that under a variety of circumstances, verbal reports may omit information that subjects use to perform a task, particularly with respect to intermediate stages in the problem solving process and/or where the task has been repeated many times. Further, if subjects are working under a heavy cognitive load they tend to stop verbalising or the content of what they do provide is rather superficial and incomplete. However although the incompleteness of reports means that some information is missing, it does seem to invalidate what information was obtained. Verbalising during problem solving was shown to affect the cognitive processes of the subjects only if the subjects were required to verbalise information they would not normally attend.

One disadvantage of protocol analysis is that having to verbalise may interfere with what the expert is doing (Berry & Broadbent, 1984). This will vary from person to person and from problem domain to problem domain. Ericsson & Simon (1980) found that interference does not normally occur to any significant extent. This will obviously not be the case where the problem solving process itself involves verbalising e.g. speaking to a client.

Observational methods involve observing and recording the behaviour of the expert as s/he works on a real problem, in their normal working environment, in as unobtrusive way as possible.

Observational techniques are probably the least used of all knowledge acquisition techniques because they tend to be extremely time consuming and require complex, indepth analysis (Welbank, 1983). However they do have the advantage that they can be used to discover what the expert actually does to solve problems in his/her normal working environment. They are also useful for extracting information about the role of the expert and the ordering of tasks.

One example of the use of observational techniques in the knowledge acquisition context is the investigation into student-advisor interactions carried out by Coombs and Alty, with a view to developing an expert system advisor (Alty & Coombs, 1980; Coombs & Alty, 1980; Coombs & Alty, 1984). In this study, random samples of conversations between advisors (experts) and students (clients) were recorded at five university computer centres. The participants were interviewed afterwards and questioned about the conversation itself and on the use of any documentation during the interaction. Participants were also questioned about the goals behind individual utterances and the success with which these goals were achieved.

Several programs have been developed which automatically derive rules form case studies and examples (Quinlan, 1973; Michalski, Carbonell & Mitchell, 1983). The first use of machine induction to derive rules for an expert system was METADENDRAL. This system generates the mass spectroscopy fragmentation rules for DENDRAL by analysing a library of cases. New rules are then incorporated into the DENDRAL knowledge base (Buchanan & Feigenbaum, 1976). The use of machine induction techniques have so far been restricted to the development of small scale expert systems. The effective use of these techniques requires a highly structured, well-defined problem domain and a large library of case studies.

The two most commonly used methods of knowledge acquisition are verbal protocol analysis and interviewing the expert. Both have their limitations. Interviews alone cannot produce a detailed picture of the context of real behaviour, and need to be backed up by observational studies. Experts are often unable to articulate their problem solving expertise and therefore direct questioning techniques may elicit little more than the terms and concepts used and a superficial overview of the domain (Welbank, 1983).

Verbal protocol analysis may prove useful in domains where a body of standard cases is readily available and where methods of problem solving are sufficiently defined and formal to be expressed verbally. Where this is not so; where there are few "typical" problems, little agreement or formalization of problem solving methods and where any particular problem solving exercise is not readily reproducible, (e.g. information interactions), verbal protocol methods may be inappropriate. Moreover, the use of artificial problems, (common in verbal protocol analysis), carries a certain amount of risk since the problem solving situation is placed at one, if not more, removes from real-life. The human's problem solving may then become theoretical and "by-the-book" rather than reflect the normal situation in which hunches, guesses and rules-of-thumb play a major role. Even when applied to real-life problem solving situations, verbal protocol analysis is obviously not appropriate where the problem solving requires the expert to engage in co-operative dialogue with a client. This has become increasingly apparent as recent research has turned to the development of intelligent advisory systems (Coombs & Alty, 1984; Kidd & Cooper, 1985).

Observational methods have the advantage that they permit the knowledge engineer to study unobtrusively the problem solving process as the expert tackles a real-life problem. Where this involves an expert working with a piece of machinery or analysing data this may not provide great insight, since what the expert is doing, and why, may not be readily apparent from just looking at the situation. However, where the problem solving process involves a co-operative interaction between human expert and a client through the medium of spoken discourse, observational methods may prove valuable and the most appropriate. Both parties must communicate their understanding of the problem being tackled, the constraints of the situation, what they themselves know about the problem domain and ways in which the problem could be solved. Usually it is also necessary for the expert to explain to the client what is being done, or could be done, and why. By recording such interactions this information becomes available to the knowledge engineer.

Reference interviews between human intermediaries and human users in information provision environments are advisory type expert-client interactions. Problem solving is carried out within a co-operative dialogue. Knowledge elicitation in this environment, therefore, seems to require the use of observational techniques and in particular the recording of the expert-client interaction. Moreover, the analysis of these recordings needs to be conducted at the discourse level since by using discourse analysis techniques, it is possible to proceed beyond an analysis of broad concepts and simple functional analysis to a more detailed, micro-level representation of expert behaviour and knowledge, which will allow complete and accurate interface implementation.

2.2. Overview.

The general method used in this research to elicit expert knowledge, is functional analysis of interactions between human experts (search intermediaries) and clients (information service users) using discourse analysis techniques. This method produces an extremely detailed analysis of what is, in any case, an extremely complex situation and in a way that is as unconstrained as possible by pre-conceived ideas.

A number of audio-recordings were made of real human user - human intermediary interactions in an information service setting. These recorded dialogues were transcribed and the transcriptions divided into utterances and foci. The tasks and functions performed by both parties were then determined with respect to each utterance. Where those tasks and functions concerned the intermediary, an attempt was made to identify the knowledge resources used. Transcription, utterance and foci division, and coding was carried out by a team of researchers - HMB, NJB, PJD and KS. The work of each transcriber/coder was independently checked by another member of the team. Coding was monitored so that a high degree of inter-coder reliability could be maintained. The basic data set consists of some seven interviews comprising in all some 1,500 individual utterances.

This method provides a relatively unobtrusive way of discovering what the participants are trying to accomplish in the dialogue, how they go about these tasks and the resources they bring to bear. There is no involvement of third parties nor is the the expert (intermediary) required to do something outside the normal events of the information interaction eg. verbalise his/her thoughts. In addition functional analysis of information interactions, using discourse analysis techniques, provides an intermediate representation of the problem solving process and expert knowledge. In this way, the problem solving process and the knowledge resources required can be described without commitment to a particular knowledge representation schema or expert system architecture.

2.3. Recording the pre-search interviews.

2.3.1. The search centres.

The raw data for the functional analysis, upon which the system design is based, consists of audio-recordings made at the University of London. London University is a federal university encompassing a large number of colleges and research institutes.

There are some sixty-five libraries of which, to date, thirty-nine offer online search services. During the period August 1983/July 1984, around 10,000 unit searches were carried out within the University (Vickery et al., 1984). The recordings were made at two search stations: the library of the Institute of Education and at the Central Information Service (CIS), Senate House. These two centres are the longest established search stations within the University. CIS has been in existence since 1974. It is the co-ordinating centre for online activities within the University.

The users of these two search services are mostly postgraduate students, academic staff and research workers. The users of the Institute of Education service include postgraduate students at all degree levels. Students are encouraged to make use of the service by the Institute's policy of subsidising student searches. In comparison, relatively few of the users of services at CIS are students. These would normally make use of the services within their own institutions. CIS carries out searches for University of London users whose college or institute library does not have a search service, or where the local search service does not cover the subject area of the search. This referral role tends to result in proportionately higher numbers of faculty users. The service is also made use of by visiting academics and researchers outside the University.

The general procedure in both institutions is for users to make a booking with the search service a day or so in advance. At the time of booking they may be asked to give a brief description of the proposed search topic (no more than a sentence at most), either written on a booking-form or given over the telephone. Neither service makes use of extensive pre-search forms (cf Smith, 1979; Meadow & Cochrane, 1981). Users come to the search service at the agreed time and discuss their problem with the intermediary. The online search is carried out by the intermediary with the user present and participating.

2.3.2. Recording methodology.

Intermediaries at CIS and Institute of Education were approached and asked for their co-operation. The permission of each user was sought before recording took place. There were no refusals. This was possibly because the users, themselves researchers, were sympathetic to the need for data collection and perhaps because academic research rarely involves the element of secrecy associated with commercial and industrial research. Another possible factor might be that simply recording the interaction did not involve the users in any additional work or involvement.

The recordings were carried out as unobtrusively as possible, using a single tape recorder and a pair of small microphones, which were clipped to the user and to the intermediary respectively. One intermediary did comment that he felt being recorded made him articulate more of the thesaurus look-up process than he would normally. However as far as is possible to judge, with the exception of the initial few utterances, the participants acted and spoke as they would if they were not being recorded.

The interaction between the intermediary and the user was recorded from the point at which the user agreed to be recorded until, at least, the beginning of the online search. Users were approached for permission to record as soon as they entered the search centre's terminal room. Given the normal social processes, this still meant that there was some dialogue between user and intermediary which was not recorded e.g. when the intermediary asked who the user was, or whether they had come for a search, or while the microphones were being attached. The beginning of the online search was taken to be the point at which the intermediary dialled up the online service. In most cases, this is the point at which recording ceased but in two of the interviews recording was continued until the search was complete. A print-out of the online search process was available for each of the searches whose pre-search interviews were recorded at CIS.

2.3.3. The recordings.

Not all the recordings made were usable. In some the quality of the sound was too poor. This proved a problem throughout, since the more powerful microphones tended to pick up background noise, telephones, printers, outside traffic, conversations etc. A total of seven recordings were analysed. These involved seven different users and four different intermediaries. Two of the intermediaries worked at CIS and two at the Institute of Education library. Of the four intermediaries recorded, two had more than 5 years search experience and one intermediary had more than nine years. The recordings analysed are listed in table 3.1.

Recording number	Descriptive Title	Inter-mediary	Search Centre
#4	community education in Africa	A	I of E
#5	Japanese learning english	B	I of E
040684HBA	activated carbon	C	CIS
120684HBA	forestry in Canada	C	CIS
190684HBA	Greek-Turkish relations	D	CIS
260684KSA	lung cancer and vitamin A	D	CIS
290684KSA	delusional thinking	D	CIS

Key: CIS ....Central Information Service.
I of E ... Institute of Education Library.

Table 3.1 : Interviews recorded and analysed

Recordings 260684KSA and 290684KSA were made by K.Stinton (see Stinton 1984), and recorded from the start of the interview through to the end of the online search.

2.4. Transcribing the recordings.

The audio recordings of the pre-search interviews were transcribed from tape using a specified format, which included means for representing breath pauses, silences, duration and extra-linguistic phenomena such as coughs and laughter. Transcription was carried out by a group consisting of HMB, NJB, PJD and KS. Each transcription was checked by at least one other member of the group to ensure accuracy and completeness.

The early recordings were initially transcribed using the transcription protocol devised by G.Jefferson (Sacks, Schegloff and Jefferson, 1974). Although the recordings were relatively easy to transcribe using this method, for the purpose of analysis the transcripts proved too cumbersome.In particular, the flow of discourse was difficult to follow because of the way in which this protocol represented the large number of overlapping speech sequences.A method of transcription which preserved the time-line sequence of the discourse was gradually developed. The stages in its evolution are recorded in Brooks & Belkin, (1983); Price, (1983); Stinton, (1984); Daniels, Brooks & Belkin, (1985). Only the final form of the protocol is described here.

The Protocol

All text is to be typed in lower case, except proper names which are to be given an upper case initial letter.

Full stops, commas and exclamation marks should not be used to indicate grammatical sentences or part sentences. (The nature of the analysis meant that it was more important to record true pauses and breath pauses that to demarcate speech sequence by grammatical rules). Question marks may be used to indicate a query where rising intonation marks a functional rather than grammatical question.

The speech of each participant, i.e. the intermediary or the user, is to be typed on alternate lines. (This enables consecutive and simultaneous speech sequences to be represented as such). The participant is indicated at the beginning of each line. I is to be used to represent the intermediary and U for the user.

The use of dialect and contracted forms are to be noted where possible. An apostrophe is used to indicate a contraction. e.g. yeah (yes); 'cause (because); an' (and); y'know (you know).

True phonetic transcription is not used. However an attempt should be made to record what is actually said rather that the correct orthographical form of the word. This particularly applies to cases of hesitation and stuttering.

Broken-off words can be indicated by using a hyphen.

Non-verbal utterances are to be transcribed using one of the following forms:
mm, mm hm, um, ah, uh huh, uh, er, err

Stress, or pronounced emphasis, is to be indicated by underscoring the word concerned. (A more detailed transcription of intonation is not necessary for the purposes of this study. Strongly emphasised words are noted because they may indicate concepts of particular concern).

The extended pronunciation of words is to be indicated by the use of a colon after the word. Each colon represents a duration of 0.5 secs. e.g. and:: represents the word 'and' extended for 1 sec.

Extra-linguistic phenomena such as laughter and coughing are to be indicated in parentheses. The duration is given by full stops. Each full stop represents 0.5 secs. e.g. (Cough ....) indicates a two-second cough.

Inaudible speech is to be marked by 'inaud' in parentheses together with the length of the sequence in 0.5 sec intervals, as denoted by the full stops. e.g. (inaud ......) represents three seconds of inaudible speech.

Breath pauses are to be indicated by a comma in parentheses e.g. but (,) are (,) you prepared to read stuff about (,) . A breath pause should only be marked when a flowing speech sequence is briefly interrupted by an audible intake of breath.

True pauses are to be marked by full stops in parentheses. Each full stop represents a silence of 0.5 secs. e.g. (....) indicates a two second pause. Silences longer than five seconds are indicated by the length of the interval, in seconds, in parentheses. e.g. (10 secs).

A sample of the transcript made from interview #4 is given in fig. 3.2:

2.5. Identifying the utterances.

2.5.1. Introduction.

In keeping with the procedures commonly used in discourse analysis, the interviews were broken down into utterances, rather than grammatical sentences. In general terms an utterance can be defined as a speech sequence by one participant during the conversation i.e. continuous speech including short silences or breath pauses. It may or may not comprise a complete grammatical entity. It is terminated by a contribution made by another participant.

Two methods for identifying utterances were employed: method U-1 and method U-2. Method U-1 is described in Brooks & Belkin, 1983 and elaborated on by Price (Price, 1983). Method U-2 was developed by K. Stinton (see Stinton, 1984). Interviews #4, #5, and 040684HBA were analysed using method U-1. The remaining interviews, (120684HBA, 190684HBA, 260684KSA and 290684KSA), were

```
--------------------------------------------------------------------
I  three countries       but (,) are (,) you prepared to read
U                  uh mm
--------------------------------------------------------------------
I  about (,) I dunno (,) ya know (,) anywhere else in the world
U
--------------------------------------------------------------------
I  (,) if::                                              if it's
U        yeah (,) if it's related to the questions I'm asking
--------------------------------------------------------------------
I  yeah (,) if it's about community education (,) but primarily
U
--------------------------------------------------------------------
I  in Africa (.) and:: (.) then those three countries
U                    uh mm mm                                then
--------------------------------------------------------------------
I
U  I'm looking at (.) a history as well (,) because these three
--------------------------------------------------------------------
I
U  (..) the reason I chose them is because they've got a
--------------------------------------------------------------------
I                                     uh mm
U  history of community education (,) which was introduced
--------------------------------------------------------------------
I                                what  (.) they actually
U  during colonial times (...)
--------------------------------------------------------------------
I    called  it community education
U
--------------------------------------------------------------------
```

Figure 3.2 : An extract from the transcript of interview #4

analysed using method U-2. A version using method U-2 also exists for interview #4. Each interview was analysed and the utterances identified by at least two members of the team, (HMB, NJB, PJD and KS), to ensure consistency and accuracy.

2.5.2. Method U-1

An utterance is defined as any continuous speech sequence by one participant in the conversation, whose end is marked by an utterance from the other participant, *which takes the conversational turn*. Utterances are indicated by an utterance number which reflects the order of the utterances in time.

e.g. xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx / [1]
xxxxxxxxxxxx /[2] xxxxxxxxxx /[3]

If a speech sequence is interrupted in midstream, for example during a breath pause or short silence, and is continued immediately after the interruption, that sequence is regarded as the same utterance. If the contribution of one participant takes the conversational turn, the previous speech sequence is regarded as a completed utterance. Such contributions may also overlap that of the previous participant. e.g. (taken from 040684HBA)

I yes that will get all the types /[51]
U will get all the types /[52]

An utterance is not <u>completed by</u>:

(i) overlapping speech that does not take the conversational turn. The overlapping sequence is to be treated as an utterance in its own right.

e.g. (taken from 040684HBA)

I right so - /[9]
U yes (...) and the consumption (..) in general /[8]

(ii) a speech sequence uttered by one participant which partially overlaps that of the first participant and which interrupts the sequence but which does so in a short breath or silence of short duration (1 - 2 secs.). The sequence of the first participant is to be treated as one utterance.

e.g. (taken from #4)

```
I                                                 uh mm /[36]
U  they've got a history of community education (,) which was
I
U  introduced during colonial times /[35]
```

Each utterance may be divided into one or more part utterances . A part utterance is a stretch of speech within the utterance which may or may not be a grammatical entity but which introduces a new function or topic. The subdivision of an utterance into two parts is often indicated by the structure of the discourse; silences, pauses and so on. Part utterances are indicated by the addition of a lower case letter to the utterance number. Letters are assigned such that the temporal sequence of the part utterance is preserved. e.g. (taken from #4)

```
I  so (,) community schools /[82a] now there are quite a lot of
U                           yeh I see /[83]
I  (,) headings (,) that (...) are (,) begin with community
U
I  /[82b]  I think perhaps if we just looked at them all
U
I  quickly /[82c] you'd be interested in something that was
U
I  just generally about community (,) and community action and
U                                   yeh (,) yes /[84]
I  presumably attitudes (.....) /[82d] it would appear to me
U
```

2.5.3. Method U-2

This is a simplified version of U-1. Simplified because of the difficulties of applying U-1 consistently. The utterance assignments under U-1 were relatively straightforward but the length and sense of the resulting divisions required that, for the purpose of this study, they be divided where necessary into smaller units. Partitioning into part-utterances however, involved some initial analysis of the underlying functions. The structural morphology of the discourse was not a sufficiently reliable indicator of the division. For this reason, the task of partitioning the utterances proved lengthy and the subdivisions assigned by different coders did not always agree.

Method U-2 requires a definition of an utterance which states that any interruption of one participant by the other, even though the interrupter does not take the turn, should be regarded as completing the previous speech sequence. Overlapping sequences which run completely in parallel and do not interrupt the first speaker, are not regarded as completing the ongoing utterance. e.g. (Taken from #4)

Compare with the same sequence coded using method U-1:

```
----------------------------------------------------------------
I so (,) community schools /[82] now there are quite alot of
U                     yeh I see /[83]
----------------------------------------------------------------
I (,) headings (,) that (...) are (,) begin with community (..)
U
----------------------------------------------------------------
I  I think perhaps if we just looked at them all quickly
U
----------------------------------------------------------------
I  (.....) you'd be interested in something that was just
U
----------------------------------------------------------------
I  generally about community (,) /[84] and community action
U                               yeh (,) yes /[85]
----------------------------------------------------------------
I  and presumably attitudes (......) it would a' appear to me
U
----------------------------------------------------------------
I  that virtually all of these headings would be /[86]
U                                                      mm (.)
----------------------------------------------------------------
```

The rule for utterance completion ensures that most utterances are of small size and this in turn ensures that each utterance, on the whole, deals with only one topic or subtopic. On the rare occasions when this is not the case, the utterances are partitioned, as before. Part utterance is indicated by a lower case letter. Since this method of coding is more dependent on the structure of the discourse, the assignment of utterances by the different coders was found to be more consistent and more in agreement.

One problem with method U-2 is that the "sense" of a particular utterance is sometimes not contained within itself but spread along several adjacent sequences. e.g. (taken from #4).

I	all of these headings would be /[86] of interest to
U	mm (.) community
I	you if they are (,) linked to Kenya /[88] Tanzania (,) Sier
U	education centres /[87] yeh /[89]

In the above extract, the sense of utterance [86] is carried through to utterance [88] and from there into utterance [90]. The reason is, of course, that the Intermediary has not completed his turn at the end of [86] and continues on through [88] and [90].

2.6. Marking the focus shifts.

Focussing in discourse refers to the narrowing of the participants' attention from the set of all possible discourse topics to the concepts and themes that are relevant to the current purpose or goals of the discourse (Grosz, 1981). Focus influences the choices among different senses of a word and the interpretation of noun phrases and actions and the overall interpretation of the utterance (Grosz, 1978). It also affects the topics to which reference can be made by the participants at any given point.

Shifts in focus occur within a task oriented dialogue whenever the subtask being carried out changes (Grosz, 1978). An attempt to shift focus by a participant may be regarded as "illegal" if s/he has not been granted "permission" to do so by the other participant. Focus shifts tend to be initiated by the participant with the higher status.

The importance of focus in the overall structuring of dialogue, particularly in task oriented dialogues, and the correspondence that has been found between focus shift and the completion of subtasks, seemed to indicate strongly that foci would prove a useful macro-level unit for the analysis. It was felt that it would be appropriate to analyse the interviews for foci and focus shifts.

A pilot analysis revealed that in the recorded interviews shifts in focus were often marked by the occurrence of "frame words" (Sinclair & Coulthard, 1975), and/or pauses of varying duration. Frame words are words or phrases which indicate that some kind of boundary has been reached within the discourse and that therefore a shift in focus is about to occur. Sinclair and Coulthard give "well", "now", "right" and "good" as instances of frame words. Examples from the recorded interviews are:

(Taken from 190684HBA)

```
------------------------------------------------------------
I  basic an' its (inaud....) /[7] right ok (,) right (laugh)
U                             yes /[8]
------------------------------------------------------------
```

(Taken from 129684HBA)

```
------------------------------------------------------------
I   yeah? /[34] um: (.) now (.) are you going to (.) look at
U       yeh /[35]
------------------------------------------------------------
```

(Taken from 040684HBA)

```
------------------------------------------------------------
I                          so (,) err the end uses (..) /[14]
U  the Dialog catalogue /[13]
------------------------------------------------------------
```

Using the frame words as indicators, the interviews were analysed for focus and focus shift. As before, each analysis was carried out by at least two members of the team independently. Foci were categorised on the basis of the functions and goals with which they appear to be concerned. (The inventory of foci types found in the interviews, and their distribution, is given in Daniels, Brooks & Belkin, (1985)).

2.7. Specifying the functions

One of the objectives of the analysis was to discover which specific activities take place and which individual goals must be achieved during the interaction between user and intermediary before the higher level goal of constructing an appropriate search formulation can be fulfilled.

Each of the interviews was subjected to a detailed utterance-by-utterance analysis to identify and categorise the subfunctions and subgoals that seem to be occurring. Two interviews (120684HBA and #4) were subjected to separate, independent analysis by three people (HMB, NJB and PJD). An extract from a transcript, taken from 120684HBA - focus 1, together with its subgoal analysis are presented in fig. 3.3 and table 3.2 respectively.

Utterance 1

This is an open question from the intermediary directed towards eliciting a problem description, in some form, from the user.

Utterance 2

The user's response to the intermediary's opening question is to describe the context of the problem, (it is connected with his research project), and the user's position within the problem, (he is just beginning the project). This utterance seems to be concerned with establishing the state of the

```
-----------------------------------------------------------------
I  emm (.) what's the (.) problem /[1]
U                                  ok I'm just beginning (.)
-----------------------------------------------------------------
I                       mm hm /[3]
U  a research project /[2] err (.) I'm a research student at
-----------------------------------------------------------------
I              yeah /[5]
U  LSE /[4a] (...) and um (...) working in the geography
-----------------------------------------------------------------
I
U  department /[4b] and I'm err (.) doing a (.) thesis (.)
-----------------------------------------------------------------
I
U  beginning a thesis on err (.) forestry (.) and err the
-----------------------------------------------------------------
I                                         interesting /[6]
U  impact of recreation (.) of conflicts of recreation and
-----------------------------------------------------------------
I           oh that's interesting (.) mm (5 sec)
U  forestry /[4c]
-----------------------------------------------------------------
I  (inaud ........) /[7]
U                    and err (...) one of the things I'm
-----------------------------------------------------------------
I
U  aiming to do (.) eventually is to look at the cost benefits
-----------------------------------------------------------------
I
U  of different err management schemes (.) for recreation /[8]
-----------------------------------------------------------------
```

Figure 3.3 : an extract from the transcript for interview 120684HBA

user's problem.

Utterance 4a

In this part utterance, the user describes his status, (a research student), and the college he is attending, (the London School of Economics).

Utterance 4b

The next part utterance is also concerned with describing the user and his status, (research student affiliated to the geography department).

<u>Utterance 4c</u>

This part utterance is a composite, containing information about the user's goals, (to produce a thesis), a re-iteration of the state of the problem, (beginning a thesis), and finally an initial description of the user's research, (recreation and the conflicts between recreation and forestry).

<u>Utterance 8</u>

The final utterance in this focus involves further description by the user of his research topic.

Utterance 3,5,6 and 7 are phatic and therefore not coded.

On the basis of this analysis, the subgoals for these utterances from 120684HBA, focus 1, were assigned.

SPEAKER	UTTERANCE	LEVEL-6 GOAL
I	1	RES
U	2	PDIM
I	3	-
U	4a 4b 4c	USER USER UGOAL .. RES
I	5	-
I	6	-
I	7	-
U	8	RES

Table 3.2 : Subgoal analysis for focus 1, interview 120684HBA

Notation.

The subgoals are referred to by their acronyms given in fig 3.4

UGOAL .. RES - indicates that two subgoals are being accomplished by the utterance, or part utterance, and that one follows on from the other in time.

UGOAL ; RES - indicates that two subgoals are present in the utterance or part utterance, and that the goals appear to be mixed, i.e. there is no clear time sequencing of the subgoals.

EXPLAIN : STRAT - indicates that the second subgoal qualifies the first. In this example, the utterance is concerned with explaining the search strategy. These qualifiers typically occur with the EXPLAIN and MATCH subgoals.

MATCH : DB : QUERY Further qualifications can occur with the MATCH goal, involving the confirmation (CONFIRM), disconfirmation (DISCONFIRM) or querying (QUERY) of particular aspects of the intermediary's or user's model of the situation.

On the basis of the analysis of the two interviews, an initial set of subgoals were drawn up. Each of the remaining five interviews were then subjected to the same analysis by at least two members of the team. Phatic utterances were left uncoded. Non-phatic utterances were coded according to the subgoal or subgoals they seemed concerned with accomplishing. The results were cumulated for all seven interactions to produce an inventory of subgoals.

An example of a full transcript, with utterances and foci boundaries marked, is given in Appendix 1. Foci and goal identification for this interview transcript is presented in Appendix 2.

Most of the subgoals appeared to contribute to a particular function and could be grouped accordingly. The pattern of interaction between the subgoals associated with the Problem Description and Retrieval Strategy functions were analysed. Information contributed by or donated to other functions was noted. These interactions were diagrammed on a focus by focus basis. An example of such an interaction "map" for both the Problem Description and Retrieval Strategy functions is given in Appendix 3. A brief example is presented below, taken from the central section of focus 5, interview 290684KSA. Fig. 3.5 is the extract from the transcript of interview 290684KSA, focus 5 and table 3.3 the subgoal analysis. Figures 3.6 and 3.7 are the interaction "maps" for the Problem Description and Retrieval Strategy functions respectively.

```
------------------------------------------------------------------
I
U  err (....) there was there's one which I've just got (.)
------------------------------------------------------------------
I
U called err delusional thinking and perceptual disorder /[46]
------------------------------------------------------------------
I yes /[47]
U   because there's a big theory at the present time (.) that
------------------------------------------------------------------
I
U  err (.) delusional thinking is an attempt (.) to find
------------------------------------------------------------------
I
U  organisation (.) in a mind that's getting disorganised
------------------------------------------------------------------
I
U  because of strange perceptions (.) so perception comes in as
------------------------------------------------------------------
I    right /[49]         is delusional thinking quite a common
U  well unfortunately /[48]
------------------------------------------------------------------
I  term y'know its um: how you /[50]
U                          no it's not that common /[51]
------------------------------------------------------------------
I  no /[52]      is it normally that sort of thinking is
U    no it's not really that /[53]
------------------------------------------------------------------
I  normally referred to as delusional thinking (.) could it be
U
------------------------------------------------------------------
I  called anything else do you think /[54]
U
------------------------------------------------------------------
```

Figure 3.5 : extract from the transcript of interview
290684KSA

SUBGOAL		DESCRIPTION OF SUBGOAL
1	CAPAB	Explain the capabilities of the system to user
2	UGOAL	Determine the user's goals
3	USER	Determine the status of the user
4	KNOW	Determine the user's knowledge of the field
5	IRS	Determine the user's familiarity with IR systems
6	PREV	Determine user's previous reference activities
7	PREVNON	Determine user's non-reference activities
8	PDIM	Determine the problem dimension
9	SUBJ	Define the background subject area of the search
10	SLIT	Determine the formal characteristics of the subject literature
11	RES	Specify the content of the user's research
12	TOPIC	Specify the topic of the search
13	DOCS	Determine the content or description of the documents the user would like to retrieve
14	TERMS	Select terms for searching
15	QUERY	Formulate the query
16	STRAT	Evolve the search strategy: how the query will be implemented
17	DB	Select the database to be searched
18	OUT	Select the output requirements
19	EXPL	Bring the user's knowledge up to minimum level necessary for user to cooperate effectively
20	DISP	Literal display of some aspect of the system
21	INFORM	Explain intermediary's intentions to user
22	PLAN	Specify the plan of the interview.
23	MATCH	Compare models that participants hold

Figure 3.4 : Subgoals

Utterance number	Subgoal of utterance
46	DOCS
47	phatic
48	SUBJ..RES;TOPIC
49	phatic
50	TERMS
51	TERMS
52	phatic
53	TERMS
54	TERMS

Table 3.2: Subgoal analysis of the extract from focus 5, interview 290684KSA

```
-----------------------------------------------------------------
T                                       Perception
O                                            ^
P                                            |
I                                            |
C                                            |
--------------------------------------------|--------------------
R                               Perception   |
E                                   ^        |
S                                   |        |
                                    |        |
                                    |        |
-----------------------------------|-----------|-----------------
S                          CURRENT THEORY: delusional thinking
U                          is an attempt to find organization
B            -------> in the mind. The mind becomes
J             |            disorganised because of strange
              |            perceptions
-------------|---------------------------------------------------
S             |
L             |
I             |
T             |
              |
-------------|---------------------------------------------------
D   KEY-PAPER-2
O   TITLE: delusional
C   thinking and
S   perceptual
    disorder
---------|-------------------------------------------------------
         V
I   RS(TERMS)
/  [select terms from
O   title of key
    documents]
-----------------------------------------------------------------
```

Figure 3.6 : Interaction map for the Problem Description function for focus 5, interview 290684KSA

```
-------------------------------------------------------------
D
B


-------------------------------------------------------------
T    CONSIDER          --------> OCCURRENCE ------> SYNONYMS
E    (delusional thinking)       (delusional        (delusional
R    (perceptual disorder)       thinking)           thinking)
M         ^                          ^     ^               ^
S         |                          |     |               |
----------|--------------------------|-----|---------------|-------------
Q         |                          |     |               |
U         |                          |     |               |
E         |                          |     |               |
R         |                          |     |               |
Y         |                          |     |               |
----------|--------------------------|-----|---------------|-------------
S         |                          |     |               |
T         |                          |     |               |
R         |                          |     |               |
A         |                          |     |               |
T         |                          |     |               |
----------|--------------------------|-----|---------------|-------------
          |                          |     |               |
I   PD (DOCS)                   PD (SUBJ)   ---UM (KNOW)---
/   [key-paper-2                PD (RES)   [user's knowledge of
O     title words]                          domain terminology]
-------------------------------------------------------------------------
```

Figure 3.7: Interaction map for the Retrieval Strategy function for focus 5, interview 290684KSA

2.8. Identifying the knowledge resources.

Having identified the subgoals associated with the functions, it was necessary to identify the knowledge resources which the intermediary needs to use in order to accomplish particular functions during the information interaction. Utterances concerned with Problem Description and Retrieval Strategies subgoals were analysed in depth. Inferences were made about the knowledge which the intermediary uses. For example: (Taken from #4)

I Yeah now does it mean when people talk of community
U

I education, do they mean ... primary, secondary, vocational
U

I technical, and universities, -or d-do they really mean only
U

I a certain level or type /[22]
U

In this extract, the intermediary's question about the educational level implies some internal knowledge of the intermediary's about the meta-structure of that subject area. That is, knowing that research topics in this area tend to focus on a particular educational level. It can also be seen that the intermediary possesses specific knowledge about the subject domain i.e. that there are primary, secondary, vocational and technical schools etc.

(Taken from 120684HBA)

I um (,) there's a social sciences (,) citation index
U

I (.....) /[95a] | | um: (.) what we've got to do there
U

I is concentrate on just <u>title</u> words /[98]
U

Here the intermediary makes a decision about the search tactic needed (a title-word search) based on her knowledge of the record structure of a particular database. Social Sciences Citation Index does not use index terms or have abstracts and therefore subject searching must be carried out on the title field.

3. RESULTS.

3.1. Problem Description Function.

3.1.1. Function specification.

Problem description can be regarded as the statement and clarification of the topic and context surrounding the user's anomalous state of knowledge. The conceptual model constructed by the the intermediary component of the information provision mechanism will be referred to as the problem description and the function concerned with constructing it as Problem Description.

Analysis of the interviews revealed that intermediaries seldom referred directly to the topic of the search. Instead problem descriptions seem to be constructed from a number of elements including:

- descriptions of the user's research
- information about the background subject domain;
- examples of useful documents in terms of either document descriptions (title, author etc.) or document content;
- statements about what is known about the literature in the subject domain.

The complex, multi-component nature of problem descriptions is obviously a reflection of the difficulties in obtaining them. Specifying the unknown can only be achieved indirectly by describing what is known - the research project with which the user's problem is concerned, a key document known to the user, information about the subject itself. For example, the problem description being evolved by the intermediary during interview 120684HBA seems to be framed largely in terms of the user's research. The user responds to the intermediary's initial question "what's the problem?" not with a statement of the topic he wishes to search on but rather with his status and affiliation and follows this with a description of his current research project:

I
U and I'm err (.) doing a (.) thesis (.) beginning a thesis

I
U on err (.) forestry and the impact of recreation (.) of

I interesting /[6] oh that's interesting
U conflicts of recreation and forestry /[4c]

I (.) mm (5 sec) (inaud ....) /[7]
U and err (...) one of

I
U the things I'm aiming to do (.) eventually is to look

I
U at the cost benefits (,) of different err (.) management

I
U schemes (.) for recreation /[8]

The intermediary then continues the negotiation with an attempt to obtain a reasonably full, unambiguous, description of this research.

I so (,) the angle you're going to take on it (,) is that
U

I you (inaud .) cost benefit and economics (.) /[13] or (.)
U mm hm /[14]

I is it looking at the area as a whole (,) what sorts of
U

I recreation goes on ? /[15]
U

Sometimes the intermediary explicitly asks the user to describe his/her problem in terms of the research they are doing or plan to do. e.g. (Taken from #4):

```
---------------------------------------------------------
I  yeah (.) yeah (..) tell me what sorts of things you're
U
---------------------------------------------------------
I  going to do in your dissertation /[5a]
U
---------------------------------------------------------
```

Descriptions of research may not always be sufficient, or clear enough, or appropriate, for an adequate description of the problem to be produced. The intermediary may request background subject information, or the user may offer it, in order to clarify some aspect of the user's research or to set the research in its context. e.g. (Taken from 120684HBA):

```
---------------------------------------------------------
I
U  but obviously I (,) want to look at um: (...) err (,) Nova
---------------------------------------------------------
I                        mm mm /[47]
U  an' in particular /[46] (.) the more North American
---------------------------------------------------------
I                           yes (,) is that because it (,)
U  literature I can look at /[48]
---------------------------------------------------------
I  it err (,) began there ? /[49]
U
---------------------------------------------------------
```

Utterance [49] can be seen as an attempt by the intermediary to elicit background subject information which could provide the rationale behind the user's research. Subject information may also be donated by the user. In the following extract from interview #4, the user offers information about the subject area in order to explain her research.

(Taken from #4)

```
-----------------------------------------------------
I       mm /[19]
U  like (,) so far they've been (,) umm (,) saying that (,)
-----------------------------------------------------
I
U  schools should serve the community (,) th' (,) its a term
-----------------------------------------------------
I                                          yeah /[20]
U  that's just been bantered around /[17c] (.) and I'm trying
-----------------------------------------------------
I                                      yes /[21]
U  to look at it in more detail /[17d] (.) to see if it really
-----------------------------------------------------
I
U  is serving the community (,) and if the stated aims and (.)
-----------------------------------------------------
I
U  objectives are really being fulfilled /[17e]
-----------------------------------------------------
```

Utterance [17c] describes a little of the subject area being researched by the user and current thinking in that field. This subject information appears to be offered by the user to "set the scene" for the description of the research goals that follows, (Utterances [17d] and [17e]).

In other instances, where the intermediary is unfamiliar with the subject area, subject information may need to be elicited from the user so that the intermediary can build up a general understanding of what the topic is about, where its boundaries lie, the concepts and relations involved and so on. e.g. (Taken from 190684HBA):

I right (,) disputes (...) (cough .) (....) other than Cy-
U

I are there any (,) particular ones /[13] (..) you know
U   um /[14]

I any /[15]
U   the Aegean dispute (........) and the: their dispute

I
U over the treatment of (....) the (,) Turkish minority in

I
U Greece (....) and the (,) Greek minority in Turkey /[16]

Here the intermediary asks the user for supplementary information to enable her to asses how the user's problem description can be developed.

If it is difficult to explain the research topic, the intermediary may ask, or be offered, a specification of the kinds of documents the user would like to read (or, alternatively, does not wish to read). e.g. (Taken from #4):

I
U and well this for example is very good (laugh) education

I   yeah /[9]
U and community in Africa /[8]   and that's related to what

I
U want to do /[10a]

----

e.g. (Taken from 040684HBA)

I as you say (,) you want to get away from the ones (.) which
U

I mention specifically (.) activated carbon // [83b]
U

----

e.g. (Taken from 120684HBA)

I
U the more North American (...) err literature I can

I
U look at /[48]

In the above example, it should be noted that "North America" refers here to the topic of the documents (forestry in North America) and not to their place of publication. Compare with the following extract where the Intermediary is referring to America as the place of publication of the documents (i.e. specifying the output requirements). e.g. (Taken from #4).

I we checked (,) I hope (,) that you're happy to read what
U

I the Americans have to say about the subject /[50a]
U

Documents may also be described in more specific terms. Key papers, or particular documents which the user feels are representative of the type of document s/he wishes to retrieve, may be named. A title only may be recalled e.g. Taken from 290684KSA

I yes /[30]
U one classic paper which I've sent off for /[29] which I've

I yeah /[32] yes /[34]
U lost /[31] which I'd like to get back (laugh .) /[33] is

I
U err alertness and clear thinking at high autonomic nervous

I
U system arousal /[35]

Sometimes the user may remember an author as well as, or instead of, the title e.g. Taken from 290684KSA

I yeah /[56]
U there's a paper on delusional disorder /[55] by Winokur (.)

I
U which is very important which I want to get hold of (.)

Not only are key papers described but also the content of a set of documents. e.g. (Taken from 190684HBA)

I (inaud.) /[249] yes /[250]
U because its its (,) its virtually impossible /[248] err

I yes /[252] not to
U for any article on Greek Turkish relations /[251]

I have Cyprus /[253]
U not to have some mention of Cyprus /[254]

Both user and intermediary may also refer to their knowledge of the literature of the subject as whole; how much they think has been written about a particular subject, the form in which it occurs (newspapers, books, magazines etc.), the treatment of the subject (general, detailed) and so on. This kind of information clearly has a bearing on any subsequent search formulation. e.g. (Taken from 120684HBA):

I I can imagine that there wouldn't be (.) a fantastic amount
U

I of literature on (.) umm (.) recreation an' and forestry
U

I linked anyway /[46b]
U

e.g. (Taken from 040684HBA)

```
------------------------------------------------------------
I  I (.) do you think that there's very much on this? /[21b]
U
------------------------------------------------------------
I  have you found very much so far ? /[21c]
U                                    yes (.) there (,) I
------------------------------------------------------------
I
U  think there's a fair amount  /[23a]
------------------------------------------------------------
```

e.g. (Taken from 260684KSA)

```
--------------------------------------------------------------
I  ok (.) have you (.) or do y- do you think there'll be a lot
U
--------------------------------------------------------------
I  of (.) references in the literature (.) that you must be (.)
U
--------------------------------------------------------------
I  fairly (.) familiar with this sort /[11]
U                                  um:: (.) th- th- there'll
--------------------------------------------------------------
I
U  be quite a lot /[12]
--------------------------------------------------------------
```

Comparatively few direct statements about the topic of the search appear to be made and in some interactions, almost none at all (see Table 3.3 below). This is perhaps not surprising in view of the difficulties users have in describing what information might resolve their problematic situation. Where the number of topic related utterances is high, this was because the user had a very clearly defined conception of what s/he wished to search on (e.g. 040684HBA), or because some preliminary conversation had taken place when booking the search and the intermediary was essentially recapping on what had been discussed (e.g. 260684KSA), or because, overall, the interaction concentrated on aspects other than problem description. For example, the greater part of the interaction in interview #5 concerns the selection of terms. Almost all the problem description oriented utterances are user initiated and the balance of subgoals reflects the user's attempt to explain his problem. Lacking direction from the intermediary, the user couches his description in terms of search "labels" (see Ingwersen, 1982) and details of his research. In interview 190684HBA, again the problem description function plays a relatively minor role in the information interaction. Both user and intermediary regard the user's problem as well-defined (perhaps a consequence of his previous search) and greater emphasis is placed on determining the user's previous reference activities, discussing the form of the output, and attempting to handle the problems with retrieval strategy development for that particular subject area.

Most utterances concerning the search topic directly seem to be made by the intermediary as a summary, or recap, at the end of a sequence of utterances concerned with problem description modelling. e.g. (Taken from 120684HBA):

I	so (.) really (,) we've got possibilities of looking at
U	
I	this from the economics side /[62] (..) but at the same time
U	mm hm /[63]
I	if we can get some information out (,) on (.) America /[64a]
U	

Sometimes topic related statements occur within the initial few utterances, where the intermediary has started the problem description process by explicitly requesting an outline of a search topic. e.g. (Taken from 190684HBA):

I	what's the subject of your query?/[9]
U	Greek Turkish relations/[10]

e.g. (Taken from 040684HBA):

I so what's the problem? /[1]
U Activated Carbon /[2a]

The intermediary may also initiate the session by referring to the topic statement on a pre-search form. e.g. (Taken from #4):

I err (,) what we've got on the form just says community
U

I education in developing countries (.) that's approximately
U yeah yeah /[2]

I (,) yeah /[1]
U

The different elements of problem description illustrated here, can be grouped into a set of five subgoals, each associated with a corresponding subfunction:

TOPIC- specifying the search topic.

RES - specifying the content of the user's research; describing the topic of the research.

SUBJ - defining aspects of the subject background to the search topic and/or research.

DOCS - determining the contents of the documents the user would like to retrieve.

SLIT - describing the literature of the subject domain.

Individual problem descriptions appear to be achieved by pursuing some or all of these subgoals, but the balance between the different elements, the extent to which particular subgoals occur, varies considerably between interviews. A comparison between the occurrence of problem description subgoals in the different interviews is given in Table 3.4

Interview number	Problem description subgoals TOPIC	RES	SUBJ	SLIT	DOCS
#4	19.1	33.8	30.8	2.9	13.2
#5	34.9	39.5	18.6	2.3	4.6
120684HBA	10.5	71.1	7.9	5.3	5.3
190684HBA	36.7	3.3	36.7	13.3	10.0
040684HBA	44.4	25.9	18.5	7.4	3.7
260684KSA	45.5	9.1	13.6	27.3	4.5
290684KSA	9.4	41.5	24.5	3.7	20.8

Table 3.4 showing the percentage contribution of subgoals to the problem description function.

An example of the way in which a problem description is constructed from topic, research subject, document and subject literature elements, is given in Appendix 3 for interview 190684HBA.

3.1.2. Knowledge Resources.

Observations of user-intermediary interaction during the process of obtaining and modelling a description of the user's problem, strongly suggest that the intermediary does not rely solely on the user's knowledge but also makes extensive use of a variety of knowledge resources of his/her own. The knowledge resources employed in problem modelling can be divided into two types therefore: the intermediary's own, internal resources and external resources supplied by the user or by printed documents which may be used as and when required. (Brooks & Belkin, 1983; Stinton, 1984).

The external resources which appear to made use of in the problem modelling process include :

(i). verbal information supplied by the user.

The information provided by the user about his/her problem is essential for problem description construction, since usually the user is the "expert" in the subject domain of the problem. On a more basic level, the whole interaction depends on the user providing some initial input about his/her problem e.g. the first few utterances from interview #5 given below.

```
I
U  OK (.) what I would like to do (.) is (....)/[2a] I have
------------------------------------------------------------
I
U  done a research project (.) /[2b] that seems to suggest
------------------------------------------------------------
I
U  that (.) a (.) community oriented task approach (.) to
------------------------------------------------------------
I
U  Japanese learning English (.) emmm (.) helps (.) reduce
------------------------------------------------------------
I
U  their anxiety and build up their self confidence (.)/[2c]
```

(ii). information supplied on pre-search forms, booking forms etc.

Many online search stations require users to fill in a booking form or pre-search form prior to coming for a search. If the booking is made over the telephone, the user may be asked to indicate the subject area of his/her problem. This information is used prior to the search to determine whether or not an online search is appropriate for this user and, if there is more than one intermediary, to delegate the search to the intermediary specialising in this area. As the example below indicates, the statement on the booking form may also be used by the intermediary to initiate the interaction, and to direct the user towards a discussion of his/her problem.

(Taken from interview #4)

```
------------------------------------------------------------
I  alright (.) right (.) the form (...) err (,) what
U
------------------------------------------------------------
I  we've got on the form just says community education in
U
------------------------------------------------------------
I  developing countries (.) that's approximately yeah /[1]
U                                 yeah      yeah     well /[2]
------------------------------------------------------------
```

Analysis of the interviews suggests the following set of **internal** resources i.e. knowledge that the intermediary possesses and makes use of to construct a problem model.

(i) classificatory knowledge.

(ii) knowledge of the structure of a subject domain.

(iii) knowledge of a subject domain

(iv) knowledge of the literature of a subject domain.

(v) knowledge of documents and their structure.

(vi) knowledge of users.

Each of these internal resources will be described in turn.

(i) Classificatory Knowledge.

Classificatory knowledge refers to the kind of knowledge embodied in knowledge classification schemes (e.g. library classification schemes such as the Dewey Decimal Classification). It is knowledge about the way in which topics might be related to each other. For instance, knowing that the subject "zoology" is a subset of the subject "biology", or that the subject "experts systems" is a subset of the topic "artificial intelligence", or that the subject "forests" could be subsumed under a variety of broader subject areas including geography, biology and even economics.

e.g. (Taken from 120684HBA):

I yeah /[92] therefore we ought to look at
U overall management of forests /[91]

I (.) um: (.) economic side as well /[93]
U

The inference of economics made by the intermediary in this example requires some knowledge that the subject "management" is , in the context of this discussion, related to economics. Even more explicit is the following sequence (Taken from 190684HBA):

I i- its basically politics isn't it it comes under this as
U

I economics (,) it's current affairs that's (,) possibly
U

I another one (.........) and (...) even a bit of social
U

I sciences (,) i- its really your your subject is spread
U mm hm /[65]

I all over /[64]
U

Classificatory knowledge also enables deductions to be made about the specificity or generality of a topic. e.g. (Taken from 190684HBA):

I what's what's the subject of y-your query /[9b]
U   Greek Turkish

I   right (......) any- anything particularly (,)
U relations /[10]

I specific /[11]
U

In this extract, the intermediary is clearly able to assess the generality/specificity of the topic and compare it with what she knows about the specificity of database coverage and document contents in this area. Having decided that the topic, as presented, is at too general a level for a search, the intermediary then requests the user to supply a more specific description. It is important to note that the making of this kind of decision rests on both a knowledge of the specificity of a topic <u>and</u> on the intermediary's knowledge of database coverage and specificity of document content. Topics for searching on MARC (which contains largely book material) could be dealt with at a more general level than topics for databases of journal articles.

Classificatory knowledge seems to be used by intermediaries as a means of placing a problem within some overall subject area context. Such context placing serves to limit the search space of the problem description i.e. constrains the possible topics that will be dealt with. It can be regarded as a means of defining the problem environment.

(ii) <u>Knowledge of the structure of a subject domain.</u>

The analysis of the interviews shows quite clearly that one of the key aspects of problem description, as far as the intermediary is concerned, is the elicitation of information about the "qualifying theme" for that subject domain. For instance, geographically related subject domains are assumed to be "qualified" by geographical location, politics by historical period, educational subjects by level of education and so on. For example:

Geographical location (Taken from 120684HBA):

I are you concentrating on (.) just Canadian (,) or or are
U

I you looking at the whole area /[36]
U

Historical Period (Taken from 190684HBA):

I um what period what time (,) span/[84]
U nineteen seventy six on/[85]

Education level (Taken from #4):

I yeah (,) now does it (.) it (.) mean (,) when people talk of
U

I community education (.) do they mean (...) primary (.)
U

I secondary (,) vocational (,) technical (,) and (.)
U

I universities (,) or d- do they really mean only a certain
U no (,) community /[23]

I level or type /[22]
U

In each case the intermediary explicitly prompted for this information, which pre-supposes some internal knowledge about the meta-structure of subject domains.

(<u>iii</u>). <u>Knowledge of a particular subject domain.</u>

Although not always the case, (and this may be related to the proportion of utterances with SUBJ goals in particular interviews), the intermediary may often know something about the subject domain with which the user's problem is concerned. The intermediary may make use of this subject knowledge to aid the construction of the problem description e.g. (Taken from 120684HBA):

```
-------------------------------------------------------------
I  or (.) is it looking at the area as a whole (,) what sort
U
-------------------------------------------------------------
I  of recreation goes on /[15] how much land is given to
U                        err: no (,) I think- /[16]
-------------------------------------------------------------
I  forestry and so on /[17]
U
-------------------------------------------------------------
```

In the following sequence, the intermediary queried the user because the user's previous statements conflict with what the intermediary knows about the subject area.

(Taken from #4):

```
-------------------------------------------------------------
I
U  the reason I chose them is because they've got a history of
-------------------------------------------------------------
I                       uh mm /[36]
U  community education (,) which was introduced during colonial
-------------------------------------------------------------
I          what (.) they actually  called  it community
U  times /[35b]
-------------------------------------------------------------
I  education? or /[37]
U         uh mm (.) they called it rural education (....)
-------------------------------------------------------------
```

(iv). Knowledge of the literature of the subject.

Intermediaries may possess quite detailed knowledge of the literature in a particular subject domain, either because they themselves have studied the subject or through long familiarity in searching within an area. Analysis suggests that this knowledge may be employed in the development of a description of the user's problem. e.g. (Taken from 120684HBA):

I I can imagine that there wouldn't be a fantastic amount of
U

I information out (,) on (.) umm (.) recreation and forestry
U

I linked anyway /[64]
U

In the example, the intermediary is making use of her knowledge of the subject literature to assist in the task of specifying the topic of the search, given the model of the user's research which she now possesses. The intermediary judges that there will be few references dealing with recreation and forestry and therefore decides that it is not worthwhile pursuing more description or specifying in more depth either the recreation or the forestry themes.

(v). Knowledge of documents and their structure.

The knowledge of documents apparent in these interviews concerns the structure of the documents rather than knowledge of key papers. Discussions with intermediaries working in other search environments suggests that where the intermediary is closely connected with a research team, for instance, the intermediary may also make use of his/her knowledge of important documents. The knowledge of documents used by the intermediaries in the recorded interviews includes knowing that documents have descriptive titles, that they may be written by more than one author, that they have a date of publication and that such attributes are represented in the document descriptions stored in the online databases. e.g. (Taken from 290684KSA):

I right you don't happen to know the author of that one (.)
U (laugh .....)

I no (..) that's the title yeah (..) do you know around when
U that's useful isn't it? (laugh ....) /[37]

I that would be published? /[36]
U

(vi). Knowledge of users.

This refers to the knowledge an intermediary possesses about the way in which the problematic situation is affected by or relates to, particular user characteristics. The range of possible problem descriptions is constrained, for example, by the college and department to which a student is attached. It would be unexpected for a student attached to a history department to want a search on some aspect of quantum mechanics.

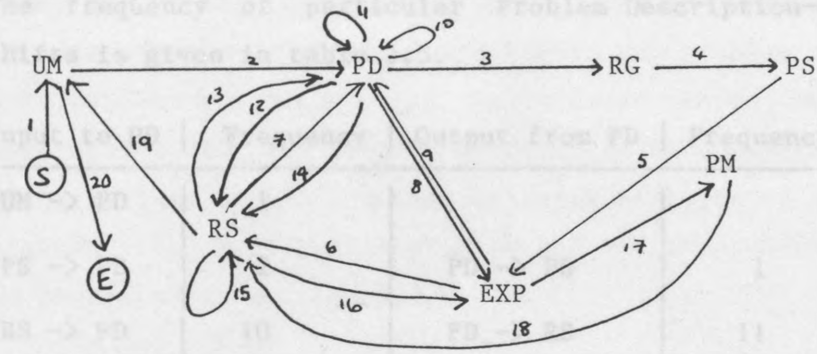
In Interview 120684HBA, the fact that the user is registered at the London School of Economics, leads the intermediary to place great emphasis on the management/economics aspects of the user's problem.

3.1.3. Interaction with other functions

The interaction between Problem Description and the other functions can be analysed at a number of levels. One level is the pattern of focus shift; looking at the functions which precede and succeed foci concerned with Problem Description. At a more detailed level is the pattern of interactions between the subgoals concerned with particular functions within individual foci.

The pattern of function change across foci is a complex one.

In each of the interviews recorded, there are at least two, and usually more, foci concerned with Problem Description. These foci are rarely adjacent to each other and may occur at opposite ends of the interview. For instance, the pattern of function change across the foci for interview 190684HBA (see fig.3.8 and 3.9) shows a considerable amount of backtracking between functions.



Key: UM=User Model PM=Problem Mode PS=Problem State
RG=Response Generator PD=Problem Description
RS=Retrieval Strategy EXP=Explanation

Figure 3.8: Function change across the foci in interview 190684HBA

[1] UM -------> [2] PD --------> [3] RS
[6] RS -------> [7] PD --------> [8] EXP
[8] EXP ------> [9] PD --------> [10] PD
[9] PD -------> [10] PD --------> [11] PD
[10] PD -------> [11] PD --------> [12] RS
[12] RS -------> [13] PD --------> [14] RS

Key: UM = User Model PD = Problem Description
RS = Retrieval Strategy EXP = Explanation

Figure 3.9: Functions of foci preceding and succeeding functions concerned with Problem Description in interview 190684HBA

Foci concerned with Problem Description were found to be preceded by foci concerned with the functions User Model, Problem State, Retrieval Strategy, Explanation, Plan and Problem Description. Foci concerned with Problem Description were followed by foci which had the functions Problem State, Retrieval Strategy, Response Generator, Explanation, Plan and Problem Description. There are 81 shifts of focus altogether in the interview corpus. The frequency of particular Problem Description-other function shifts is given in table 3.5.

Input to PD	Frequency	Output from PD	Frequency
UM -> PD	1		
PS -> PD	2	PD -> PS	1
RS -> PD	10	PD -> RS	11
		PD -> RG	1
EXP-> PD	1	PD -> EXP	3
PLAN->PD	1	PD -> PLAN	1
PD -> PD	6	PD -> PD	6

Key: UM = User Model PD = Problem Description
PS = Problem State RG = Response Generator
EXP = Explanation PLAN = Plan
RS = Retrieval Strategy

Table 3.5: Frequency of occurrence of the functions of foci preceding and succeeding Problem Description foci

There are a relatively large number of shifts to and from foci concerned with Problem Description. These shifts reflect stages in the development of the problem description, with the focus of the discourse passing from one aspect of problem description to another as each problem description element is built-up or as attempts are made to resolve any anomalies or ambiguities that have arisen.

By far the greatest number of shifts are from Problem Description to Retrieval Strategy functions and vice-versa. The Problem Description-Retrieval Strategy shift may result from the intermediary having decided that the problem description was sufficiently complete to begin formulation of a retrieval strategy. The shift from Retrieval Strategy to Problem Description seems to occur if the intermediary and/or user realises either that there is some error or misunderstanding in the problem description the intermediary has constructed or because there is insufficient information in the problem description to complete the retrieval strategy task in hand. A more detailed analysis of the interaction between the Problem Description and Retrieval Strategy functions is presented in section 3.3.

There are a number of cases in which there is a single occurrence of a particular shift into Problem Description. For instance, a Plan oriented focus precedes a Problem Description focus in interview 290684KSA. The intermediary explained to the user what would be discussed in the interview before starting to ask the user about his research. Similarly, an Explanation oriented focus precedes a Problem Description focus in interview 190684HBA. Here the intermediary explains to the user a strategy they might use to search for documents about a particular historical period, underlining the difficulties involved. The user, perhaps because he thinks the intermediary is trying to justify <u>not</u> searching for historical period, shifts focus and outlines why historical period is so important.

(Taken from 190684HBA) @@ denotes a shift in focus

I other than that I think the only way is t- is to look at a
U

I database that hopefully will deal with recent material /[96]
U

I we don't want (,) say ancient history coming out (laugh)
U ok /[97]

I and this is the danger w- /[98]
U @@ th- the reason why I've cho- I've chosen nineteen

I mm /[102]
U seventy four /[100] there was a war in on Cyprus and that

I yes /[106]
U dominated their relationships /[103]

There is also an instance of Problem Description foci being preceded by a focus concerned with the function User Model. In 190684HBA, the intermediary - detecting that the user seems to be a visitor - tries to establish the user's status before enquiring about the user's problem.

In two cases Problem Description foci were preceded by foci concerned with Problem State and in particular, with the user's previous reference activities. Here the user and intermediary first of all discuss what steps the user has taken to find information so far, whether anything has been found and if so how much. In one case (290684KSA) this leads to a discussion about the documents the user has found. In the other (120684HBA) the user indicates that he has not found a great deal so far because he has been looking for material about another aspect of his research. The lack of any previous reference activities, coupled with the mention of research, leads the intermediary to switch focus to the user's research.

(Taken from interview 120684HBA) @@ denotes a change in focus

---

I
U I've really not looked in the area so far /[10a] cos I've-

I
U I've been looking more in err (.) just the err (.)

I right (.) mm hm /[11]
U theoretical benefits of cost benefit /[10b] and conflict

I @@ yeah (..) so th- the angle that you're going
U (.) those sorts of things /[12]

I to take on it is it is it more the (.) say cost benefit
U

I economics /[13]
U

The pattern of focus change from Problem Description foci is not markedly different from the pattern of change into Problem Description (although the underlying reasons for the shifts will differ of course). Again there are a large number of shifts to other Problem Description foci, reflecting the complex process of developing problem descriptions and the fact that this process is made up of a number of subprocesses. Shifts to Retrieval Strategy foci (again a large number) are dealt with in section 3.3.

There are three shifts from Problem Description to Explanation. In two of these cases, the Explanation focus appears to buffer a shift from obtaining a description of the user's problem to constructing a retrieval strategy. When the intermediary decides that there is a sufficiently developed problem description to begin to construct a retrieval strategy, since the user's participation may be required during the Retrieval Strategy process, the intermediary must ensure that the user knows enough about the task and how it will be done, to be able to participate effectively. For example in interview #4 the intermediary passes from a discussion of the search topic in focus 6 to an explanation in focus 7 of how the terms will be selected (using a thesaurus) before going on in focus 8 to actually selecting the terms.

There a number of single instances of types of shift. One such shift is from Problem Description to Response Generator. In 190684HBA, the user explains that much of the subject literature is in newspapers. This leads to a discussion about the type of document the user wishes to retrieve. The shift from Problem Description to Plan in interview #5 occurs because the intermediary has judged the problem description to be sufficient and wants to move on to retrieval strategy formulation. This move is buffered by a focus in which the intermediary explains to the user how he intends to proceed and the plan for the remainder of the interview.

(Taken from interview #5)   @@ denotes a shift in focus

I
U but I don't want it just restricted to just Japanese students

---

I   right /[7] @@ so::
U or we'll get nothing out of the computer /[6c] emm /[8]

I (.) I think the first thing to do (,) is to have a look at
U

I the ERIC database /[9a]
U

There is also a change from Problem Description to Problem State functions in interview 290684KSA. The intermediary appears to have decided that she has sufficient information about the user's research and moves on to enquiring whether the user has found any useful documents so far.

The pattern of focus shifts is obviously of great importance to the control of the discourse and reflects the intermediary's need to achieve certain goals and, to some extent, the judgements made as to whether those goals have been attained (and therefore a move can be made to perform some other function). In order to investigate what information is passed to Problem Description from other functions and vice versa and to investigate how the Problem Description process is influenced by the state of the other functions and the models they have built, plans they have formulated etc., it is necessary to subject the interviews to a more micro-level of analysis. In particular, the interactions between the individual utterances and the subgoals each attempts to achieve, can be investigated. Interaction maps were drawn up for the interviews (see section 3.2.6). These maps represent the interaction between the subfunctions of Problem Description and between Problem Description and the other functions. An example of an interaction map is presented in Appendix 3.

The maps showed that the majority of interactions at the utterance level were concerned with the information being passed to and from Problem Description and Retrieval Strategy. This reflects the use of the problem description in formulating a retrieval strategy. However there was also some flow in the opposite direction when, for instance, the selection of a particular search strategy had consequences for the type of information that must be modelled in the problem descriptions. The issue of Problem Description - Retrieval Strategy interactions is dealt with in detail in section 3.3.

The two main contributors of information to Problem Description, if Retrieval Strategy is excluded, were found to be the User Model and Problem State functions.

<u>User Model to Problem Description</u>

The user model constructed by the intermediary appears to comprise a number of different user characteristics, attributes and beliefs (see Daniels, 1985). The user characteristics which were observed to have a bearing on Problem Description include:-

- User's state of knowledge in the subject domain;

- User's background;
- User's goals;
- User's status;

The user's knowledge in the subject domain

The state of the user's knowledge in the subject domain may be used to make inferences about whether or not a search should be made for some topic theme. In general there is the assumption that if the user is familiar with some topic theme, then s/he will not want a search to be carried out on that theme. Sometimes such inferences are made explicit by the intermediary and/or user e.g.

(Taken from 190684HBA)

I mm /[111]
U I wanna cut all that out because I have (...) /[109] a lot

I yes ok
U of material on th- innumerable situations like this /[112]

I so (,) so you jus want you don't want (inaud.) include
U (inaud.) so I I just wanted to exclude that really /[114]

I that would you if you've got it already (,) I see
U I've got /[115]

In interview 040684HBA there are several instances where the user's lack of knowledge about some aspect of the topic inhibits further discussion. If the user admits s/he does not know about a topic theme, then the intermediary can deduce that there will be no point in prompting the user for additional background subject information on this matter, nor will such information be volunteered. Information about the user's lack of knowledge may also serve to prompt intermediary to "fill-in" if s/he is familiar with the subject area and can use his/her own knowledge.

The user's background

Information about the user's background appears to be used by Problem Description in a similar way to information about the state of the user's knowledge of the subject area. In interview 120684HBA, information is volunteered by the user about his background to show that he is very familiar with some aspects of the topic and therefore it is not necessary to search for these.

The user's goals

The short term goals of the user - what s/he intends to do with the outcome of the search, how the information is to be used - has a bearing on the intermediary's understanding of the user's research. Typically the information about user goals involves specification of the kind and level of research the user is carrying out e.g. M.Sc dissertation, Ph.D. As far as Problem Description is concerned, such information appears to be used as an indication of how detailed and specific the intermediary expects to be the user's description of his/her research.

The user's status

The user's status may set expectations about the nature and depth of the user's research. For example, in interview 190684HBA, the fact that the user is a lecturer may have influenced the intermediary to press the user to describe his research in more detail. The institution to which the user is affiliated may be used to determine the range of topics with which the user is likely to be concerned.

<u>Problem State to Problem Description</u>

The two elements of Problem State observed to influence the Problem Description are the user's previous reference activities and the stage reached by the user in his/her problem solving process.

The user's previous reference activities

In all the interviews recorded, at some stage - and usually quite an early one - the intermediary asks the user whether s/he has searched for information already and if so how much material was found and how relevant was it. This information seems to be used in Problem Description as an initial indication of the amount of literature in the subject area. For instance, in the excerpt from 190684HBA presented above, the user's remark in utterance 112 enables the intermediary to infer that there is likely to be a great deal of literature on that particular theme.

The stage reached by the user in the problem-solving process

In some cases, the user specifies whether s/he is beginning or completing a research project. This information is used by Problem Description as an indication of how well-formed the problem is likely to be, and therefore how concrete and accurate will be the problem description. Sometimes the user describes research that has been done in the past or that s/he intends to do in the future which will also have a bearing on the topic that is to be searched for now.

Output from the Problem Description function that is not directed to the Retrieval Strategy functions seems to go to either the Response Generator or Explanation functions.

Problem Description to Response Generator

Information from the problem description is used to help determine the kind of output the user requires. In interview 040684HBA, for example, the user's wish to avoid documents about specific uses of activated carbon leads to a discussion about restricting the output to review articles.

Problem Description-Explanation

Explanation

In the case of the Explanation function, the problem description does not so much donate information but rather prompts the Explanation function to generate an explanation to the user e.g. about some difficulty that has arisen during the course of developing the problem description. For example, in interview #4, the user supplies background information that the term "community education" has a different usage in the USA. This prompts the intermediary to explain that the way in which the retrieval strategy has been constructed means that documents about alternative interpretations of this concept are unlikely to be retrieved.

Inform

As with the explanation subfunction, the problem description sometimes results in a prompt to the inform subfunction, and the intermediary explains to the user what s/he intends to do or not do.

3.2. Retrieval Strategy Function

3.2.1. Function specification

The Retrieval Strategy function is concerned with the selection and application of appropriate retrieval strategies to the information system's knowledge resource. In the context of document retrieval systems, these strategies consist of one or more queries, a list of databases to search, an overall search plan, strategies for achieving the desired outcome, and a number of tactics which might be employed to further the search. "Retrieval strategy" is <u>not</u> used here to refer to the retrieval techniques used by the knowledge resource's access mechanisms e.g. boolean search or weighted retrieval techniques.

In the interactions recorded, the information system's knowledge resources are online bibliographic databases held by a variety of processors. Currently, all the commercial vendors use access software which is based on boolean retrieval techniques. The retrieval strategies human intermediaries attempt to construct during the pre-search interview are plans of search: indicating the databases to be searched and the order of search; and containing one or more queries, arranged in some specified way.

Each query comprises a command and one or more terms. If more than one term is used, these may be linked with boolean operators (AND, OR, NOT) or by adjacency indicators. Most host query languages support truncation of terms (so that a word stem may be searched for) and the limiting of the search to a particular field or fields in a record. There are large differences between individual databases, not only in obvious characteristics such as content, coverage and recency, but also in database structure and in the use of controlled indexing languages. Thus different search formulations may need to be constructed for the same problem on different databases even when those databases are held by the same host. The number and variety of databases available means that both the choice of databases to search and the order in which they are to be accessed, are an integral part of the overall strategy.

The negotiation between user and intermediary to determine an appropriate and effective retrieval strategy is to a large extent intermediary initiated and intermediary controlled. This is not surprising since the construction of retrieval strategies is part of the intermediary's expertise and an area in which s/he is expected to make use of his/her specialised knowledge of the information system, its knowledge resources and access mechanisms. The intermediary initiates the retrieval strategy construction process when s/he believes that sufficient information is available - in the form of his/her models of the user and the user's problem. For example, in the extract below by the end of utterance 6 the intermediary has decided he has sufficient information about the user and her problem and therefore indicates that he is going to start constructing a search strategy.

(Taken from interview #5)

I	
U	but I do not want it just restricted to just Japanese
I	right / [7]
U	students or we'll get nothing out of the computer /[6c]
I	so:: (.) I think the first thing to do is to have a look
U	emm /[8]
I	at the (,) thesaurus for the ERIC database /[9]
U	

However the Retrieval Strategy function is not completely intermediary dominated. Particularly in the task of selecting appropriate search terms, users may have an important contribution to make, given their knowledge of the terminology of the domain and the way it is used in the literature. The intermediary may actively seek the user's involvement in such matters. In the following extract, the intermediary prompts the user for appropriate terms to describe a particular concept in the search topic.

(Taken from 190684HBA)

I we've got to have Greek (,) or (,) or Greece (,) and we've
U

I got to have Turkish (,) or Turkey (....) or Turkish (,) do
U

I they ever refer to them as Turko? /[229] Greek or (,)
U w- yeah (,) Graeco Turkish

I something (laugh.) /[230]
U Graeco-Turkish /[231]

Where the user has some knowledge of the information retrieval system, perhaps through frequent use of a search service, s/he may also contribute towards other aspects of retrieval strategy development such as database selection e.g. (Taken from 040684HBA)

I right so - /[9] mm /[10]
U yes (.) /[8a] and the consumption in general /[8b] emm (,)

I
U I had a look through a book and I thought Chemical Industry

---

I yes (,) that would
U Notes (,) /[11a] we could start by that (,) /[11b] 'cos

I be a good one mm (.) /[12a]
U that's a good one /[11c]

Since only certain aspects of the Retrieval Strategy function are negotiated and others are determined by the intermediary on the basis of his/her experience and knowledge alone, the process of constructing a retrieval strategy is not as "visible" in the discourse as model construction processes in other functions. However, in order to maintain the user's role as a participant in the interaction, it is necessary for the intermediary to inform the user of whatever decisions have been made or plans adopted. These have to be explained to the user in sufficient detail so that the user can continue to contribute. In the example below, the intermediary explains to the user about the search tactic she proposes to use; indicating its implications for the user and for the interaction during the remainder of that focus (and perhaps the succeeding focus). The intermediary explains as much as is necessary to enable the user to contribute effectively to the processes of the function. What is not explained, nor made explicit, is the basis on which the intermediary made the decision to adopt that particular tactic i.e. what the factors are that influenced that choice of tactic.

(Taken from 290684KSA)

I what we could do is (.) um (.) its quite a useful start off
U

I point for a search is to actually try and see if they've got
U

I that article in the database /[39] if they have we can look
U yeah /[40]

I at (.) the words that they've used to index it /[41] um and
U oh yeah (.)

I then we could start using those terms for our search /[43]
U that's a good idea yeah /[42]

The search tactic - searching for a key document, studying its indexing, and then incorporating those terms into a query - is explained to the user presumably so that the user understands the need for him to describe the titles and abstracts of key documents This appears to achieve its objective because in the sequence following this explanation the user recalls information about other key papers.

(Taken from 290684KSA)

I um (.) t /[45]
U yes (.) yes I see yeah /[44] err yeah they err (.) let me

I
U see now any more (.) um (......) err (....) there was

I
U there's one which I've just got (.) called err Delusional

I
U Thinking and Perceptual Disorder /[46]

The Retrieval Strategy function therefore, encompasses two different kinds of interaction. One is a negotiated process: whereby the user contributes on an equal or near equal basis to the performance of the task being carried out. The other is an intermediary directed process: wherein the intermediary makes some decision, on the basis of his/her knowledge and experience. The user is informed of this decision and its implications, rather than how or why it was arrived at, so that s/he will be able to participate in the next phase of the interaction. Contrast the example taken from interview 040684HBA, utterances 9-12a, (example A) in which intermediary and user negotiate over database selection and in which the user plays an active role in the selection process with examples B and C where the user's role is almost non-existent.

Example B is from interview 120684HBA and illustrates database selection in which the Intermediary attempts to involve the user but because of the user's lack of knowledge is unable to do so. The intermediary makes an initial choice and then tries to assess the extent to which the user is knowledgable enough to negotiate on the issue. Since the user does not know of the secondary source in question, obviously he would be unable to contribute to discussions about its coverage and so on. The intermediary therefore proceeds to select databases but maintains contact with the user by outlining the main features of the databases selected.

Example C is an extremely terse fragment from interview #5. The intermediary does not expect any user participation in database selection. Either the intermediary has decided that the user does not know anything about databases and their subject coverage or he has decided that database selection is the task of the intermediary alone.

Example A - (Taken from 040684HBA)

I   right so - /[9]   mm /[10]
U yes (.) /[8a] and the consumption in general /[8b] emm (,)

I
U I had a look through a book and I thought Chemical Industry

---

I   yes (,) that would
U Notes (,) /[11a] we could start by that (,) /[11b] 'cos

I be a good one mm (.) /[12a]
U that's a good one /[11c]

Example B - (Taken from 120684HBA)

I you know if you start to think (,) of sources really of
U

I information then the (,) Commonwealth Agricultural
U

I Bureaux /[70] are you familiar with their publications?
U

I /[73] they've got an abstracting journal that looks at
U no /[74]

Example C - (Taken from interview #5)

```
-----------------------------------------------------------------
I  so:: (.) I think the first thing to do is to have a look at
U
-----------------------------------------------------------------
I  the (,) thesaurus for the ERIC database /[9b] which is the
U
-----------------------------------------------------------------
I   educational database
U
-----------------------------------------------------------------
```

The Retrieval Strategy function in the interviews analysed seemed to be achieved through the attainment of several subgoals e.g.

DB: selection of one or more appropriate databases, given a particular user with a particular problem.

TERMS: selection of one or more terms, free text or controlled vocabulary, sometimes with the aid of a thesaurus.

QUERY: construction of one or more individual queries i.e the combining of particular terms using boolean operators, usually in conjunction with a query language command.

STRAT: the formation of a plan for applying the queries to the database, often in consideration of a desired outcome.

The database selection and term selection subgoals are similar in that their attainment involves the selection of one or more items from a set of such items, on the basis of the intermediary's models of the user and the user's problem, and the intermediary's own knowledge of information retrieval systems. Analysis of the database and term selection processes reveals a number of subprocesses. These subprocesses may take one or more associated arguments. e.g. SELECT (anxiety) where SELECT is the TERM subprocess and "anxiety" the argument. The subprocesses (or subfunctions) can be thought of as equivalent to the "choosing" aspect of the Retrieval Strategy function and the argument as the component (or potential component) of the strategy itself.

Database selection is usually, but not always, intermediary directed. It is a process in which the intermediary is usually the knowledgable "expert" and the user is not. The subfunctions found associated with database selection include:

- CONSIDER
- REJECT
- SELECT
- MATCH-SUBJ
- COVERAGE

The arguments taken by these subfunctions may be a single database, several named databases or a set of databases defined by some group characteristic such as subject coverage or host (e.g. historical databases or Dialog databases).

<u>CONSIDER</u>: the process in which a possible candidate database is assessed by the intermediary, and possibly the user, for its appropriateness, given the user and his/her problem.

e.g. CONSIDER:(DB-set=historical databases) Taken from 190684HBA

I ah some of them are the historical ones /[93] and that's
U

I a possibility (,)
U

---

e.g. CONSIDER:(DB=PAIS International) Taken from 190684HBA

I um Public ah this is another possibility Public Affairs (,)
U

I International (,) its its a little bit (,) like Magazine
U

I Index (.) its perhaps a little bit (inaud) /[175]
U

REJECT: the process of rejecting a candidate database or set of databases as unsuitable for use during this search

e.g. REJECT:(DB-SET=Dialog databases) Taken from 190684HBA

I sounds like Dialog (.) yeah um can you remember which
U

I particular files (,) you you searched (......) ’cos i- if you
U

I wanted we could try and avoid those (,) or was it that you
U   well I I

I wanted an update? /[35]
U would /[37]   yes actually no I would want us to please

I
U avoid those (.) I would (laugh .) /[38]

SELECT: the process of finalising the choice of a database or set of databases to search.

e.g. SELECT:(DB-SET=Dialog databases) Taken from 190684HBA

I I’m jus’ really looking through the Dialog (.) /[140] guide
U   mm hm /[141]

I because I think that Dialog is going to be the one that gives
U

I the widest range of databases /[142]
U   ok /[143]

---

e.g.SELECT:(DB=CAB abstracts & SSCI) Taken from 120684HBA

I um (,) I think if we look at C.A.B. and we have a look at
U

I the Social Sciences (,) /[101]
U

MATCH-SUBJ: the process of matching the subject area of the problem description against descriptions or experience of the subject coverage of a particular database.

e.g. MATCH-SUBJ (subject=water treatment) Taken from 040684HBA

```
-------------------------------------------------------------
I
U  err: I don't know if there are any special databases for
-------------------------------------------------------------
I
U water treatment? /2f
-------------------------------------------------------------
```

COVERAGE: the process of determining whether non-subject aspects of the databases' coverage (e.g. time period) meet the requirements of the user and the user's problem.

e.g.COVERAGE:(DB=magazine index) Taken from 190684HBA

```
-------------------------------------------------------------
I  right (11 sec) ok an that goes (.) right back to nineteen
U
-------------------------------------------------------------
I  seventy -w -
U
-------------------------------------------------------------
```

Term selection is the subtask of the Retrieval Strategy function in which the user is the most involved. Intermediaries actively seek the user's participation. This is particularly so if a free text search is to be carried out e.g. (Taken from 290684KSA)

I	is it normally that sort of thinking (.) is it normally
U	
I	referred to as delusional thinking (.) could it be called
U	
I	anything else do you think? /[54]
U	

Although the role of the intermediary may be greater during the selection of controlled vocabulary terms from a thesaurus, it is rarely carried out without any consultation with the user - quite the reverse. Furthermore, if the user is familiar with the indexing system used on the database (perhaps through use of the printed source), s/he may play quite an active role in this process, e.g. (Taken from 260684KSA)

I	
U	how about laryngeal? laryngeal's not in there (.....) can we
I	
U	put laryngeal in as a separate one then if that's - /[46]
I	let's have a look (......) /[47]
U	it isn't under larynx I suppose?
I	laryngeal neoplasms (.) that comes slightly
U	(.) no (17 sec) /[48]
I	down (.) /[49]
U	yeah /[50]

In the search sessions from which the interviews were recorded, a range of topics were searched for using a variety of databases. Consequently, examples of both free text and controlled vocabulary searching may be found - sometimes within the same session, depending on the databases to be searched, and the extent to which the intermediary decides to rely on the database's indexing language. Table 3.6 illustrating this.

Interview no.	Databases searched
#4	ERIC*, British Education Index
#5	ERIC*, PSYCINFO*, Language & Language Behavior Abstracts
040684HBA	Chemical Industry Notes, PTS PROMPT
120684HBA	CAB Abstracts, Social SciSearch, Comprehensive Dissertation Index, Agricola
190684HBA	Social SciSearch, PAIS International, Magazine Index
260684KSA	Medline*
290684KSA	PSYCINFO*

* indicates where the database uses a controlled vocabulary to index documents <u>and</u> where the intermediary explicitly made use of this when constructing the search queries.

Table 3.6 : databases and thesauri used in the interviews

Differences in the selection process itself were found, depending on whether free-text terms or controlled vocabulary terms were being selected. Involving the user in thesaural look-up requires careful guidance and explanation by the intermediary. The user is presented in the thesaurus with a large array of possible terms and the intermediary must use his/her knowledge of how the indexers apply those terms to guide the selection process. Conversely, if free-text searching is to be used because the intermediary may well be unfamiliar with the terminology of the domain, s/he must prompt the user for synonyms and alternative words, check spellings, attempt to ascertain term usage in the subject literature and so on.

The term selection subfunctions found in the recorded interactions include:

- SELECT
- CONSIDER
- REJECT
- OCCURRENCE
- SYNONYMS
- SPELL
- LOOKUP
- SCOPE-NOTE
- MAP

SELECT, CONSIDER, REJECT and MAP are common to both types of term selection process. LOOKUP and SCOPE-NOTE are only found where thesaural consultation for controlled term selection occurs. OCCURRENCE, SYNONYMS and SPELL are typically found where there is free-text term selection. The arguments taken by these subfunctions may be terms (words or phrases), several terms or a term set (such as a group of narrower terms or all terms beginning with the word "paranoia").

<u>SELECT</u>: deciding to use a term in one of the search queries.

e.g. SELECT: (respiratory tract neoplasms, laryngeal neoplasms) Taken from 260684KSA

I	right (.) let's just pick out those two headings then
U	
I	(.........) so we want (6sec) respiratory tract neoplasms
U	
I	(......) laryngeal neoplasms (......) and (inaud .) /[58]
U	

e.g. SELECT: (international communication) Taken from interview #5

I   umm /[278]
U interpersonal communication which is (.) ok you can add

I
U that one /[277]

REJECT: decide against the use of a particular term or group of terms in one of the search queries.

e.g. REJECT:(language) Taken from interview #5

I
U no! I don't like the lang- (.) I don't like language in

I
U there /[264]

---

e.g. REJECT: (test anxiety) Taken from interview #5

I would you like test anxiety? /[23d] no (.) right /[26]
U   no /[25]

CONSIDER: deciding whether or not a term is appropriate and should be used in one of the search queries.

e.g. CONSIDER: (paranoia psychosis, paranoid personality, paranoid schizophrenia) Taken from 290684KSA

I they've got paranoia psychosis /[115] and they've also got
U yeah /[116]

I paranoid personality /[117] paranoid schizophrenia /[119]
U (inaud .) /[118]

I are they? /[121]
U oh god all those are relevant /[120] yeah /[122]

---

e.g. CONSIDER: (neighbourhood schools, non traditional education) Taken from interview #4

I emm (..) what about a concept of neighbourhood schools (,)
U

I or non (,) traditional education? yeah /[96e]
U non traditional education yeh /[98a]

The number of terms under consideration at any one time is usually substantially larger for controlled terms selection that for free-text term selection. This is because the listings in thesauri provide a ready-made set of candidate terms for selection e.g.(Taken from interview #5)

I yeah (,) we'd also want second language programs (.) /[125a]
U

I so we've got (.) english second language (.) english
U

I instruction (,) english curriculum (.) immersion programs (.)
U mm /[126]

I second language learning (,) second language programs /[125b]
U

OCCURRENCE: ascertain how common the usage of the term is i.e. the frequency of the term in the subject literature

e.g. OCCURRENCE:(delusional thinking) Taken from 290684KSA

```
------------------------------------------------------------
I  is delusional thinking quite a common (.) term y'know its
U
------------------------------------------------------------
I um: how you /[50]                            no /[52]
U         no its not that common /[51]
------------------------------------------------------------
```

SYNONYMS: try to find synonyms or word equivalents for a particular term

e.g. SYNONYMS: (delusional thinking) Taken from 290684KSA

```
---------------------------------------------------------------
I  is it normally that sort of thinking is it normally referred
U
---------------------------------------------------------------
I  to as delusional thinking (.) could it be called anything
U
---------------------------------------------------------------
I  else do you think /[54]
U
---------------------------------------------------------------
```

e.g. SYNONYMS: (greek-turkish) Taken from 190684HBA

```
---------------------------------------------------------------
I  do they ever (,) refer to them as Turko? (.) /[228]  greek
U                                           w- yeah (,) graeco
---------------------------------------------------------------
I  or (,) graeco turkey or something (laugh .) /[230]
U  turkish /[229]                      graeco turkish /[231]
---------------------------------------------------------------
```

SPELL: determine the correct spelling for a term

e.g. SPELL:(graeco turkish) Taken from 190684HBA

I and tha- that's how they spell it normally (,) yeah (,) 's
U

I ok? /[236] (inaud .) /[238] yes /[240]
U yes /[237] uh (,) actually without the a /[239]

e.g. SPELL: (Winokur) Taken from 290684KSA

I
U see oh I think its simple delusional disorder (.) by Winokur

I w i n (..) o k e u r (.) right /[59]
U w i n o k u r / [58] o k u u r /[60]

LOOKUP: look-up a term in the thesaurus

e.g. LOOKUP (anxiety) Taken from interview #5

I
U ok now to the Psychological A'Abstracts (.)/ [189a] what

I well (,) we'll start with anxiety
U shall I look under first ? /[109b]

I again /[190]
U ok an-xi-e-ty /[191]

e.g.LOOKUP: (community education)
Taken from interview #4

I emm (,) let's look under community and hope that they
U

I actually use (.) that heading (.) yeah /[64e] community
U they usually do /[69]

I education they actually use as a heading /[70a]
U

SCOPE-NOTE: check the scope-note in the thesaurus for a term - usually to ensure that the term is used by the indexers to express some concept(s)

e.g. SCOPE-NOTE: (non traditional education) Taken from interview #4

```
-----------------------------------------------------------------
I  I'm not sure what they mean by that /[99a] but if we look
U
-----------------------------------------------------------------
I it up (,) they ought to give us a little definition which
U
-----------------------------------------------------------------
I  might be helpful /[99b]                      (inaud .) non
U                       unless they mean nonformal (...) (inaud ..)
-----------------------------------------------------------------
I  traditional that are offered as alternatives within or
U   formal /[100]
-----------------------------------------------------------------
I  without (inaud .) /[101a] emm (6sec) I think (.) you
U                                                  oh I see
-----------------------------------------------------------------
I  might - /[101b]
U  now /[102] no! I think you might need to just leave it out
-----------------------------------------------------------------
```

e.g. SCOPE-NOTE: (delusions) Taken from 290684KSA

```
-----------------------------------------------------------------
I  they've got a heading for delusions but their sort of (,)
U
-----------------------------------------------------------------
I  explanation of the way they use that is false personal
U
-----------------------------------------------------------------
I  beliefs /[79]
U
-----------------------------------------------------------------
```

MAP: determine what the word under consideration "means" in terms of the underlying concepts it expresses. (This a more general function than SCOPE-NOTE and applies to both free-text and controlled vocabulary selection processes).

e.g. MAP: (anxiety) Taken from interview #5

I so they use anxiety as a descriptor in its (own) right
U

I so we can put that (,) in first /[21b]
U and it covers fear and

I right /[23a]
U inhibition (,) right? /[22]

---

e.g. MAP: (fear of success) Taken from interview #5

I
U I'm not sure what they mean by fear of success /[202b]

A query can be regarded as made up of one or more search statements, where each statement comprises one or more terms, often in conjunction with a system command, organised to conform with the syntactic and logical requirements of the host query language. The process of organising selected terms into queries for applying to a particular database, is a task in which the intermediary must employ his/her expert skill. Relatively little consultation with the user occurs and very little of the process is explicit in the transcripts. Table 3.7 gives the proportion of utterances concerned with developing queries as a percentage of the total number of utterances in that interview.

Interview number	No.query-directed utterances	% query-directed utterances
040684HBA	1	1.04
120684HBA	0	0.0
190684HBA	9	2.3
260684KSA	7	6.2
290684KSA	2	1.1
#4	24	11.0
#5	13	3.7

Table 3.7: the number and percentage of query directed utterances in each interview.

On average, only 3.62% of all utterances concern query construction (cf 12.25% for term selection).

Where the query is mentioned, the utterances tend to be concerned with explaining rather than specifying. Explanations about queries occur when it is important that the user understands what is going on or where there is the possibility that the user might raise questions during the course of the online session. The expense of online and telecommunications connect costs means that intermediaries try to minimise non-task directed dialogue during this stage.

User involvement, and direct query specification, seems to centre around the organisation of the selected terms into concept groups e.g. (Taken from interview #5)

I well (.) what we could do in fact (,) emm (,) is to add
U

I these (,) together so we get one set of terms for anxiety
U

I (,) fear (,) inhibition (,) self esteem (,) self concept
U

I (.) and another set for these terms for english /[133a]
U

Sometimes the way in which these terms are to be linked i.e. the internal organisation of the queries, is referred to e.g. Taken from 290684KSA).

I we've got delusions or delusional (.) um or paranoid (.)
U

I or paranoia /[137] linked with thinking perception (.)
U (inaud .) yeah /[137]

I perceptual (..) arousal /[138]
U

The intermediary seems to present this information as a summary of the term selection process. Users may interpret such a summary as an opportunity to check that no concepts or terms have been omitted. For example, following on from the extract presented above the user says:

(Taken from 290684KSA)

I   yeah /[140]
U what about (.) personality? /[139] because that's (..) you

I
U know that's relevant isn't it? /[141]

All the commercially available online services use search software based on boolean retrieval techniques. Since boolean operators are often misunderstood, the intermediary, if s/he assesses the user is unfamiliar with information retrieval systems, may take care to explain to the user the consequence of employing boolean operators to combine particular terms e.g. (Taken from 190684HBA).

```
----------------------------------------------------------
I  and we want not Cyprus /[242] I think the only way
U                                yes /[243]
----------------------------------------------------------
I  we can do it is to say not Cyp- the only problem is i-
U
----------------------------------------------------------
I  (,) if they mention a really good one (,) and they
U
----------------------------------------------------------
I  err a good paper and it happens to have Cyprus in it (,)
U
----------------------------------------------------------
I  we lose it this is the only problem /[244]
U
----------------------------------------------------------
```

If the user is sufficiently aware or interested s/he may question the intermediary's formulation of the queries. In focus 13 in interview #4 there is a long sequence in which the user does exactly that. For example:

```
------------------------------------------------------------
I  we'll have these two (,) combined with all these /[175]
U                      mm /[176]                   so (.)
------------------------------------------------------------
I                                                   umm
U  you know like you showed me the overlap from those / [177]
------------------------------------------------------------
I  /[178]
U  you're not going to put them in so they all overlap
------------------------------------------------------------
I         no /[180a]
U  are you? /[179]
------------------------------------------------------------
```

Other aspects of query formulation involve the use of limiting devices and truncation. These are explained to the user in advance so that s/he will understand and be able to help decide whether this approach should be used or, more usually, so that s/he will comprehend what is being done online without needing to question the intermediary during this stage e.g.

(Taken from interview #5).

I	that would bring all that in /[178a] (.) so if you don't
U	ahh /[179]
I	want to do that I can (,) stick to having just attitudes
U	
I	DF which will just handle references which have got that
U	
I	descriptor but not (,) all these other ones (.) more
U	
I	specific ones /[178b]
U	

In the above case, the intermediary explains the consequences of both courses of action. Note the user is asked to select one or other option and is not expected to suggest alternatives.

The planning element of the retrieval strategy is also largely the intermediary's concern. Very little user consultation over strategies or tactics occurs and therefore relatively few utterances in the interviews are concerned with development of a search plan.

Interview number	strategy directed utterances number	%	explanation of strategy utterances number	%
040684HBA	5	5.2	0	0.0
120684HBA	1	1.0	1	1.0
190684HBA	17	4.3	17	4.3
260684KSA	11	9.7	0	0.0
290684KSA	0	0.0	1	0.5
#4	5	2.3	0	0.0
#5	6	1.7	3	0.9

Table 3.8: utterances concerned with devising a strategy or explaining the strategy (as a % of the total number of utterances).

On average, 3.5% of all utterances are concerned with strategy development. There are few explanatory statements about the strategies selected either, an average of 1.0% of utterances in each interview. Since this is a task carried out by the intermediary alone, there is little need to verbalise - at least not in the presearch interaction.

The planning element of the retrieval strategy was found to occur at a number of levels. At the top there are the strategies which are concerned with the overall search plan. These are used to inform or determine search formulation decisions. Strategies include approaches to search formulation such as "Pearl-growing", found in 290684KSA, and "block-building", found in #4, (see Hawkins & Wagers, 1981), or significant constraints such as "single-database search" and "multi-database search", or plan components such as "do a pilot search first". Little of this is explicit and hardly ever discussed with the user.

Most of the strategy oriented utterances that do occur concern the proposal or discussion of particular tactics i.e. short term plans or manoeuvres (see Bates, 1979a) rather than with the overall plan for the search. Tactics discussed include ways of narrowing the search should the initial formulation retrieve too many references. In such cases the discussion may be user initiated and may be based on the user's assessment of the number of references a particular query will retrieve e.g. Utterances 154-156 in from interview #5 in which the proposed means for narrowing the search is to make the concept being searched for more specific.

(Taken from interview #5)

I and Japanese /[152] and not specifying any particular
U yeah /[153]

I subject area if you prefer to do it that way /[154]
U yes (.)

I
U and then if there's too much there you can add this in

I yes (.) then we can always have that as an option/ [156]
U /[155]

A similar exchange occurs in interview 190684HBA.

I
U and I'm wondering whether there's any c- connection between

---

I
U paranoia and creativity (.) that would cut the search down a

---

I
U bit wouldn't it /[183]

In extract below, only the <u>need</u> to narrow the search is mentioned and not the means for doing so (although in effect this is discussed later on when the user indicates that only review articles are required - since one possible means of narrowing a search is to restrict the output to documents of a particular type).

(Taken from 040684HBA)

I
U yes (.) there (,) I think there's a fair amount /[23a] (,)

I
U it's better than the ones I've brought (laughter) to you

I laughter /[24]
U before certainly /[23b] though we will have to narrow

I
U down the search (,) from there /[25]

The intermediary, and sometimes the user, often makes the (not unreasonable) assumption that the initial search formulation may not retrieve the desired set of references, particularly as far as the <u>number</u> of retrieved items is concerned. The need for a "backup" plan or alternative option is seen as necessary e.g. (Taken from #5).

I
U see if- you think I might come up with too many? if we

I well I just want to (,) make some
U have Asians and Japan? /[246]

I provision in case that happens (,) so that we (,) we've got
U

I our (,) em (,) strategy worked out already (,) so we don't
U

I have to stop in the middle /[248]
U

A number of utterances are concerned with the most effective order for searching databases in a multiple database search e.g. utterance 83d from interview 040684HBA

I well maybe if (,) if we do it with- try our Chemical
U

I Industry Notes first (...) eh mm (.) and perhaps (,) do
U

I Chem Abs (,) just putting review (.) and (inaud .) reviews
U

I and again the recent material (.) and (,) then we can try
U

I the Water Net
U

In two cases, mention is made of the use of tactics to monitor the retrieved items. In one case, this is to ensure that relevant items have not been omitted (by checking a set which will be discarded) e.g. (Taken from 190684HBA)

I and they err a good paper and it happens to have Cyprus in
U

I it (,) we lose it this is the only problem (,) but we'll
U

I have a we'll sample as we go we won't (,) w- we'll see
U

I what we would have missed by taking out Cyprus /[244]
U

In the other case, taken from interview 290684KSA, the items to be retrieved are known key papers and the purpose of the monitoring is to discover how the database (Psycinfo) has indexed these article.

I what we could do is um its quite a useful starting point
U

I for a search us actually to try and see if they've got
U

I that article in the database/[39] (.) if they have we can
U yeah /[40]

I look at (.) the words that they've used to index it /[41]
U

I (.) um and then we could start using those terms for our
U oh yeah (.) that's a good idea /[42]

I search /[43]
U

The Retrieval Strategy subfunctions: TERM, DB, QUERY and STRAT, are directed towards the same goal - constructing an effective retrieval strategy. Nonetheless they appear in some respects to be less homogeneous that the subfunctions of, say, Problem Description. Retrieval Strategy subfunctions seem to fall into two groups: one concerned with selection from a finite list of possibilities, and in which the user may, if knowledgable enough, actively participate (DB and TERM); and the other group, concerned with organisation and planning (QUERY and STRAT) which are largely the intermediary's concern. The spread of Retrieval Strategy subfunctions, as indicated by the corresponding subgoals towards which utterances are directed, is given in table 3.9.

Interview number	DB	TERMS	QUERY	STRAT	total no. utterances
040684HBA	19.8	1.0	1.04	5.2	96
120684HBA	4.9	0.0	0.0	1.0	103
190684HBA	4.8	4.8	2.3	4.3	392
260684KSA	0.8	24.8	6.2	9.7	113
290684KSA	0.6	14.2	1.1	0.0	183
#4	0.5	11.9	11.0	2.3	214
#5	4.3	41.6	3.7	1.7	346

Table 3.9: percentage utterances concerned with each Retrieval Strategy subfunction.

This spread may be accounted for by a number of factors including the intermediary's and/or user's search experience, the nature of the databases selected, the choice of free-text or controlled vocabulary terms.

3.2.2. Knowledge resources

The Retrieval Strategy function is largely intermediary directed and initiated. It is in this area particularly that the intermediary's expert skills and knowledge are called into play and where the user is unlikely to be sufficiently knowledgable to be able to make an equal contribution. Intermediaries must rely on their own internal knowledge resources, supplemented with information from external printed resources and, where possible, from the user. Most important, however, is the role of the internal model constructed by the intermediary about the user and his/her problem. The use of these models in retrieval strategy formulation is dealt with in section 3.3 (problem description) and 3.2.3 (user model).

The external resources which are made use of in the process of formulating a search include:

(i) Verbal information supplied by the user

(ii) Thesauri

(iii) Database catalogs

(iv) Database documentation

(i). Verbal information supplied by the user.

Use may be made of the user's knowledge of the terminology of a particular subject, indexing languages, and the literature of the subject area as well as the user's experience from previous searches.

e.g. User's previous search experience: (Taken from 040684HBA).

I yes that's - /[45]
U Prompt (.) I've (,) I have used before (,) it's quite

I
U good /[44]

---

e.g. User's knowledge of the terminology of the subject area : (Taken from 260684KSA)

I now vitamin A (....) /[71] they sometimes (inaud .) /[73]
U um (.) I think its um- I think its under

I
U retinol (.) and (.) there's a retinoin (.) which is a

I
U chain molecule (inaud .) which is a separate heading and

I
U (.........) and perhaps retinoic acid (...) ah well its

I
U got - (.) um /[72]

(ii). Thesauri

Several of the databases used in the searches recorded are indexed with a controlled vocabulary and this vocabulary and its usage is documented in a printed thesaurus. Even in environments where the intermediary searches a wide range of databases, only a selection of the thesauri produced for online databases will be used. In the interactions recorded, mention is made of MeSH (the Medical Subject Headings used to index the Medline database), the thesaurus for the PsycInfo database and the thesaurus for the Eric database e.g. (Taken from interview #4)

I ok emm /[64a] (....) if we start with the American (,)
U

I one /[64b] (.) this thing (,) th- this thesaurus is a list
U

I of the subject headings /[64c]
U

I emm (,) let's look under community and hope that they
U

I actually use that heading (.) yeah /[64e]
U

The reasons for non-use are the unavailability of the thesaurus, unfamiliarity with the indexing language and/or thesaurus, or poor opinion of the usefulness of the controlled vocabulary (either in general or for the topic of the search).

(iii). Database catalogs

Lists of available databases may be used to suggest appropriate databases for searching or to check on the coverage of a particular database. The catalog that appears most often used is that produced by Dialog e.g. (Taken from 190684HBA).

I so if we look through we've got this one (,) current
U

I affairs /[137] I'm just really looking through the
U oh yes /[139]

I Dialog /[140] (.) guide
U mm hm /[141]

Each entry in the Dialog catalog contains information about the producer, cost, subject and time-span of the database. Indexes group particular databases according to the subject discipline they cover e.g. social sciences, current affairs.

(iv). Database documentation

Database documentation includes host produced documentation e.g. Dialog blue sheets and producer developed documentation e.g. manuals produced as guides for particular databases (excluding thesauri, glosseries, word lists etc.). This type of printed documentation was not used in the recorded interviews but is commonly used as a back-up resource when searching unfamiliar databases or fields.

The **internal** resources that the intermediary employs (excluding internal models of the user and his/her problem - for which see sections 3.2.3 and 3.3), include knowledge of:

(i) online information retrieval systems (in general)

(ii) the query languages of one or more online processors

(iii) individual online bibliographic databases

(iv) the indexing languages used on particular databases

(v) free-text searching

(vi) thesauri - in general and of specific works

(vii) search tactics and strategies

(i) General knowledge of online IR systems

Intermediaries possess, as part of their expert knowledge, general background knowledge about online bibliographic retrieval systems. This includes knowledge of the way online bibliographic files are organised, indexed and searched i.e. knowledge about features applicable to all commercially available online information retrieval systems. In the context of retrieval strategy construction, this general knowledge is most readily apparent where it concerns boolean search formulation and the consequences of using particular boolean operators e.g. explaining the consequences of using boolean NOT (taken from 190684HBA)

I and we want not Cyprus /[242] I think the only way
U yes /[243]

I we can do it is to say not Cyp- the only problem is in (,)
U

I if they mention a really good one (,) and they err
U

I a good paper and it happens to have Cyprus in it (,) we
U

I lose it this is the only problem /[244]
U

The intermediary's knowledge of boolean logic is in any case implicit in the way s/he constructs queries - since this knowledge is a necessary pre-requisite for the task.

Another example of general information retrieval system knowledge which is made use of during retrieval strategy construction, is knowledge of the structure of online services e.g. knowing that there are several online hosts.

(Taken from 190684HBA)

I which't was using the Dialog system wast it ? (....) can
U

I you remember what um databases (,) you searched /[30]
U

(ii) Knowledge of query languages

All intermediaries, in order to carry out a search at all, must be familiar with the query language of at least one, and frequently several, host systems. This knowledge is used to formulate syntactically correct queries which can then be applied to databases on the selected host. With the exception of truncation and field limiting, little mention is made during the interviews of query language commands - yet clearly the intermediary must possess this knowledge to be able to construct operable queries. In interview 040684HBA, the search that is eventually carried out contains the statements:-

B19 - log onto Dialog file 19 (Chemical Industry Notes)

S ACTIVATED (w) CARBON - search through all the fields of all the records and select records which contain both terms where these terms are immediately adjacent to each other

S PRODUCTION - select all records containing the term, wherever it appears

C 1 * 2 - combine retrieved sets 1 and 2 using boolean operator AND i.e. select records which contain both "activated carbon" as well as "production"

Neither BEGIN (B), SELECT (S) nor COMBINE (C) are referred to in the presearch interview. It may be that the system commands are not assigned pre-online and only "added" when the search statements are entered. This is particularly likely in the case of experienced intermediaries. The other possibility is that the search commands are assigned pre-online but because this process rarely involves the user, there is no need to verbalise it. Truncation is a rare example of a query language feature that is explicitly referred to in the presearch discussion e.g. (Taken from interview #4)

I so if I put in (....) that little (,) symbol there just
U

I tells the computer to pick up everything (.) every (.)
U that

I heading with co' community in it /[82g]
U would be (inaud .) wanted /[87]

(iii) Knowledge of online databases

Many hundreds of databases are now available for access online. Dialog alone hosts some 200 databases (Hoskinson Camp, 1985). Intermediaries do not, and cannot, know the structure of each and every database it is possible to search. They are familiar with the few that they search regularly, relying on their knowledge of information retrieval systems in general and on printed documentation to access unfamiliar databases. The database knowledge the intermediaries possess may encompass:

The structure of the database: e.g. Taken from 120684HBA

I um (,) there's a social sciences (,) citation index
U

I (......) /[95a] || um: (.) what we've got to do there
U ||

I is concentrate on just title words /[98]
U

The intermediary knows that the records in SSCI (social sciences citation index), are not subject indexed and therefore the main field for subject searching is the title. On other databases index, abstract and titles fields may be available for subject searching. In the following extract, the intermediary informs the user that although they will try to use descriptors (controlled vocabulary terms) for searching, there are other possibilities for subject searching on the database. (Taken from 290684KSA)

I um it doesn't matter if we can't because as I say we can
U

I always search on the titles and abstracts too /[77]
U

The relationship between the database and the corresponding printed sources:

This knowledge can be used to ascertain the user's familiarity with the retrieval source. This enables the intermediary to assess the extent to which the user can contribute to the search formulation process e.g. (Taken from 120684HBA)

I you know if you start to think of (,) of sources really
U

I of information then the (,) Commonwealth Agricultural
U

I Bureaux /[70] are you familiar with their (.) publications?
U mm hm /[72]

I /[73] they've got an abstracting journal
U no /[74]

Another reason for explicitly mentioning the hard-copy version of the source is that the user may have searched the printed source already and not wish to repeat this search online (and therefore the database should not be selected). This reasoning probably underlies the intermediary's explanation in the following extract.

e.g. (Taken from 290684KSA)

I um (.) the best database to use for this sort of area is
U

I probably Psychological Abstracts /[66] (.) which you can
U (cough .) yeah /[67]

I find in sort of hard printed copy in the library but which
U

I we have access to on the on the computer /[68]
U

The nationality of authors of articles held by a database:
If the literature from a particular community of researchers is required, it is necessary to assess whether or not members of that community are likely to be represented on the database under consideration. Intermediaries displayed an awareness of the main nationalities of authors indexed on a database. e.g. (Taken from interview #4)

I because one of the databases we use is (,) very heavily
U

I American (.) oriented (.) it is (,) primarily Americans
U mm /[52]

I writing in American journals (.)
U

---

e.g. discussing the contributors to Comprehensive Dissertation Abstracts (Taken from 190684HBA)

I its mainly American with a few Canadian (,) I think it would
U

I be worth trying /[191] no unfortunately (.)
U what about British? /[192]

I you have to slog through it by hand at the moment /[193]
U

Here the intermediary proposes Comprehensive Dissertation Abstracts as a suitable source to search and in the process, describes its coverage. The nationality of the contributors is important because the user is primarily interested in material produced in the United Kingdom (and therefore Dissertation Abstracts is rejected as unsuitable for searching).

Subject coverage:
Intermediaries may posses knowledge of database coverage in great detail and to a far greater depth than appears in printed descriptions of database coverage. The intermediaries recorded had substantial search experience, particularly with a core set of databases. Their knowledge of the subject coverage of these databases is extensive, concerns not only what is covered but also how well it is covered and may also include document titles and/or content remembered from previous searches. The following extracts show examples of such experience-based knowledge:

(Taken from 190684HBA)

I I think that's (inaud.) the only other possibility although
U

I it sounds like social social sciences (,) sometimes that can
U

I have /[218] its its its got such a wide coverage /[220]
U mm /[219] ok /[221]

I that it can sometimes have some political
U

In the following extract the intermediary justifies her choice of database, given the user's problem, by describing the kind of subjects she has encountered on this database during previous searches. e.g. (Taken from 120684HBA) - showing the intermediary's knowledge of CAB

I they've for an abstracting journal (,) that looks at leisure
U

I and recreation /[75] and (.) um: (,) I can remember from
U mm hm /[77]

I previous searches that we have had things like (,) for
U

I example (.) footpath (inaud.) wear and tear /[78] (inaud.)
U mm hm /[79]

I comes from having loads of people trotting round on those
U

I /[80] (.) popular footpaths on (.) um: Scafell Pike (,) and
U yes /[81]

I and that sort of thing /[82]
U

Whether or not a controlled vocabulary is used:

e.g. (Taken from 290684KSA) - the database is PsycInfo

```
------------------------------------------------------------
I  When they put the articles in they index them with terms
U
------------------------------------------------------------
I  taken from this list (.) its sort of a thesaurus /[73]
U                                                     oh I
------------------------------------------------------------
I  um and its quite helpful if we can start of finding words
U  see /[74]
------------------------------------------------------------
I  in this list to search on /[75]
U
------------------------------------------------------------
```

Knowledge of whether or not a database uses a controlled vocabulary to index documents is obviously of use when selecting terms. The extract from 290684KSA shows that Intermediaries use not only their knowledge of whether or not an indexing language is used but also make a judgement as to whether or not it will be helpful to use it. Some controlled vocabularies are not particularly useful - overall or for certain topics.

The vocabulary profile for a database:

Intermediaries cannot, and will not, know the whole vocabulary profile i.e. term postings, for any one database. However they may know the postings for key terms and may be able to guess the likely magnitude of postings for a term on a particular database e.g. the term "education" is likely to be very highly posted on ERIC, whereas a specific, non-educational term e.g. PROSPECTOR, will have only a few postings, if any.

e.g. (Taken from Interview #5) - postings on PsycInfo

I
U how about - education would be too broad a term? /[264]

I well (.) in this particular database it might not be
U

I because (,) emm (,) you won't have (,) so many references
U

I (,) so we can try that (,) as a matching term /[266]
U

The time period of database coverage:
The time period coverage of a database refers to the span of dates of publication of the articles it represents. Intermediaries are usually aware of the start date (the earliest material covered) and the frequency with which the database is updated. This knowledge can be used to determine whether or not a particular database is appropriate i.e. will meet the user's output requirements.

(iv) Knowledge of indexing languages

Many databases index documents with a controlled vocabulary. Intermediaries do not always make use of these indexing languages but for certain topics and databases their use is essential for effective searching. The intermediary must know how the terms are applied and how their meanings have been interpreted by the indexers. Some thesauri provide scope notes to assist with term selection but there are many "grey" areas e.g. differences between UK and US usage, where no definitive guide will be found and the intermediary must rely on his/her own knowledge.

Even when the databases uses a controlled vocabulary to index documents, the intermediary may decide to mix controlled terms with free-text terms in the query. Such decisions will be based on the intermediary's judgement of how well the controlled vocabulary covers a concept.

Intermediaries also often develop, with experience, knowledge of which terms occur in a controlled vocabulary e.g. (Taken from 260684KSA)

I I think we can probably have a heading for err (..)
U

I respiratory tract (7sec) respiratory tract neoplasms /[37]
U

In the next example, the intermediary knows that the concept "deficiency" is not covered by a MeSH term but must be expressed by using the subheading "deficiency" to qualify the term to which it applies.

e.g. (Taken from 260684KSA)

I you can use the subheading for deficiency /[74]
U

The intermediary may also be aware of concepts which cannot be covered by the indexing language. e.g. (Taken from 260684KSA) - selecting terms for searching Medline

---
I (cough.) what is quite <u>difficult</u> to say is something
U

I like low level because what we've got to do there is just
U

I look for (.) the word low level or something meaning low
U

I level in the title and abstract (.....) /[77]
U mmmm /[78]

Another aspect of controlled vocabularies the intermediary must know about is whether the selection of a broader term automatically includes any narrower terms (or at least that this can be achieved). Many users misunderstand the purpose of the listings of narrower terms in thesauri and assume that the broader term is inclusive of narrower terms. e.g. (Taken from interview #5)

I
U narrower terms (.) adjectives and adverbs (.) I don't

I ah well (,) you won't em (.) they're more specific
U want that /[254]

I (.) you see (.) you <u>won't</u> get them if you (,) put in
U

I (,) language in general /[255]
U but that - aren't (,) aren't these subsumed under that

I no (,) no (,) they're only suggestions for other terms that
U ? /[256]

I are used in the thesaurus (,) which (,) emm (,) you can use
U

I as well but em (,) they're not subsumed under it /[257]
U

(v) Knowledge of free-text searching

In order to search effectively with free-text terms, Intermediaries must understand that all word-forms for a term must be taken into account. Often this results in the root of the word being searched for but truncation is not always possible or suitable. Therefore, at least in the initial stages, all possible variations must be considered. Similarly, all possible synonyms for a term must also be taken into account. For example:

(Taken from 120684HBA)

I what we've got to do there (,) is to concentrate on just
U

I on title words /[98] so it's not so easy (,) you
U

I know so if you say forestry and they talk about (.) a
U

I large (inaud.) plantation or something (,) you wouldn't
U

I actually get it /[99]
U

The terms themselves may have been derived from the problem description but knowing to use them is part of the intermediary's expertise.

(vi) Knowledge of thesauri

In order to select controlled vocabulary terms, and to guide the user in this process, the intermediary must understand the construction of thesauri, the abbreviations used and what implications the notes and pointers have for term selection. e.g. (Taken from interview #4)

I this (,) SN (,) stands for scope note and this is
U

I what they mean by it (5sec) that is their (.) definition
U umm /[72]

I of it /[70]
U

---

e.g. (Taken from 260684KSA - utterance 52)

I yeah if we just look in the classification (.) numbers
U

I that's the quickest way to find out really (.)
U

(vii) Knowledge of search tactics and strategies

Knowledge of search tactics and strategies refers to the knowledge intermediaries use to plan the search, to deal with any problems that are likely to be encountered during the search, and to maximise the effectiveness and efficiency of their search formulations. Further, the intermediary must not only know a range of strategies, tactics and search heuristics, but must also know when to apply them, given a particular user and his/her problem.

The selection of strategies is usually not negotiated and is made explicit only in so far as it is necessary for the user to understand, in general terms, what will be done. Thus the decision to opt for a multi-database search is not discussed but because the choice of the individual databases will be negotiated, the user may be told that a number of databases will be searched. The choice of a multi or single database approach is probably conditional on the subject area, with the intermediary knowing which subject areas imply which strategy e.g. social sciences will require a multi-database search whereas medical searches are likely to require only one. Other strategies involving overall approaches such as "pearl-growing" or "building-block" are part of search "lore" which is taught to trainee searchers, and is part of the intermediary's professional knowledge. Their use may be to some extent a matter of individual search style but will also depend on the problem description and on the intermediary's knowledge of which strategy should be used given a particular type of problem.

One set of tactics that was evident in the interviews concerned the back-up plans devised to correct a search which retrieved too many, too few, or irrelevant references. not that desired. The intermediaries recorded knew a range of tactics which could be used to make a search broader, or more narrow, or more specific e.g. replace free-text terms with controlled vocabulary terms, drop "qualifying" concepts first, use synonyms. Other tactics concerned database use e.g. search the most promising database first, express broad concepts by searching a database with the appropriate coverage.

The sample set of interviews analysed is too small produce a comprehensive listing of all the strategies and tactics that intermediaries use, and hence know. Moreover, since the use of these strategies and tactics is rarely discussed with the user, it seems that to produce an extensive cataloguing knowledge elicitation using some other method is required e.g. interviewing the intermediary. Any catalog of strategies and tactics must be also be accompanied by information about the circumstances in the strategies or tactics should be applied and their likely effect on the search.

3.2.3. Interaction with other functions.

The Retrieval Strategy function was not usually found to be active at the beginning of the interview. In four of the interviews, the first Retrieval Strategy focus did not occur before the fifth focus. This is because in order to carry out its tasks, the Retrieval Strategy function needs some minimal input from the other functions, notably Problem Description. The two exceptions, in which Retrieval Strategy occurs relatively early on, are interviews 040684HBA and #5. In interview 040684HBA the user knew a great deal about online searching, had had a number of searches carried out and in the first focus presented a concise description of the problem, its background and her goals. Both user and intermediary seem to have concluded that this concentrated description was sufficient and the focus shifts to a discussion about suitable databases. In interview #5 the intermediary forces the conversation towards consideration of retrieval strategies at a very early stage and throughout the dialogue, concentrates overwhelmingly on the task of term selection. This has a very negative effect on the interview and it becomes clear that the intermediary's lack of understanding of the user and her goals seriously hinders the process of term selection. The user attempts to explain why the words under consideration are unsuitable, in terms of her research goals and the background subject but these remarks are not taken up by the intermediary, and little discussion about the topic or the user and her research occurs. Consequently, the pattern of within focus and between foci interaction is atypical of the interview corpus as a whole.

Retrieval Strategy foci were preceded by foci concerned with Problem State, Problem Mode, Problem Description, Explanation, Plan and Retrieval Strategy, and were succeeded by foci concerned with Problem State, Problem Mode, Problem Description, Explanation and Retrieval Strategy. The type and frequency of Retrieval Strategy - other function shifts is given in table ********.

Input to RS	Frequency	Output from RS	Frequency
PS -> RS	5	RS -> PS	1
PM -> RS	1	RS -> PM	1
PD -> RS	11	RS -> PD	10
PLAN->RS	1		
EXP ->RS	7	RS -> EXP	7
RS -> RS	16	RS -> RS	16

Key: PS = Problem State PD = Problem Description
RS = Retrieval Strategy EXP = Explanation
PLAN = Plan

Table 3.10: Frequency of occurrence of the functions of foci preceding and succeeding Retrieval Strategy foci.

Over the whole corpus of interviews, the largest number of shifts are between foci concerned with Retrieval Strategy e.g. database selection to term selection, or term selection to query formulation. However it should be noted that this result was heavily skewed by interview #5. The concentration on term selection in interview #5 resulted in a total of 11 shifts between Retrieval Strategy foci. If the results of interview #5 are disregarded, there are an average of 2 such shifts in the remaining interviews. The most common shift is from a focus in which terms for one particular concept are discussed to a focus in which terms for another concept are discussed. However, all eight cases of this shift occur in interview #5. The next most common shift is from a focus concerned with term selection to one concerned with query formulation.

There are also a substantial number of shifts from Problem Description foci to Retrieval Strategy foci and the reverse. This confirms similar findings for the Problem Description function (see 3.2.3). The issue is treated in more detail in section 3.3.

Many Retrieval Strategy foci are preceded by foci concerned with Explanation. This often indicates an attempt by the intermediary to ensure that the user is sufficiently aware of what Retrieval Strategy tasks are to be carried out and how, and of tools such as thesauri, so that the user is able to participate. Such a process can be seen, for example, in interview #4 where the intermediary prefixes the term selection process with a brief explanation of how the controlled vocabulary is used in Eric and a description of the Eric thesaurus.

(Taken from interview #4)   @@ denotes a change in focus

I	you see (.) /[78a] so they (,) have to use this
U	
I	<u>same</u> set of subject headings emm (,) even
U	umm /[79]
I	though they're talk they're indexing articles about
U	mm mm /[80]
I	(..) Africa (.) or Britain (,) or where ever /[78b]
U	mm mm mm /[81]
I	@@ so (,) community schools /[82a] now there are quite
U	yeh I see /[83]
I	a lot of (,) headings (,) that (...) are (,) begin
U	
I	with community /[82b]
U	

There is also an instance of a Retrieval Strategy focus being preceded by one concerned with Plan. This has much same purpose as preceding Retrieval Strategy with Explanation. The intermediary indicates to the user how the interview will proceed so that the user knows what to expect and can contribute effectively.

Interestingly, there are also shifts from Retrieval Strategy to Explanation foci. This may be because the intermediary needs to explain to the user some potential difficulty with the terms or database selected. Often though, the preceding Retrieval Strategy focus is concerned with the final formulation (offline) of the retrieval strategy and the following focus explains the search procedures to the user e.g.

(Taken from interview #4)   @@ denotes a change of focus

I   I don′t think we′ll need to specify particularly if it
U mm /[202]

I turns out that there are (,) an awful lot of references that
U

I we can get out (.)/[203a] @@ ok (.) there′s no need to write
U   mm /[204]

I anything down during the search because if you like the look
U

I of any (,) references (,) we print them /[203c]
U   mm mm /[204]

Sometimes the sequence of focus shifts to and from the Retrieval Strategy focus can be seen as part of a larger sequence. There is a sequence of focus shifts in interview #4 in which a Retrieval Strategy focus moves into a focus concerned with Problem State which then shifts back into a Retrieval Strategy focus. The intermediary introduces another database for consideration and then asks if the user has already searched the printed version. There then ensues a discussion about the user′s experiences searching a particular printed secondary source before proceeding to term selection. The purpose of the manoeuvre, interposing a discussion of the user′s previous reference activities between database selection and term selection, seems to be to enable the intermediary to assess how substantive a contribution from the user to the Retrieval Strategy task is likely. In interview 190684HBA, the Retrieval Strategy focus is sandwiched between two Problem Mode foci. Here the intermediary, is deciding whether or not an online search is appropriate, hence the lengthy discussions about system capabilities. The difficulty concerns the user′s wish for British publications only and the Retrieval Strategy focus deals with possible ways of entering a request to one of the databases which might achieve this outcome. Since this may fail, the intermediary tries to suggest alternative methods of locating

British material.

The interaction between Retrieval Strategy and the other subfunctions,within foci is also complex. As in the case of Problem Description, interaction maps were constructed to show the pattern of interaction. An example is given in Appendix 3. With the exception of interview #5, most of the interactions within foci were between Retrieval Strategy subfunctions themselves. The exceptions were interactions with Problem Description (for which see section 3.3) and Explanation.

Retrieval Strategy subfunctions are frequently preceded and followed by utterances concerned with Explain, Inform, Plan or Match. Sometimes the purpose of these utterances is to enable the user to contribute to the task to be carried out next or simply to understand what is to be done next e.g.

(Taken from interview 040684HBA)

I mm (...) let's have another look (.) err for the business
U

I side I think (,) Prompt (.) comes into that (....) (inaud.)
U

I (..) Managing and Marketing (........) there's more than
U

I the sort of- /[78b]
U

Most commonly though, the explanatory utterances are used to justify or explain the intermediary's selection of some database or term or query formulation. The selection/construction and explanation utterances may weave alternately through the focus e.g.

(Taken from interview 120684HBA)

I um (,) there's a Social Sciences (,) Citation Index
U

I (......) /[95a] it sounds (,) it sounds as though (,) as
U

I if its very much social sciences (,) but it's (,) it's
U

I a very broad based (,) um: (.) indexing (,) source /[95b]
U

I   and I think it would be worth trying /[97]
U mm hm /[96]

(95a DB, 95b Explain, 98 DB)

Since the user must know, in general terms, what the retrieval strategy is so that s/he can contribute to the online search itself, the intermediary takes steps to ascertain whether or not the user has understood what has been decided. Thus at the end of a Retrieval Strategy focus, or the completion of some process, utterances concerned with the strategy construction process are often followed by utterances concerned with checking that the intermediary's model of the retrieval strategy matches that of the user and vice versa e.g.

(Taken from interview #4)

I oh true /[76a] but (,) err (,) we will be cutting all that
U

I out /[76b] you see /[78a]
U   umm /[77]

(Taken from interview 120684HBA)

```
------------------------------------------------------------
I  and we have a look at the Social Sciences (,) and I think
U
------------------------------------------------------------
I  it's worth looking at some of the economics ones as
U
------------------------------------------------------------
I  well /[101]
U       right /[102]
------------------------------------------------------------
```

If term selection is done with the aid of a thesaurus, both user and intermediary may quote from the thesaurus. Thus utterances aimed at term selection are frequently preceded or followed by utterances in which portions of the thesaurus are read out e.g. in the extract below, the user quotes from the thesaurus and then states that these terms are not appropriate.

(Taken from interview #5)

```
-------------------------------------------------------------
I
U  narrower terms (.) adjectives and adverbs (..) I don't
-------------------------------------------------------------
I
U  want that! /[254]
-------------------------------------------------------------
```

In addition there is evidence of interaction at the subfunction level with Problem Mode, User Model and Response Generator functions. This interaction may be the consequence of the "triggering" of the function by a word or phrase in an utterance essentially aimed at something else. For example, in 190684HBA, the mention of the database Magazine Index initiates a discussion about information from magazines. Alternatively, information from one of the other functions may be required in order to determine how to proceed with the Retrieval Strategy task. In 120684HBA the intermediary first names a database for consideration and then asks whether or not the user is familiar with it, before proceeding to describe the details of the database to the user. The user, on receipt of this information, should then be in a position to assess whether the intermediary's suggested database seems reasonable, in the light of his/her understanding of the problem.

3.3. Interaction between the Problem Description and Retrieval Strategy functions

The Retrieval Strategy function must use the model of the user's problem developed by the Problem Description function in order to formulate a retrieval strategy. It could be predicted therefore, that the two functions would be highly interrelated in the interviews and this is bourne out by the analysis. Of the 81 focus shifts found in the interview corpus, 22 are shifts from Problem Description to Retrieval Strategy and vice versa. What is important is not so much the amount of interaction, but why it occurs. That is, at what point do the participants decide that the problem description is sufficient and switch to developing a retrieval strategy and what causes a shift from discussing an aspect of the retrieval strategy back to talking about the user's problem. It is also important to investigate how problem description is made use of during the formulation of a retrieval strategy and which elements of the problem description affect which subfunctions of Retrieval Strategy.

The dynamics of a particular interview are relatively specific. Therefore, a detailed analysis of the interaction between Problem Description and Retrieval Strategy functions will be given for one interview (190684HBA) and this will be supported by extracts from additional interviews. A brief outline of the foci in 190684HBA will place the interview in context.

Focus	Goal	Description
1	(UM) USER	Establish the status of the user - is he a member of the university?
2	(PD) TOPIC	Develop an initial description of the search topic - Greek-Turkish relations
3	(RG) OUTPUT	Establish whether the user wants a particular type of document
4	(PS) PREV	Find out details about the user's previous search

5	EXPLAIN	User wants UK journals only. Explain how difficult it would be to to restrict the output in this way.
6	(RS) DB	Explore the range of possible databases to search
7	(PD) TOPIC	Define the topic further - particularly with respect to historical period.
8	EXPLAIN	Explain the difficulties of searching for historical period
9	(PD) TOPIC	Discuss the importance of historical period. Determine topics NOT of interest to user.
10	(PD) SLIT	Determine how much literature exists on topic aspects.
11	(PD) TOPIC	Further description of topic - the types of relation between Greece and Turkey
12	(RS) DB	Discuss which databases might be suitable to search - including Middle East Index
13	(PD) SUBJ	Find out if the Middle East conflicts also cover Greek-Turkish relations
14	(RS) DB	Discuss suitable databases further
15	(RS) TERM	Select free-text terms
16	EXPLAIN	Search procedure - what will happen online
17	(PM) CAPAB	Do any databases contain only British publications?
18	(RS) QUERY	Formulate a query for restricting output to UK publications
19	(UM) UGOAL	Why British publications are so important to the user - in terms of the user's research goals

Key: UM = User Model — PD = Problem Description
PM = Problem Mode — PS = Problem State
RS = Retrieval Strategy

Figure 3.10: A description of the foci in interview 190684HBA

The first occurrence of a Retrieval Strategy subgoal is in focus 6. The discussion about potential databases for searching in focus 6 follows an explanation by the intermediary about the difficulties of restricting output to UK journals. This could only be easily effected if the database indexed the place of publication. The intermediary takes the Dialog catalogue and scans the list of databases. The intermediary has classified the user's topic (perhaps as "international relations") and goes to the subject index to the database looking for the broad group which encompasses this subject. Thus the search for likely databases is guided not by the topic itself but rather the intermediary's classification of that topic. The search for databases is further affected by the way in which the intermediary's classification of the subject maps into the broad subject categories used in the Dialog catalog to group together databases.

FOCUS 2		FOCUS 6
(PD) TOPIC: Determine search topic: Greek-Turkish relations (= International Relations)	---->	(RS) DB: MATCH-SUBJ (international relations) CONSIDER DB-SETs economics, current affairs & social sciences

Figure 3.11 : Interaction map for Interview 190684HBA, foci 2-6

From this initial glance at the database provision in this area (and perhaps also because of the subject area itself) the intermediary concludes that a multi-database search is appropriate. The "scatter" of the topic among a number of different subject areas suggests to the intermediary that a database with a broad subject coverage might be appropriate as well e.g. MARC files. The MARC files cover books rather than periodical articles, so the intermediary inquires whether the user wants books but this turns out not to be the case, the user knows the main ones, and so

MARC databases are rejected.

```
                    FOCUS 6

                                      (RS) DB:
                                      MATCH-SUBJ
                                           |
                                           |
                                           V
(PD) TOPIC:                           (RS) STRAT:
= International relations  --------->  multi-database
                           |          search
                           |
                           |
                           -------->  (RS) STRAT: search tactic
                                      select a database
                                      with broad subject
                                      coverage
                                             |
                                             |
                                             V
                                       (RS) DB:
                                       CONSIDER (Marc files)
```

Figure 3.12 : Interaction map for interview 190684HBA, focus 6.

Further on in the interview, the topic of the search is discussed further. The user does not want material about the Cyprus dispute. The intermediary formulates an initial query "Greek Turkish Relations NOT Cyprus". In addition, the user wants to restrict the retrieved items to those concerning the historical period 1976 to 1984. The intermediary explains the difficulties of achieving this, since few databases index the time period covered by the documents. One possible way of doing it (i.e. a possible search tactic) is to search on the couple of historical databases which do index historical period. Alternatively the intermediary suggests that they might select databases which only cover recent material, thereby avoiding items published before 1976 (and hopefully, therefore avoiding items about pre-1976). This latter tactic implies consideration of current affairs/newspaper databases.

```
                       FOCUS 7

(PD) TOPIC:              ---------->    (RS) QUERY:
Greek-Turkish relations                  Greek Turkish Relations
excluding Cyprus                         NOT Cyprus

(PD) TOPIC:      ------------------>     (RS) QUERY:
1976-1984                                1976, 1977, 1978, 1979
   |                                     1980, 1981, 1982, 1983,
   |                                     1984
   |
   |                    FOCUS 8
   |
   |
   ------>(RS) STRAT:      ----------->       (RS) DB:
          search tactics                      CONSIDER [DB set]
          (a) restrict to databases           (a) historical
          covering history                    databases
          (b) search databases only           (b) current
          holding recent material             affairs/news
                                              databases
```

Figure 3.13 : Interaction map for interview 190684HBA, focus 7-8

The intermediary's emphasis on the difficulties of achieving the restriction to historical period lead the user to explain in more detail why it is important for him to not to retrieve documents which deal with Greek-Turkish relations before 1976. This is because documents dealing with the period before 1976 will almost certainly be concerned with Cyprus and the Cyprus war (1974). Thus restricting to a particular historical period effectively excludes much of the material on Cyprus.

FOCUS 9

```
(PD) SUBJ:
Cyprus war in 1974
     |
     |
     V
(PD) DOCS:               --------------->  (RS) STRAT: - Search tactic
Items about pre-1976                       Exclude Cyprus by
period will also be                        limiting to items about
about Cyprus                               1976-1984
```

Figure 3.14 : Interaction map for interview 190684HBA, focus 9

The intermediary decides to use the Dialog system because this offers the widest range of databases and it is obviously more expedient to be able to constrain a search to a single vendor. The Dialog catalog is used to assist with selection. The intermediary locates a suitable group of databases by mapping the search topic to the broad subject headings labelling the database groups. The first group to be checked is "current affairs", and the databases in it are considered in turn. One of these is Middle East: Abstracts and Index. The intermediary wonders whether this would be relevant (perhaps because Cyprus and Turkey are in the Eastern Mediterranean). Intermediary and user proceed to discuss whether Cyprus could be regarded as a Middle Eastern country and about what is meant by the Middle East conflict.

FOCUS 13

```
 ------------------------------------    (RS) DB:
 |                                       CONSIDER (Middle East
 V                                         Abstracts & Index)
(PD) SUBJ:
Is Cyprus part of the Middle
East?
Discussion of Middle East
Conflict
    |
    |
    V
(PD) SUBJ:                ------------>   (RS) DB:
Cyprus is not considered                  REJECT (Middle East
part of Middle East                              Abstracts &
                                                 Index)
```

Figure 3.15: Interaction map for interview 190684HBA, focus 13

The database selection process continues, with databases being assessed primarily on the likelihood of their covering the subject and then whether they only deal with recent publications or whether they index historical period.

The term selection process involves extracting words from the problem description which express the concepts mentioned in the description and which are "good" words to search (i.e. they occur in a particular context, are uniquely defined and concept specific (Fidel, 1986)). The intermediary's initial approach is to select appropriate terms for the "Greek" and "Turkey" concepts separately and then as a single, co-ordinated concept. The user does not want information on Cyprus so this concept must be excluded. However since using NOT can create problems, resulting in relevant documents being rejected, a tactic for monitoring the situation is proposed. The user comments that it is unlikely that any document on Greek-Turkish relations will not mention Cyprus.

```
                                FOCUS 15
(PD) TOPIC:
Greek-Turkish
relations -
excluding Cyprus
    :
    :
|theme-1
| greek ............ (RS) TERMS: ........... (RS) QUERY
|                    SELECT (greek)          (greek OR
|                    SELECT (greece)           greece)
|theme-2
| turkey ........... (RS) TERMS:  ........... (RS) QUERY
|                    SELECT (turkey)         (turkey OR
|                    SELECT (turkish)         turkish)
|theme-1&2
| greek-turkish ...... (RS) TERMS:
|                    SYNONYMS (greek-turkish)
|                    SELECT (graeco-turkish)
|                    SPELL (greco-turkish)
| theme-3
|  cyprus  .......... (RS) TERMS:  ...........  (RS) QUERY
| (exclude)          SELECT (cyprus)          (NOT cyprus)
                                                   |
                                                   V
                                              (RS) STRAT:
                                           search tactic
                                           monitor effect
                                           of using
                                           NOT - check relevant
                                                   |
DOCS:   <------------------------------------------
all articles of
greek-turkish
relations will
mention cyprus
    |
    ----------------------------------------> (RS) STRAT
                                            check how data
                                            -base indexes
                                             country & use
                                             descriptor for
                                             cyprus conflict
                                             if possible
```

Figure 3.16 : Interaction map for Interview 190684HBA, focus 15.

The intermediary does not select terms to express the "relations" theme. "Relations" (or its synonyms) is not a "good" seach term - it is too broad and non-specific. The intermediary hopes that by selecting current affairs databases, any references on both Greece and Turkey will refer to relations between them. That is, the concept is expressed by restricting the search space. Thus the final negotiated query is:

```
(((greece or greek) and (turkey or turkish))
   or (greco turkish))
   not cyprus)
```

The final foci in the conversation return to the problem of finding UK material. The intermediary had not realised the importance of this to the user. The first search, therefore, is on Social Sciences Citation Index where the authors' country of residence can be searched for in the corporate source field. An outline of the search strategy used online is given in Appendix 4.

In interview 190684HBA, the problem description is used to assist with database selection and to suggest terms for selection. It also influences the choice of search tactics. The retrieval strategy being developed also has some bearing on the construction of the problem description. Problems encountered constructing a retrieval strategy may result in aspect of the problem description being expanded (e.g. see figs. 3.14 and 3 .15). This two-way interaction between Problem Description and Retrieval Strategy is very evident in interview 290684KSA. In focus 5 of interview 290684KSA, for example, the intermediary tries to elicit from the user more about the terminology of his subject by asking him if he can remember the titles of documents he has read. The user cites a key paper, whose title words become potential search terms. The terms themselves trigger off further background information from the user. The intermediary decides that since the user can remember specific key papers a good strategy might be to search for these key papers and then use the terms they are indexed with to refine the query. The user is therefore encouraged to describe other key papers.

```
                    FOCUS 5

                         PLAN:
                      elicit terminology
                      from user by asking
                      for words in titles
                      or abstracts of known
                      documents
  -------------------------|
  |
  V
(PD) DOC:  ............................ (RS)  TERMS:
key-paper-1                             CONSIDER [title-words]
Title: alertness and clear              (thinking)
thinking at high autonomic              (autonomic nervous
nervous system arousal                       system)
Author: unknown                         (arousal)
Date: 1977
     |-----------------------------------> (RS) STRAT:
                                           search tactic
                                           search for key
                                           documents
                                           and use their index
                                           terms in query
                             --------------------|
                             |
                             V
                          PLAN:
                         ask the user for
                         descriptions of key
                         papers
      ---------------------|
      |
      V
(PD)  DOCS:  ........................  (RS) TERMS:
key-paper-2                            CONSIDER [title-words]
title: delusional                        (delusional
thinking and perceptual                       thinking)
disorder                                  (perceptual
     |                                          disorder)
     V
(PD) SUBJ:
involvement of perception
in delusional thinking
     |
     V
 (PD) TOPIC:   ...........................(RS) TERMS:
 theme: perception                      CONSIDER (perception)
```

Figure 3.17 : Interaction map for interview 290684KSA, focus 5

The intermediary then takes up "delusional thinking", one of the themes in the topic, which occurs as a phrase in the title of one of the key papers. Not being familiar with the field, the intermediary asks the user if the phrase is a common one. This is to establish the specificity of the term, and therefore whether it will be a "good" term for searching. Once it has been ascertained that the term is not common, the user is asked to suggest synonyms. The user recalls "delusional disorder" and remembers that this is used in the title of a key paper by a key author, Winokur. The mention of a key author enables the intermediary to select as a possible tactic an author search. The user also notes that many simply refer to the disorder as paranoia.

```
 FOCUS 3
(PD) RES: ...................................
delusional thinking and                     :
                                            :
                      FOCUS 5     :
                                            :
(PD) DOCS: ............................. (RS) TERMS:
key paper 2                             CONSIDER (delusional
title: delusional                                thinking)
                                        OCCURRENCE
                                        (delusional thinking)
                              .......... SYNONYMS
                              :         (delusional thinking)
                              :                |
                              :                V
(PD) DOCS: < -----------------:----------- (RS) TERMS:
key-paper-3                   :           CONSIDER (delusional
title: simple delusional      :                     disorder)
disorder                      :
author: winokur               :
    |                         :
    --------------------------:---------> (RS)  TERMS:
                              :          SPELL (winokur)
                              :          SELECT (winokur)
                              :              |
                              :              |-----> (RS) STRAT:
                              :                 search tactic:
                              :                 author search
SUBJ:
this disorder is
usually called paranoia
    |
    ---------------------------------> (RS) TERMS:
                                       SELECT (paranoia)
```

Figure 3.18: Interaction map for interview 290684KSA,
focus 5

In interview #5 the Problem Description and Retrieval Strategy processes are closely inter-related because of the intermediary's failure to develop the problem description sufficiently. Formulating the retrieval strategy continually runs into difficulties, particularly with term selection, because the intermediary does not know enough about the user and her problem to be able to guide the process effectively. A pattern emerges whereby terms are rejected as inappropriate by the user, who tries to explain why by offering more information about the subject and about her research.

The interesting point about interview #5 is that in this interaction it seems to be the user who is assessing the terms under consideration against her own internal problem description rather than the intermediary. In the other interactions, the intermediary uses the concepts in his/her internal model of the user's problem to suggest terms and assesses how well these terms express the concept. The user's opinion is sought and if this differs from the intermediary's, the intermediary analyses the reasons, extracts additional information from the user if necessary, and updates the problem description.

```
                                                FOCUS 3

                                               (RS) TERMS:
                                               LOOK-UP (anxiety)
                                               CONSIDER (anxiety)
(PD) SUBJ:        <------------------------------|
Japanese people
are more likely to
feel anxiety when
interacting with
strangers
   | ------------------------------->   (RS) TERMS:
                                         SELECT (anxiety)
                                         MAP (anxiety)
                               ---------------|
                               |
                               V
                         DISPLAY:
                       (maths anxiety)
                       (test anxiety)
                               |-------------------
                                                  |
                                                  V
                                           (RS) TERMS:
                                           CONSIDER (test anxiety)
                                           REJECT (test anxiety)
(PD) RES:   <----------------------------------|
subjects are not being
tested
interested in fear
and inhibition
        |---------------------------->   (RS) TERMS:
                                          CONSIDER [broader term
                                                    for anxiety]
                                              (psychological
                                                    patterns)
```

Figure 3.19 : Interaction map for interview #5, focus 3

Probably the most common category of Retrieval Strategy to Problem Description is one in which the need to expand on some aspect of the problem description is indicated by difficulties experienced when constructing a retrieval strategy. During the process of chosing a database or selecting terms, either the user or the intermediary becomes aware that the intermediary's model of the problem is anomalous. This can be seen, for instance, in foci 6 and 7 of interview 120684HBA. The intermediary selects a database and justifies the choice by outlining some items found on a previous search which the intermediary thinks similar to the current one. The intermediary's explanations are sufficient to alert the user to the fact that the intermediary's problem description is faulty in some respects and he proceeds to rectify this by describing further his research. This provided the intermediary with an insight into what might be the main emphases of the topic.

```
                              FOCUS 6
                                          DB:
                                          SELECT (CAB)
                        --------------------|
                        |
                        V
                      EXPLAIN:DB
                      describe what
                      has been found
                      before e.g.
                      erosion of paths
                      on Scafell Pike
     -------------------|
     |                       FOCUS 7
     V
  RES:
Not interested
in specific sites.
Looking at overall
management of
forests
     |
     V
TOPIC:
economic aspects
of forests
     |--------------------------------> DB:
                                          SELECT (SSCI)
```

Figure 3.20 : Interaction map for interview 120684HBA, focus 6-7

Problem description plays a major role in Retrieval Strategy formulation. Certain aspects of the problem description seem of more significance in this process than others, although all the elements seem to be made use of in one or other of the interviews.

The subject-area of the topic is very significant in determining which database to search. Background subject information may be used to help assess whether the coverage of a database under consideration is appropriate. Term selection seems drawn from the concepts which make up the topic. The arrangement of terms within a query is affected by the structuring of the topic concepts in the problem description but also, on occasions, by what is known about the subject literature. For example, if it is known that little has been written about a particular topic then it is inadvisable to use too many "AND"s in the queries. The selection of strategies and tactics is conditional on the topic itself, the subject-area (social sciences topics imply a multi-database search) and information about key-papers (do a citation search).

It is also evident that Retrieval Strategy may have a reciprocal effect on problem description development. Problems encountered trying to develop a retrieval strategy may lead to further discussion about the user's research or of the background subject. Alternatively, the choice of a particular strategy may require the user to provide more information on a particular element in the problem description (e.g. see fig.3.17). The effect of Retrieval Strategy on Problem Description is more marked in some interviews than in others and may reflect how well-structured the user's problem is, the depth and accuracy to which the intermediary has modelled that problem and the ease with which that problem description can be transformed into a request to the knowledge resources.

4. DISCUSSION

4.1. Problem Description

Users come to an information provision mechanism such as an online search service, because they have some problem which they hope may be managed more effectively by obtaining appropriate information from the system. Since this problematic situation arises because the user has decided that his/her own knowledge is insufficient for some particular purpose, the user will probably be unable to specify precisely what it is that s/he needs to know in order to resolve the gap or anomaly in their knowledge. It is usually possible though for the user to describe the anomaly itself; to outline the topics the problem concerns and its context. Problem Description can be regarded therefore as:

> the statement and clarification of the topic and context surrounding the user's anomalous state of knowledge.

One of the primary functions of an intelligent interface for document retrieval systems must clearly be the development of a description of the user's problematic situation, because without it no effective search formulation can be constructed. Further, the description of the user's problem must be constructed in such a way that it can indeed be made use of to generate effective search formulations and to guide any necessary subsequent reformulation.

Most designers of intelligent interfaces for document retrieval systems, (with the exception of systems using retrieval techniques which do not require an explicit query), have treated problem description as the elicitation of the user's search topic. The user may be asked to input a description of their search topic and the information culled from the one or two lines that the user enters forms the basis for the generation of a search strategy e.g. (Taken from Yip, 1981)

EXPERT-1: Please describe your search topic clearly in one sentence or two.
Terminate your input with a space followed by the dollar sign ($) and a carriage return.

USER ---> RELEVANCE OF TRANSFORMATIONAL GRAMMAR TO LANGUAGE LEARNING

Sometimes the user's statement may be supplemented by asking the user to supply additional concepts or terms. The onus is on the user to produce a short statement/request and it is unlikely that the user, in these circumstances, will deliver more than a topic "label". No reference is made to the context of the problem or even to the relative importance of the concepts. The over-simplification of the problem description process in the design of these interfaces has been also noted by Fidel (1986).

The analysis of the interactions that has been carried out in the study presented here, shows that human intermediaries do <u>not</u> confine the task of constructing a problem description to asking the user for a one or two utterance statement about their search topic. Details about the search topic may be explicitly requested from the user, (though not always and seldom at a detailed level), but this statement is supplemented by a great deal of additional information. Substantial amounts of information are acquired by asking, or by being told, about the user's research, about the subject background to the research, about the literature in that subject area, and about the kind of documents the user thinks might prove useful or specifically does not want. All this information seems to be incorporated into the problem description. Since the information interactions are purposeful, goal-directed dialogues, it is unlikely that intermediaries would ask, or users volunteer, information which was not thought useful to the task in hand (Grice, 1975). Therefore it must be concluded that the information is necessary to a well-developed problem description, is required for the formulation of an retrieval search strategy and contributes towards the achievement of a satisfactory outcome to the interaction. Moreover it is also clear that human intermediaries know that the information is necessary (otherwise they would neither try to elicit it from the user nor listen attentively when the user volunteered it).

Problem Description was found, in an academic search environment at least, to comprise five subfunctions:

TOPIC: specifying the topic of the user's search;

RES: specifying the user's research;

SUBJ: describing the subject background to the research and search topic;

DOCS: obtaining descriptions of useful documents;

SLIT: specifying the literature of the subjet area;

Each subfunction is responsible for the construction of the corresponding element of the problem description model. The complex, multi-component nature of problem descriptions is obviously a reflection of the difficulty in trying to determine what sort of documents would help the user to resolve his/her problem. Specifying the unknown can only be achieved indirectly by specifying what is known - the research project with which the user is concerned, a key document known to the user, information about the subject itself.

The extent to which the Problem Description subfunctions occur in particular interviews varies. In some interviews one or other subfunction seems to play a much greater role than the rest whereas in other interviews the spread was more even. The small number of interviews in the data collection means that quantitative analysis of this phenomena is not possible. It is likely though, that the observed differences are significant and may be accounted for by underlying differences in how well the user's problem is structured, the type of problem, the subject domain of the problem, the nature of the available knowledge resources (e.g. database content and structure) and the extent of both intermediary's and user's knowledge of the subject area. For example, the relatively large number of DOCS directed utterances in interview 290684KSA are a direct consequence of both parties' perception of the topic of the user's research as vague and nebulous. Feeling unable to secure a sufficiently concrete model of the user's problem through descriptions of the user's research and search topic, the intermediary requests descriptions of useful documents.

The figures show that whilst all five Problem Description subfunctions occur in each of the interviews recorded, the major subfunctions of Problem Description (at least in the environment studied), are RES, SUBJ and TOPIC. The subfunctions RES, SUBJ and TOPIC and the elements of the problem description they construct, are highly interrelated For example: background subject information is used to justify research and to explain aspects of both research and search topic that the intermediary has not understood or finds anomalous. Research is used to explain the users' search topic. However, intermediaries are careful to distinguish between these three aspects and to develop retrieval strategies based on search topic and not research or subject background.

The interactions recorded and analysed in this study were all taken from similar environments (university search services), concern similar types of problem (problems concerned with academic research) and deal with a relatively homogeneous group of users (postgraduate students and academic faculty). It is unlikely that the detailed findings, particularly with respect to problem description, can be totally generalised to information interactions in other settings dealing with different users and different types of problem. It would be useful, therefore, to compare the findings from this study with comparable work carried out in other environments. Unfortunately very little research has been done in this area. Most online research concentrates on the online search itself and/or the formulation of the retrieval strategy. Studies of reference interviews have tended to concentrate on other aspects of the interview or have treated Problem Description as a unit function e.g. "clarifying and negotiating the information need and search objectives " (Meadow & Cochrane, 1981). Comparison is made with two studies; one concerned with information systems as a whole and the other with fiction collections in public libraries.

In the experiments simulating the MONSTRAT model, (Belkin, Hennings & Seeger, 1984), the Problem Description function was broken down into five subfunctions:

(i) PROBLEM TYPE: e.g. procedural, decision making, cooking etc.

(ii) PROBLEM STRUCTURE: e.g. well or ill structured, well or poorly understood, few or many concepts, with or without connections etc.

(iii) PROBLEM TOPIC:

(iv) PROBLEM ENVIRONMENT: e.g. restrictions on the topic in connection with time, extent of desired answer, formulation of questions etc.

(v) PROBLEM CONTEXT: i.e. the relationship between the specific problem, as understood at the moment, and the user's life situation; why the question has been posed and why the problem is a problem.

This function specification was constructed in the context of a much wider set of problem types than those encountered in the present study e.g. How to make Bismarck herring (Belkin, Seeger & Wersig, 1983). Much of the MONSTRAT function specification is compatible however. For instance, PROBLEM TOPIC corresponds to the TOPIC subfunction. PROBLEM CONTEXT can be related to RES and possibly to SUBJ. In the current study, the RES and SUBJ subfunctions are differentiated. This is because the parts of the problem description each constructs are used differently. The element of the problem description relating to the user's research has a much closer relationship and influence on the search topic (and therefore retrieval strategy) than background subject information.

Problem Description subfunctions not explicitly mentioned in the MONSTRAT specifications are DOCS and SLIT. These subfunctions are wholly concerned with aspects of documents and document collections. Their non-appearance in the MONSTRAT function is probably because the remit of MONSTRAT was wider than bibliographic retrieval systems.

The MONSTRAT subfunction PROBLEM ENVIRONMENT seems to correspond closely to the idea of qualifying theme (see 3.1.2 (iii)). Since the elicitation of the qualifying theme was found in the recorded interactions always to be embedded within attempts to specify both the user's search topic and the user's research, it did not warrant being considered as a subfunction in its own right.

No subfunction was found in the recorded interviews corresponding to the MONSTRAT subfunction PROBLEM TYPE. This is probably because the problems in the recorded interactions were all similar in type i.e. problems concerned with carrying out a piece of academic research. The exception is interview 040684HBA where the user was compiling a report (rather than doing research) but, because the intermediary knew the user well and had carried out a number of searches for her, the intermediary was already aware of the problem type (confirmed by the intermediary after the event). It would seem that in the interactions recorded, the intermediaries had expectations about the kind of problem the users would come with and therefore did not need to acquire this information directly. Knowing the type of user problem may enable the intermediary to make a number of assumptions e.g. the type of knowledge resource to be searched, the type of information the user requires (periodical articles or non-fiction books). If the information system encompasses a very heterogeneous set of users with a wide variety of problems, then Problem Description may need to take more account of PROBLEM TYPE. The intermediary (human or machine) may need both to differentiate between problem types and to determine into which category the current problem falls.

PROBLEM STRUCTURE is another aspect about which the intermediaries recorded did not seem to need to elicit information, although it is sometimes referred to by users. For instance, the user in 290684KSA comments that his research topic is nebulous. The user in 040684HBA says "I think its on a very much more general level than an awful lot of things, that's the problem". The intermediary may infer the extent to which the user's problem is well or ill structured on the basis of how the user describes his/her research and the supplementary background information. Thus although there was little direct evidence of a PROBLEM STRUCTURE subfunction it it possible, and indeed probable, that intermediaries make such assessments with respect to the user's problem and there is some evidence that user's donate such information explicitly. What is not clear is whether or not a PROBLEM STRUCTURE subfunction is required or whether it should be treated as a meta-function, pertaining to the problem description as a whole.

Pejtersen & Austin (1983) carried out a study concerned with developing a faceted classification scheme for fiction. The classification scheme was developed from an analysis of the way in which public library users ask for fiction books. It comprises four "independent" dimensions:

SUBJECT MATTER: what the story is about;

FRAME: the time and place chosen by the author as the scenario of his/her work;

AUTHOR'S INTENTION: the author's attitude towards the work;

ACCESSIBILITY: the level of communication.

These dimensions can be correlated with the Problem Description subfunctions. That is, SUBJECT MATTER corresponds to the topics of the research (RES) and the search (TOPIC). FRAME relates to the "qualifying" theme of both RES and TOPIC. ACCESSIBILITY is not regarded as a Problem Description subfunction but as a Response Generator subfunction. No subfunction corresponding to INTENTION was found. This may be because both user and intermediary assume that the authors of the documents to be retrieved, wrote those documents within the framework of scientific and technical report writing. INTENTION may well be a significant aspect of problem description in other information system settings. Pejtersen and Austin also mention users asking for "something like ..... (book title)". This correlates well with the findings here of a DOCS subfunction.

The two studies quoted for comparative purposes indicate that the Problem Description specification developed here corresponds in general to ones derived from analysis of information interactions in other settings. It seems probable that the Problem Description specification might need to be expanded to deal with interactions in which the information retrieval systems do not access bibliographic databases and where the users are not academic researchers with research-type problems. There is also some suggestion that meta-function concepts e.g. the structure of the problem, might be a necessary addition.

Besides the specification of the Problem description function itself, the knowledge resources it will need to employ must also be considered. The results of the analysis show that the intermediary makes use of both his/her own knowledge and the user's knowledge, to construct a model of the user's problem. This finding concurs with that of Ingwersen (1982) who states:

> "In the construction of a picture which matches with the real need, the librarian continuously combines relevant concepts recalled from memory with new information linked to recognised user concepts".

The balance between external knowledge, supplied by the user directly in speech or indirectly on a pre-search booking form, and the intermediary's own internal knowledge, will depend very much on the states of knowledge of both parties. What is certain is that the user's contribution is essential and cannot be substituted by any amount of knowledge on the intermediary's part. The intermediary may choose not to attend to a significant proportion of the user's input or to minimise it, as in interview #5, but this is likely to have serious consequences with respect to the success of the search. The intermediaries seem to use their knowledge to organise the information supplied by the user and to help resolve and ambiguities or to fill any gaps (directly or by prompting the user for more information). In order to do this, the intermediaries were found to make use of a wide range of internal knowledge resources including:

- classificatory knowledge;
- knowledge of the structure of a subject domain;
- knowledge of a subject domain;
- knowledge of the literature in the subject area;
- knowledge of documents;
- knowledge of users.

Knowledge about users seems to enable the intermediary to constrain and set limits on the possible range of search topics. In other environments, where perhaps the users are better known to the intermediary, this knowledge may be of even greater significance, allowing the intermediary to infer that because the user is a certain individual then the search (in general terms) is likely to concern some particular topic.

The knowledge of the literature of the subject area, gained through search experience or study of the subject, enables the intermediary to estimate the possible outcome of a search (and therefore develop the strategies necessary to cope with too great or too few retrieved items). It may also help the intermediary in assessing the level of specificity of the search topic. Knowledge of documents, both in terms of document structure, the likely content of titles of desired documents, key authors and papers, enables the intermediary to prompt the user for information about known, useful documents.

Classificatory knowledge, knowing how subjects relate to each other, can be used to define the overall boundaries of the topic and to provide a larger, meta-structure, for the model of the user's research and search topic. The knowledge of the structure of subject domains serves to integrate the "qualifying" themes for that subject. Possible types of qualifier e.g. time or location, can be considered. If the user has failed to volunteer information on this aspect, the intermediary can (and will) prompt for it. Intermediaries use their knowledge of the subject domain to resolve ambiguities, as a check on the meaningfulness of what the user has said and to elicit more detailed information from the user. It has long been known that searches are often "easier" and more successful if the intermediary is knowledgeable in the subject area to be searched. The analysis carried out here suggests why this might be the case.

The Problem Description function interacts with other functions. From some of these function it will receive information and the other functions may need to make use of the problem description. The function with which Problem Description interacts most heavily in Retrieval Strategy. This is to be expected since the problem description is essential to the construction of the search formulation. If during the process of constructing a search strategy it becomes apparent that there are inconsistencies in the problem description which require resolving or filling, the intermediary must switch from Retrieval Strategy back to Problem Description.

Problem Description also interacts with the User Model and Problem State functions. Information from both functions is used to develop the problem description e.g. information about what the user has already found (from the Problem State function) may determine the topics <u>not</u> to be searched. The state of the problem description being modelled may prompt some explanation from the intermediary, either as a prelude to a different function or to encourage the user to co-operate in some way. In addition the problem description may provide the Response Generator function with information about the desired outcome and its form.

What is most evident from the information interactions analysed is the co-operative nature of the Problem Description function. Both parties contribute to the model building process. Intermediaries do not apply just their knowledge of information retrieval systems but must also make use what knowledge they have of subject domains - in terms of relations with other subject areas, internal structure, facts, terminology, the literature and key documents. Unfortunately this suggests that the knowledge resources of the Problem Description function of an intelligent interface would need to be substantial and, to a large extent, subject domain dependent.

The other finding that should be stressed is the multi-component nature of problem descriptions. These models are constructed through the achievement of a number of subgoals which involve obtaining descriptions of the user's research, the literature in the subject, useful documents, the subject background, and the search topic itself. It is not enough to consider problem description as a single, unit entity and any design for an intelligent interface for document retrieval systems should take this into account.

4.2. Retrieval Strategy

The goal of the Retrieval Strategy function is the formulation of an effective retrieval strategy i.e. a complete plan for searching the information system's knowledge resources. In the current context, these knowledge resources are online bibliographic databases hosted by a variety of system vendors. A a retrieval strategy comprises a number of elements. There will be a list of databases to be searched and a list of queries to be applied. These lists are constructed and arranged according to the dictates of a search strategy, and may be achieved through the use of particular search tactics. Other, subsidiary, tactics may explicitly noted for use if the initial strategy is unsuccessful. Each query consists of a command to the knowledge resource's access mechanism and one or more terms (or the equivalent sets of retrieved items). Multiple terms may be linked by boolean operators or adjacency indicators. Individual terms may be qualified by a role indicator, a field delimiter or truncated.

Search strategy formulation has been one of the most extensively investigated of all the aspects of online searching. Bates (1981) reviewing the literature on search techniques for the period 1976-1980, and disallowing items referring to pre-search interviews, cites some 163 papers. The majority of the studies in this area have been carried out with one of two objectives in mind: educating trainee searchers (e.g. Meadow & Cochrane, 1981; Hawkins & Wagers, 1982), or defining the parameters of the process which have a particular influence on search effectiveness (e.g. Fidel & Soergel, 1983; Fairhall, 1985).

Where the use of search techniques by intermediaries has been studied for its own sake, this has usually been restricted to consideration of online events and particularly reformulation of initial strategies (e.g. Fidel, 1985). Bates, in two seminal papers, covers the area of search strategy and search tactics, both on and offline and both in library and online environments (Bates, 1979a; 1979b). Harter & Peters (1985) present a typology of search heuristics i.e. rules-of-thumb or guidelines for searching. These are divided into six classes:

(i) philosophical attitudes and overall approach;

(ii) language of problem description;

(iii) record and file structure;

(iv) concept formulation and reformulation;

(v) recall and precision;

(vi) cost/efficiency;

Each class may then be further subdivided. For instance, the class "language of problem description" is broken down into: general development of search terms; acronyms and abbreviations; spelling and usage variations; compound terms; codes. Examples of the search heuristics presented are:

- don't assume perfect indexing;
- watch for differences in UK and US usage and spelling;
- use truncation to pick up word variations;
- know the stop words defined by a search system;
- always question null sets.

These heuristics were compiled from a study of the literature of searching. They do not deal at all with database selection nor the presearch interaction. Nor are questions of when and why a heuristic might be used tackled.

Some of the very few studies of search strategy formulation that are based on empirical research are those of Fidel (1984; 1985; 1986). One study investigated differences in online searching styles between intermediaries (Fidel, 1984). Two categories of search style were defined, operationalist and conceptualist, based on observation of 5 intermediaries each carrying out 10-13 searches, backed-up with additional interviews with these intermediaries. Operationalists are searchers whose search tactics generally involve modification of a retrieved set without changing its conceptual meaning whereas the conceptualists modify by changing the meaning of the concept a set represents.

A more recent study by Fidel concerns the decision making processes behind the selection of free-text terms as opposed to descriptors, by human intermediaries (Fidel, 1986). Eight searchers were asked to verbalise their though processes during searching. Some 100 protocols were analysed and a set of decision rules formulated. An example of such a rule is:

If the term is a single term meaning, and
   the term is mapped to a descriptor, and
   the concept is not "trustworthy" as an index term, and
   the descriptor is a broader term
Then use free-text terms

Fidel found that intermediaries made decisions about whether or not to use a free-text term or a descriptor to express a word on the basis of:

(a) whether or not the word can be expressed by a descriptor

(b) whether or not the word is a "good" word for free-text retrieval - where a "good" word is one that usually occurs in a particular context, is uniquely defined and is specific to the concept it represents.

The investigation into Retrieval Strategy function carried out in this research, differs from most other work on search strategy formulation in that it concerns presearch <u>negotiation</u> of retrieval tactics. Thus what is being examined is not solely the strategies, tactics and terms selected by the intermediary but also how these processes are negotiated with the user and the extent and nature of the user's involvement. It is not enough to specify a list of tactics or heuristics; it is also necessary to understand when, how and why such tactics and heuristics are used. Lists of tactics and heuristics can be formulated by interviewing intermediaries or analysis of the literature or through self-introspection but the when and how requires analysis of real human-human interactions.

The analysis of the Retrieval Strategy function in the interviews revealed that while this function tends to be Intermediary initiated and directed it is not completely intermediary dominated. The user's involvement and assistance is actively sought at all stages, particularly with respect to term selection. Intermediaries go to great lengths to find out to what extent and in what ways, the user will be able to contribute, and to provide the user with sufficient explanation to ensure their continued involvement. The least negotiated aspects are those which require an extensive knowledge of the access mechanisms of the information system's knowledge resources.

Four Retrieval Strategy subfunctions have been identified:

DB: database selection

TERM: term selection

QUERY: query formulation

STRAT: construction of search strategies and selection of search tactics.

The activation of these subfunctions did not appear to follow any temporal sequencing (e.g. DB did not automatically precede TERMS), although on the whole DB and TERMS foci occur before STRAT and QUERY. Often the activation of these subfunctions was triggered by other functions, particularly Problem Description (see also section 3.3).

The subfunctions listed have been identified by other research. Meadow and Cochrane (1981), for example, present a list of steps in the pre-online search sequence which includes:

(a) clarifying and negotiating the information need and search objectives;

(b) identifying the relevant online systems and databases;

(c) formulating the basic logic and planning search strategies;

(d) completing the search terms;

(e) making the output choices;

(f) conceptualising the search as input into the retrieval system.

DB corresponds to (b), TERMS to (d), QUERY to (c) and STRAT to (f). Similarly Smith (1979) identified six subproblems that intermediary's tackle in the presearch interview:

(a) identifying the database to be searched;

(b) identifying the concepts;

(c) developing the content search terms;

(d) developing the context search terms;

(e) refining the terms;

(f) developing the search logic.

In terms of the current analysis, there is not a one-to-one match although DB corresponds to (a), QUERY to (f) and TERMS to (c) and (d).

The Retrieval Strategy subfunctions seem to divide into two categories. One group, TERMS and DB, involves the activation of well defined subsubfunctions e.g. CONSIDER. The other, QUERY and STRAT, are less well defined and negotiated to a lesser extent.

The subsubfunctions associated with TERMS and DB involve choice-processes e.g. CONSIDER, REJECT, or assessing the appropriateness of a particular choice e.g. COVERAGE or SCOPE-NOTE, or obtaining additional information from the user e.g. SPELL or SYNONYM. Differences were found in the term selection process (TERMS) depending on whether or not a controlled vocabulary could be used. From the analysis it is possible to specify a minimal list of TERM and DB subfunctions.

An equivalent list of heuristics for query formulation or search tactics could not be produced. Comparatively little negotiation took place under QUERY or STRAT. Choices of tactic or boolean operator are seldom made explicit, let alone the underlying rationale behind the choice. There is no discussion over which system commands to use and it is probable that these are not assigned by most intermediaries until the search is actually entered online. Explanations of the decisions made with respect to the search plan, and the consequences of these decisions, are given to the user. These explanations seem to serve to maintain the user's interest or to preempt long discussions online.

It is in the Retrieval Strategy function that specialist knowledge of the information system and its knowledge resources is of particular importance. The number of databases and online systems means that one individual cannot know all that is required to search every database on every system. Therefore, the intermediary's own, internal knowledge is backed-up by information from printed guides, directories, manuals and thesauri. The balance between the use of internal and external knowledge depends largely on the intermediary's familiarity with the databases and systems selected and, to a lesser extent, on the user's familiarity with the systems.

The internal resources that the Intermediary uses to construct a retrieval strategy are primarily the Intermediary's models of the user and his/her problem. In addition, there is evidence that intermediaries also make use of their own knowledge of:

- online Information Retrieval systems in general;
- query languages;
- databases;
- indexing languages;

- searching free-text;
- thesauri;
- search tactics and strategies.

Some of this knowledge could be categorised as "professional expertise", which one might expect any trained intermediary to possess e.g. knowledge of an online system's query language. Other knowledge is clearly the product of long searching experience, particularly that relating to knowledge of a database which could not have been obtained in any other way. Work by Fenichel (1980, 1982) tends to confirm the significance of database experience. She found that the most cost-effective searches are done by intermediaries who are both experienced searchers in general and experienced on the database being searched. If the experienced searcher is unfamiliar with the database, the searches are more costly. This work and the analysis carried out here, suggests that only through experience with searching a particular database can intermediaries gain sufficient knowledge to be able to construct effective retrieval strategies for these databases.

The list of internal knowledge resources elicited from the analysis differs in some respects from those produced by other researchers (e.g. Smith, 1979; Harter & Peters, 1985). This may be accounted for by the fact that the knowledge elicited from the analysis applies only to those aspects of the Retrieval Strategy function which are negotiated or explained to the user. Thus there may be types of knowledge used by the intermediary to make non-negotiated decisions, or to assist with the realisation of the retrieval strategy online. Such knowledge cannot be elicited by the analysis methods employed here. To this end, it may be necessary to supplement functional discourse analysis or interactions with additional interviews with intermediaries and perhaps with verbal protocol methods (as has been done by Fidel, 1986). The disadvantage with interviewing is that it requires the intermediary to articulate internalised knowledge and this may be difficult. Verbal protocols suffer from the disadvantage of being applicable to the online phase only (since an intermediary could not both verbalise and interview the user simultaneously).

Whereas the goal of the Problem Description function is to develop an adequate description of the user's problem, and is therefore primarily concerned with modelling, the Retrieval Strategy function's goal is the construction of a search plan and therefore the emphasis is very much on action (planned, projected and executed) rather than descriptive modelling.

The Retrieval Strategy function was found to include at least two types of process: a selection/decision making process (essentially the DB and TERM subfunctions); and an organisation and planning process (the QUERY and STRAT subfunctions). In addition to the knowledge resources indicated previously, therefore, the Retrieval Strategy function must also make use of meta-knowledge about decision-making processes and about planning. Knowing a list of possible search tactics, for instance, is insufficient. The function needs to know when to select such a tactic, what effect it will have, what consequences that choice will result in and what the trade-offs are with other tactics. Similarly, the function must not only have knowledge resources to use as a basis for any decision making (e.g. knowledge about a database and its coverage), but also what steps should be taken in the decision-making process and in what order, and again, what are the trade-offs between alternatives and the consequences of particular decisions.

Since the decision-making processes were more extensively negotiated, and therefore open to the analysis, these factors are better defined. For instance, the subfunctions in the term selection and database selection processes have been outlined. The order in which these steps are carried out depends in part on the way in which the Retrieval Strategy function interacts with the problem description and on the knowledge of both intermediary and user and so there is no rigid sequencing. However, some possible sequences are more likely than others e.g. CONSIDER usually precedes either SELECT or REJECT in the term selection process.

In conclusion, it must be reiterated that this analysis did not and could not extract the lists of search tactics mentioned by Bates (1979a; 1979b) nor the search heuristics mentioned by Harter and Peters (1985). Instead the analysis has enabled the tasks carried out during the process of constructing a retrieval strategy to be specified, particularly with respect to database and term selection. The analysis also indicated some of the knowledge resources, internal and external, that the intermediary uses during this process and when and how this knowledge is applied.

4.3. Interaction between the Problem Description and Retrieval Strategy functions

In order to formulate a retrieval strategy, the intermediary (human or interface) must have a description of the user's problem, however minimal. The depth of problem description and the emphasis given to particular themes within that description depends, to some extent, on the intermediaries initial choice of strategy, tactics and terms. Thus the two functions, Problem Description and Retrieval Strategy are closely interconnected and interdependent. The conventional image of the presearch interviews separates out these two functions and although lip-service is paid to complex sequencing and backtracking of tasks, Problem Description is generally viewed as a unit task which is completed before Retrieval Strategy begins.

The Retrieval Strategy function uses information provided by several functions including User Model and Problem State. However, Problem Description provides the core information required by Retrieval Strategy. It is on the basis of information in the problem description that terms chosen and databases selected. The element of the problem description which is most used for term selection is TOPIC but the DOCS element may also be involved. For instance, terms may be selected through consideration of the words in the title of key documents. Database selection seems influenced by the intermediary's classification of the topic and the subject domain. Assessment of the suitability of a term or database may require the more detailed information in the RES and SUBJ elements.

The problem description also influences the strategies and search tactics that are employed. For instance, a description of a key document which includes the author's name may result in the intermediary considering the possibility of an author search. Information in the SLIT element may affect how complex and detailed a strategy is constructed.

The main role of the SUBJ element may be to serve as a context for the problem description and to provide the intermediary with enough background that s/he can negotiate a sufficiently detailed problem description. Over time, the remembered aspects of problem description SUBJ elements may serve to increase and enhance the intermediary's own subject knowledge. The RES element outlines possible search topics, (because in the academic environment it is unusual for a user to want a search wholly unconnected with his/her research), and place these in context. Both the RES and SUBJ elements may be used by the intermediary later on, during the online phase, as a resource against which the relevance of the retrieved items can be assessed.

The flow of interaction is not solely from Problem Description to Retrieval Strategy. There is evidence that the retrieval strategy affects the development of the problem description. Difficulties with constructing a retrieval strategy may make it apparent that certain aspects of the problem description need further development or are incorrect. In most cases, (see interview #5 for an exception), the intermediary will pause in the Retrieval Strategy process and attempt to develop the problem description by filling the gaps or correcting an erroneous impression.

The return from Retrieval Strategy to Problem Description need not always indicate a "faulty" problem description. It may be that the elements emphasised in the problem description are not those required by the Retrieval Strategy function. The selection of a particular search strategy or tactic may call for particular types of information from the problem description and if that information is not already available it must be acquired. For example if the intermediary decides to search for key papers, the titles and authors of those papers must be ascertained.

The complexity is intensified by the non-unitary nature of both Retrieval Strategy and Problem Description. Both functions are composed of separate but interacting subfunctions. The balance of subfunction activity within each function is dependent on a number of factors including the problem type, the retrieval systems available, the intermediary's knowledge of the subject, of the retrieval system and so on. However, the analysis indicates that the balance of subfunction activity in the Retrieval Strategy process may be influenced by the activities of the Problem Description function and vice versa. For example, in interview 040684HBA the problem is well-defined, involving one key concept with a specific name (activated carbon) and a number of broader concepts (use, production etc.). Consequently, the negotiations in the interview concentrate on expressing the broader concepts through database selection i.e. choosing databases whose coverage will restrict the context. Almost all the Retrieval Strategy activity that is negotiated concerns database selection (18 utterances cf 1 utterance concerned with term selection). Alternatively if the search topic is covered by one or two databases only but is a less well defined topic and the terminology less specific, most effort will be put into term selection. This is the case in interview #4 where a considerable amount of time is spent selecting terms to express the concept "community education" which is the very concept that the user is uncertain about and is hoping that information from the search will help to define.

The <u>negotiated</u> retrieval strategy may not correspond exactly with the initial retrieval strategy used online. Typically the negotiated retrieval strategy may be incomplete, lacking those aspects the intermediary thinks unnecessary to discuss with the user. The negotiated retrieval strategy in interview 040684HBA, for example, consists of a sequence of databases selected for searching. With the exception of "activated carbon" itself, search terms are not mentioned. The actual online search however, includes quite a number of search terms: activated carbon, production, water treatment, sugar decolourisation. These appear to have been derived directly from the problem description. One of the purposes of negotiating the term selection process may be to enable the intermediary to select "good" terms, given a complex problem description encompassing a great number of concepts, and to gain an understanding of the relative importance of the concepts and their relation to each other.

The complex sequencing of tasks suggests that processing cannot be linear and the pattern of interaction between functions and subfunctions suggests a highly interactive process. The simple, linear model, in which Problem Description (negotiation of need etc) was completed before Retrieval Strategy (transformation of the compromised need, etc.) does not correspond to what occurs in real human-human interactions. Further, research has tended to become restricted to investigating how problem descriptions are used to generate retrieval strategies and not explore how the activities of the Retrieval Strategy function influence those of the Problem Description function and vice-versa.

IV. FORMALISMS FOR THE PROBLEM DESCRIPTION FUNCTION.

1. INTRODUCTION.

Implementation of the Problem Description and Retrieval Strategy functions within an intelligent interface requires not only that the knowledge resources each function employs be specified but also that a formalism be selected or developed to represent these knowledge resources. This is no simple task because, as the analysis of the information interactions has indicated, a wide variety of types of knowledge are used; ranging from detailed subject knowledge through to classificatory knowledge, or knowledge about a specific database through to knowledge about online systems in general.

Implementation of the Problem Description function requires, in addition, that a formalism be devised for representing the model of the user's problem which the system needs to construct. This model is constructed from a framework derived from internal knowledge resources (for that particular function), supplemented to a greater or lesser extent with information obtained directly, or inferred, from the user's utterances.

Similarly, implementation of the Retrieval Strategy function requires a formalism for representing the retrieval strategy. A number of quite different elements would need to be incorporated into this representation including "declarative" elements, such as terms and lists of databases, organisational factors, such as the order in which to search the databases, heuristics and tactics to use when carrying out the search and the goals and subgoals of the overall search plan.

Of the two functions investigated, Problem Description is the more fully specified. The problem description it constructs could be examined in detail using the analysis techniques. Therefore emphasis has been placed on developing formalisms to represent the problem description and suggesting representations suitable for the knowledge resources of the function. Unfortunately, much of the "planning" knowledge used to construct a retrieval strategy could not be elicited from the interview transcripts (see chapter III, section 4.2) and representational formalisms for the Retrieval Strategy function cannot be discussed at this stage. Sufficient information is available, though, to assess the extent to which problem descriptions constructed using the formalism outlined here could be employed by the Retrieval Strategy function to develop the kinds of retrieval strategy observed in the interactions.

In terms of overall system design, it is important to ascertain whether there should be a single representation scheme for all the interface models and for all the knowledge resources. An initial analysis and comparison of the models developed by the User Model and Problem Description functions indicated that there are substantial differences. The two functions are distinct entities, making use of different types of knowledge resource and they appear to construct dissimilar models which differ not only in content but also in manner and scope. Thus it appears that a uniform method of representation for the models may be neither appropriate nor efficient (Brooks, Daniels & Belkin, 1985).

The same appears true as far as function knowledge resources are concerned. Given the very wide range of categories of knowledge which seem to employed, even within one function, it seems unlikely that a single representational formalism could adequately represent them all. Although most knowledge-based systems employ only one type of representation formalism it has long been recognised that implementation of many applications requires, ideally, the use of several knowledge representation schemes e.g. Bobrow, (1975):

> " It is often convenient and sometimes necessary to use several different representations within a single system."

What is lacking is any clear guidelines on how to select an appropriate representation formalism. Sloman (1985) has emphasised the need to explore different formalisms and evaluate which formalisms are best suited for which purposes.

Advantageous as mixed representations may be on theoretical grounds, they add to the complexity of the system and the use of mixed representations on a large scale is difficult to achieve. Employing more than one representation method means, for example, using more than one inferencing mechanism. Nonetheless it it is clearly disadvantageous to "force" knowledge into inappropriate representations. Therefore, representations for the Problem Description knowledge resources have been suggested on the basis of their appropriateness to the category of knowledge to be represented and on the way in which this knowledge is used in the problem description.

2. REPRESENTING PROBLEM DESCRIPTIONS.

2.1. Overview.

Problem descriptions are descriptive in nature, employing diverse knowledge resources and embodying different kinds of information. In general, a substantial proportion of the information represented in this model has been contributed, directly or indirectly, by the user. The intermediary mechanism (human or machine) must use its knowledge to provide a structure for the model within which to place the information donated by the user. If the knowledge of the intermediary mechanism encompasses some of the background to the user's search topic, then the framework may be more defined and detailed than if the subject area is unfamiliar. The structure, and to some extent "context", is provided by the intermediary mechanism whilst the basic "content" of the problem description is derived from the user's utterances.

Representing the model of the user's problem requires a formalism which permits an explicit structured description and which will allow the intermediary's knowledge to be represented and the user's knowledge to be incorporated. Of the main types of knowledge representation formalism - frames, semantic networks, production rules and predicate calculus (Barr & Feigenbaum, 1981), only frames and semantic networks would seem to meet these criteria.

Frames are structures which specify the objects and events associate with a particular concept. Each frame comprises a number of labelled slots which may be filled with values taken from a particular situation (Minsky, 1975; Bobrow, 1975; Kuipers, 1975). Associated with each slot is usually a default value and an indication of the range of acceptable values with which that slot might be filled. Frames provide a framework within which to interpret a situation. Incomplete or missing knowledge is clearly indicated by the absence of values for slots. Required information is indicated by the slots themselves and the range of values that may be used to fill them. Using frames as a method of knowledge representation facilitates expectation driven processing i.e. looking for things that are based on the context (Barr & Feigenbaum, 1981; Aikins, 1984).

Semantic networks are a set of knowledge representation methods based on a common notation of labelled noses and arcs (or links) which connect the nodes together. Nodes usually represent objects, concepts or situations in the domain and the arcs represent the relations between them (Woods, 1975; Brachman, 1979; Barr & Feigenbaum, 1981). The idea of using semantic networks originated in the work of Quillian (Quillian, 1968). Since then semantic network systems have been developed in a number of areas including psychology, linguistics and artificial intelligence. Unfortunately, the fact that semantic networks have been exploited in so many domains, each with their own research goals, has meant that apart from the commonality of the notation, different semantic network systems bear little resemblance to each other, neither in what is represented, how it is represented or how it is used (Brachman, 1979).

The great variety in problem descriptions, both in scope and type (even within quite narrow subject domains), means that there is rarely a typical problem description and certainly none is suggested in the information interactions recorded. Although there is evidence that problem descriptions are structured this seems to occur at quite an abstract level and the range of entities represented is almost indefinable. Since there appears to be no typical "framework", and an indefinable range of acceptable values with no clear defaults, the use of frames to represent the problem description does not seem to be appropriate.

The representational formalism which seems offer the flexibility and descriptive power required, is that of semantic networks. Semantic networks are a relatively well understood method of knowledge representation and have been made use of in several intelligent knowledge-based systems e.g. PROSPECTOR (Duda, Gaschnig & Hart, 1979). They have also been exploited as a method of representing aspects of human discourse (e.g. Grosz, 1978).

A formalism for representing problem descriptions, using partitioned semantic networks, has been developed. Each element of the problem description: search topic, research, subject background, literature in the subject area and document descriptions, is represented in a different plane. These planes may be regarded as analogous to the planes in Quillian's semantic networks (Quillian, 1968) - except that here the planes refer to different elements of the problem description whereas Quillian's relate to semantic information about a particular concept. Each of the subfunctions involved in problem description; TOPIC, RES, SUBJ, SLIT and DOCS, operates on the corresponding element ("plane") in the model. The representation formalism specifies and defines the permissible node and link types for each problem description element. It is assumed that the knowledge resources for the Problem Description function contain meta-knowledge about the structures for modelling the problem description elements, and for modelling the various concepts, themes and topics within each element. The knowledge resources must also embody knowledge about the way in which propositions derived from the user's utterances, may be incorporated into the model.

Representations of problem descriptions have been drawn up manually for every foci and for the interaction as a whole, for each of the recorded interviews. An example, using interview 190684HBA, is presented in Appendix 5.

2.2. Partitioned Semantic Networks.

There are a number of formalisms for semantic networks. The formalism proposed to represent problem descriptions is based on **partitioned semantic networks** which were developed by Hendrix (1978a; 1978b; 1979). Partitioned semantic networks make use of constructs from set theory and predicate calculus so that the networks can readily be used to generate statements in formal logic.

They possess all the expressive power of predicate calculus. Quantification, conjunction, disjunction, and negation can be represented.

There is no restriction on the types of objects that may be represented by the nodes. Arcs however are restricted to the encoding of formal binary relationships such as element, subset and deep case relationships. Element arcs denote the relationship between a particular element and the set of such elements. For example:

$$\text{GIRLS} \xleftarrow[\text{e}]{\qquad\qquad\qquad} \text{MARY}$$

Mary is an instance, or element, of the set girls. This is represented as two object nodes, GIRLS and MARY, linked with an element arc (labelled "e"). Subset arcs link together nodes where what is represented by one node is a subset of what is represented by the other. For example:

$$\text{GIRLS} \xrightarrow[\text{s}]{\qquad\qquad\qquad} \text{HUMANS}$$

Girls are a subset of the set of all humans. This can be represented as two object nodes, GIRLS and HUMANS, linked with a subset arc (labelled "s").

Arcs may also represent deep case relations i.e. the deep cases proposed by case grammars (e.g. Fillmore, 1968). Deep cases specify the "semantically relevant syntactic relations" by which noun phrases relate to verbs. In the sentence "Mary hit the computer with a brick" - the verb "hit" has associated with it an **agent** (Mary), and **object** (computer) and an **instrument** (brick). In the context of partitioned semantic networks, the deep cases must be functions and must be constant over time and circumstance. In his specification of this network formalism, Hendrix does not employ a fixed set of case relationships but indicates that case relations should be assigned where ever possible, since this results in a more detailed indexing of the network and therefore increases efficiency (Hendrix, 1979).

Relationships that are not represented by arcs are represented by nodes which have outgoing case arcs pointing to the participants in the relationship. The outgoing arcs may be labelled with the case name (Hendrix, 1979) or with the name of the argument (Duda et al., 1978). For example, the statement "BIOSIS is hosted by Dialog" could be represented as a relation node HOSTED-BY which has arcs pointing to two object nodes, BIOSIS and DIALOG. The arcs would be labelled OBJECT and AGENT respectively i.e.

```
              BIOSIS      DIALOG
                 ^           ^
                 |           |
        object   |           |  agent
                 |           |
                 |           |
                  HOSTED-BY
```

Each particular instance of a relation is an element of the set of all relations of that type. Thus the particular HOSTED-BY relation depicted above, and labelled H1 below, is one element of the set of all such relations i.e.

```
      BIOSIS     DIALOG
         ^         ^
         |         |
  object |         | agent
         |         |
     [   (H 1)     ]--------------> HOSTED-BY
                          e         RELATIONS
```

Often each member of a particular set of relations will take the same arguments e.g. all HOSTED-BY relations take the arguments AGENT and OBJECT. It may also be possible to define which types of object node should be linked by these case arcs to the relation e.g. for the HOSTED-BY relation, the AGENT arc should be linked to nodes representing online vendors, and the OBJECT arc to nodes representing particular databases. A "typical" structure can be denoted for a particular set of relations, and this structure is called a delineation. A delineation for the HOSTED-BY relation is presented in fig. 4.1.

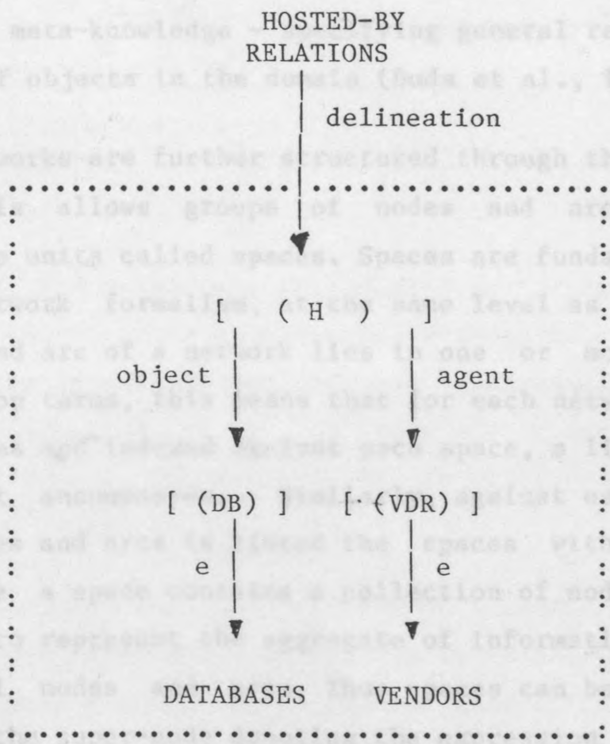


Figure 4.1 : a delineation for HOSTED-BY relations

More formally, set delineation names and restricts the participants of situations in the set i.e. names the case relations that are to be associated with the situation being delineated, and indicates a possible set of values for each case. The delineation for HOSTED-OF relations could be formally denoted as:

V x { member (x, HOSTED-BY
    => d,v [member (d, DATABASES) & object (x,d)
        & member (v, VENDORS) & agent (x,v)]}

i.e. all HOSTED-BY situations have an object and an agent, where the agent is a member of the set of vendors and the object is a member of the set of databases. Delineation in effect provides a framework within which particular situations may be represented. Hendrix notes that delineations may be used to resolve anomalies and ambiguities in candidate representation of natural language statements (Hendrix, 1979). PROSPECTOR appears to use delineation as a form of meta-knowledge - specifying general relations between categories of objects in the domain (Duda et al., 1979).

The networks are further structured through the use of **partitioning**. This allows groups of nodes and arcs to be bundled together into units called **spaces**. Spaces are fundamental entities in this network formalism, at the same level as nodes and arcs. Every node and arc of a network lies in one or more spaces. In implementation terms, this means that for each network, there is a list of spaces and indexed against each space, a list of the nodes and arcs it encompasses. Similarly, against each entry in the index of nodes and arcs is listed the spaces within which they fall. Since a space contains a collection of nodes and arcs, it can be used to represent the aggregate of information encoded by its internal nodes and arcs. Thus spaces can be used as super-nodes, with the super-node denoting the expression of the information represented by the space.

Several spaces can be grouped together to form composite bundles of nodes called **vistas**. Vistas are usually created in a hierarchical manner. A new vista inherits a view of (or access to) the information in the "parent" vista and the newly added space can be used for extending local information without altering the view of the parent. When new vistas are created, they form a partial ordering of viewing capability. This ordering is denoted by heavy arrows which point to parent vistas. Information is thus visible in any space that may be may be reached by following the arrows (see fig.4.2).

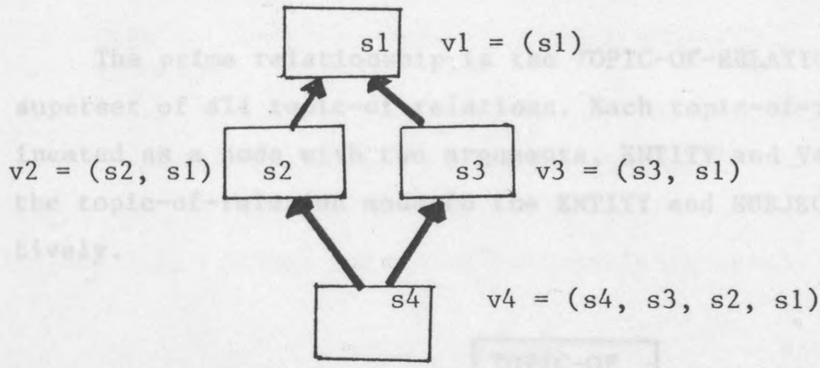


Figure 4.2: Spaces and vistas:
(adapted from Hendrix (1979))

In figure 4.2, the spaces are represented as rectangles. The list notation to the right of each rectangle indicates the vista associated with that space. Arrows indicate the inheritance of viewing capability i.e. from v4 it is possible to see information in spaces s3, s2 and s1, but from v3 it is only possible to see information in s3 and s1.

Conventions based on the partitioned semantic network formalism, have been developed to enable representations to be constructed for each element of the problem description. These conventions specify the structure of "internal" knowledge into which information supplied by the user can be incorporated. This organisational framework is relatively well defined for the DOCUMENT and SUBJECT-LITERATURE elements, less so for the RESEARCH and TOPIC elements and least for SUBJECT elements.

2.3. Conventions for RESEARCH and TOPIC elements.

For the TOPIC and RESEARCH components, a common set of conventions have been devised. A basic structure specifies at a very abstract level the kind of information and relations anticipated and required. The information supplied verbally by the user is used to instantiate values in this framework or to generate inferences through which such values would be instantiated. The model may be further extended through the use of knowledge supplied from the intermediary's own knowledge resources.

The prime relationship is the TOPIC-OF-RELATION. This is the superset of all topic-of-relations. Each topic-of-relation is delineated as a node with two arguments, ENTITY and VALUE. These link the topic-of-relation node to the ENTITY and SUBJECT nodes respectively.

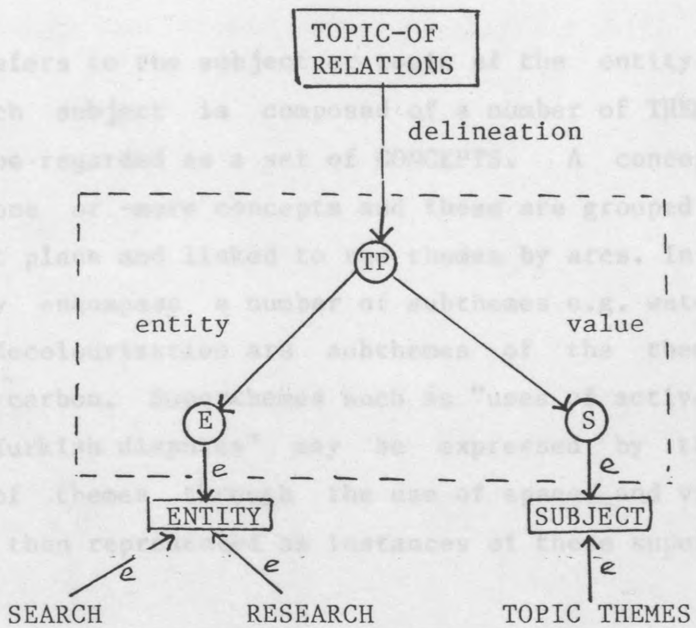


Figure 4.3: the delineation for Topic-of-Relations

ENTITY: refers to the object whose topic is the point of concern. In the environment from which the recordings have been taken, there are only two classes of entity: search topic and the user's research. The user's research can be further defined by type e.g. doctoral, masters, project, and/or by time period e.g. past, present, future.

```
USER'S    <-------------- DOCTORAL RESEARCH
RESEARCH         e
   ^
   |
   |---------------------- CURRENT RESEARCH
                 e
```

Figure 4.4 : Subset relations for user's research.

SUBJECT: refers to the subject or topic of the entity (see fig. 4.5). Each subject is composed of a number of THEMES and each theme can be regarded as a set of CONCEPTS. A concept set may comprise one or more concepts and these are grouped together in the concept plane and linked to the themes by arcs. In some cases, themes may encompass a number of subthemes e.g. water treatment and sugar decolourisation are subthemes of the theme uses of activated carbon. Superthemes such as "uses of activated carbon" or "Greek-Turkish disputes" may be expressed by the grouping together of themes through the use of spaces and vistas. Subthemes are then represented as instances of these superthemes.

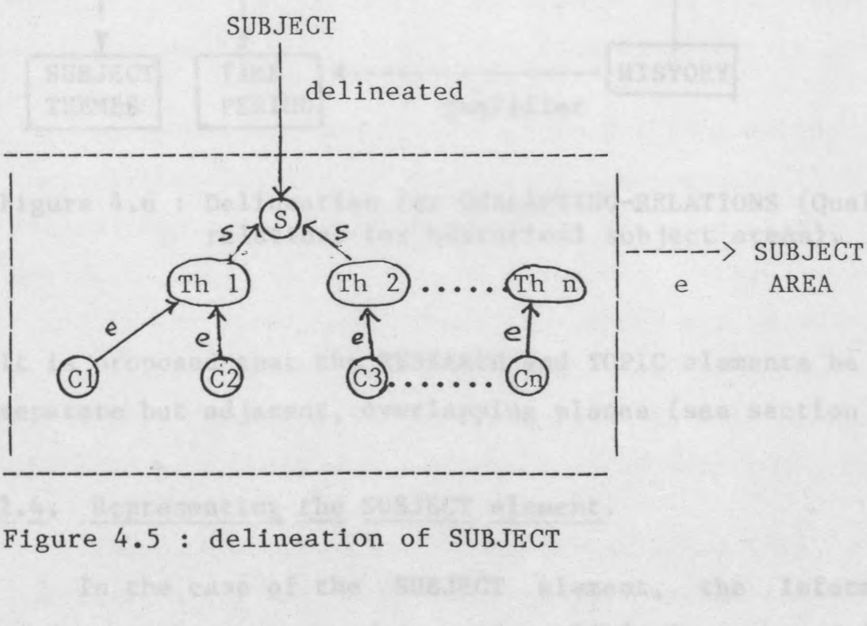


Figure 4.5 : delineation of SUBJECT

The subject, as a whole, is a subset of a broader SUBJECT-AREA e.g. "Greek-Turkish relations" might be regarded as a subset of

"International Relations". Each SUBJECT-AREA has associated with it one or more QUALIFYING-RELATIONS. Qualifying-relations take two arguments: VALUE and OBJECT. These two nodes are linked by arcs labelled with the argument they connect. OBJECT refers to the subject which is qualified, usually the subjects of the search topic. VALUE refers to the qualifying aspect e.g. time-period, educational level, geographical location (see fig. 4.6). The particular qualifying aspect if determined by the subject-area. Thus the subject-area history delineates a qualifier which is a time period.

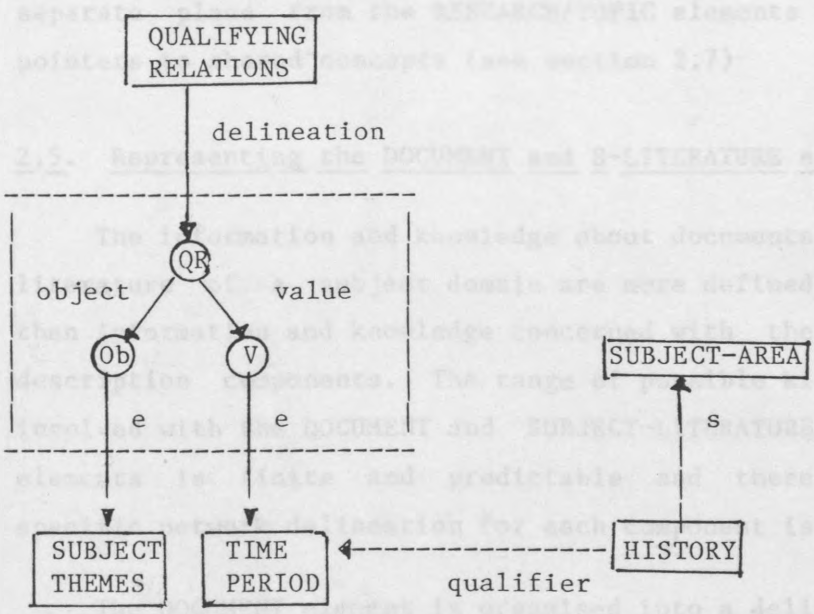


Figure 4.6 : Delineation for QUALIFYING-RELATIONS (Qualifying relations for historical subject areas).

It is proposed that the RESEARCH and TOPIC elements be modelled on separate but adjacent, overlapping planes (see section 2.7).

2.4. Representing the SUBJECT element.

In the case of the SUBJECT element, the information that needs to be represented varies widely in type and form. Background subject information is usually donated by the user and is probably organised by the intermediary on the basis of what the intermediary already knows about the subject domain. Ingwersen, for instance, concluded from his analysis of librarian-user negotiations that librarians combine their own internal knowledge with new information provided by the user where that information contained recognised concepts. Concepts in the user's statements which were not recognised by the librarian seemed not to be perceived (Ingwersen, 1982). It is proposed that the SUBJECT element of the problem description should be represented using partitioned networks and the standard case relationships (i.e.comparable to the nets constructed from natural language utterances by Hendrix, (1978a; 1978b; 1979). The SUBJECT component is modelled on a separate plane from the RESEARCH/TOPIC elements but is linked by pointers to shared concepts (see section 2.7)

2.5. Representing the DOCUMENT and S-LITERATURE elements.

The information and knowledge about documents and about the literature of a subject domain are more defined and catagorical than information and knowledge concerned with the other problem description components. The range of possible kinds of knowledge involved with the DOCUMENT and SUBJECT-LITERATURE (S-LITERATURE) elements is finite and predictable and therefore a single, specific network delineation for each component is proposed.

The DOCUMENT element is organised into a delineation in which the relation DESCRIPTION-OF takes two arguments: VALUE and OBJECT. The object of the description may be either a single document or a document set. Descriptions for both categories were found to occur in the interviews. The object types may be so arranged that specific documents are represented as instances of a group of documents, thereby allowing information in more general descriptions to be inherited by descriptions of individual documents.

Value refers to the document description itself. A document description (or value) is composed of a number of entities including CITATION, ABOUTNESS, and WORDS (see fig. 4.7). Each of these entities can be broken down further. CITATION comprises standard document description elements such as TITLE, AUTHOR, DATE-OF-PUBLICATION, DOCUMENT TYPE. ABOUTNESS refers to the overall topic which the document is "about" and to the broader subject areas into which the document content might be classified. There are obvious links between this entity and the subject-area and subject-theme nodes in the RESEARCH/TOPIC plane. WORDS refers to terms which are known, or thought likely, to be associated with this document and each term is composed of one or more concepts.

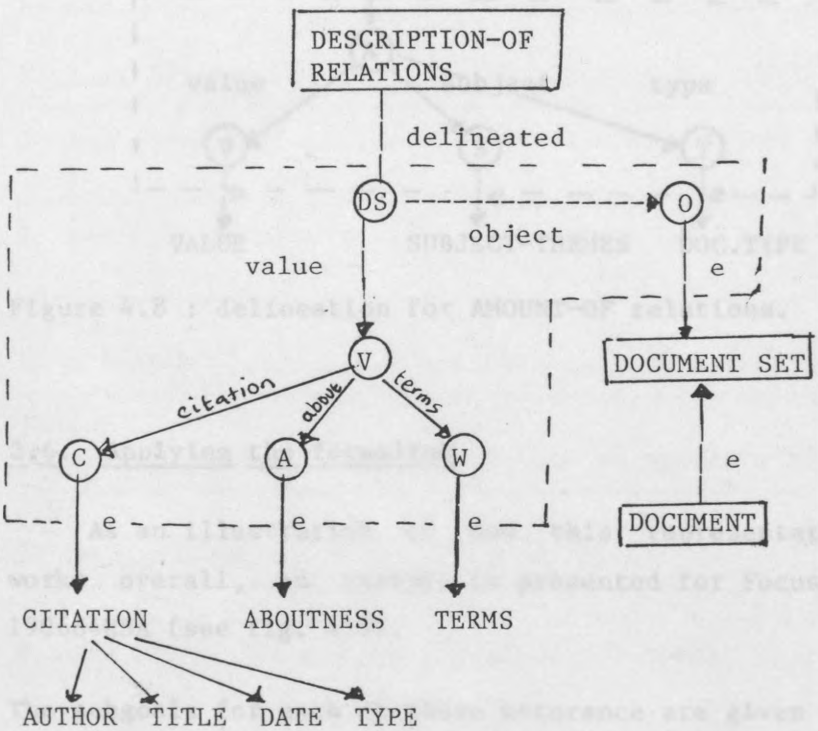


Figure 4.7: Delineation of DOCUMENT.

Knowledge about the literature of the subject seems at a less detailed level than any of the other elements. A very simple delineation is probably sufficient to represent this aspect of problem description in the S-LITERATURE plane. The proposed delineation consists of a relation AMOUNT-OF which takes three arguments, SUBJECT, VALUE and TYPE. The SUBJECT node may point to one or more subject-themes or part subject-themes, which may correspond to subject themes in the TOPIC plane. The VALUE node represents what is known or guessed about how much is written on the topic. This value could be weighted by a certainty factor based on factors such as an assessment of how well the user knows the literature and so on. TYPE represents the document type e.g. book, periodical, newspaper.

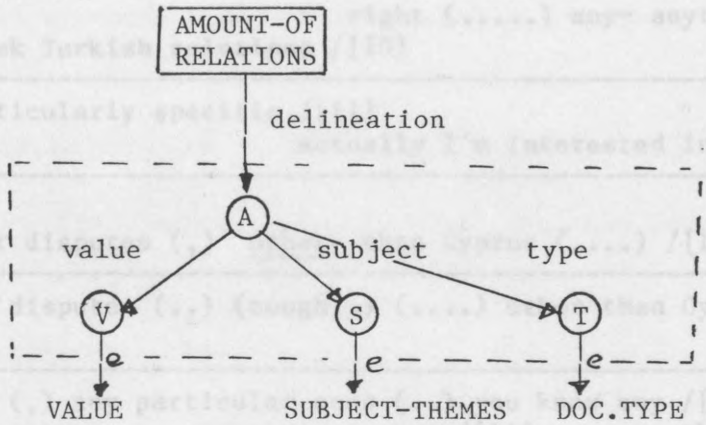


Figure 4.8 : delineation for AMOUNT-OF relations.

2.6. Applying the formalism

As an illustration of how this representation convention works overall, an example is presented for Focus 2 of interview 190684HBA (see fig. 4.9).

The subgoals for each of these utterance are given in table 4.1.

```
--------------------------------------------------------------------
I  right ok (,) right (laugh ...) we've got that out of the way
U                     (laugh ..)                       (laugh ..)
--------------------------------------------------------------------
I (laugh ..) what's what's the subject of y- your query /[9]
U
--------------------------------------------------------------------
I                      right (.....) any- anything
U  Greek Turkish relations /[10]
--------------------------------------------------------------------
I  particularly specific /[11]
U                        actually I'm interested in their (,)
--------------------------------------------------------------------
I                                                       right
U their disputes (,)  other  than Cyprus (....) /[12]
--------------------------------------------------------------------
I  (,) disputes (..) (cough .) (....) other than Cy- are there
U
--------------------------------------------------------------------
I  any (,) any particular ones (..) you know any /[15]
U                          um /[14]              the Aegean (,)
--------------------------------------------------------------------
I
U  dispute? (.........) and the: their disputes over the
--------------------------------------------------------------------
I
U  treatment of (....) the (,) Turkish minority in Greece and
--------------------------------------------------------------------
I                             right (.........) /[17a]
U the (,) Greek minority in Turkey /[16]
--------------------------------------------------------------------
```

Figure 4.9: Extract from interview 190684HBA, focus 2.

Focus	Utterance no.	Participant	P.D. subgoal
2	9	I	TOPIC
	10	U	TOPIC
	11	I	TOPIC
	12	U	RES
	13	I	SUBJ
	14	U	-
	15	I	SUBJ
	16	U	SUBJ
	17a	I	-

Key: I=intermediary, U=user

Table 4. : subgoals for utterances in focus 2, interview 190684HBA

Fig. 4.11 represents the problem description constructed from the utterance pair 9-10 i.e.

I: What's the subject of your query?

U: Greek Turkish relations.

Fig.4.12 represents the problem description following utterances 11-12 i.e.

I: Right, anything in particular?

U: actually I'm interested in their disputes other than Cyprus

Figure 4.13 is the representation of the problem description following utterances 15-16 i.e.

I: Right disputes other than Cyprus. Do you know any?

U: the Aegean dispute and their disputes over the treatment of the Turkish minority in Greece and the Greek minority in Turkey.

The RESEARCH and TOPIC elements are grouped together on separate but overlapping planes and the the SUBJECT is on another. Concepts are represented once, in the concept-plane, and are linked to themes or terms in the other planes as required. Planes are indicated in the figures by dotted lines

Within the TOPIC plane, several of the nodes have been grouped or partitioned into concept-groups or spaces (Hendrix 1979 (a), 1979 (b)), which can then act as fundamental entities on the same level as nodes and arcs. Thus "Greek Turkish Relations" can be treated as one whole theme as well as three individual themes. Spaces are indicated in the figures by dashed lines.

This process can be extended one stage further. Several spaces can be grouped together to form vistas (Hendrix 1979 (a), 1979 (b)). Thus vistas can be superimposed over the network structure to group together themes, to suggest the ranking of

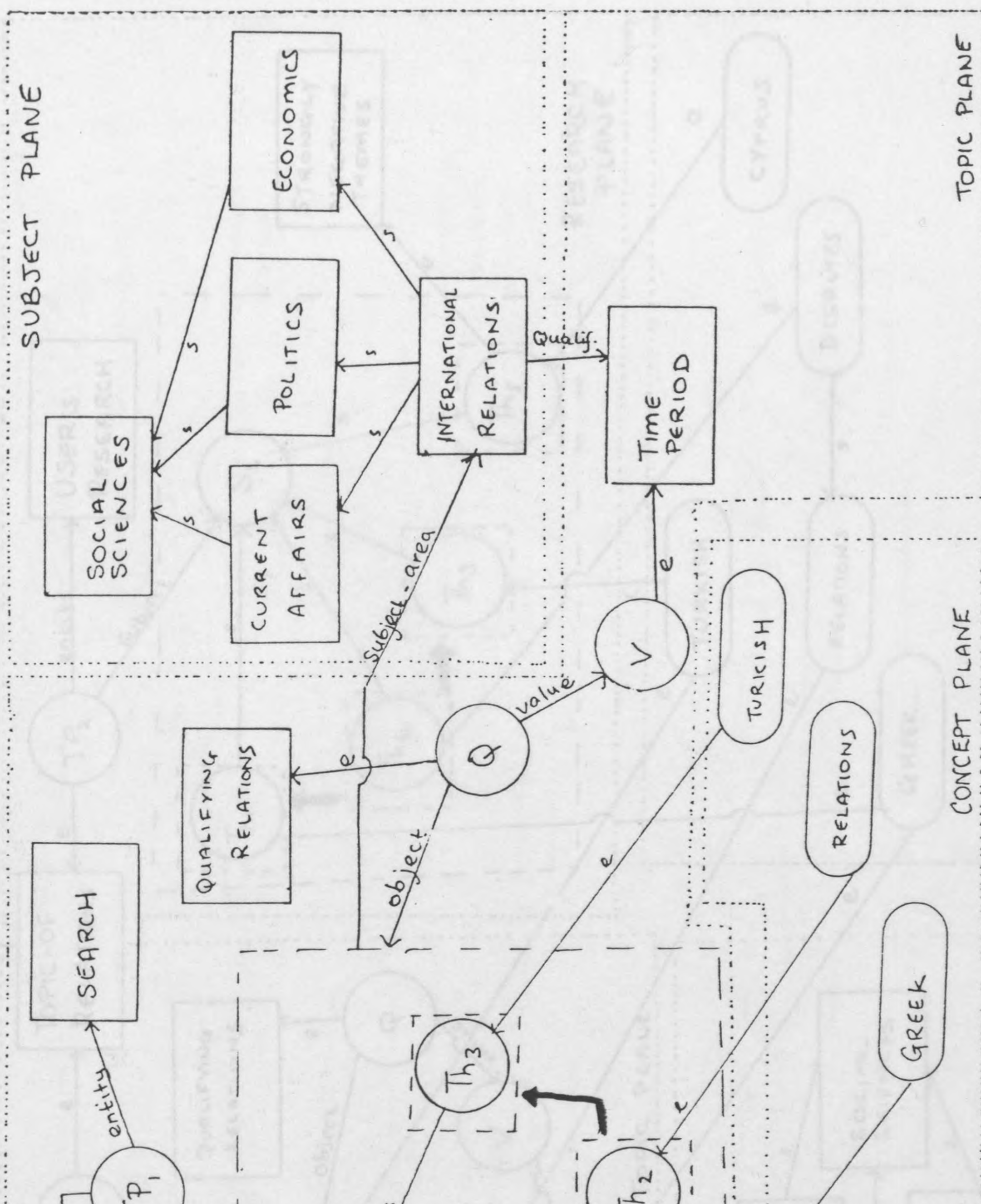


Figure 4.11: problem description for Interview 190684HBA, utt 9-10

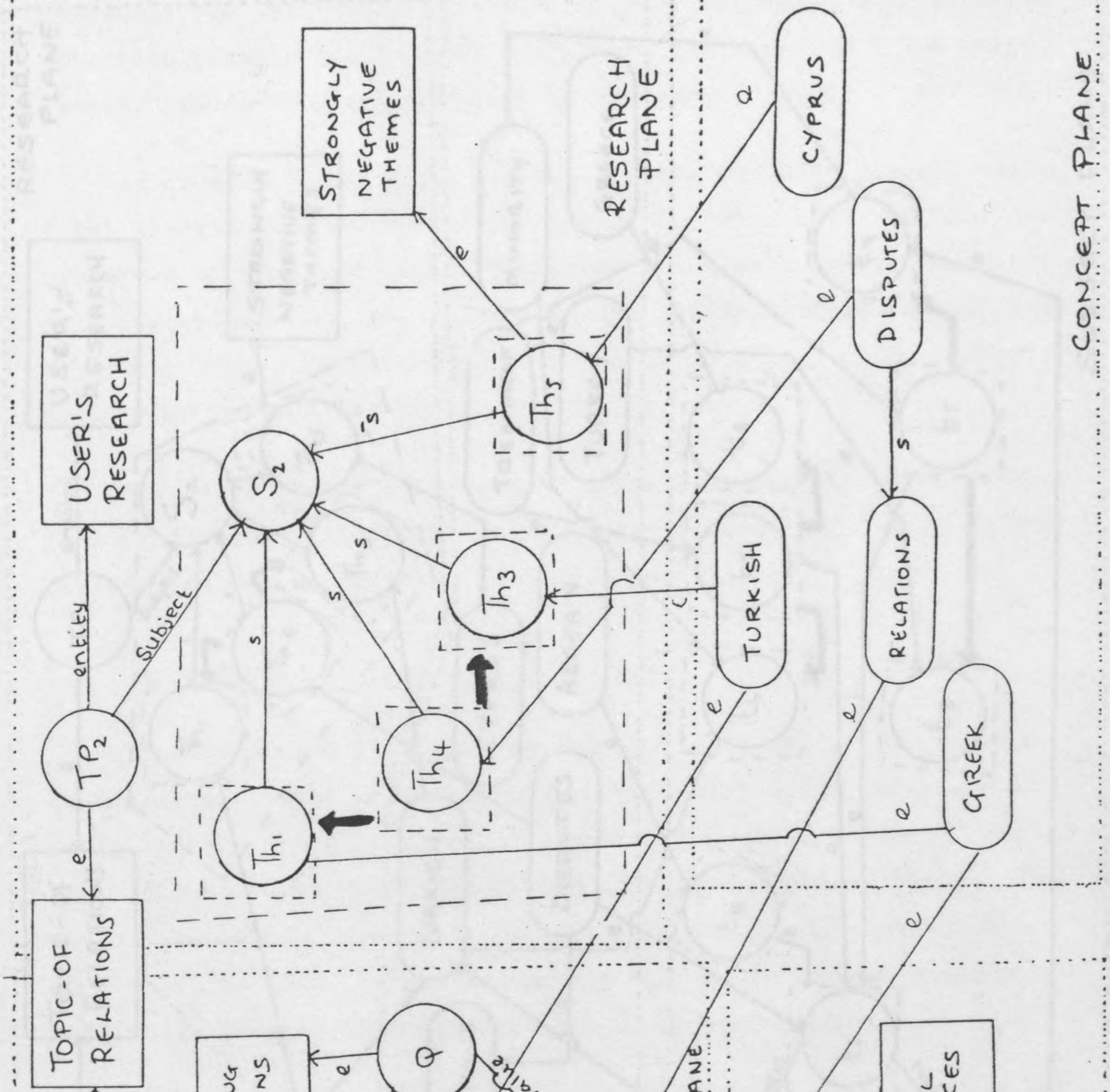


Figure 4.12 : problem description for Interview 190684HBA, utt.11-12

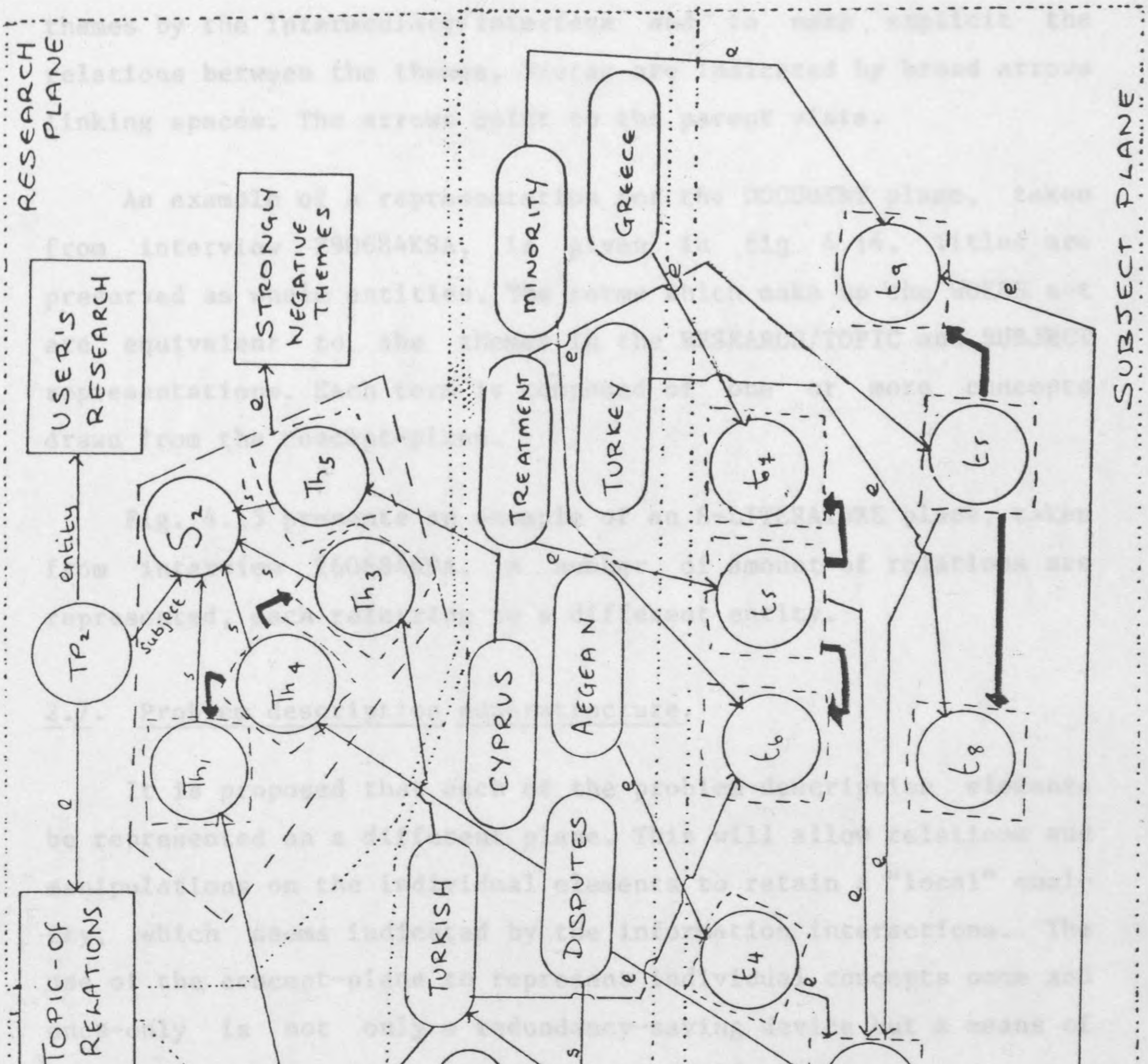


Figure 4.13 : problem description for Interview 190684HBA, utt.15-16

themes by the intermediary/interface and to make explicit the relations between the themes. Vistas are indicated by broad arrows linking spaces. The arrows point to the parent vista.

An example of a representation for the DOCUMENT plane, taken from interview 290684KSA, is given in fig 4.14. Titles are preserved as whole entities. The terms which make up the WORDS set are equivalent to the themes in the RESEARCH/TOPIC and SUBJECT representations. Each term is composed of one or more concepts drawn from the concept-plane.

Fig. 4.15 presents an example of an S-LITERATURE plane, taken from interview 260684KSA. A number of amount-of relations are represented, each referring to a different entity.

2.7. Problem description superstructure.

It is proposed that each of the problem description elements be represented on a different plane. This will allow relations and manipulations on the individual elements to retain a "local" quality, which seems indicated by the information interactions. The use of the concept-plane to represent individual concepts once and once-only is not only a redundancy-saving device but a means of indirectly forging links between the individual planes.

The planes represent one level of model superstructure. On another level, the use of spaces and vistas allows the parts of the model to be abstracted. Thus it is possible to represent the subject of the search topic as a whole and to represent each of the themes in that topic either individually or as larger groupings. Not only does this permit consideration of the problem description at varying levels of abstraction by the Problem Description function but also enables the other functions to interact with the problem description at different levels of abstraction. This is significant because the other functions may often require only an abstract perception of the problem description without any of the underlying detail. For instance, the User Model may require information about the subject of the user's

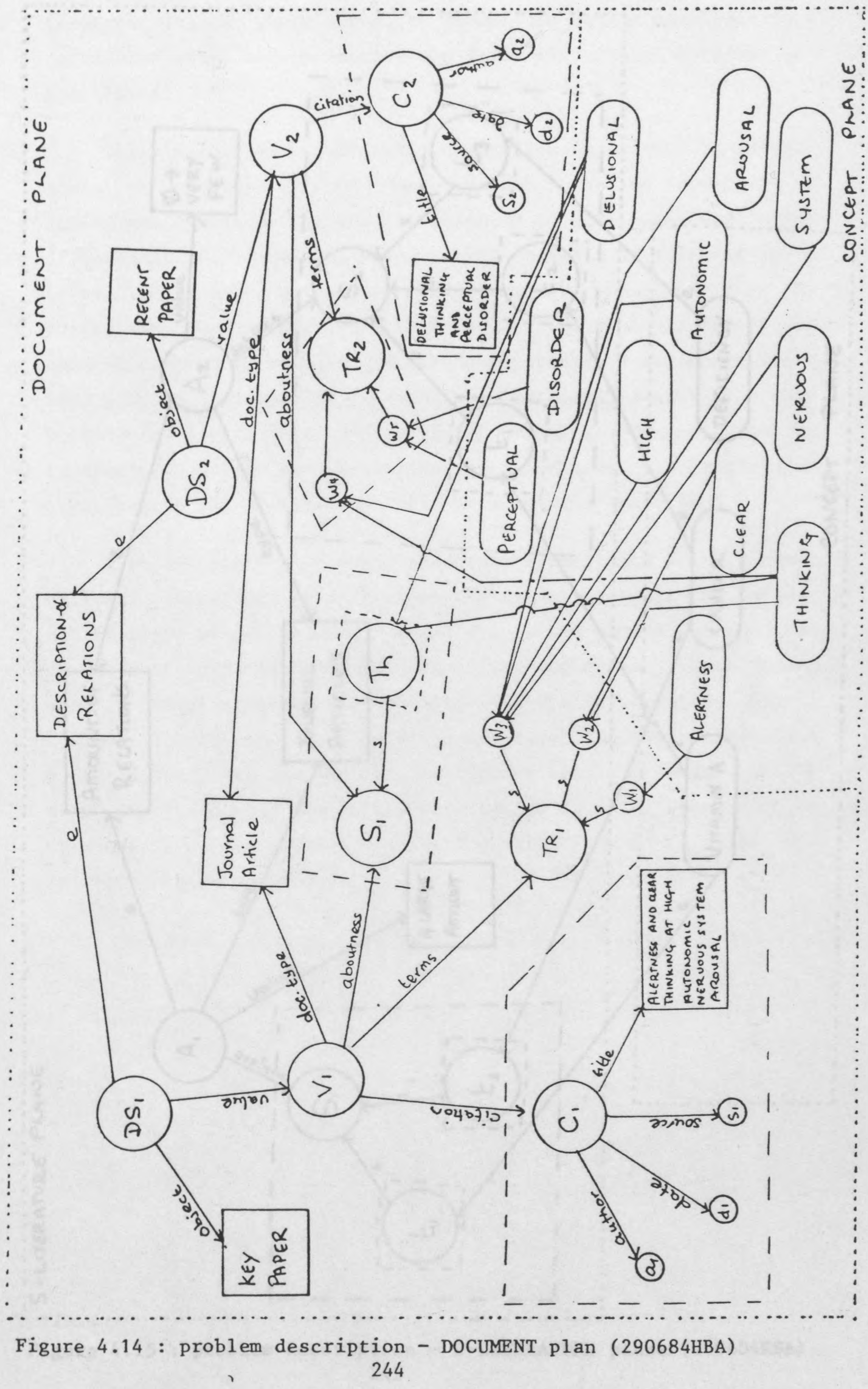


Figure 4.14 : problem description - DOCUMENT plan (290684HBA)

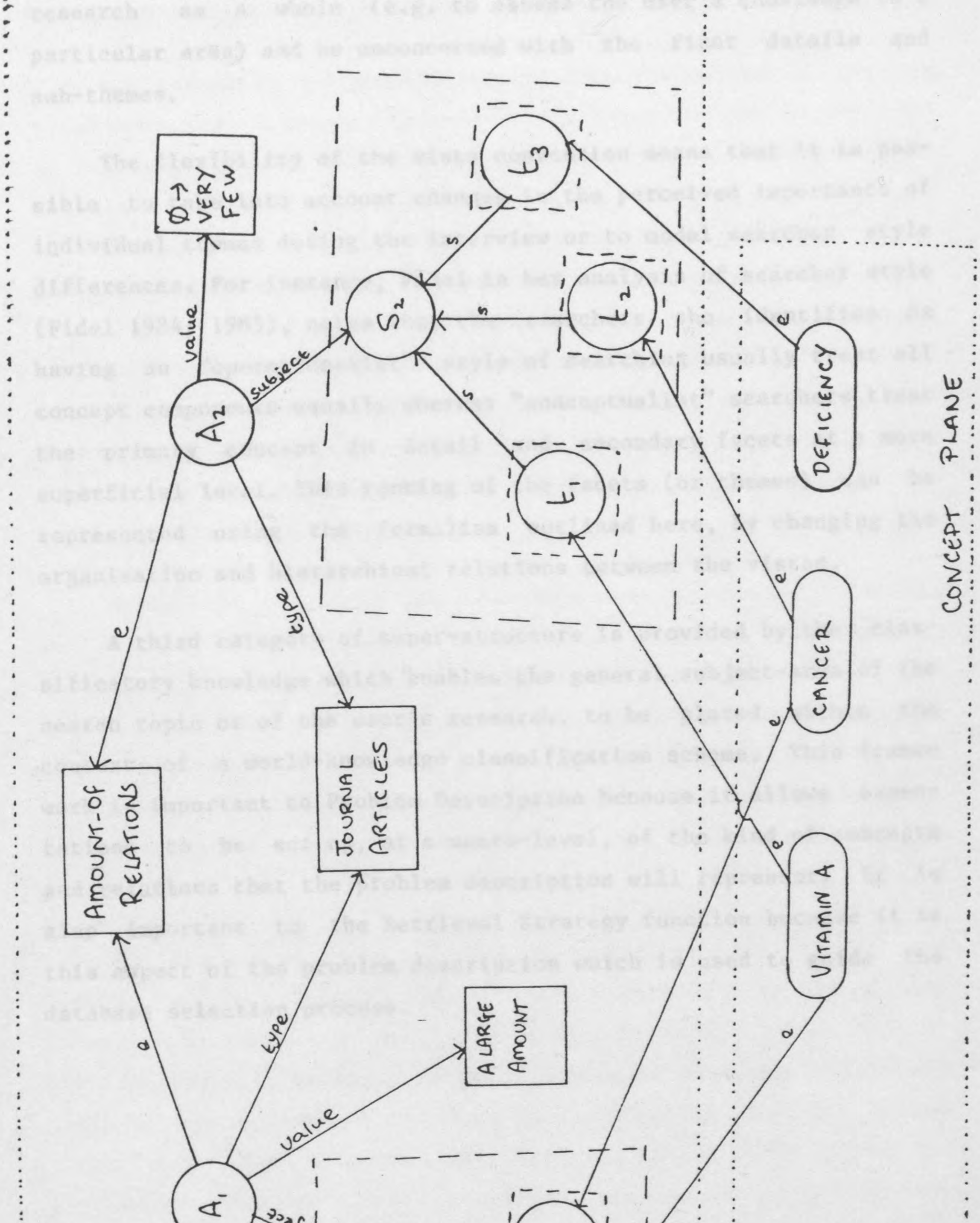


Figure 4.15 : problem description - S-LITERATURE plane (260684KSA)

research as a whole (e.g. to assess the user's knowledge in a particular area) and be unconcerned with the finer details and sub-themes.

The flexibility of the vista convention means that it is possible to take into account changes in the perceived importance of individual themes during the interview or to model searcher style differences. For instance, Fidel in her analysis of searcher style (Fidel 1984; 1985), notes that the searchers she identifies as having an "operationalist" style of searching usually treat all concept components equally whereas "conceptualist" searchers treat the primary concept in detail and secondary facets at a more superficial level. This ranking of the facets (or themes) can be represented using the formalism outlined here, by changing the organisation and hierarchical relations between the vistas.

A third category of super-structure is provided by the classificatory knowledge which enables the general subject-area of the search topic or of the user's research, to be placed within the context of a world-knowledge classification scheme. This framework is important to Problem Description because it allows expectations to be set up, at a macro-level, of the kind of concepts and relations that the problem description will represent. It is also important to the Retrieval Strategy function because it is this aspect of the problem description which is used to guide the database selection process.

3. USING PROBLEM DESCRIPTIONS

Any formalism for problem descriptions can be judged on its epistemological adequacy (how well it represents its "world") and on pragmatic grounds (how useful is it in enabling other functions to achieve their goals) (McCarthy & Hayes, 1969). Within the scope of this thesis it is not possible to assess the formalism fully and indeed this could only be done within the context of an operational interface. However, it is possible to examine the formalism in the light of what the analysis revealed about the Retrieval Strategy function and attempt to evaluate whether the problem description models constructed by the formalism could be used to provide the Retrieval Strategy with the information required for it to carry out its tasks.

Using the analysis of the interaction between Problem Description and Retrieval Strategy as a basis, the problem descriptions constructed for each interview were examined to see whether they provided the necessary information at the appropriate time. They were also examined to see whether the inferences that the analysis suggested had been made from the intermediary's problem description could be made from the constructed problem descriptions.

The ability to group together network nodes into spaces and relate the spaces within vistas, seems very useful as far as term selection is concerned. Most problem descriptions, by the end of the interview, contained a great number of concepts. Obviously there were more concepts relating to the RESEARCH plane but often there were also a large number involved in the TOPIC plane. The pattern of space and vista assignments enabled accurate predictions to be made about the concepts for which terms would be selected.

Concepts concerned with themes at the top of the vista hierarchy (the most local) were more likely to be selected than those at the root. If the theme comprises more than one concept, and these concepts are not synonymous, then often only the most specific concept will be used in the query. A concept which is expressed as a compound of two or more other concepts, (and represented by linking subconcepts to an intermediate node and then linking that node to the theme), may appear in the query as two or more terms linked with adjacency indicators.

Consider, for example, interview 040684HBA. Fig. 4.16 gives the TOPIC plane for the network in the final focus of that interview. The terms selected for searching were:

ACTIVATED CARBON
PRODUCTION
SUGAR
WATER

"Sugar" and "water" are the most specific concepts in their respective themes. Activated carbon appears as a compound concept. With the exception of "production", for which neither subthemes nor appropriate database coverage were found (see below), terms to express the broad themes such as "use" and "consumption" were not selected. Term selection in the case of interview 040684HBA is particularly interesting because, with the exception of an extremely brief discussion about activated carbon, no explicit negotiated term selection took place in the interview.

Database selection seems to be guided by the superstructure provided by the classificatory knowledge i.e. the subject-area of the topic and subject-themes. In interview 040684HBA, for example, database selection appears to have been carried out purely on that basis. Fig: 4.17 gives the TOPIC and SUBJ planes for the network by focus 7 (concerned with database selection). Databases covering market analysis, water and industrial chemistry are considered in turn. These are the subject-areas of the subject-themes "consumption of activated carbon" and "water treatment as a use for activated carbon" and for the supertheme "activated carbon" respectively. Later, databases covering food technology, the subject-area of the other main use of activated carbon, "sugar decolourising", are dealt with. The user indicated that the third use of activated carbon, "evaporative emission control", was not important so this aspect is largely ignored and does not appear in the online search formulation. Thus the problem description formalism would seem to permit all the necessary inferences to carry out the database selection task.

Interview 040684HBA is unusual in that the subject-areas of the subthemes were considered as well as the subject area for the subject overall. The intermediary seems to have decided that general concepts such as "use" and "consumption" would be difficult to express as terms in a search formulation and that selecting databases which covered the area of use, or of consumption, might be more appropriate. Searching for activated carbon on a database covering water, for example, would almost certainly restrict output to documents about the use of activated carbon in water treatment. This is another example of the way in which choice of strategy or tactics can affect the structure of the problem description. The depth to which the subject-area aspects of the description were developed depended on the initial selection of a tactic to express broad concepts through the selection of appropriate databases.

The organisation of the query may also be derived from the problem description representations. Concepts forming a particular themes are usually OR′d. Concepts expressing subthemes which are related to the same supertheme are also OR′d. Sets of terms expressing different themes are, on the whole, AND′ed. If there is a key concept at the bottom of the vista hierarchy, expressed with specific terms, and a number of other themes all standing at the same level in the vista hierarchy, above the key concept, then these themes will be OR′d and the product AND′ed to the main theme. For example in interview 040684HBA (see figure 4.16 for the problem description), the formulation used online was:

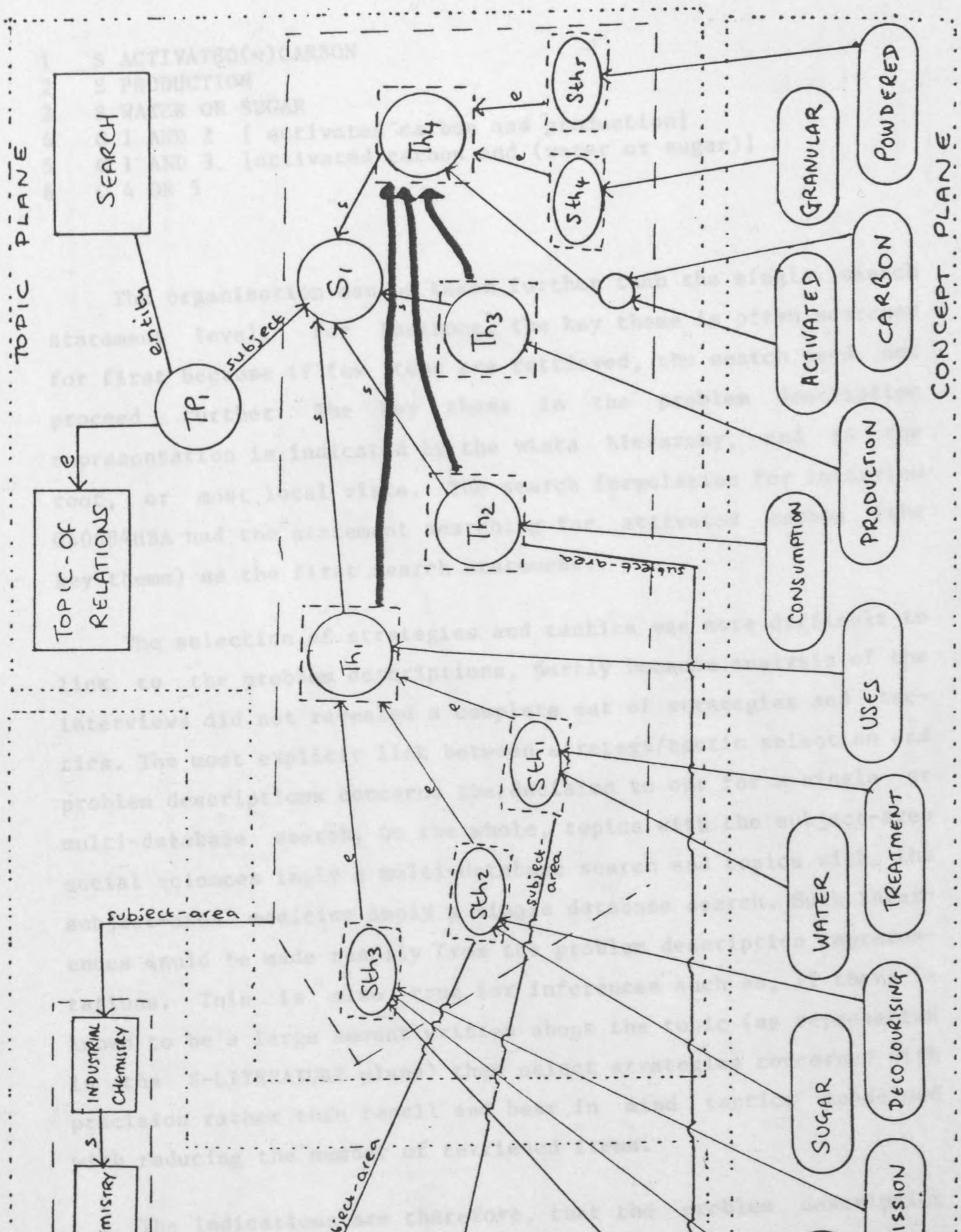


Figure 4.16 : problem description for Interview 040684HBA

1 S ACTIVATED(w)CARBON
2 S PRODUCTION
3 S WATER OR SUGAR
4 C 1 AND 2 [activated carbon and production]
5 C 1 AND 3 [activated carbon and (water or sugar)]
6 C 4 OR 5

The organisation can be taken further than the single search statement level. For instance, the key theme is often searched for first because if few items are retrieved, the search need not proceed further. The key theme in the problem description representation is indicated by the vista hierarchy, and is the root, or most local vista. The search formulation for interview 040684HBA had the statement searching for activated carbon (the key theme) as the first search statement.

The selection of strategies and tactics was more difficult to link to the problem descriptions, partly because analysis of the interviews did not revealed a complete set of strategies and tactics. The most explicit link between strategy/tactic selection and problem descriptions concerns the decision to opt for a single or multi-database search. On the whole, topics with the subject-area social sciences imply a multi-database search and topics with the subject-area medicine imply a single database search. Such inferences could be made readily from the problem description representations. This is also true for inferences such as, if there is known to be a large amount written about the topic (as represented in the S-LITERATURE plane) then select strategies concerned with precision rather than recall and bear in mind tactics concerned with reducing the number of retrieved items.

The indications are therefore, that the problem description formalism devised is expressive enough to provide the information required for the Retrieval Strategy function. The latter function seems to require a model of the user's problem which can be abstracted at various levels in different places, and will permit topic themes to be structured in relation to each other. The use

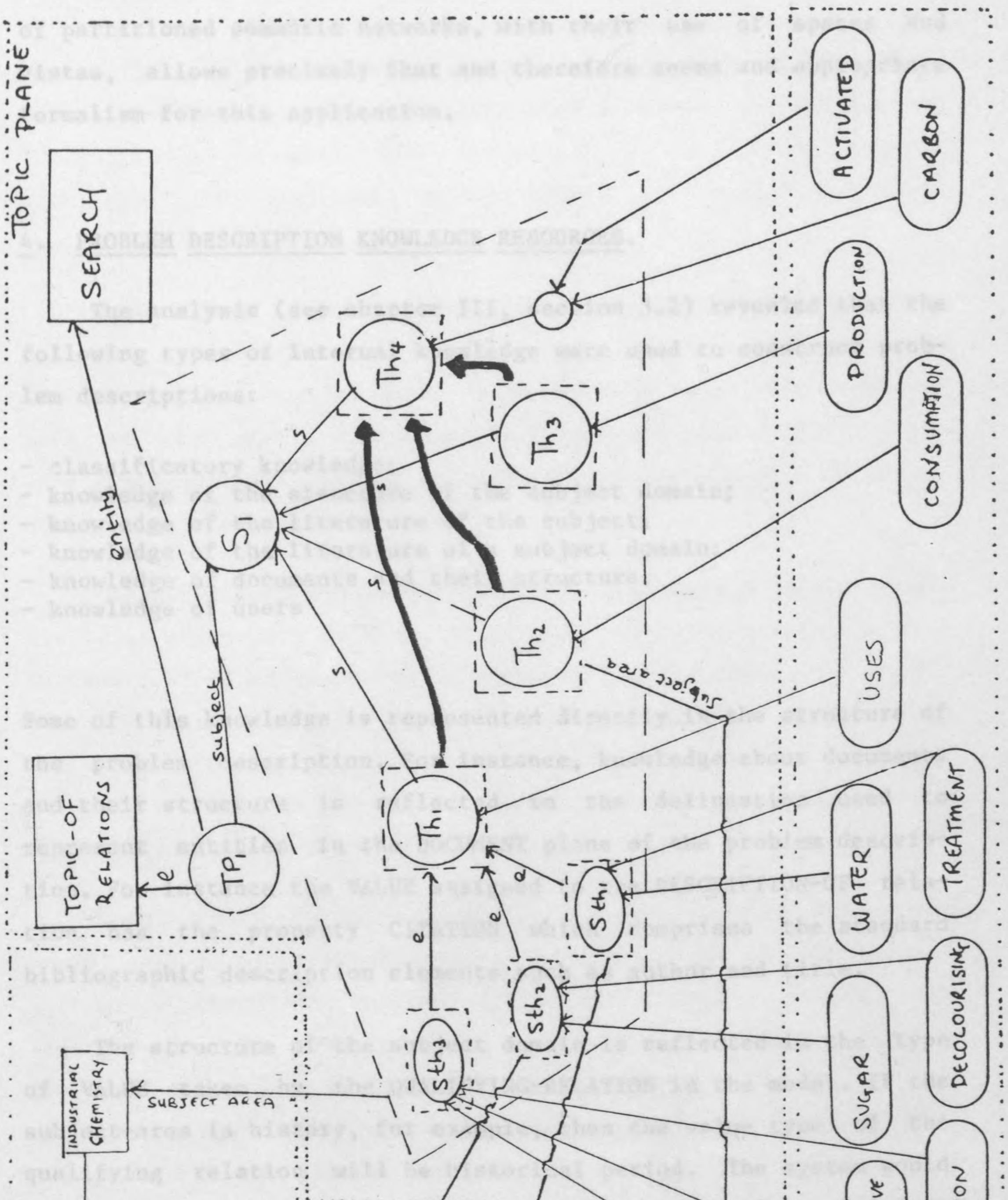


Figure 4.17 : problem description at focus 7, Interview 040684HBA

of partitioned semantic networks, with their use of spaces and vistas, allows precisely that and therefore seems and appropriate formalism for this application.

4. PROBLEM DESCRIPTION KNOWLEDGE RESOURCES.

The analysis (see chapter III, section 3.2) revealed that the following types of internal knowledge were used to construct problem descriptions:

- classificatory knowledge;
- knowledge of the structure of the subject domain;
- knowledge of the literature of the subject;
- knowledge of the literature of a subject domain;
- knowledge of documents and their structure;
- knowledge of users

Some of this knowledge is represented directly in the structure of the problem description. For instance, knowledge about documents and their structure is reflected in the delineation used to represent entities in the DOCUMENT plane of the problem description. For instance the VALUE assigned to the DESCRIPTION-OF relation has the property CITATION which comprises the standard bibliographic description elements such as author and title.

The structure of the subject domain is reflected in the type of VALUE taken by the QUALIFYING-RELATION in the model. If the subject-area is history, for example, then the value type of the qualifying relation will be historical period. The system would need to store, in addition to the basic structure of qualifying-relations, information about the values relevant to particular subject-areas. For example, not only must there be a qualifying the value "ANIMAL" but there must also be stored knowledge which indicates that such a qualifying value is applicable if the subject area is biology or medicine. This would require probably require no more than a simple "look-up" table.

Knowledge of the subject literature could be represented in a manner similar to the delineation used in the S-LITERATURE plane of the problem description. Information about the amount of literature of a particular document type, in a certain subject area could be compiled, on the basis of "experience" and represented as a simple network.

Knowledge of users refers to the way in which knowledge about the user constrains the kind of problem description that is developed. The knowledge would be of the form " if the subject-area of the user's department is X then the subject-area of the user's research will be likely to be X". A rule-based formalism might be most appropriate here.

Classificatory knowledge could be represented most effectively as a semantic network, displaying the relationships between various subject areas. Semantic networks are the formalism traditionally used to represent this type of knowledge and seem the most appropriate given the use to which classificatory knowledge is put to use in the problem description and given the problem description formalism outlined here.

Knowledge of the subject itself is a more complex. This obviously varies considerably between intermediaries and within subject areas. Therefore, whatever formalism is chosen, it must be flexible and be able to handle descriptive knowledge of varying levels of detail. A semantic network formalism would seem the most appropriate but there is no requirement that the formalism used should be the same as that used in the problem description. The system's internal subject knowledge should probably be represented at the most detailed and expressive level possible and abstracted for the purposes of constructing a problem description. Therefore a semantic network formalism along the lines of the conceptual nets proposed by Sowa, might be appropriate (Sowa, 1984). Recent research has resulted in the development of a Prolog-like resolution method by which large amounts of background knowledge can be represented as conceptual graphs, thereby enabling deductions to be performed on very large scale domains (Fargues et al., 1986). This work would seem to have potential for applications such as the one proposed here.

V. CONCLUSIONS

1. OVERVIEW

Two of the key functions of any information provision mechanism are the development of a description of the user's problem and the construction of a plan to help the user deal with that problem. This is true whether the information provision mechanism is the oracle at Delphi or an online search service. The aim of this research was to investigate these functions, within the context of an online search service. Here, users with problems are seeking documents within which they hope to find information to help them manage their problem. In some cases problem management will mean problem resolution but in others it may serve to define more clearly the problem, thereby indicating what further steps need to be taken to resolve it. Plan construction, in the online context, involves the formulation of an effective retrieval strategy for searching bibliographic databases.

The investigation of the Problem Description and Retrieval Strategy functions requires the specification of these functions, the identification and specification of the knowledge resources each function uses, and the identification and specification of the interactions between these two functions and between the other functions of the information provision mechanism. This investigation was carried out by analysing real human-human interactions i.e.the presearch interview between user and intermediary. The analysis enabled both functions to be specified, an outline of the categories and specific knowledge each function requires to be produced and the interactions between the functions to be described. The model building activity of the intermediary with respect to the problem description was analysed and using this analysis a formalism for representing problem descriptions was developed.

The ultimate objective is to use these results to design and implement the corresponding functions in an intelligent interface to document retrieval systems. The interface is based on a functional model composed of a number of independent but interacting experts, each carrying out a particular function but co-operating to achieve the goal of assisting the user with his/her problem. Problem Description and Retrieval Strategy are two of the functional experts in this interface and designing and developing these two interface functions depends on finding out how human intermediaries carry out the tasks of modelling the user's problem and constructing a retrieval strategy. This is what the functional discourse analysis of the presearch interviews achieved and on the basis of this analysis, it should be possible to design and implement the Problem Description and Retrieval Strategy functional experts.

The analysis of what human intermediaries do, and how they do it, because it is independent of any commitment to system design or knowledge representation, can be used in its own right. It can be used to enhance our understanding of the presearch information interaction between user and intermediary, and of the role and tasks of the intermediary in that interaction.

One area where there may be positive benefits arising from this type of analysis is in the training of intermediaries. Training of intermediaries has usually involved the teaching of search mechanics, database structure and coverage and indexing languages. While this knowledge plays a part in the construction of effective retrieval strategies, it is insufficient in terms of the overall process. The basis for the formulation of a retrieval strategy, for instance, is not the intermediary's knowledge of databases nor of search mechanics but rather the intermediary's model of the user's problem. Thus it is important that intermediaries be adept at carrying out this kind of cognitive modelling and are able to negotiate with users to elicit the information to do so. Initial observations suggest that analysed transcripts, used as case studies, present a useful approach to teaching students about presearch interviewing.

Interestingly, the analysis of the interviews also seems to provide an indication of where interviews "fail". The pattern of occurrence of the functions within the interview, both in terms of extent and in terms of which party is involved, provides one indication (see Brooks & Belkin, 1983; Belkin, 1985), but there is also some evidence that "pathological" patterns of interaction appear at the subfunction level e.g. in interview #5 where the task of assessing how well a term expressed the underlying concepts was by-and-large carried out by the user (see chapter III, section 3.3). Developing some objective indication of interaction failure would be useful since it would enable the interface to monitor for the occurrence of such patterns and take appropriate corrective measures. It could also be used to teach trainee intermediaries how to and how not to conduct a presearch interview.

2. KNOWLEDGE ELICITATION USING DISCOURSE ANALYSIS

> "The studies that appear to us most promising for the understanding of conversation share one feature: They examine meticulously at least some details of recorded verbal behaviour." (Labov & Fanshel, 1977, p19)

> "If we are to be accountable to the events of the therapeutic session, or even to 15 minutes of that session, we will be faced with an extraordinary amount of detail and the problem of making that information accessible and intelligible to the reader." (Labov & Fanshel, 1977, p27)

The method used to specify the functions, their knowledge resources and interactions, was functional discourse analysis of real human-human interactions between intermediary and user in an online search centre. This method was found to produce a wealth of detailed information about each interview both with respect to the tasks of the intermediary and the knowledge s/he employs and also with respect to the process of negotiation between user and intermediary. This latter aspect, the interaction between expert and client, is of considerable importance to the design of any interactive interface (e.g. see Kidd, 1985).

The advantage of the method used is that it allows the interaction to be investigated at the level of granularity which allows the negotiation between user and intermediary to be explored and elicits the expert knowledge and problem solving skills at a level of detail sufficient to consider implementation of the functions within an intelligent interface. It also enables the elicitation of knowledge about the way in which the participants in the information interaction use discourse to carry out their tasks and it is possible to take into account the co-operative nature of the interaction.

The disadvantage of using functional discourse analysis is that it is a lengthy and complex process. The whole procedure - recording the interview, transcription and analysis, took several months. For each individual interview, the intricate negotiation, goals, tasks and knowledge employed can be investigated in detail. However, comparison between interviews, or any quantitative analysis, is difficult to achieve because of the effort required to compile a sufficiently large corpus of analysed interviews.

3. THE PROBLEM DESCRIPTION AND RETRIEVAL STRATEGY FUNCTIONS

> "The major thrust of the work on MYCIN was fundamentally to the knowledge level in capturing the knowledge used by medical experts. The processing, an adaptation of well understood notions of backward chaining, played a much smaller role."
>
> (Newell, 1982)

The Problem Description and Retrieval Strategy functions were seen to be very different, not only in respect of their goals but also in the way these goals were achieved. Problem Description concerns modelling by the intermediary of the user's problem, which in turn involves eliciting from the user a description of his/her problem. The user is the expert as far as the problem and its subject domain is concerned while the intermediary is the expert with respect to knowing what it is necessary to understand about the user's problem in order to be able to construct an effective retrieval strategy. Both parties need to co-operate to enable the intermediary to develop a sufficiently detailed model of the user's problem, suitable for using to derive a retrieval strategy. That is, Problem Description is essentially a negotiated process.

In contrast, the formulation of a retrieval strategy is more intermediary dependent. Although intermediaries are at pains to involve the user as much as possible in developing a retrieval strategy, the expertise necessary to do this lies largely with the intermediary and not with the user. There is much less need to negotiate with the user and the intermediary simply makes the decisions s/he considers most appropriate. The user is normally involved only because aspects of the terms and database selection processes require his/her expert subject knowledge. Otherwise the intermediary uses the dialogue not, as in Problem Description, to elicit information from the user but donate information to the user i.e. justify the choices that have been made, explain the consequences of particular decisions and explain any features of the retrieval strategy which might give rise to questions from the user when employed online.

Both Problem Description and Retrieval Strategy are composite functions. They are made up of a number of subfunctions, each with a corresponding subgoal. The subfunctions of Problem Description are concerned with describing the various elements of the user's problem. In academic online search environments problem components include the topic of the search, the user's research, the subject background to the research and the topic, the literature of the subject and documents of interest to the user. These elements were apparent in all the interviews recorded although there was variation in the extent to which each occurred, with some elements more pronounced in some interviews that in others. Variation between interviews seems partly a consequence of problem type, partly related to the intermediary's knowledge of the subject domain and to the user's knowledge of the subject domain and partly related perhaps to the communicative abilities of both parties. In other information provision environments, with different types of users and user problems, it is probable that a somewhat different set of subfunctions might be found.

The subfunctions that make up the Retrieval Strategy function are concerned with different aspects of the plan construction process. Two are concerned with selection, of terms and of databases, and two are concerned with query formulation and the construction of an effective search strategy. The selection processes are to some extent negotiated which meant that it was possible, through analysis of the interviews, to draw up a list of subtasks involved in term selection and in database selection. The relatively non-negotiated nature of query formulation and search strategy organisation meant that it was not possible to do the same for these subfunctions although an outline of the processes involved was established. There were differences between the interviews with respect to the number of utterances devoted to each subfunction within Retrieval Strategy. This appeared to depend on the problem, and in particular on the subject area of the problem, on the intermediary's skill and expertise, how well the user's problem was understood, and on the user's ability to follow and to participate in the process.

The range of categories of knowledge employed by Problem Description and Retrieval Strategy functions appears from the analysis, substantial. Human intermediaries constructing problem descriptions were found to make use of external sources such as booking forms and the user him/herself and at least six categories of internal knowledge. In retrieval strategy formulation, the number of external sources is greater, some four categories were recorded, and at least seven categories of internal knowledge. In each case, the categories of internal knowledge may encompass non-trivial amounts of specific knowledge e.g. knowledge of online IR systems including basic theory and infrastructure, or knowledge of the subject domain itself. One of the advantages of using a knowledge elicitation method based on real interactions is that it was possible not merely to produce a listing of the knowledge used but also to study the context in which this knowledge was employed. For example, intermediaries may use their own subject knowledge to disambiguate information provided by the user, or the intermediaries' knowledge of the structure of subject areas enables intermediaries to prompt users for "missing" information. Knowledge by itself is not enough, it is also necessary to know how this knowledge can be used during the problem solving process.

The interaction between the other functions and the Problem Description and Retrieval Strategy functions, are relatively interview specific. In general, interactions with the Problem Description function seem the most varied and the most complex. Problem Description seems to require information from the User Model and Problem State functions and to be required by them. There is also a large amount of interaction within Problem Description, between the individual Problem Description subfunctions. This is also true of Retrieval Strategy, where term selection may be followed by query formulation and query formulation by strategy construction. Retrieval Strategy was found to interact strongly with Explanation. Sometimes the explanation would "buffer" a move from some other function to Retrieval Strategy. In other cases, Explanation followed Retrieval Strategy and represented an attempt by the intermediary to justify what s/he has decided or to explain the consequences of this decision.

Problem Description and Retrieval Strategy were found to interact very strongly. The problem description forms the basis for many Retrieval Strategy processes, particularly database selection (the SUBJ component) and term selection (the TOPIC component). The formulation of the query depends on the internal structuring of the TOPIC component but is affected by other features, e.g, the qualifying theme which carries less weight than the main themes. Choice of strategy and tactics is less obvious. There is some evidence for the influence of subject area e.g. in the decision to opt for a multi or single database search, but there seem to be other factors at work such as the nature of the desired output. Most importantly, the interaction between Problem Description and Retrieval Strategy does not seem to be a one-way process. It was shown that the selection of a preliminary strategy or set of tactics influences the way in which the problem description is developed and the extent to which the elements of the problem description are specified.

A formalism for the problem description based on partitioned semantic networks has been proposed. Using this model it is possible to predict some of the events in the interactions and to explain some of the observations about strategy formulation, particularly about the way in which database selection and term selection is carried out.

4. FURTHER RESEARCH

The research presented in this thesis has been limited both in the number of functions it has considered and in the type of information interaction investigated. Work on the User Model function is underway (see Daniels, 1985) but the remaining functions are largely unexplored. The overall dialogue structure must also be investigated before a complete interface design can be attempted.

The information provision mechanism investigated was an online search service in an academic institution. Keeping within the online context, it would be useful to carry out a similar investigation of interactions between user and intermediary in a public library online search service setting and in an industrial online search service. In the latter case, the intermediary is often closely associated with a research team and knows both users and the subject background of the research very well. This is in contrast to the academic search service where users often come for one-off searches and there may be a very wide range of subjects to be searched, so that it is unlikely that the intermediary has the opportunity to build up detailed knowledge of one small domain but rather must have a general understanding of one or more disciplines.

The present research has examined a number of issues qualitatively. Some aspects cannot be investigated in this way, e.g. comparative analysis of the interaction patterns between interviews, where the qualitative analysis needs to be supported by a quantitative approach. Using quantitative methods and a large corpus of analysed interviews, it might be possible, for instance, to relate the balance of Problem Description subfunctions or Retrieval Strategy subfunctions to problem type or subject area of the topic.

Further work need to be done to explore more fully the strategies and tactics used by intermediaries. Most of this is not negotiated and would need to be elicited by interviewing the expert or asking him/her to analysed transcripts of interviews (case studies). It is important to discover how, why and when, tactics or strategies are used as well as listing those employed.

Another area for future research would be a reappraisal of the whole issue of retrieval strategy formulation, from the negotiated strategy through to the strategy employed online and any reformultions. On the whole, retrieval strategies are largely seen as static lists of queries which can be manipulated by applying particular tactics or strategies. The whole strategy may be seen as encompassing both queries and tactics/strategies but it is not usually explicit how these two elements might be co-ordinated. It might be a more fruitful approach to look at a retrieval strategy as a search plan. The goal of the search plan is to retrieve a document set with particular characteristics and the tactics and queries provide the means for achieving this goal. De-emphasising the search statements and re-emphasising the planning and goal elements may prove more insightful and enable a coherent theory of retrieval strategy formulation to be developed.

5. IMPLEMENTING AN INTELLIGENT INTERFACE

The goal towards which this project is aimed, is the design and development of a distributed expert based intelligent interface for document retrieval systems. The analysis of real-live interactions indicated some factors which must be taken into account when designing such an interface. Firstly, although the MONSTRAT functional model for an intelligent interface is complex, the analysis would suggest that another layer of complexity needs to be added. At least two of the functional experts, Problem Description and Retrieval Strategy are not "unit" functions but instead seem to comprise a number of subfunctions which interact to achieve the functions' goals. There are serious implications here for system design. Current techniques for constructing intelligent knowledge-based systems are quite limited in the extent to which they can handle multi-functional, highly interactive, distributed problem solving (Hayes-Roth, Waterman & Lenat, 1983a; Stefik et al., 1983). This situation may change as developments in knowledge-based processes and machines are forthcoming e.g. parallel processing or the development of new problem-solving and inference systems (Gaines, 1984).

Secondly, a wide range and large amount of knowledge seems required by the two functions investigated. Scaled-up this would suggest that the amount of knowledge the whole interface would require is substantial and again, given current intelligent knowledge-based system/expert system techniques, this would present a non-trivial problem for any implementation (Lenat, Prakesh & Shepherd, 1986).

The interface design specifies each function as an independent expert. Therefore it should be possible to develop the functional experts as independent modules to be combined within some overall system architecture at a later stage. The Problem Description function could be constructed from a design based on the research presented here, even if it was possible to incorporate only a limited amount of knowledge at the early stages. Similarly a prototype Retrieval Strategy function might be developed to explore the interaction between Problem Description and Retrieval Strategy. To implement these four functions would involve:

(i) Developing software for carrying out the tasks each function must perform. Mechanisms must be developed which will simulate the problem structures observed in terms of goals and subgoals, and means for achieving these goals must be provided.

(ii) Implementing the knowledge resources required by each function. This would mean developing the means for representing the knowledge in the way most appropriate for that type of knowledge. Software for easy entry, verification and consistency checking of specific components of the knowledge resources must be built.

(iii) Software to enable the function to use and apply their knowledge resources must be constructed. This would involve implementing the inferencing mechanisms associated with each representation formalism.

(iv) Developing mechanisms for carrying out the model building activities of the Problem Description function and for the planning construction activities of Retrieval Strategy.

(v) Defining and developing the exact nature of the communications to and from each function and in particular from Retrieval Strategy to Problem Description.

6. CONCLUDING REMARKS

It was the intention of this research to investigate information interactions in sufficient detail to permit the eventual design and implementation of two of the functions of an intelligent interface for document retrieval systems. Such work has been, and is, hindered by our lack of understanding of the problem solving processes involved and of the tasks carried out and knowledge used, by the human intermediary. Presented in this thesis is a very detailed analysis of the tasks and knowledge resources involved in developing descriptions of the user's problem and in constructing an effective retrieval strategy. The analysis suggests that the functions, as carried out by a human intermediary, are extremely complex. Too complex perhaps to permit direct implementation with existing techniques. Simplifying assumptions will of necessity need to be made since computers cannot yet equal human thought processes. However, having a theoretical basis enables such simplifying assumptions to be made appropriately and on a rational basis. Moreover if the resulting system should prove ineffective, there is an underlying body of theory upon which to fall back and use to analyse the failure and correct faults. The pressure on researchers to produce working systems has a negative effect in application domains where there is little theoretical underpinning.

Finally, I would hope any attempt to develop an intelligent interface to document retrieval systems is not seen as an attempt to get rid of human intermediaries. Even when the interface is realised, human intermediaries will still be required to deal with complex problems and to counsel the users. It is unlikely that the interface would be able to tackle more than routine enquiries, simply because of the extent of the knowledge resources and expertise required. In fact one of the most important conclusions from this research should be that the task of the human intermediary is <u>not</u> a simple one. It requires a wide variety of tasks to be carried out including the construction of cognitive models, the ability to communicate effectively with the user, and a great deal of knowledge. Hopefully one of the outcomes of this research will be a greater appreciation of the skills, expertise and knowledge of human intermediaries.

VI. REFERENCES.

ADDIS, T. (1981). Definition of an expert system. AISB Quarterly, Issue 40-41, 1981, p.39.

AIKINS, J.S. (1984). A representation scheme using both frames and rules. In: Rule-based expert systems: the MYCIN experiments. Edited by B.G. Buchanan and E.H. Shortliffe. Reading, Mass., Addison-Wesley, 1984, pp.424-440.

ALTY, J.L. and COOMBS, M.J. (1980). Face-to-face guidance of university computer users - I: a study of advisory services. International Journal of Man-Machine Studies, vol.12, 1980, pp.390-406.

AUSTER, E. (1983). User satisfaction with the online negotiation interview: contemporary concern in traditional perspective. Reference Quarterly, vol.23, Fall 1983, pp.47-59.

BARR, A. and FEIGENBAUM, E.A. (eds). (1981). The handbook of Artificial Intelligence. Vols. 1-3. Los Altos, CA., William Kaufmann, 1981

BATES, M.J. (1979a). Information search tactics. Journal of the American Society for Information Science, vol.30, no.4, 1979, pp.205-214.

BATES, M.J. (1979b). Idea tactics. Journal of the American Society for Information Science, vol.30, no.5, 1979, pp.280-289.

BATES, M.J. (1981). Search techniques. In: Annual Review of Information Science and Technology, vol.18, 1981. Edited by M.E. Williams. White Plains, N.Y., Knowledge Industry Publications, 1981, pp.139-169.

BELKIN, N.J. (1977). A concept of information for information science. Ph.D Thesis. University of London, 1977.

BELKIN, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science, vol.5, 1980, pp.133-143.

BELKIN, N.J. (1984). Cognitive models and information transfer. Social Science Information Studies, vol.4, 1984, pp.111-129.

BELKIN, N.J. (1985). Modelling the user in information interaction. In: Proceedings of the 12th AIOPI Conference, Torquay, July 4-12th 1985 (in press).

BELKIN, N.J. and HAPESHI, K. Representation and matching of anomalous states of knowledge and document texts. Final report on BLR&D department grant no. SI/G/566. London, The City University, Department of Information Science (in preparation).

BELKIN, N.J., HENNINGS, R-D. and SEEGER, T. (1984). Simulation of a distributed expert-based information provision mechanism. Information Technology: Research Development Applications, vol.3, no.3, 1984, pp.122-141.

BELKIN, N.J. and KWASNIK, B.H. (1986). Using structural representations of anomalous states of knowledge for choosing document retrieval strategies. In: Proceedings of the joint ACM/BCS International Conference on R&D in IR. Pisa, 1986 (in press).

BELKIN, N.J., ODDY, R.N. and BROOKS, H.M. (1982a). ASK for information retrieval, part I. Journal of Documentation, vol.38, 1982, pp.61-71.

BELKIN, N.J., ODDY, R.N. and BROOKS, H.M. (1982b). ASK for information retrieval, part II. Journal of Documentation, vol.38, 1982, pp.145-164.

BELKIN, N.J., SEEGER, T. and WERSIG, G. (1983). Distributed expert problem treatment as a model for information system analysis and design. Journal of Information Science, vol.5, 1983, pp.153-167.

BELKIN, N.J. and VICKERY, A. (1985). Interaction in information systems: a review of research from document retrieval to knowledge-based systems. LIR report no.35. London, The British Libraru, 1986.

BELKIN, N.J. and WINDEL, G. (1984). Using MONSTRAT for the analysis of information interaction. In: Representation and Exchange of Knowledge as a Basis of Information Processes. Edited by H.J. Dietshmann. Amsterdam, Elsevier, 1984, pp.359-382.

BERRY, D.C. and BROADBENT, D.E. (1984). On the relationship between task performance and associated verbal knowledge. Quarterly Journal of Experimental Psychology, vol 36A, no. 2, 1984, pp.209-291.

BOBROW, D.G. (1975). Dimensions of representation. In: Representation and Understanding: studies in cognitive science. Edited by D.G. Bobrow and A. Collins. New York, Academic Press, 1975, pp.1-34.

BOGURAEV, B.K. and SPARCK JONES, K. (1982). A natural language analyser for database access. Information Technology: Research and Development, vol.1, no.1, Jan 1982, pp.23-40.

BORGMAN, C.L., CASE, D. and MEADOW, C.T. (1985). Incorporating user's information seeking styles into the design of an information retrieval interface. In: *Proceedings of the 48th Annual Meeting of the American Society for Information Science*, vol.22. White Plains, N.Y., Knowledge Industry Publications Inc., 1985, pp.324-330.

BRACHMAN, R.J. (1979). On the epistemological status of semantic networks. In: *Associative Networks*. Edited by N.V. Findler. New York, Academic Press, 1979, pp.3-49.

BRACHMAN, R.J. et al. (1983). What are expert systems? In: *Building Expert Systems*. Edited by F. Hayes-Roth, D.A. Watermann and D.B. Lenat. Reading, Mass., Addison-Wesley, 1983, pp.31-57.

BROOKS, H.M. and BELKIN, N.J. (1983). Using discourse analysis for the design of information retrieval mechanisms. In: *Research and Development in Information Retrieval. Proceedings* of the Sixth Annual International ACM SIGIR conference, Washington D.C., 1983. New York, ACM, 1983, pp.31-47.

BROOKS, H.M., DANIELS, P.J. and BELKIN, N.J. (1985). Problem descriptions and user models: developing an intelligent interface for document retrieval systems. In: *Informatics 8: Advances in Intelligent Retrieval*. London, ASLIB, 1986, pp. 191-214.

BUCHANAN, B.G. (1982). New research on expert systems. In: *Machine Intelligence 10*. Edited by J.E. Hayes, D. Michie and Y.H. Pao. Edinburgh, Edinburgh University Press, 1982, pp.269-299.

BUCHANAN, B.G. et al. (1983). Constructing an expert system. In: *Building Expert Systems*. Edited by F. Hayes-Roth, D.A. Waterman and D.B. Lenat. Reading, Mass. Addison-Wesley, 1983, pp.127-168.

BUCHANAN, B.G. and FEIGENBAUM, E.A. (1978). DENDRAL and METADENDRAL: their applications dimension. *Artificial Intelligence*, vol.11, 1978, pp.5-24.

BUCHANAN, B.G. and SHORTLIFFE, E.H. *Rule-based expert systems. The MYCIN experiments*. Reading, Mass., Addison-Wesley, 1984.

CLANCEY, W.J. (1981). *Methodology for building an intelligent tutoring system*. Report no. STAN-CS-81-894 (HPP-81-18). Department of Computer Science, Stanford University, CA., October 1981.

CLANCEY, W.J. (1985). Heuristic classification. *Artificial Intelligence*, vol.27, 1985, pp.289-350.

CLARKE, A. and CRONIN, B. (1983). Expert systems and library/information work. *Journal of Librarianship*, vol.15, no.4, Oct 1983, pp.277-292.

COCHRANE, P. (1981). *Presearch interview project: executive summary, project bibliography and sample data*. Report no. ED 205 190, Syracuse N.Y., School of Information Studies, Syracuse University, 1981.

COOMBS. M.J. and ALTY, J.L. (1980). Face-to-face guidance of university computer users - II: characterizing advisory interactions. International Journal of Man-Machine Studies, Vol.12, 1980, pp. 407-429.

COOMBS, M.J. and ALTY, J.L. (1984). Expert systems: an alternative paradigm. International Journal of Man-Machine Studies, vol.20, 1984, pp.21-43.

CROFT, W.B. (1985). An expert assistant for a document retrieval system. In: RIAO-85. Actes of the conference: Recherche d'Informations Assistee par Ordinateur, Grenoble, March 1985. Grenoble, I.M.A.G., 1985, pp.131-149.

CROFT, W.B. and THOMPSON, R.H. (1985). An expert system assistant for document retrieval systems. COINS Technical Report 85-05. Computer and Information Science Department, University of Massachusetts, 1985.

CRYSTAL, M.I. and JAKOBSON, G.E. (1982). FRED, a front-end for databases. Online, Sept.1982, pp.27-30.

DAMERAU, F.J. (1980). The transformational question answering (TQA) system: description, operating experience and implications. In: Proceedings of the 2cnd International Conference on Databases in the Humanities and Social Sciences. Report RC 8287, IBM, Thomas J. Watson Research Centre, Yorktown Heights, New York. 1980.

DANIELS, P.J. (1985). The user modelling function of an intelligent interface for document retrieval systems. In: IRFIS 6: Intelligent Information systems for the information society. Frascati, Italy, Sept. 1985 (in press)

DANIELS, P.J., BROOKS, H.M. and BELKIN, N.J. (1985). Using problem structures for driving human-computer dialogues. In: RIAO 85. Actes of the conference: Recherche d'Informations Assistee par Ordinateur, Grenoble, March 1985. Grenoble, I.M.A.G., 1985, pp.645-660.

DOSZKOCS, T.E. (1983). Automatic vocabulary mapping in online searching. International Classification, vol.10, no.2, 1983, pp.78-83.

DUDA, R. et al. (1978). Semantic network representations in rule-based inference systems. In: Pattern Directed Inference Systems. Edited by D. Waterman and F. Hayes-Roth. New York, Academic Press, 1978, pp.203-221.

DUDA, R.O., GASCHNIG, J.G. and HART, P.E. (1979). Model design in the PROSPECTOR consultant system for mineral exploration. In: Expert systems in the micor-electronic age. Edited by D. Michie. Edinburgh, Edinburgh University Press, 1979, pp. 153-167.

ERICSSON, K.A. and SIMON, H.A. (1980). Verbal reports as data. Psychological Review, vol.87, no.3, 1980, pp.215-251.

ERMAN, L.D. and LESSER, V.R. (1979). The Hearsay-II system: a tutorial. In: Trends in Speech Recognition. Edited by W.A. Lea. U.S.A., Prenitice-Hall, 1979, pp.361-381.

EVANS, N. and PISCIOTTA, H. (1985). Search Helper: testing acceptance of a gateway software system. In: Proceedings of the 5th National Online Meeting. New York, April 1985. Medford, N.J., Learned Information Inc. 1985, pp.131-136.

FAIRHALL, D. (1985). In search of searching skills. Journal of Information Science, vol.10, 1985, pp.111-123.

FARGUES, J. et al. (1986). Conceptual graphs for semantics and knowledge processing. IBM Journal of Research and Development, vol.30, no.1, Jan.1986, pp.70-79.

FEIGENBAUM, E.A. (1979). Themes and case studies of knowledge engineering. In: Expert Systems in the Microelectronic Age. Edited by D. Michie. Edinburgh, Edinburgh University Press, 1979, pp. 3-25.

FEIGENBAUM, E.G., BUCHANAN, B.G. and LEDERBERG, J. (1971). On generality and problem solving. A case-study using the Dendral program. In: Machine Intelligence 6 Edited by B. Meltzer and D. Michie. Edinburgh, Edinburgh University Press, pp. 165-190.

FIDEL, R. (1984). Online searching styles: a case-study based model of searching behavior. Journal of the American Society for Information Science, vol.35, no.4, 1984, pp.211-221

FIDEL, R. (1985). Moves in online searching. Online Review, vol.9, no.1, 1985, pp.61-74.

FIDEL, R. (1986). Towards expert systems for the selection of search keys. Journal of the American Society for Information Science, vol.37, no.1, 1986, pp.37-44.

FIDEL, R. and SOERGEL, D. (1983). Factors affecting online bibliographic retrieval: a conceptual framework for researchi. Journal of the American Society for Information Science, vol.34, no.3, 1983, pp.163-180.

FILLMORE, C.J. (1968). The case for case. In: Universals in Linguistic Theory. Edited bu E. Bach and R. Harms. New York, Holt, Rinhart & Winston, 1968, pp.1-88.

GAMMACK, J.G. and YOUNG, R.M. (1985). Psychological techniques for eliciting expert system knowledge. In: Research and Development in Expert Systems. Proceedings of the 4th Technical Conference of the B.C.S. Specialist Group on Expert Systems, University of Warwick, 18-20th December 1984. Edited by M.A. Bramer. Cambridge, Cambridge

University Press, 1985, pp. 105-112.

GRICE, H.P. (1975). Logic and conversation. In: Syntax and Semantics. Vol.3 Speech Acts. Edited by P.Cole and J.L. Morgan. New York, Academic Press, 1975, pp.41-58.

GRISHMAN, R. (1984). Natural language processing. Journal of the American Society for Information Science, vol.35, no.5, 1984, pp.291-296.

GROSZ, B.J. (1978). Discourse knowledge. In: Understanding Spoken Language. Edited by D.E.Walker. New York, Elsevier-North Holland, 1978, pp.229-346.

GROSZ, B.J. (1981). Focusing and description in natural language dialogues. In: Elements of discourse understanding. Edited by A. Joshi, B. Weber and I.Sag. Cambridge, CUP, 1981 pp.84-105.

GROVER, M.D. (1983). A pragmatic knowledge acquisition methodology. In: Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany, Aug.8-12th 1983. Los Altos, William Kaufmann, 1983, pp.436-438.

GUIDA, G. and TASSO, C. (1983). IR-NLI: an expert natural language interface to online databases. In: Proceedings of the Conference on Applied Natural Language Processing, Santa Monica, California, February 1-3, 1983. U.S.A, ACL, 1983, pp.31-38.

HARRIS, L. (1979). Experience with ROBOT in 12 commercial naturla language database applications. In: Proceedings of the 6th International Joint Conference on Artificial Intelligence. Tokyo, 1979, pp.365-368.

HARRIS, L.R. (1984). Experience with INTELLECT. AI Magazine, vol.5, no.2, 1984, pp.43-50.

HARTER, S.P. and PETERS, A.R. (1985). Heuristics for online information retrieval: a typology and preliminary listing. Online Review, vol.9, no.5, 1985, pp.407-424.

HAWKINS, D.T. and WAGERS, R. (1982). Online bibliographic search strategy development. Online, vol.6, no.3, May 1982, pp.12-19.

HAYES-ROTH, F., WATERMAN, D.A. and LENAT, D.B. (Eds). (1983a). Building Expert Systems. Reading, Mass., Addison-Wesley, 1983.

HAYES-ROTH, F., WATERMAN, D.A. and LENAT, D.B. (1983b). An overview of expert systems. In: Building Expert Systems. Edited by F. Hayes-Roth, D.A. Watermann and D.B. Lenat. Reading, Mass., Addison-Wesley, 1983, pp.3-30.

HENDRIX, G.G. (1978a). Semantic aspects of translation. In: Understanding Spoken Language. Edited by D.E. Walker. New York, Elsevier, 1978, pp.193-226.

HENDRIX, G.G. (1978b). The representation os semantic knowledge. In: Understanding Spoken Language. Edited by D.E. Walker. New York, Elsevier, 1978, pp.121-181.

HENDRIX, G.G. (1979). Encoding knowledge in partitioned networks. In: Associative Networks. Edited by N.V. Findler. New York, Academic Press, 1979, pp.51-92.

HITCHINGHAM, E.E. (1979). Online interviews: Charting user and searcher interaction patterns. In: Proceedings of the 42cnd Annual ASIS Meeting, vol.16. White Plains, N.Y., Knowledge Industry Publications, 1979, pp.66-74.

HOSKINSON CAMP, J. (1985). Development at Dialog. Online Review, vol.9, no.5, 1985, pp.377-381.

HURT, C.D. (1983). Intermediaries, self-searching and satisfaction. In: Proceedings of the 4th National Online Meeting, New York, 1983. Medford, N.J., Learned Information Inc., 1983, pp.231-237.

INGWERSEN, P. (1982). Search procedures in the library analysed from the cognitive point of view. Journal of Documentation, vol.38., no.3, Sept.1982, pp.165-191.

JANKE, R. (1984). Online after six: End user searching comes of age. Online, vol.8, no.6, 1984, pp.15-29.

JOHNSON, T. (1984). The commercial application of expert systems technology. Knowledge Engineering Review, vol.1, no.1, 1984, pp.15-25.

KAPLAN, S.G. (1982). Special section on natural language. ACM SIGART Newsletter, Jan. 1982, pp.27-109.

KEHOE, C.A. (1985). Interfaces and expert systems for online retrieval. Online Review, Vol.9, no.6, 1985, pp.489-505

KELLY, G.A. (1955). The psychology of personal constructs. New York, Norton, 1955.

KIDD, A.L. (1985). The consultative role of an expert system. In: People and Computers: Designing the Interface. Proceedings of the B.C.S. HCI Specialist Group Conference, University of East Anglia, 17-20 Sept. 1985. Edited by P. Johnson and S.Cook. Cambridge, Cambridge University Press, 1985, pp.248-254.

KIDD, A.L. and COOPER, M.P. Man-machine interface issues in the construction and use of an expert system. International Journal of Man-Machine Studies. Vol.22, 1985, pp.105-112.

KIRBY, M. and MILLER, N. (1985). Medline searches on BRS colleague: success of untrained users in a medical school and hospital. In: Proceedings of the 6th National Online Meeting, New York,

1985. Edited by M.E. Williams and T.H. Hogan. Medford, N.J., Learned Information Inc., 1985, pp.255-263.

KUIPERS, B.J. (1975). A frame for frames. In: Representation and Understanding: studies in cognitive science. Edited by D.G. Bobrow and A. Collins. New York, Academic Press, 1975, pp.151-184.

KUIPERS, B. and KASSIRER, J.P. (1983). How to discover a knowledge representation for causal reasoning by studying an expert physician. In: Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, 18-21 Aug. 1983, pp.49-56.

LABOV, W. and FANSHEL, D. (1977). Therapeutic Discourse. New York, Academic Press, 1977.

LAMB, M.R., AUSTER, E.W. and WESTEL, E.R. (1985). A friendly front-end for bibliographic retrieval: the implementation of a flexible front-end. In: Proceedings of the 48th ASIS annual meeting, vol.22. Las Vegas, Oct.20-24th, 1985. Edited by C.A. Parkhurst. New York, Knowledge Industry Publications, 1985.

LEIPZIG, N, KOZAK, M.G. and SCHWARTZ, R.A. (1983). Experiences with end user searching at a pharmaceutical company In: Proceedings of the 4th National Online Meeting. Medford, N.J., Learned Information Inc., 1983, pp.325-332.

LENAT, D., PRAKASH, M. and SHEPHERD, M. (1986). CYC: using commonsense knowledge to overcome brittleness and knowledge acquisition bottlenecks. AI Magazine, vol.6, no.4, 1986, pp.65-85.

LEVY, L.R. (1984). Gateway software: Is it for you? Online, vol.8, Nov.1984, pp.67-79.

McCARTHY, J. and HAYES, P.J. (1969). Some philosophical problems from the standpoint of Artificial Intelligence. In: Machine Intelligence 4. Edited by B. Melzer and D. Michie. Edinburgh, Edinburgh University Press, 1969, pp.463-502.

MARCUS, R.S. (1981). An automated expert system for information retrieval. In: Proceedings of the 44th Annual ASIS meeting, vol.18. White Plains, N.Y., Knowledge Industry Publications, 1981, pp.270-273.

MARCUS, R.S. (1983). An experimental comparison of the effectiveness of computers and humans as search intermediaries. Journal of the American Society for Information Science, vol.34, no.6, 1983, pp.381-404.

MARCUS, R.S. (1985). Development and testing of expert systems for retrieval assistance. In: Proceedings of the 48th Annual ASIS meeting, vol.22. White Plains, N.Y., Knowledge Industry Publications, 1985, pp.289-292.

MARCUS, R.S. and REINTJES, J.F. (1981a). A translating computer interface for end-user operation of heterogeneous retrieval systems. I. Design. Journal of the American Society for Information Science, vol.32, no.4, 1981, pp.287-303.

MARCUS, R.S. and REINTJES, J.F. (1981b). A translating computer interface for end-user operation of heterogeneous retrieval systems. II. Evaluations. Journal of the American Society for Information Science, vol.32, no.4, 1981, pp.304-317.

MEADOW, C.T. and COCHRANE, P. (1981). Basics of online searching. New York, John Wiley, 1981.

MEADOW, C.T., HEWETT, T.T. and AVERSA, E.S. (1982a). A computer intermediary for interactive database searching. I. Design. Journal of the American Society for Information Science, vol.33, no.5, 1982, pp.325-332.

MEADOW, C.T., HEWETT, T.T. and AVERSA, E.S. (1982b). A computer intermediary for interactive database searching. II. Evaluation. Journal of the American Society for Information Science, vol.33, no.6, 1982, pp.357-364.

MICHALSKI, R.S., CARBONELL, J. and MITCHELL, T. (eds). (1983). Machine learning: and artificial intelligence approach. Palo Alto, Tioga, 1983.

MINSKY, M. (1975). A framework for representing knowledge. In: The Psychology of Computer Vision. Edited by P. Winston. New York, McGraw-Hill, 1975.

MULLER, R. (1985). A "standard" for expert system terms. BCS Specialist Group on Expert Systems Newsletter, No.13, June 1985, p.7.

NEWELL, A. (1982). The knowledge level. Artificial Intelligence, vol.18, 1982, pp.87-127.

OBERMEIER, K.K. and COOPER, L.E. (1984). Information network facility organizing system (INFOS): an expert system for information retrieval. In: Proceedings of the 47th Annual ASIS meeting, vol.21, 1984. White Plains, N.Y., Knowledge Industry Publications, 1984, pp.95-98.

ODDY, R.N. (1977a). Information retrieval through man-machine dialogue. Journal of Documentation, vol.33, 1977, pp.1-14.

ODDY, R.N. (1977b). Retrieving references by dialogue rather than by query formulation. Journal of Informatics, vol.1, 1977, pp.37-53.

OJALA, M. (1985). End user searching and its implications for librarians. Special Libraries, vol.76, no.2, 1985, pp.93-99.

PEART, P.A. (1985). Online retrieval: Intermediaries vs end-users. In: Proceedings of the 6th National Online Meeting. New York, 1985. Edited by M.E. Williams and T.H. Hogan. Medford, N.J., Learned Information Inc., 1985 pp.357-361.

PEJTERSEN, A.M. and AUSTIN, J. (1983). Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria: part I Journal of Documentation, vol.39, 1983, pp.230-246.

PLATH, W. J. (1976). REQUEST: a natural language question answering system. IBM Journal of research and development, vol.20, no.4, 1976, pp.326-335.

POLLITT, A.S. (1981). An expert system as an ouline search intermediary. In: Proceedings of the 5th International Online Information Meeting. London, 1981, Learned Information Inc., pp.25-32.

POLLITT, A.S. (1984). A "front-end" system: an expert system as an online search intermediary. ASLIB Proceedings, vol.36, no.5, 1984, pp.229-234.

POLLITT, A.S. (1985). The CANSEARCH approach to end user information retrieval. In: Proceedings of IRFIS 6: Intelligent Information Systems for the Information Society, Frascati, September, 1985. Amsterdam, North-Holland (in press).

PRICE, L.E.T. (1983). Functional and satisfaction analyses of information interaction dialogues. MSc Thesis. Department of Information Science, The City University, London, 1983.

QUILLIAN, M.R. (1968). Semantic memory. In: Semantic Information Processing. Edited M. Minsky. Cambridge, Mass., M.I.T. Press, 1968, pp.227-270.

QUINLAN, J.R. (1979). Discovering rules by induction from large collections of examples. In: Expert systems in the microelectronic age. Edited by D. Michie. Edinburgh, Edinburgh University Press, 1979, pp.33-46.

RICH, E. (1979). User modeling via sterotypes. Cognitive Science, vol.3, 1979, pp.329-354.

RICH, E. (1983). Users are individuals. International Journal of Man-Machine Studies, vol.18, no.3, 1983, pp.199-214.

RICH, E. (1984). Natural language interfaces. Computer, vol.17, Sept. 1984, pp.39-47

ROBERTSON, S.E. et al. (1986). Weighting ranking and relevance feedback in a front-end system. Journal of Information Science, (in press)

SACKS, H., SCHEGLOFF, E. and JEFFERSON, G. (1974). A simplest systematics for the organization of turn-taking in conversation. Language, vol.50, 1974, pp.731-734.

SHAW, M.L.G. (1984). Knowledge engineering for expert systems. In: Interact '84. Proceedings of the IFIP Task Group on Human-Computer Interaction. Imperial College, London, 4-7 Sept. 1984, pp.328-332.

SHOVAL, P. (1985). Principles, procedures and rules in an expert system for information retrieval. Information Processing and Management, vol.21, no.6, 1985, pp.475-487.

SINCLAIR, J.McH. and COULTHARD, R.M. (1975). Towards an analysis of discourse. The English used by teachers and pupils. Oxford, OUP, 1975.

SLOMAN, A. (1985). Why we need many knowledge representation formalisms. In: Research and Development in Expert Systems. Proceedings of the 4th Technical Conference of the BCS Expert System Specialist Group. Edited by M.A. Bramer. University of Warwick, Dec.18-20 1984. Cambridge, CUP, 1985, pp.154-163.

SMITH, L.C. (1979). Selected artificial intelligence techniques in information retrieval systems research. Ph.D Thesis. School of Information Studies, Syracuse University, N.Y., 1979.

SMITH, P.J. and CHIGNELL, M. (1984). Development of an expert system to aid in searches of Chemical Abstracts. In: Proceedings of the 47th Annual ASIS meeting, vol.21, 1984. White Plains, N.Y., Knowledge Industry Publications, 1984, pp.99-102.

SOWA, J.F. (1984). Conceptual Structures: Information processing in mind and machine. Reading, Mass., Addison Wesley, 1984.

SPARCK JONES, K. (1982). Natural language access to databases: some questions and a specific approach. Journal of Information Science, vol.4, 1982, pp.41-48.

SPARCK JONES, K. (1983a). Report on a visit to the U.S., 27.1.83 - 4.2.83 Report to BLR&D on grant SI/V/79. March 1983.

SPARCK-JONES, K. (1983b). Intelligent Retrieval. In: Informatics 7 : Intelligent Information Retrieval. Edited by K.P. Jones. London, Aslib, 1983, pp.136-142.

SPARCK JONES, K. (1984). User models and expert systems. Technical report no.61. Computer Laboratory, University of Cambridge, Cambridge, 1984.

STEFIK, M. et al. (1983). The organization of expert systems: a tutorial. Artificial Intelligence, vol.18, 1983, pp.135-173.

STINTON, C. (1984). Intermediary knowledge resources in information interaction. M.Sc.Thesis, Department of Information Science, The City University, London, 1984.

STOUT, C. and MARCINKO, T. (1983). Sci-Mate: a menu-driven universal online searcher and personal data manager. *Online*, vol.7, no.5, Sept 1983, pp.112-116.

TAIT, J.L. (1983). Automatic request parsing and varient generation. In: *Information 7: intelligent information retrieval.* Edited by K.P. Jones. London, ASLIB, 1983, pp.53-63.

TATALIAS, J. (1985). Attitudes and expectations of potential end user online searchers. In: *Proceedings of the 6th National Online Meeting*, New York, 1985. Edited by M.E. Williams and T.H. Hogan. Medford, N.J., Learned Information Inc., 1985, pp.457-462.

TAYLOR, R. S. (1968). Question negotiation and information seeking in libraries. *College and Research Libraries*, vol.29, no.3, May 1968, p.178-193.

TOLIVER, D. (1981). A program for machine mediated searching. *Information Processing and Management*, vol.17, 1981, pp.61-68.

VICKERY, A. and BROOKS, H.M. (1983). *An expert system for referral*. Project proposal to the British Library Research and Development Department. July, 1983.

VICKERY, A. et al. (1984). *Annual Report of the activities of the Central Information Service of the University of London, Aug.1983-July 1984*. London, CIS, University of London, Aug.1984.

VICKERY, A. et al. (1986). *An expert system for referral*. Final report to BLR&D on project SI/G/625. February 1986. London, Central Information Service, University of London, 1986.

WALKER, G. and JANES, J.W. (1984). Expert systems as intermediaries. In: Challenges to an Information Society. *Proceedings of the 47th Annual ASIS Meeting*, vol 21. White Plains, N.Y., Knowledge Industry Publications, pp.103-105.

WALTZ, D.L. (1982). The state of the art in natural language understanding. In: *Strategies for Natural Language Processing.* Edited by W. G. Lehnet and M. H. Ringle. Hillsdale N.J., Lawrence Erlbaum Associates, 1983, pp.3-34.

WELBANK, M. (1983). A review of knowledge acquisition techniques for expert systems. *British Telecom Research Laboratories Report.* Martlesham Heath, Martlesham Consultancy Services, December 1983.

WHITE, M.D. (1985). Evaluation of the reference interview. *Research Quarterly*, vol.25, Fall 1985, pp.76-84.

WILLIAMS, M.E., LANNOM, L. and ROBINS, C. (eds). (1984). *Computer readable databases: a dictionary and data source book*. Chicago, American Library Association, 1984.

WILENSKY, R., ARENS, Y. and CHIN,D. (1984). Talking to UNIX in English: an overview of UC. Communications of the ACM, vol.27, no.6, June 1984, pp. 574-593.

WINOGRAD, T. (1975). Frame representations. In: Representation and Understanding. Edited by D.G. Bobrow and A. Collins. New York, Academic Press, 1975, pp.185-210.

WOODS, W.A. (1975). What's in a link: foundations for semantic networks. In: Representation and Understanding: studies in cognitive science. Edited bu D.G. Bobrow and A. Collins. New York, Academic Press, 1975, pp.35-82.

WOODS, W.A., KAPLAN, R.M. and NASH-WEBBER, B. (1972). The lunar sciences natural language information system: final report. Report 2878, BBN, 1972.

YAGHMAI, N.S. and MAXIN, J.A. (1984). Expert systems: a tutorial. Journal of the American Society for Information Science, vol.35, no.5, 1984, pp.297-305.

YIP, M.K. (1981). An expert system for document retrieval. M.S. Thesis. M.I.T., Cambridge, Mass. February, 1981.

VII. APPENDICES.

APPENDIX 1:
Transcript of Interview 190684HBA

TRANSCRIPT FROM INTERVIEW 190684HBA

I right (,) would you like to (,) turn your chair round
U

I (laugh .) we- we'll begin (...) just get some paper (....)/[1]
U (inaud.) yeah sure /[2]

I now I gather you're (cough.) excuse me
U (cough ....) (.......)

I you're a visitor (.) um /[3] yes are you part of the
U yes I am /[4]

I university or /[5] ya (,)
U well I teach at a Canadian university /[6]

I um I I just (,) we ask you this because i- its awful to
U

I bring up charges straight away (laugh .) but just so
U

I that you know (,) you know that its a
U

I ten pound basic and its (inaud ....) /[7] right ok
U yes /[8]

I (,) right (laugh ...) we've got that out of the way
U (laugh..) (laugh ..)

I (laugh ..) /[9a] what's what's the subject of of y- your
U

I query /[9b] right (.....) any- anything
U Greek Turkish relations /[10]

I particularly (,) specific /[11]
U actually I'm interested in their

I
U (,) disputes (,) other than Cyprus (....) /[12]

I right (,) disputes (..) (cough .) (....) other than
U

I Cy- are there any (,) any particular ones /[13]
U

I   you know any /[15]
U um /[14]   the Aegean (,) dispute? (........)

I
U and the: their disputes over the treatment of

I
U (....) the (,) Turkish minority in Greece (....) and the

I   right (..........) /[17a]
U (,) Greek minority in Turkey /[16]   (cough.)

I   now have you found very much published on
U

I on this so far (....) (inaud.) /[17b]   mm /[19]
U   's a good (cough.) 's a good deal

I   yes are you are you interested
U   in newspapers /[18]

I newspapers or (,) or you are you 's really articles /[20]
U   no (,) I'm not /[21]   no /[22] I I've

I   good (laugh ....) that's fine /[24]
U covered the newspapers /[23]   (laugh ..)   I

I   mm /[26]
U might also add that I've done a similar search down in

I   ah (,) I wondered about that
U Canada (,) at my University /[25]

I when you said you were from Canada /[27] which 't was using
U yes /[28]   yes /[29]

I the Dialog system was it ? (....) ca- can you remember
U

I what um database (,) you searched? /[30]
U   was an American based

I   yeah /[32]   sounds like Dialog /[33]
U   err database/[31]   yes /[34]

I   yeah um can you remember which particular
U

I files (,) you you searched (.....) 'cos i- if you
U

I   wanted we could try and avoid those (,)
U   I don't /[36]

I or was it that you wanted an update?
U well I I would /[37]

I /[35]
U yes actually no I would want us please to avoid those (.) I

I (laugh ..) I tell you what /[39]
U would (laugh ..) /[38] cos there's no need for duplication

I yes /[41] let me get the the the book which
U I think /[40] I think that /[42] ok /[44]

I describes the files (,) that I think you probably searched
U

I 'scuse me jus take just take off this mike /[43] (20 secs)
U

I right (.) now um /[45]
U I mentioned to the gentleman that I (,) spoke

I yeah /[47]
U to yesterday /[46] about the possibility of limiting the

I yeah (laugh .)
U search to British (,) journals and magazines /[48]

I that's that's not as easy as it seems (laugh..) /[49]
U (inaud .) /[50]

I no um (....) th- the only way
U is that right (,) yes /[51]

I you can do that really i- it'd be nice if you could sit sort
U

I of do do it in one single step but /[52] the only
U mm /[53]

I way would be literally to put in every title (,) of
U

I every British journal that you wanted included
U

I which is (,) probably unless you've got
U

I say three or four (,) specific ones (,) its
U

I such a- an incredible task (,) th- th- that (inaud.) /[54]
U I have no idea what

I (laugh ..) no /[56] no its a problem (,)
U British journal or or magazine anyway (inaud.) /[55]

I um the only thing we can do is t- is once we've decided on
U

I our database see if there was any um (,) any k- key as to to
U

I say place of publication or country of publication to get it
U

I that way /[57] but not <u>all</u> the databases do that so
U right /[58]

I we have to see /[59] how it went /[62] right um (,)
U I see /[60] ok /[61] ok (....) /[63]

I if we have a look through (.....) i- its basically its
U

I politics isn't it it comes under this as economics (,) its
U

I current affairs that's (,) possibly another one (.........)
U

I and (...) even a bit of social sciences (,) i- its really
U

I your your subject is spread all over /[64] lots of different
U mm hm /[65] mm /[66]

I (.) areas /[67] what about books ? (...) which we
U mm hm (cough ..) /[68]

I haven't talked much (,) but its not /[69] good s' right (,)
U I /[70] I am fully conversant with /[71] cos

I i- its really /[72] yeah /[74]
U there's so few of them really that come up /[73] on my

I yeah (.) I can imagine /[76] mm (...) t so
U on the subjects that I'm interested in th- /[75]

I wh- what re- its (,) really Greek Turkish relations /[77]
U other than

I yes /[79]
U Cyprus because /[78] I've <u>tons</u> of material on Cyprus

---

I mm hm /[83]
U an' I (.) any more w- (...) would (inaud ..) (..........)

I so really if if we could say anything on (,) um wh-
U /[82]

I what period what time (,) span /[84]
U nineteen seventy six on /[85]

I uh huh (,) which is going to be difficult yes (,) seventy six
U

I onwards (.....) woops so really if we could look at anything
U

I on (..) Greek Turkish relations (,) and say exclude Cyprus
U

I /[86] from this period onwards /[88] yeah /[90a]
U mm hm /[87] mm hm (...) /[89]

I so what we gotta do now then (,) the har- in these
U

I databases the hardest thing really is to get this time span
U

I in (,) because they don't automatically all (,)
U

I they're not all indexed neatly where you can
U I see /[91]

I just say /[90] nineteen seventy six on
U uh huh /[94] uh huh /[94]

I (,) ah some of them are the historical ones /[93] an
U I see /[95]

I that's a possibility (,) other than that I think the only
U

I way is t- is to look at a databse that hopefully will deal
U

I with recent material /[96] we don't want (,) say ancient
U ok /[97]

I history coming out (laugh .) and this is the
U no /[99] th- the

I danger wi- /[98]
U reason why I've cho- I've chosen nineteen

I mm /[101]
U seventy six is that in nineteen seventy

I mm /[102]
U seventy four /[100] there was a war in on Cyprus and that

I yes /[104] yes
U dominated their relationships /[103] you see /[105]

I /[106] yes of course mm /[108] mm /[110]
U so /[107] I I want to cut all that out because I

I mm /[111]
U have (...) /[109] I have a lot of material on th-

I yes ok so (,) so you jus
U innumerable situations like this /[112] (inaud) so I I

I want you don't want (inaud.) include that would you if
U jus wanted to exclude that really /[114]

I you've got it already (,) I see (,) right (,) um (........)
U I've got /[115]

I Turkish minority in Greece (,) (inaud.) Turkey (........)
U

I /[113] mm
U I can predict there won't be many (,) by the way /[116]

I yes /[118] so its /[120]
U I have already talked to people who are um /[119] I dunno

I yeah /[122] so its really it
U maybe interested in in the subject (..) /[121]

I its (,) political relations /[123] of any sort /[125]
U (cough.) mm hm /[124] mm hm /[126]

I err or any (,) relations really /[127]
U absolutely I mean I I'm primarily interested in their

I yeah /[129]
U disputes but I /[128] I'm dying to find some harmonious

I yeah yes
U aspects of their relationships (inaud.) I mean if there are

---

I (laugh ....) /[131]
U articles that deal with other (,) with with a relationship

I yes right
U that don't warrant a dispute (,) that's fine by me too /[130]

I an' it really any kind of (,) contact in a way /[132] an' it
U absolutely

I can be cultural exchanges or whatever /[134] yes (,) so its
U /[133] yes /[135] yes /[136]

I really our our problem now to find the (,) the right sources
U

I t- t- to tuck in (,) um (...) t (,) let me think about your
U uh huh /[138]

I main (inaud.) books (..........) so if we look through we've got
U

I this one (,) Current Affairs /[137] I'm just really looking
U oh yes /[139]

I through the Dialog /[140] guide because I think Dialog is
U mm hm /[141]

I going to be the one that gives us the (,) widest range of
U

I databases /[142] (....) ah (....) now I've got /[144]
U ok /[143] (cough..)

I this Magazine Index it tends to be more (,) sort
U excuse me /[145]

I of mega activities kind of thing /[146] how about this one
U mm /[147]

I Middle ah Middle East /[148] (inaud .) /[150]
U that's traced in the index (,) what

I I think possibly (,) there
U about (,) magazines magazines y- you have some magazines

I are some /[151] yes /[152]
U that I respect (,) for instance err (.) New New Society's

I I think yes I think it (,) it might
U considered (inaud.) and err /[149]

I be worth (inaud.) /[153] yes (,) yes /[155]
U the Spectator (,) and /[154] magazines like that

I right (11 sec) /[157] ok and that goes (,) right back to
U /[156]

I nineteen well (,) we only want nineteen seventy w- you said
U

I the most recent one which is (inaud.) and we can put
U ok /[159]

I not (,) Cyprus /[158] Middle East not really (,) no
U ok /[160]

I no I think or (,) would it (,) would it come in? /[161]
U it it it

I it could come in yeah /[163] (inaud.) /[164]
U might you see what I find that (,) I 'appen I jus' had

I yes /[165]
U additional course in Middle Eastern politics (,) when people

---

I
U when people say Middle East they generally exclude (,)

I yes (....) /[166]
U Cyprus /[162] much to my chagrin they usually

I yes cos if you think of it
U just mean Arabs and Israelis (....) /[167]

I then its /[168] (inaud.) very much like it /[170]
U yes /[169] people say the Middle

I
U East conflict an an and you know (,) refer to the Arab Israeli

---

I mm (,) lets try those anyway (,) um /[173]
U dispute by that /[171] ok /[174]

I you say you've covered newspapers haven't you (inaud.) won't
U oh yes /[174]

I worry about that (,) um Public ah this is another possibility
U

I Public Affairs (,) International (,) its its a little bit like
U

I the Magazine Index its perhaps a little bit (inaud.) /[175]
U I think that's

I ah /[178] right
U what we tapped into in Canada but I'm not sure /[177]

I (laugh ..) /[179]
U I can't be certain I am merely guessing (,) I am merely

I if you (,) well maybe when you see the sort of
U guessing /[180]

I things coming out (,) because if you see things that you
U (laugh ..) I'll pray /[182]

I know we can hop onto another database (,) um t- (...)
U yes /[183]

I th- that's news again (.....) this is another one World
U

I Affairs Reporter (,) it sounds (,) as if there should be
U

I something there again (,) nineteen seventy onwards (...)
U

I /[181] (cough .)
U mm hm (....) /[184] can you instruct them to (,) give you

I yes you can (.) you can limit them
U nineteen seventy six? /[185] you can ok /[187]

I /[186] yes (.) if you want me to (,) err I
U ok /[188]

I don't think directories would be much use (......)
U

I nor: (......) di- well Dissertations this is
U

I another possibility /[189] its mainly its
U oh yes (inaud ..) it'd be very

I American with a few Canadian (,) I think it
U be very nice to have that /[190]

I would be worth trying /[191] no unfortunately
U what about British? /[192]

I (.) you have to slog through that by hand at the moment
U (laugh ....)

I (,) err until they get it /[193] in machine readable form
U see the /[194]

I and up on on a /[195] a on a host /[197]
U mm hm /[196] because I'm here I my main

I I know /[199] yes (.) the only w-
U interest really is to find British /[198] material /[200]

I way that you could if if you had the time y- th-
U

I they've got them in the library downstairs they've got uh
U

I indexes to them (,) um I think its Aslib produces (,)
U I see /[202]

I produces (,) they could route you to the
U

I right source (,) but you'ld have to go through
U

I it manually /[201] but at least its its (,)
U ok /[226] I see /[203]

I (laugh.) you know /[204] its not too far away (.) /[206a]
U (cough.) mm hm /[205]

I um (,) the only other possibility (7sec) /[206b]
U

I is Historical Abstracts but it it /[207] it is fairly (,)
U no /[208]

I they can include some recent material (.....) /[209]
U

I we'll think about it we'll see we'll put a query by that
U well (.) maybe (.) maybe ok maybe /[210]

I one /[211] mm /[213] its the only database which has
U ok ok alright (..) /[212]

I really (,) obviously because it deals with history tried to
U

I (,) cope with this time limitation /[214] err /[216]
U mm hm (...) /[215] ok

I
U (,) maybe I'm (,) cos I'm in political science I generally

I (inaud.) International we've already had that that's ok
U /[217]

I (.....) I think that's (inaud.) the only other possiblility
U

I although it sounds like social social sciences (,)
U

I sometimes that can have /[218] (....) its its its got such
U mm /[219]

I a wide coverage /[220] that can sometimes have political (,)
U ok /[221]

```
--------------------------------------------------------------------
I  um (inaud.) (....) we've already got World Affairs /[222]
U  (cough .)                                              mm hm
--------------------------------------------------------------------
I   right so if we've got Magazine Index I think these are
U  /[223]
--------------------------------------------------------------------
I  probably the most (.) the ones we ought to start with /[224]
U                (inaud .) /[225]
--------------------------------------------------------------------
I     now the next thing is to get our (,) our strategy (,)
U  ok /[226]
--------------------------------------------------------------------
I  sorted out (inaud.) (11 secs) /[227] (cough.) we've got to
U
--------------------------------------------------------------------
I  have Greek (,) or (,) or Greece (,) and we've got to have
U
--------------------------------------------------------------------
I  Turkish (,) or Turkey (....) or Turkish (,) do they ever (,)
U
--------------------------------------------------------------------
I  refer to them as Turko? /[228] Greek or (,) Graeco Turkey or
U                               w- yeah (,) Greco Turkish /[229]
--------------------------------------------------------------------
I  something (laugh.) /[230]    yes (6sec) that way round
U        Greco Turkish /[231]
--------------------------------------------------------------------
I  /[232]  do they do it the other way round? /[234]
U   mm hm /[233]                                 uh huh (,) no
--------------------------------------------------------------------
I  no    's ok (,) and tha- that's how they spell it normally
U  not really no (,) inaud (.) /[235]
--------------------------------------------------------------------
I  /[236]  yeah (,) 's ok (...) (inaud.) /[238]
U   yes /[237]                    uh (,) actually without the
--------------------------------------------------------------------
I    yes /[240]  yes (inaud.) (......) and we want not
U  a /[239] right /[241]
--------------------------------------------------------------------
I  Cyprus /[242]  I think the only way we can do it is to say
U      yes /[243]
--------------------------------------------------------------------
I  not Cyp- the only problem is i- (,) if they mention a really
U
--------------------------------------------------------------------
I  good one (,) and they err a good paper and it happens
U
--------------------------------------------------------------------
I  to have Cyprus in it (,) we lose it this is the only problem
U
--------------------------------------------------------------------
I  (,) but we'll have a we'll sample as we go we won't (,) w-
U                     w- (inaud.) /[245]
```

I we'll see what we would have missed by taking out
U

I Cyprus (inaud.) /[244] mm (,) we ca- we can try
U oh (,) we will? /[246]

I that either way (,) but what we'll do is put everything
U because its its (,)

I /[247] (inaud.) /[249] yes /[250] yes /[252]
U its virtually impossible /[248] err for any article on Greek

I not to have Cyprus /[253]
U Turkish relations /[251] not to have some

I yes /[255] well
U mention of Cyprus /[254] see it is the (inaud.)

I what we can do (,) according to each database /[257]
U /[256] ok /[258]

I hopefully they will have (,) indexed in some way (,) or they
U

I may have indexed in in some way whether its a conflict
U

I or whether its just mention of the country /[259] um
U (inaud.) /[261]

I otherwise yes it is crude if we just take out Cyprus
U yes /[263]

I (inaud.) but um (inaud.) properly (inaud.) so anything with
U

I Greek and Turkey (,) or Greece and Turkey /[262] you know
U mm hm /[264]

I we can have any combination of those two (,) or Greco
U mm hm /[266] I see /[267]

I Turkish /[265] and (,) hopefully because we're in
U ok /[268] (cough.)

I (,) these current affairs type databases /[269]
U

I which should cover things like
U mm hm /[270]

I relations politics and so on /[271] cos if we were to
U mm hm /[272]

I out in specific key words we might well miss papers /[273]
U I see

I its only if we get a lot of information /[275] like
U /[274] I see /[276]

I cookery or something like that (laugh.) /[277] that we'd
U uh huh /[278]

I have to try putting (,) um (,) politic and you can search on
U

I the root of the words and get political /[279] err relations
U ok /[280]

I (...) relationship (....) exchange i- if you could have a
U

I huge list of words that w- we keep those in reserve
U

I /[281] so that if we need to put those in /[283] um again
U ok /[282] (cough.) mm hm /[284]

I we can look at each as each as we go through each database
U

I we can see (.) if there's indexing /[285] and if so um is
U mm /[286]

I there a (,) very good term that we could make our search
U

I more specific /[287] right ha- have you oh you were
U ok /[288]

I (inaud.) were you present at the searches (,) when you did
U

I them in (,) Canada? /[289a] (,) um can you remember (inaud.)
U I was actually /[290]

I ok so you know what goes on (,) we get (,) /[289b] first
U

I thing is we we l- err (,) tap into each database in turn
U

I /[289c] and then we get the number of references
U uh huh /[291]

I /[292] um (,) we can then (.) sample those by just looking
U uh huh /[293]

I at the title (.) if they're a- any good /[294]

U I'd forgotten that

I ah right ok if they're any good we

U this was a while ago /[295] yes /[297]

I can print then there and then /[296] um /[299]

U ok /[298] so I have the option

I yeah the thing is how long are you staying now? /[301]

U of um /[300]

I ah right so there's not much point I

U I'm leaving on Friday /[302]

I was going to say if we get a lot of (,) you don't think

U

I there'll be a lot but is we get a lot of references we could

U

I print them offline but (,) that's not going to be any good

U I could be surprised (,)

I to you yeah because they take a week to come from the

U y'know (,) one doesn't know /[304]

I States (,) we're using the Dialog (,) same computer that you

U I see

I you used from Canada /[303]

U /[305] really basically (,) um anything that

I

U (,) you're going to be obtaining from the States I'll get as

---

I yes /[307] but um we're not (inaud..)

U well when I go back /[306] so /[308]

I the documents its just the references yes w- we're not

U uh huh /[310]

I offering anything that you can't get (,) in Canad- no

U (cough.) you're not?

I (laugh ....) does that alter things? /[309] it could do

U (laugh ....) /[311] well i- /[312] it

I /[313] yes (,) because we (,) y'see we we're

U disappoints me /[314]

I dependent (,) we we don't do any input here we're not (,) um

U

I (,) we're not a sort of in-house system where where we
U

I actually collect and input because its all (,) it'd be
U

I pointless for us to collect information when its all
U

I available on Dialog anyway /[315] um /[317]
U aha (,) but do they include all

I n- no wh- whatever we get is
U the British stuff too? /[316] mm hm /[319]

I purely what Dialog makes available or what (,) um th- they
U

I (,) the British stuff that you wouldn't be able to get
U

I would be (.) um on a British system like Blaise (,) but
U

I again that (,) ahaa that's books (,) in the Marc files
U

I /[320] (inaud ...) book book catalogue (,) its a
U that is books? /[312] uh huh (cough .)

I catalogue of books that have been (,) deposited in the in
U /[323]

I the err copyright /[322] agreement in this country /[326]
U mm /[324] mm hm /[325] see I

I err and they have files on education and conferences and
U want a ref- /[327]

I on (,) but (,) um (,) y'know i- it isn't it c- it wouldn't
U

I it no means by no means comprehensive for your search
U

I /[328] so thi- this obviously makes a difference /[330] mm
U mm hm see my /[329] my primary interest is is to

I /[331] mm /[333]
U get (.) references that are (,) that are British (,) that I

I
U am (,) not very likely to be able to get (,) while I'm in

I   right /[335]
U Canada /[334] you see (.) um the other thing is because I

I   mm /[337]   mm
U still have some time here I have access to the (,) LSE

I /[338]   yes /[339]   yes /[340]
U library and and you know (,) other libraries (,) because I I

I   yeah /[341]
U used to be a a student here (,) some years back /[336]

I   mm ah so in this case (,) mm /[343]
U um /[342]   so things that need to be looked up I can

I   yes /[345]   and
U <u>easily</u> (,) you know get the list an and look up the

I then you'll get (,) yes /[346]   which is
U brochure and Xerox them and (,) you know buy them or whatever

---

I sensible (,) yeah (...) so well the only thing I I can
U /[344]

I suggest is that we did something but tried to get it just to
U

I (,) wh- where they indexed it as place of publication being
U

I British (,) but its not always /[347] terribly easy /[349]
U   yes /[348]   well

I that's the only thing (,) um I can check (,) /[351a] I'm
U /[350]   (inaud ..) /[352]

I just trying to think with Blaise (,) but the Marc files are
U

I the main ones I can think of (,) um they've got one on
U

I conferences and research (,) um no I think Dialog is still
U

I err (,) still the best bet (......) um (,) /[351b] no wh-
U

I what you really want (,) i- is some kind of in-house system
U   (cough.)

I where just the British information (,) say (,) the holdings
U

```
-------------------------------------------------------------------
I  of a university or whatever or certain collections of um
U
-------------------------------------------------------------------
I  (inaud ..) /[351c]                                       no no
U  mm hm yeah its not as if I'm not interested in others (,)
-------------------------------------------------------------------
I  I know while you're here you want to take advantage (,) yes
U  its just that /[353]     while I'm here (.) um /[355]
-------------------------------------------------------------------
I  of the British ones /[354]      right its up to you really
U                          mm hm /[356]
-------------------------------------------------------------------
I  (laugh ..) /[357a] um (,) we could we could use some
U
-------------------------------------------------------------------
I  of these and and hopefully try and get out (,) stuff
U
-------------------------------------------------------------------
I  that's published (,) in British journals /[357b]  the o-
U                                            mm hm ok /[358]
-------------------------------------------------------------------
I  the one I I do not know for certain Social Science Citation
U
-------------------------------------------------------------------
I  Index has got where the work was done /[359] so we could put
U                                             ok /[360]
-------------------------------------------------------------------
I  England UK and so on (,) as (,) corporate source /[361] as a
U                                                      ok /[362]
-------------------------------------------------------------------
I possibility /[363]     this we know (,) is not going to
U              alright /[364]
-------------------------------------------------------------------
I  be any good for the dissertations because its American
U
-------------------------------------------------------------------
I  anyway /[365]   um (....) these (,) we can try again
U        yes /[366]
-------------------------------------------------------------------
I  looking at (,) 's the public affairs of the world (,)
U
-------------------------------------------------------------------
I affairs ones /[367]  we can see if we can just get it down
U               mm hm /[368]
-------------------------------------------------------------------
I  just to British journals /[369]       lets tr- lets do a a
U                            mm hm (...) /[370]
-------------------------------------------------------------------
I  little pilot one if you like on this little subset see how
U                                                    sure /[373]
-------------------------------------------------------------------
I  it works /[371]  and then stop and think /[375]  ok
U          ok /[374]                              ok /[376]
```

I (laugh....) /[377] I'm sorry about th- its a good job
U fair enough /[378]

I we discovered before we got (inaud.) /[379] (laugh.) /[381]
U well I /[380] I think

I yes yes
U that was important for me to convey to you (inaud.) /[384]

I well you did say that when you first cane actually
U

I (laugh..) um (,) an then then I think possibly you
U (laugh .)

I though we were accessing (,) sources other than (,) purely
U

I Dialog or or that we had some (inaud.) /[383]
U yeah (,) I mentioned it to the gentleman yesterday

I yes (,) yes (...) no a- as far as the online searches go I
U (inaud.) /[385]

I mean obviously (,) you you could (inaud.) to the library you
U

I could find other things (,) but th- the one that's
U

I peculiarly British is this the one British Library one
U

I /[385] um but s- I don't think if you don't want books I
U uh huh /[387]

I don't think that's going to to really satisy you (,) ok so
U

I we'll go online? /[388] right /[391]
U ok /[389] very good /[390] yeah ok /[392]

APPENDIX 2:
Analysis of Interview 190684HBA

ANALYSIS OF INTERVIEW 190684HBA (GREEK-TURKISH RELATIONS).

Focus	utterance	speaker	subgoal analysis
1	3	I	USER
[USER]	4	U	USER
	5	I	USER
	6	U	USER
	7	I	EXPLAIN
	8	U	ph
2	9	I	TOPIC
[TOPIC]	10	U	TOPIC
	11	I	TOPIC
	12	U	RES
	13	I	SUBJ
	14	U	ph
	15	I	SUBJ
	16	U	SUBJ
	17a	I	ph
3	17b	I	PREV
[OUTPUT]	18	U	SLIT
	19	I	ph
	20	I	OUTPUT
	21	U	OUTPUT
	22	U	OUTPUT
	23	U	PREV
	24	I	MATCH:OUTPUT
4	25	U	PREV
[PREV]	26	I	ph
	27	I	MATCH:PREV
	28	U	ph
	29	U	ph
	30	I	PREV;IRS
	31	U	PREV
	32	I	ph
	33	I	PREV
	34	U	ph
	35	I	PREV..EXPLAIN:DB..OUTPUT
	36	U	(failed)
	37	U	(failed)
	38	U	DB
	39	I	ph
	40	U	OUTPUT
	41	I	MATCH:OUTPUT
	42	U	(failed)
	43	I	INFORM
	44	U	ph
	45	I	ph

5	46	U	OUTPUT
[EXPLAIN]	47	I	ph
	48	U	OUTPUT
	49	I	CAPAB
	50	U	ph
	51	U	ph
	52	I	EXPLAIN:STRAT
	53	U	ph
	54	I	EXPLAIN:STRAT
	55	U	KNOW
	56	I	ph
	57	I	PLAN..EXPLAIN:STRAT
	58	U	ph
	59	I	EXPLAIN:STRAT
	60	U	ph
	61	U	ph
	62	I	EXPLAIN:STRAT
	63	U	ph
6	64a	I	focus-shift
[DB]	64b	I	DB..SUBJ
	65	U	ph
	66	U	ph
	67	I	SUBJ
	68	U	ph
	69	I	OUTPUT
	70	U	(failed)
	71	U	KNOW
	72	I	MATCH:OUTPUT..(failed)
	73	U	SLIT
	74	I	ph
	75	U	SLIT
	76	I	ph
7	77	I	MATCH:TOPIC
[TOPIC]	78	U	TOPIC
	79	I	ph
	80	U	KNOW
	81	I	ph
	82	U	(failed)
	83	I	ph
	84	U	TOPIC
	85	U	TOPIC
	86	I	STRAT..QUERY..TOPIC
	87	U	ph
	88	I	TOPIC
	89	U	ph
8	90a	I	focus-shift
[EXPLAIN]	90b	U	EXPLAIN:STRAT
	91	U	ph
	92	U	ph
	93	I	EXPLAIN..STRAT..DB

	94	U	ph
	95	U	ph
	96	I	EXPLAIN..STRAT..DB
	97	U	ph
	98	I	EXPLAIN:STRAT.. EXPLAIN:DB
	99	U	ph
9 [TOPIC]	100	U	SUBJ
	101	I	ph
	102	I	ph
	103	U	SUBJ
	104	I	ph
	105	U	ph
	106	U	ph
	107	U	ph
	108	I	MATCH:SUBJ
	109	U	TOPIC
	110	I	ph
	111	I	ph
	112	U	KNOW
	113	I	MATCH:TOPIC
	114	U	TOPIC
	115	U	(failed)
10 [SLIT]	116	U	SLIT
	117	I	ph
	118	I	ph
	119	U	PREVNON
	120	I	(failed)
	121	U	PREVNON
	122	I	ph
11 [TOPIC]	123	I	MATCH:TOPIC
	124	U	ph
	125	I	MATCH:TOPIC
	126	U	ph
	127	I	MATCH:TOPIC
	128	U	MATCH:TOPIC:CONFIRM ..RES
	129	I	ph
	130	U	UGOAL..DOCS
	131	I	ph
	132	I	MATCH:TOPIC
	133	U	MATCH:TOPIC:CONFIRM
	134	I	TOPIC
	135	U	ph
	136	U	ph
12 [DB]	137	I	DB
	138	U	ph
	139	U	ph
	140	I	DB
	141	U	ph

	142	I	DB
	143	U	ph
	144	I	(failed)
	145	U	ph
	146	I	EXPLAIN:DB
	147	U	ph
	148	I	DB
	149	U	OUTPUT
	150	I	(inaud)
	151	I	(failed)
	152	I	ph
	153	I	(failed)
	154	U	OUTPUT
	155	I	ph
	156	U	OUTPUT
	157	I	ph
13	158	I	DISPLAY..TOPIC
[SUBJ]			..QUERY
	159	U	ph
	160	U	ph
	161	I	DB..MATCH:DB:QUERY
	162	U	MATCH:DB .. SUBJ
	163	I	MATCH:DB
	164	I	(inaud)
	165	I	ph
	166	I	ph
	167	U	SUBJ
	168	I	SUBJ
	169	U	ph
	170	I	SUBJ
	171	U	SUBJ
	172	I	ph
	173	I	DB
	174	U	ph
14	175	I	PREV..DB..
[DB]			EXPLAIN:DB
	176	U	ph
	177	U	PREV
	178	I	ph
	179	I	ph
	180	U	PREV
	181	I	PLAN..DB
	182	U	ph
	183	U	ph
	184	I	ph
	185	U	QUERY
	186	I	QUERY
	187	U	MATCH:QUERY:CONFIRM
	188	U	ph
	189	I	MATCH:QUERY..DB
	190	U	DB
	191	I	EXPLAIN:DB

	192	U	MATCH:OUTPUT:QUERY
	193	I	DB..EXPLAIN:DB
	194	U	ph
	195	I	EXPLAIN:DB
	196	U	ph
	197	I	EXPLAIN:DB
	198	U	UGOAL
	199	I	MATCH:UGOAL:CONFIRM
	200	U	UGOAL
	201	I	CAPAB..EXPLAIN:CAPAB
	202	U	ph
	203	U	ph
	204	I	CAPAB
	205	U	ph
	206a	I	CAPAB
	206b	I	DB
	207	I	DB
	208	U	ph
	209	I	EXPLAIN:DB
	210	U	MATCH:DB:CONFIRM
	211	I	INFORM
	212	U	MATCH
	213	I	ph
	214	I	EXPLAIN:DB
	215	U	ph
	216	I	ph
	217	U	USER..(failed)
	218	I	DB..EXPLAIN:DB
	219	U	ph
	220	I	EXPLAIN:DB
	221	U	ph
	222	I	EXPLAIN:DB..DB
	223	U	ph
	224	I	DB
	225	U	ph
	226	U	MATCH:DB:CONFIRM
15	227	I	PLAN
[TERM]	228	I	TERM
	229	U	TERM
	230	I	TERM
	231	U	TERM
	232	I	TERM
	233	U	ph
	234	I	TERM
	235	U	TERM
	236	I	TERM..MATCH:TERM:QUERY
	237	U	MATCH:TERM:CONFIRM
	238	I	ph
	239	U	TERM
	240	I	MATCH:TERM
	241	U	MATCH:TERM:CONFIRM
	242	I	QUERY
	243	U	MATCH:QUERY:CONFIRM

	244	I	QUERY..EXPLAIN:QUERY.. STRAT
	245	U	MATCH:STRAT:QUERY
	246	U	ph
	247	I	STRAT
	248	U	DOCS
	249	I	ph
	250	I	ph
	251	U	DOCS
	252	I	ph
	253	I	DOCS
	254	U	MATCH:DOCS
	255	I	MATCH:DOCS:CONFIRM
	256	U	(failed)
	257	I	STRAT
	258	U	ph
	259	I	STRAT
	260	U	ph
	261	U	ph
	262	I	EXPLAIN:STRAT.. QUERY
	263	U	ph
	264	U	ph
	265	I	EXPLAIN:QUERY..QUERY
	266	U	ph
	267	U	ph
	268	U	ph
	269	I	STRAT
	270	U	ph
	271	I	STRAT
	272	U	ph
	273	I	EXPLAIN:STRAT
	274	U	ph
	275	I	EXPLAIN:STRAT
	276	U	ph
	277	I	EXPLAIN:STRAT
	278	U	ph
	279	I	STRAT..TERM.. EXPLAIN:QUERY
	280	U	ph
	281	I	TERM..EXPLAIN:TERM ..STRAT
	282	U	ph
	283	I	STRAT
	284	U	ph
	285	I	STRAT
	286	U	ph
	287	I	STRAT
	288	U	MATCH:STRAT
16	289a	I	PREV;IRS
[EXPLAIN]	289b	I	IRS
	289c	I	EXPLAIN:STRAT
	290	U	PREV

	291	U	ph
	292	I	EXPLAIN:STRAT
	293	U	ph
	294	I	EXPLAIN:STRAT
	295	U	IRS..PREV
	296	I	EXPLAIN:STRAT
	297	U	ph
	298	U	ph
	299	I	ph
	300	U	(failed)
	301	I	UGOAL
	302	U	UGOAL
	303	I	EXPLAIN:STRAT..SLIT EXPLAIN:STRAT
	304	U	ph
	305	U	ph
17	306	U	CAPAB
[CAPAB]	307	I	MATCH:CAPAB:CONFIRM
	308	U	ph
	309	I	CAPAB
	310	U	ph
	311	U	MATCH:CAPAB:QUERY
	312	U	ph
	313	I	MATCH:CAPAB:QUERY
	314	U	MATCH:CAPAB
	315	I	EXPLAIN:CAPAB
	316	U	CAPAB
	317	I	ph
	318	I	EXPLAIN:CAPAB..DB
	319	U	ph
	320	I	(inaud)
	321	U	OUTPUT
	322	I	EXPLAIN:DB
	323	U	ph
	324	U	ph
	325	U	ph
	326	I	EXPLAIN:DB
	327	U	ph
	328	I	EXPLAIN:DB..DB
	329	U	(failed)
	330	I	MATCH:CAPAB:QUERY
	331	U	UGOAL
	332	I	ph
	333	I	ph
	334	U	UGOAL
	335	I	MATCH:UGOAL:CONFIRM
	336	U	EXPLAIN:UGOAL..USER
	337	I	ph
	338	I	ph
	339	I	ph
	340	I	ph
	341	I	ph
	342	U	ph

	343	I	(failed)
	344	U	UGOAL
	345	I	ph
	346	I	MATCH:UGOAL
18 [QUERY]	347a	I	change-of-focus
	347b	I	STRAT
	348	U	ph
	349	I	STRAT
	350	U	ph
	351a	I	STRAT
	351b	I	DB
	351c	I	CAPAB
	352	U	ph
	353	U	MATCH:CAPAB
	354	I	MATCH:CAPAB
	355	U	UGOAL
	356	U	ph
	357a	I	EXPLAIN
	357b	I	STRAT
	358	U	MATCH:STRAT
	359	I	DB..EXPLAIN:DB
	360	U	ph
	361	I	QUERY
	362	U	ph
	363	I	QUERY
	364	U	ph
	365	I	EXPLAIN:DB
	366	U	ph
	367	I	DB
	368	U	ph
	369	I	STRAT
	370	U	ph
	371	I	INFORM
	372	U	ph
	373	U	ph
	374	U	ph
	375	I	INFORM
	376	U	MATCH
	377	I	MATCH:INFORM:QUERY
	378	U	MATCH:INFORM:CONFIRM
19 [UGOAL]	379	I	PLAN
	380	U	(failed)
	381	I	ph
	382	U	PLAN;UGOAL
	383	I	MATCH
	384	U	ph
	385	U	PREVNON
	386	I	CAPAB
	387	U	ph
	388	I	MATCH:CAPAB..INFORM
	389	U	ph
	390	U	MATCH:INFORM

391	I	MATCH:INFORM
392	U	MATCH:INFORM

APPENDIX 3:
Interaction maps of the Problem Description and Retrieval Strategy functions for Interview 190684HBA

Interaction map for Problem Description
Interview 190684HBA

FOCUS	TOPIC	RES	SUBJ	SLIT	DOCS	I/O
Focus 2 [TOPIC]	GREEK TURKISH RELATIONS	DISPUTES BETWEEN GREECE AND TURKEY ↓ EXCLUDE DISPUTE OVER CYPRUS	Subject area = INTERNATIONAL RELATIONS AEGEAN DISPUTE ↓ DISPUTE OVER TREATMENT OF TURKISH MINORITY IN GREECE ↓ DISPUTE OVER TREATMENT OF GREEK MINORITY IN TURKEY			RS (DB) FOCUS 6 RS (STRAT) FOCUS 6
Focus 3 [OUTPUT]				LARGE NUMBER OF ITEMS ON GREEK-TURKISH RELATIONS IN NEWSPAPERS		PS (PREV) [What user has found so far] RG (OUTPUT) [Ascertain if user wants newspapers]
FOCUS 4 [PREV]						
FOCUS 5 [EXPLAIN]						

Interview 190684HBA (2·)

FOCUS	TOPIC	RES	SUBJ	SLIT	DOCS	I/O
Focus 6 [DB]			[focus2 → International Relations]			RS (DB) MATCH-SUBJ
						RS (STRAT) [Multi-database Search]
						RS (STRAT) TACTIC [Select a database with broad subject coverage]
						RS (DB) CONSIDER (MARC)
				FEW BOOKS ON GREEK-TURKISH RELATIONS		UM (KNOW) [Knows books]
						RG (OUTPUT) [Do not restrict to books]
Focus 7 [TOPIC]	(FOCUS 2) → GREEK-TURKISH RELATIONS					RS (QUERY) [GREEK + TURKISH + RELATIONS]
	EXCLUDE CYPRUS					RS (QUERY) [NOT CYPRUS]
				VERY MANY ITEMS ABOUT CYPRUS		
	QUALIFIER: HISTORICAL PERIOD = 1976-1984					RS (QUERY) [1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984]
						RS (STRAT) TACTIC [Select historical databases]
						RS (STRAT) TACTIC [Select databases covering recent material]
Focus 8 [EXPLAIN]						

Interview 190684HBA (3)

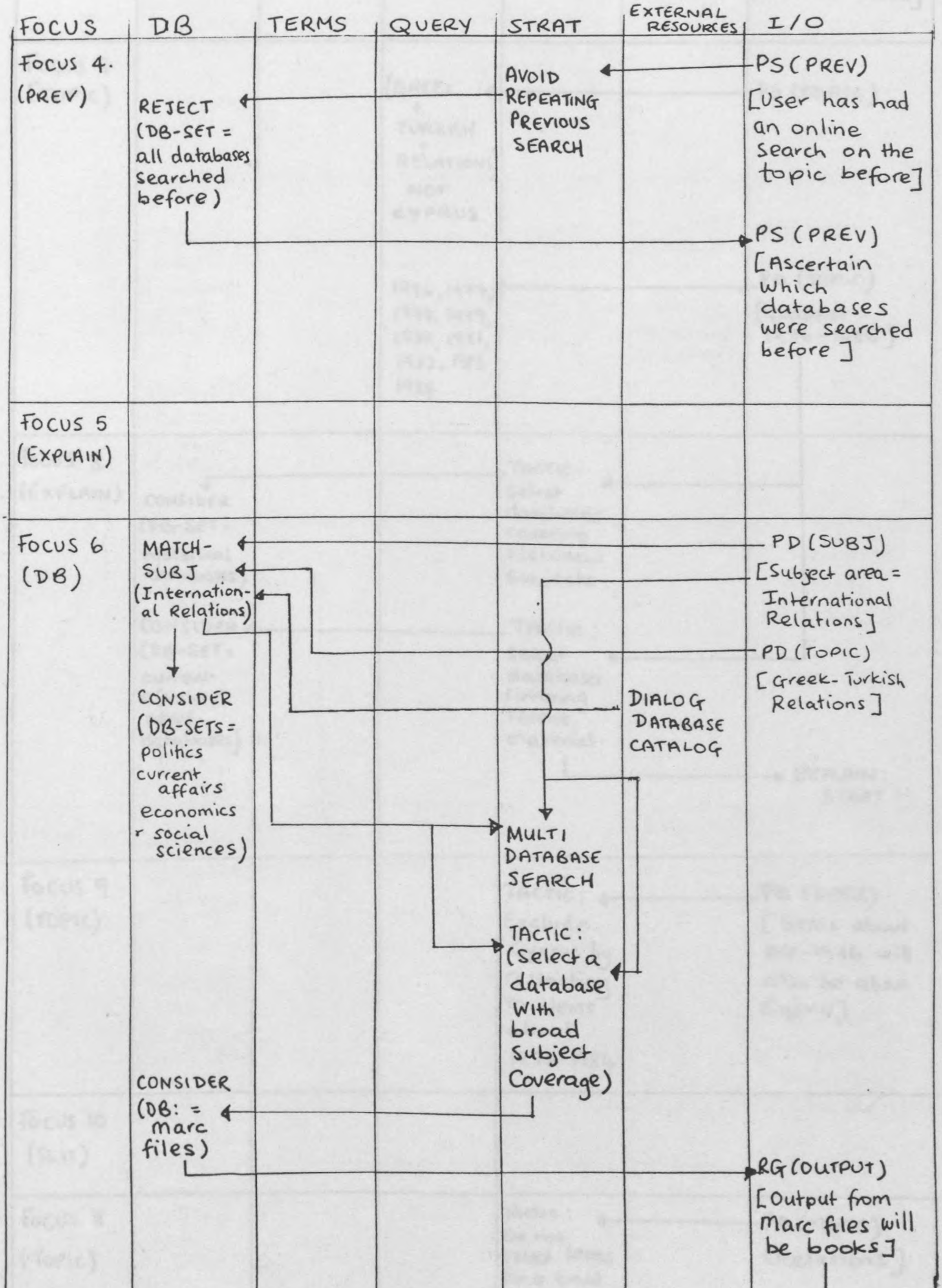
FOCUS	TOPIC	RES	SUBJ	SLIT	DOCS	I/O
Focus 9 [TOPIC]	(Focus 7) ↓ 1976 – 1984 NOTHING ON CYPRUS EXCLUDE ALL ITEMS ON CYPRUS TURKISH MINORITY IN GREECE GREEK MINORITY IN TURKEY		CYPRUS WAR IN 1974 ↓ CYPRUS WAR DOMINATED GREEK-TURKISH RELATIONS [Focus 2 → Turkish Minority in Greece Greek minority in Turkey]	VERY MANY ITEMS ON CYPRUS DISPUTE	DOCUMENTS BEFORE 1976 ON GREEK-TURKISH RELATIONS CONCENTRATE ON CYPRUS	RS (STRAT) TACTIC [Exclude Cyprus by limiting to documents about 1976 –] UM (KNOW) [User knows enough about Cyprus] ↓ PS (PREV) [User has got alot on Cyprus]
Focus 10 [SLIT]				NOT VERY MUCH LITERATURE ON TOPIC		PS (PREVNON) [Talked to others about topic]
Focus 11 [TOPIC]	[Focus 2 → Greek-Turkish Relations] ↓ POLITICAL RELATIONS AND/OR NON-POLITICAL RELATIONS	PRIMARILY DISPUTES ↓ HARMONIOUS RELATIONS (IF ANY) ↓				

Interview 190684HBA (4)

FOCUS	TOPIC	RES	SUBJ	SLIT	DOCS	I/O
Focus 11 continued	ANY TYPE OF RELATION E.G. CULTURAL EXCHANGES		↓		DOCUMENTS ABOUT RELATIONS WHICH ARE NOT DISPUTES ARE ACCEPTABLE	RS (STRAT) [Do not need to select terms for RELATION]
Focus 12 [DB]	$\boxed{\text{TOPIC}}$		$\boxed{\text{SUBJECT AREA}}$			RS(DB)
Focus 13 [SUBJ]			IS CYPRUS PART OF MIDDLE EAST? ↓ MIDDLE EAST CONFLICT IS ARAB-ISRAELI DISPUTE ↓ MIDDLE EAST DOES NOT INCLUDE CYPRUS			RS (DB) CONSIDER [Middle East Index or Abstracts] RS (DB) REJECT [Middle East Index]
Focus 14 [DB]	$\boxed{\text{TOPIC}}$		$\boxed{\text{SUBJECT AREA}}$			RS (DB)
Focus 15 [TERMS]	$\boxed{\text{TOPIC}}$ GREEK TURKISH GREEK-TURKISH CYPRUS				ALL ITEMS ON GREEK-TURKISH RELATIONS WILL MENTION CYPRUS	RS (TERMS) GREEK/GREECE TURKEY/TURKISH GRECO TURKISH CYPRUS RS (QUERY) [NOT CYPRUS] RS (STRAT) TACTIC [Monitor effect of NOT]

Interaction map for Retrieval Strategy

Interview 190684HBA



Interview 190684HBA (2)

FOCUS	DB	TERMS	QUERY	STRAT	EXTERNAL RESOURCES	I/O
Focus 6 Continued	REJECT (DB = Marc)					UM (KNOW) [User knows all the books]
Focus 7 (TOPIC)			(GREEK + TURKISH + RELATIONS) NOT CYPRUS 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983 1984			PD (TOPIC) PD (TOPIC) [Qualifier— 1976-1984]
Focus 8 (EXPLAIN)	CONSIDER (DB-SET = historical databases) CONSIDER (DB-SET = current affairs / news databases)			TACTIC: Select databases covering historical Subjects TACTIC: Select databases covering recent material		EXPLAIN: STRAT
Focus 9 (TOPIC)				TACTIC: Exclude Cyprus by restricting to items about 1976-1984		PD (DOCS) [Items about pre-1976 will also be about Cyprus]
Focus 10 (SLIT)						
Focus 11 (TOPIC)				TACTIC: Do not select terms for a broad concept		PD (TOPIC) [Relations]

Interview 190684HBA (3)

FOCUS	DB	TERMS	QUERY	STRAT	EXTERNAL RESOURCES	I/O
Focus 12 (DB)	CONSIDER (DB SET= Dialog databases)			(Focus 6 → MULTI-DATABASE SEARCH) ↓ TACTIC: Select host with widest range of databases		EXPLAIN
	MATCH-SUBJ (International Relations)				Dialog database catalog	PD (TOPIC) PD (SUBJ)
	↓ CONSIDER (DB SET = Current Affairs)					
	↓ CONSIDER (DB = Magazine Index)					RG (OUTPUT) [user wants Magazine Articles]
	↓ CONSIDER (DB = Middle East Index)					
Focus 13 (SUBJ)	COVERAGE (Magazine Index)					PD (TOPIC) [1976-1984]
			NOT CYPRUS			
	CONSIDER (DB = Middle East Index)					PD (SUBJ) [is Cyprus part of the Middle East?]
	REJECT (DB = Middle East Index)					↓ PD (SUBJ) [Cyprus is not part of Middle East]

Interview 190684HBA (4)

FOCUS	DB	TERMS	QUERY	STRAT	EXTERNAL RESOURCES	I/O
FOCUS 14 (DB)	REJECT (DB-SET= newspaper databases)					PS (PREV) [User has covered newspapers]
	CONSIDER (DB = PAIS)					EXPLAIN: DB
						PS (PREV) [User thinks he searched this before]
	SELECT (DB = PAIS)			TACTIC: Check PAIS quickly & Exit if user recognises output		
	CONSIDER (DB = World Affairs Reporter) ↓					
	COVERAGE (World Affairs Reporter)					PD (TOPIC) [1976-1984] ↓ EXPLAIN: STRAT
	CONSIDER (DB = Compreh. Dissertations Abstracts) ↓					EXPLAIN: DB
	COVERAGE (Diss. Abs.)					PM [Sources for UK dissertations]
	CONSIDER (DB-SET= historical databases) ↓			(Focus 8: TACTIC - select historical databases)		
	CONSIDER (DB = Historical Abstracts)					EXPLAIN: DB

Interview 190684HBA (5)

FOCUS	DB	TERMS	QUERY	STRAT	EXTERNAL RESOURCES	I/O
Focus 14 Continued	CONSIDER					Um (USER) [User is in Political Science]
	(DB-SET = Political Science databases) ↓				Dialog database catalog	
	CONSIDER (PAIS) ↓					
	CONSIDER (DB = Social Sciences Citation Index)					EXPLAIN: DB
	SELECT (DB = World Affairs) (DB = magazine Index) (DB = SSCI)			TACTIC: Start search with most promising databases		
Focus 15 [TERMS]		SELECT (greek) (greece)				PD (TOPIC) Theme: Greek
		SELECT (turkey) (turkish)				PD (TOPIC) Theme: Turkey
		SYNONYMS (Greek-Turkish) ↓				PD (TOPIC) Theme: Greek-Turkish
		SELECT (greco-turkish) ↓				
		SPELL (greco-turkish) ↓				
		SYNONYMS (Turkish-Greek)				

Interview 190684HBA
(6)

FOCUS	DB	TERMS	QUERY	STRAT	EXTERNAL RESOURCES	I/O
Focus 15 Continued		SELECT (CYPRUS)				PD (TOPIC) Theme: Exclude Cyprus
			NOT (CYPRUS)			EXPLAIN (Problems of NOT-ing)
				TACTIC: monitor output to see what was missed by excluding Cyprus		
				TACTIC: check indexing- try to use descriptor to express concept		PD (DOC) [All docs. will mention Cyprus]
			(GREEK AND TURKISH) OR (GREECE AND TURKEY) OR (GRECO- TURKISH)	(Focus-11 → Do not select terms for a broad concept)		PD (TOPIC) Theme: Relations
				TACTIC: Use data- base coverage to express concept		EXPLAIN: STRAT
				BACK-UP PLAN: If too many irrelevant items THEN add in stem 'politic' 'relation'		

Interview 190684HBA (T)

FOCUS	DB	TERMS	QUERY	STRAT	EXTERNAL RESOURCES	I/O
Focus 15 continued				TACTIC: Check indexing in retrieved items - use to reformulate		
Focus 16 (EXPLAIN)						
Focus 17 (CAPAB)						RG (OUTPUT) [UK produced material]
Focus 18 (QUERY)	CONSIDER (DB-SET = databases with place of publication field)					EXPLAIN
	CONSIDER (DB-SET = Blaise databases)					
	MATCH-SUBJ					PD (TOPIC) PD (SUBJ)
	REJECT (DB-SET = Blaise databases)					
	SELECT (DB-SET = Dialog databases)					RG (OUTPUT) [UK journals]
	SELECT (DB = SSCI)					EXPLAIN: DB
	REJECT (DB = Diss. Abstracts)					EXPLAIN: DB
	SELECT (DB = PAIS)					
	SELECT (DB = WORLD AFFAIRS)			PILOT SEARCH		

APPENDIX 4:
The online search strategy for Interview 190684HBA

THE ONLINE SEARCH STRATEGY (INTERVIEW 190684HBA)

HOST = DIALOG

DATABASES USED (IN ORDER OF SEARCHING):

1. SOCIAL SCIENCES CITATION INDEX (SSCI)
2. PUBLIC AFFAIRS INFORMATION SERVICE (PAIS)
3. MAGAZINE INDEX

SEARCH STRATEGY

DATABASE 1 (SSCI)

1 SELECT GREECE/TI OR GREEK/TI
2 SELECT TURKEY/TI IR TURKISH/TI
3 SELECT GRECO(W)TURK?
4 COMBINE (1 AND 2) OR 3
5 SELECT ENGLAND/CS OR UK/CS OR ENGLISH/CS OR BRITAIN/CS OR BRITISH/CS OR LONDON/CS
6 COMBINE 4 AND 5
7 PRINT SET-6
8 COMBINE 4 NOT 5
9 SAVETEMP (1) [save as a temporary search strategy]

DATABASE 2 (PAIS)

1 EXECUTE SAVETEMP (1) [failed because no title field on this database]
2 SELECT GREECE OR GREEK
3 SELECT TURKEY OR TURKISH
4 SELECT GRECO(W)TURK?
5 COMBINE (2 AND 3) OR 4
6 SELECT CS=ENGLAND OR CS=BRITAIN OR CS=BRITISH OR CS=GREAT(W)BRITAIN
7 COMBINE 5 AND 6
8 PRINT SET-5
9 SAVETEMP (2)

DATABASE 3 (MAGAZINE INDEX)

1 EXECUTE SAVETEMP (2)

7 PRINT SET-6 [greek-turkish combined with england]
8 SL=ENGLAND
9 COMBINE 8 AND 4

APPENDIX 5:
Problem descriptions for Interview 190684HBA

Key:

```
:'''''''':
:        :   =  problem description plane
:........:

:''''''
:            = problem description plane in which some entities
........       diagram clearer

 --------
|        |   = a network "space"
|        |
 --------
```

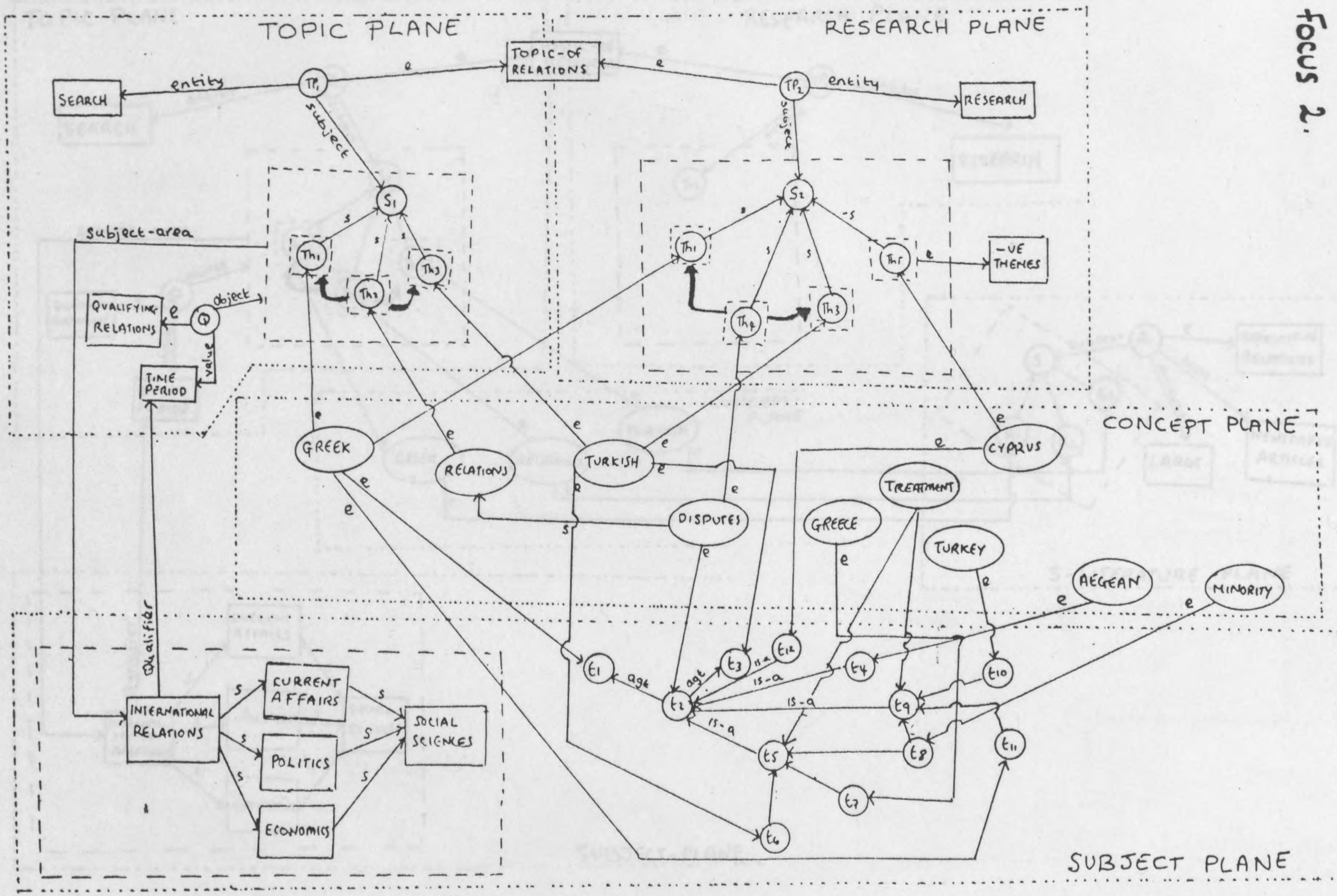
➡ = indicates a network "vista". The arrow points towards the parent vista

——— = a link

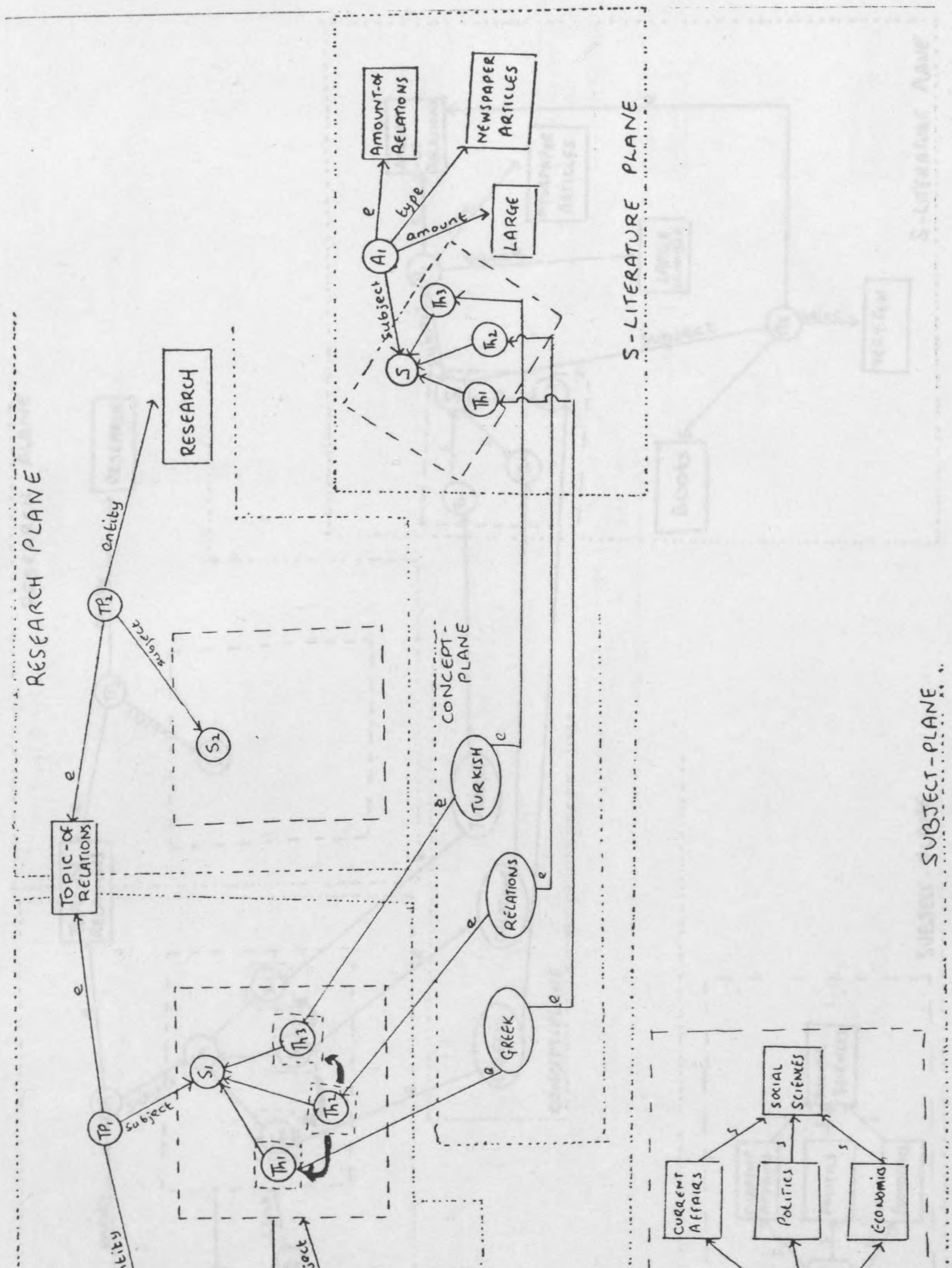
○ ▭ } = nodes

Interview 190684HBA

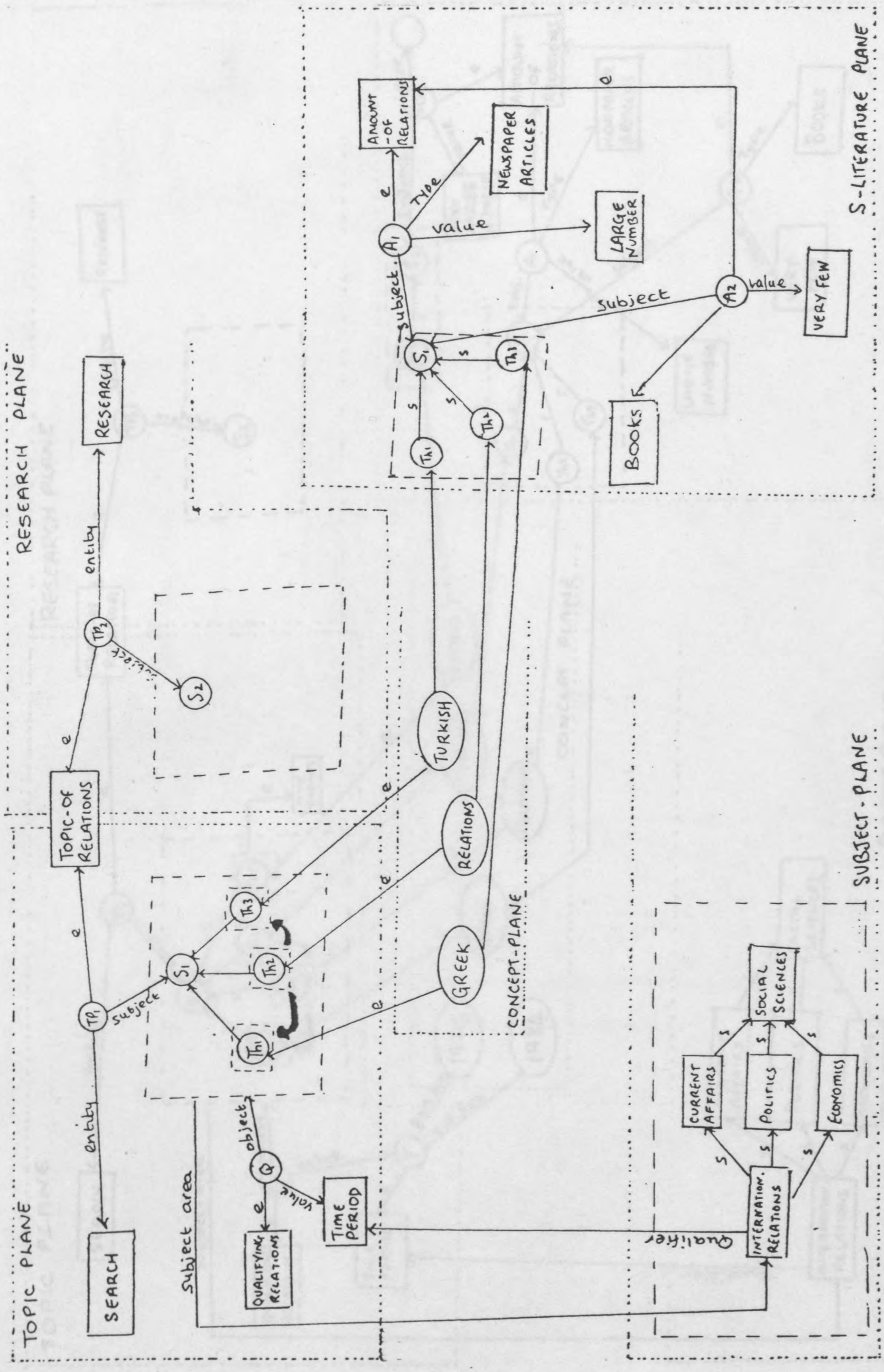
Focus 2.



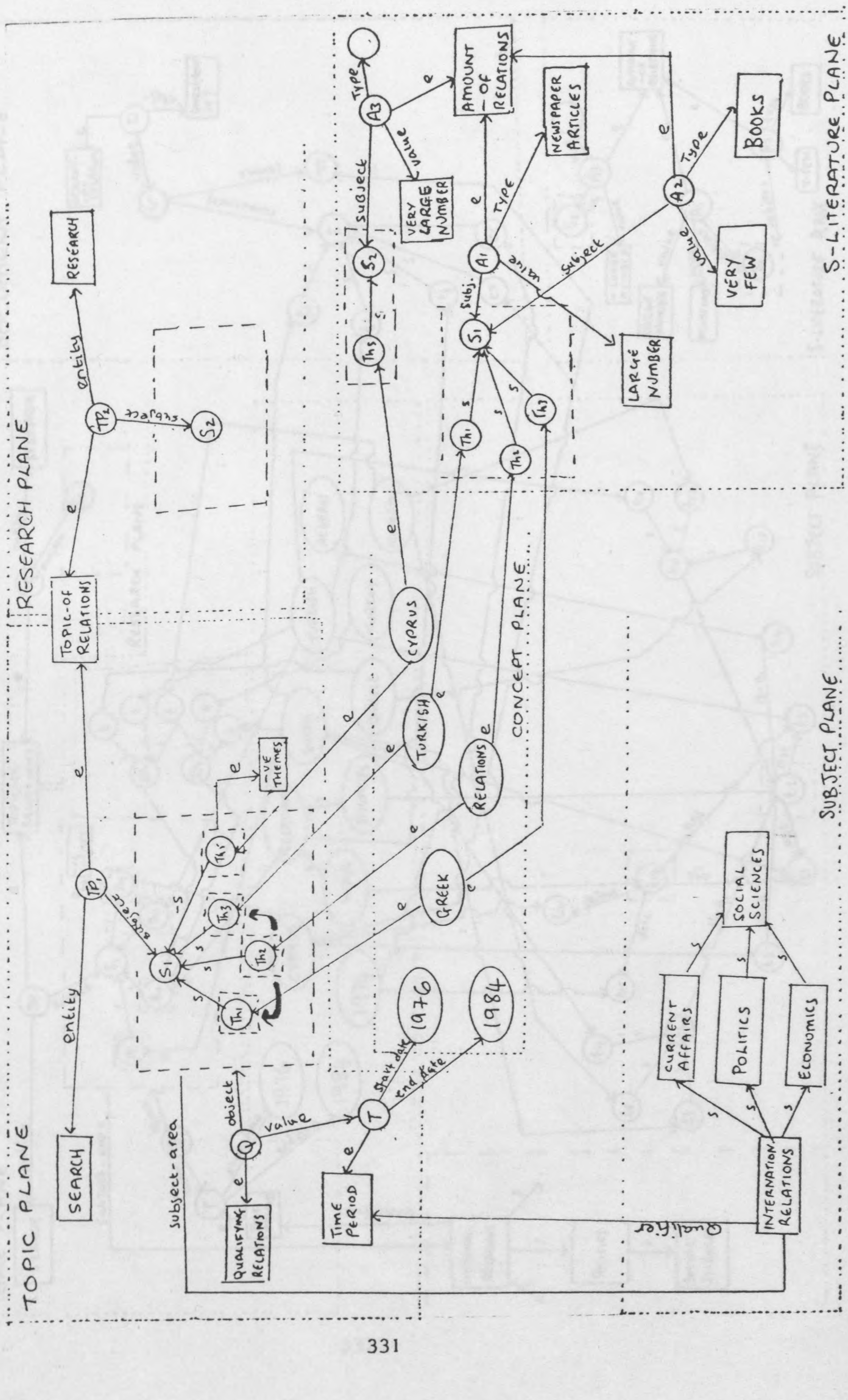
Interview 190684HBA Focus 3.

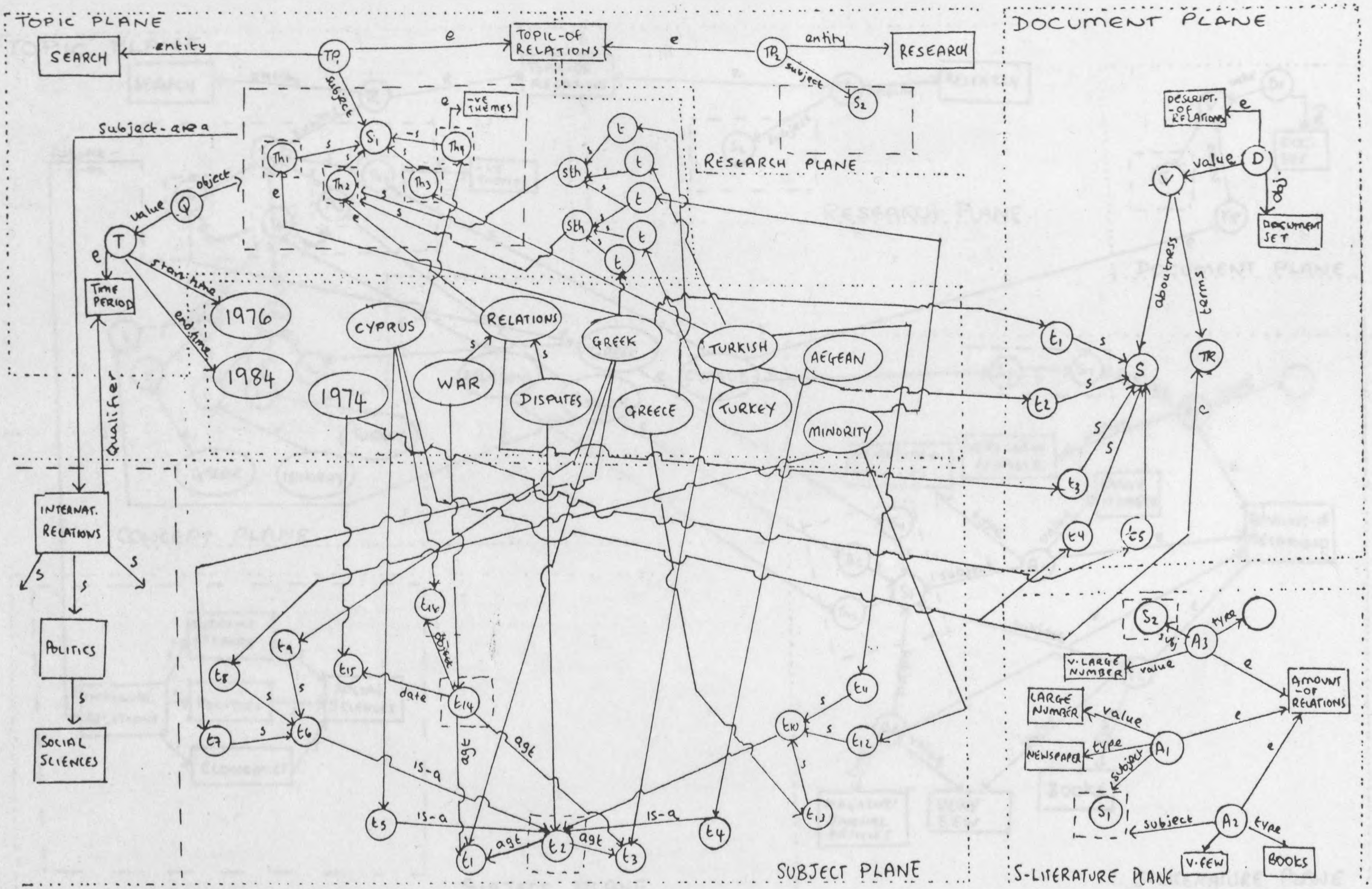


Interview 190684HBA Focus 6.



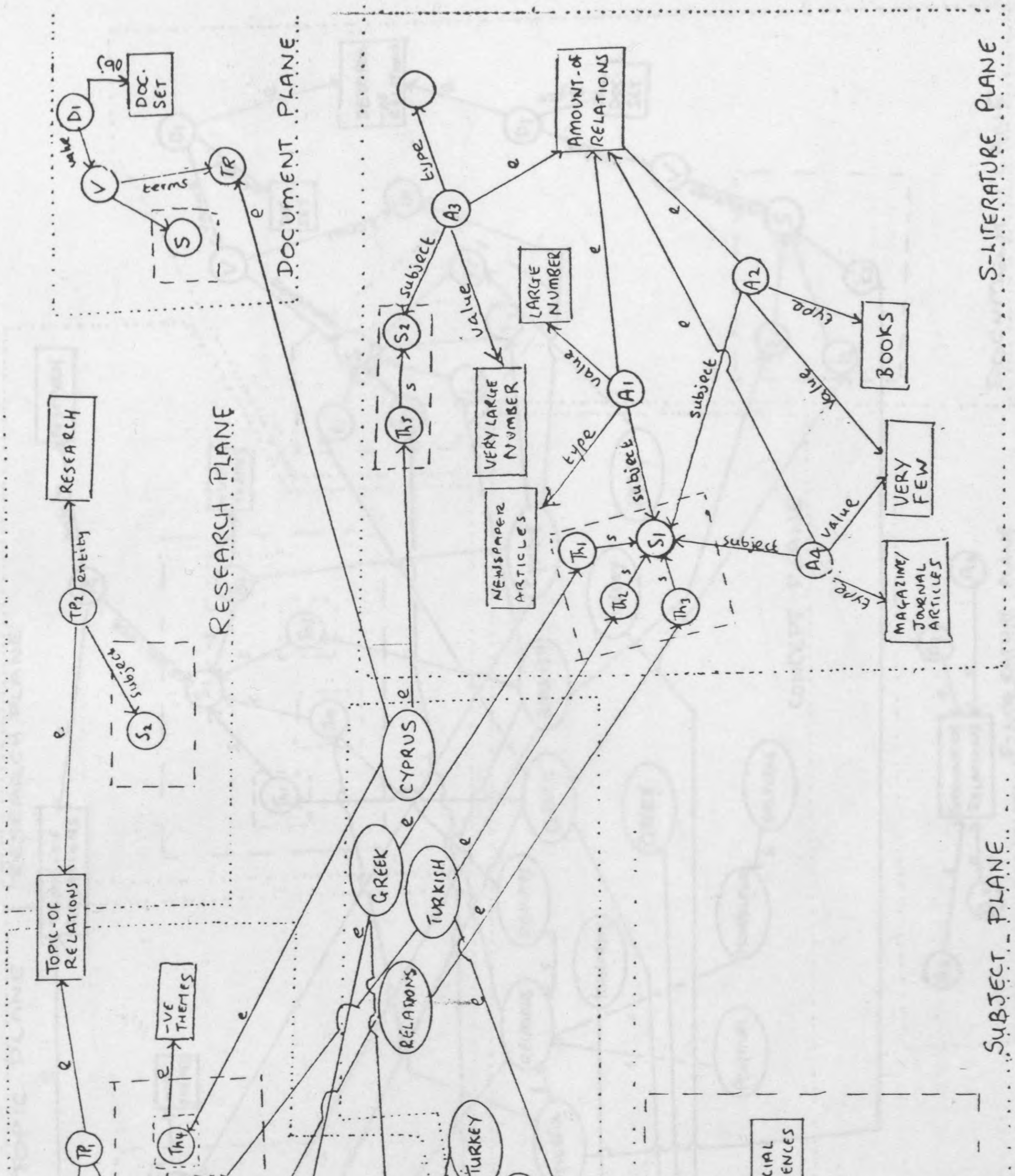
Interview 190684HBA Focus 7.



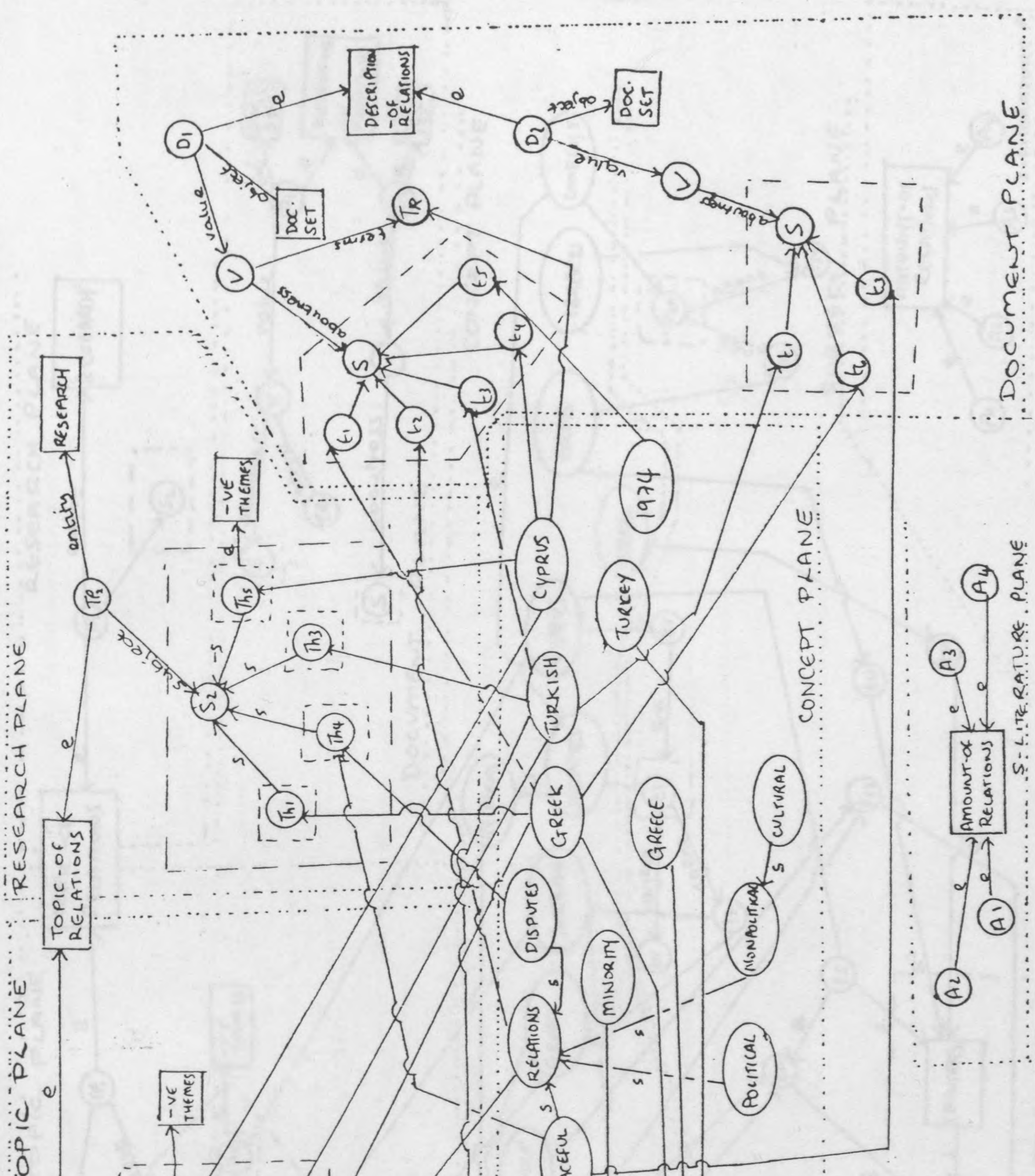


Interview 90684 HBA Focus 9.

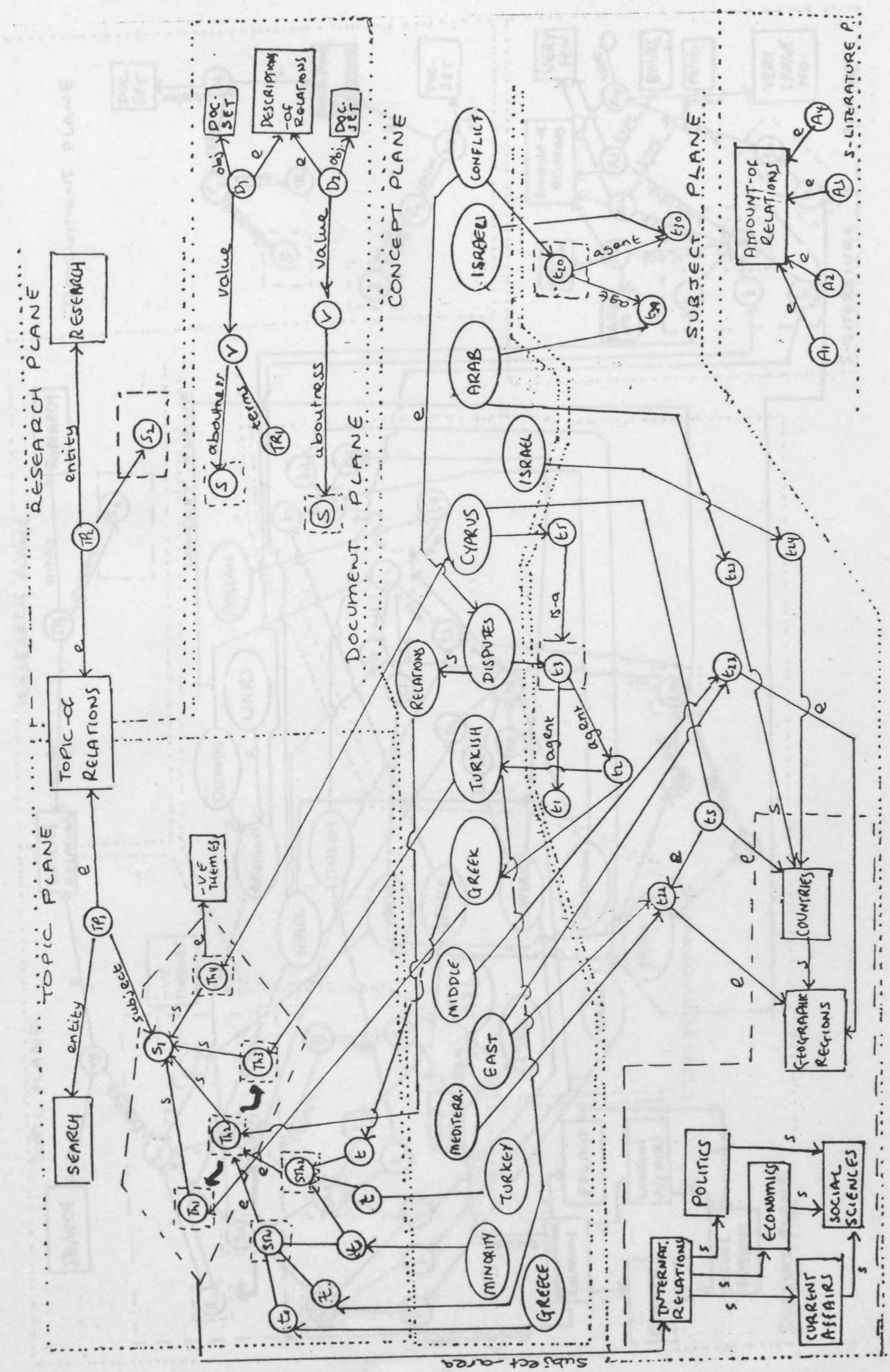
Interview 190684HBA FOCUS 10.



Interview 190684 HBA   Focus 11.



Interview 190684HBA Focus 13.



Interview 190684HBA Focus 15.