EFFECTIVE KNOWLEDGE TRANSFER:
A TERMINOLOGICAL PERSPECTIVE

Dismantling the jargon barrier to
knowledge about computer security

Agnes Maria Kukulska-Hulme

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

November 1993

1

SUMMARY

The research presented in this thesis is concerned with the terminological problems that computer users experience when they try to formulate their knowledge needs and attempt to access information contained in computer manuals or on-line information systems in the process of building up their knowledge. This is the recognised but unresolved problem of communication between the specialist and the layman.

The initial hypothesis was that computer users, through their knowledge of language, have some prior knowledge of the subdomain of computing they are trying to come to terms with, and that language can be a facilitating mechanism, or an obstacle, in the development of that knowledge. Related to this is the supposition that users have a conceptual apparatus based on both theoretical knowledge and experience of the world, and of several domains of special reference related to the environment in which they operate.

The theoretical argument was developed by exploring the relationship between knowledge and language, and considering the efficacy of terms as agents of special subject knowledge representation. Having charted in a systematic way the territory of knowledge sources and types, we were able to establish that there are many aspects of knowledge which cannot be represented by terms. This submission is important, as it leads to the realisation that significant elements of knowledge are being disregarded in retrieval systems because they are normally expressed by language elements which do not enjoy the status of terms. Furthermore, we introduced the notion of "linguistic ease of retrieval" as a challenge to more conventional thinking which focuses on retrieval results.

The empirical part of the research was carried out by means of a survey whose primary aim was to collect user queries for analysis. The domain of computer security provided the focus for the data collection, and IBM's AS/400 computer system the documentation to which the queries could subsequently be related. Several global categories of knowledge needs were distilled from the query data, along with a catalogue of specific needs. A number of new principles were formulated for the creation of indexing tools.

Key words:- indexing, terminology, information retrieval, adult learning, knowledge representation

2

Dedicated to Shrimp and Kitty

Acknowledgements

Contents

Chapter V:
Knowledge transfer – establishing a mapping

Appendices

List of Figures and Tables

Introduction

When one considers the century's most significant developments with regard to the written word, the electronic encoding of texts has to be classed amongst the most portentous. One striking effect of the move to site texts - and indeed the act of writing itself - within a computer environment can be seen in the way that the latent open-ended nature of texts has been released, with greater emphasis and scope being given to the activities involved in their elaboration and decoding. As texts become more flexible and interactive, so one of the major barriers which has traditionally separated spoken and written discourse becomes eroded. We are witnessing a significant change in the correlation between texts and other communication media, as well as a redefinition of the relationship between writer, reader, and the intermediary of the written word.

It may be supposed that the medium which carries any text might colour one's perception of a text's essential nature, and a computer environment can indeed have this effect. Depending on one's perspective, a text can be: the 'end product' of a process of thought; an externalised fragment of personal or collective knowledge; an interplay of intentions, choices, linguistic resources, conventions, and rhetorical forms, shaped in the process of interaction between a writer and a reader. It is not only a means of self-expression (the predominant feature of childrens' writing; Bereiter & Scardamalia, 1982), a means of communication (Beaugrande, 1980), and a "constituent of the context in which it is produced" (Lyons,

1981), but for some, a "social construct" (Barrett, 1989), which, given the right online networked environment, can be a focus for an active, social construction of knowledge through collaborative writing. It has, of late, been the object of detailed investigation (eg. van Dijk & Petöfi, 1977; Aarts & Meijs, 1986), has become the focus of text theory and text linguistics (Dressler, 1978; Beaugrande & Dressler, 1980; Turney, 1988; Schröder, 1991), and has been given special attention in a number of text-centred disciplines, including advanced and specialist language teaching (Nuttall, 1982; Crombie, 1985; Swales, 1990) and information retrieval (Kay, 1985; van Vliet, 1986). It has come to be analysed as a source of knowledge for artificial intelligence applications: natural language understanding (Schank and Abelson, 1977; Wilensky, 1978), question-answering (Katz, 1988), automatic translation (Sadler, 1989), and knowledge bases (Moulin & Rousseau, 1990).

As many an individual elevated to a higher level of appreciation and prominence, the text has been attracting an entourage of aides, which would enhance it or act as 'stand-ins' for the text. This has resulted in methods and tools for concordancing (Hockey, 1980; Sinclair, 1991), planning (Friedman, 1987, 1989), indexing and abstracting (Sparck Jones, 1971; Borko, 1978; Sharp, 1989), editing and readability analysis (Kincaid et al., 1981; Macdonald et al., 1982; Kieras, 1989).

We can also observe a re-appraisal of the space occupied by text, as its woven fabric becomes stretchable, so that one can peer through windows into 'hyperspace', or follow threads to deeper layers (Barrett, 1988; McAleese, 1989). It is

11

susceptible to repeated alterations in the course of
its creation, being revised iteratively on a
graphological, syntactical, semantic or pragmatic
level; it can be reconfigured to fit a chosen
structure, or reformed to suit requirements on a
screen or page.

Text-based on-line help is providing timely support
for computer system users (Burrill, 1986; Duffy et
al., 1989). In this way, on-line documentation,
front-ended by an appropriate interface, conceals
from view the forbidding volume of information which
users may have to access, and can provide a body of
reusable text to accelerate technical writing
(Buchanan, 1990). On-line, interactive text is a
source of expertise when experts are unavailable,
and a source of document building blocks when time
is lacking. It has the potential to help change the
traditionally poor image of technical documentation.

Technical terminologies are, in parallel, undergoing
significant advancements in respect of sheer growth
in volume and diversity, and in the development of
both the science of terminology and terminographical
analysis methods (Wüster, 1974; Felber, 1984). In
the course of the last twenty years, the task of
analysing and representing relationships between
specialised terms has been pursued with great
precision and vigour. Terminology is making a
contribution to knowledge based systems (Czap &
Galinski, 1987; Czap & Nedobity, 1990; Schmitz,
1993), and is itself benefiting from tools and
methods being developed in the field of knowledge
engineering (Skuce & Meyer, 1990; Skuce, 1993).
Recognising the importance of specialised
terminologies in the development of science and

technology, and in the international exchange of information and documentation, standardisation bodies have intensified their efforts to establish and propagate standard terminology collections. This work has acquired even greater importance with the realisation that "the changeover from information bases to knowledge bases requires a new approach in information, making use of systematic terminologies" (Felber, 1984;1).

The proliferation of specialised vocabularies has drawn attention to the problem of communication between specialists and lay people. Mapping correspondences between the specialised terms used by members of a given trade or profession, and everyday language, has been the subject of study in areas as diverse as medicine and religion. An early example of this is the study made by Louis Marie Raymondis and Michel Le Guern in the seventies (Raymondis & Le Guern, 1976) of the language of penal law; environmental law is the subject of more recent studies (eg. Hansjörg, 1992). Today, the rapid growth of knowledge in all specialisms and its wider dissemination through computerised information systems affords this type of research even greater urgency. It calls for an adaptation of methodological approches in the science and practice of terminology (Kukulska-Hulme, 1991). A more informal approach to the explanation of specialised vocabularies is also emerging (eg. computer dictionary by Williams & Cummings, 1993).

Progress in cognitive studies has yielded tentative descriptions of the complex cognitive processes of reading and writing, which go some way towards explaining the shape of their products: mental representations and written texts. Cognitive

psychologists have been exploring the psychology of
language (Hörmann, 1977; Johnson-Laird, 1987), and
process models of writing have emerged from
psychological research (Hayes & Flower, 1980; Cooper
& Matsuhashi, 1983; Bereiter & Scardamalia, 1987),
as well as from investigations concerned with
improving teaching methods (Calkins, 1980; Bridwell,
Nancarrow & Ross, 1984). Reading and text
comprehension, too, have become the object of
research and modelling (Kintsch & van Dijk, 1978;
Kieras, 1982; Reiser & Black, 1982). Some of these
process models have been applied in the creation of
computer software for the teaching of writing
(Sharples, 1985, Friedman & Rand, 1989) and reading
(eg. Geoffrion & Geoffrion, 1983).

One problem which writers and readers share alike is
the logistical impossibility of satisfying the
information needs of all eventual readers by the
production of a single text. This basic problem of
the explosive relationship of 'one to many' is
waiting to be counterbalanced by a means of
transforming a ubiquitous text into an
interpretation by, or for, an individual. A
computerised environment promises to be able to
accommodate this need. The patient individual may be
prepared to spend time interpreting texts, but
overall the cost of the interpretation effort is
running high. Training in new domains of knowledge,
particularly related to new technologies, is
continually on the increase, with high formal
training costs and the scarcity of suitable manpower
swaying decisions in favour of self-tutoring and on-
the-job learning. In this scenario, specialist
information which may be obtained from technical
manuals increases in value, and manuals come to play
a dual - reference and training - role. In

14

particular, knowledge about computers is no longer
solely in the hands of specialists, as computer
systems pervade all areas of life. Learning through
reading becomes a critical faculty, and technical
writing becomes a valued skill.

As writers' skills become augmented by new
computerised tools which can analyse and provide
assistance with their writing, so too readers'
expectations of a higher standard of documentation
are growing. If a higher standard is not delivered,
or if a reader has personal misgivings about the
precise meaning of parts of a text, there is still
potentially the opportunity to look at an on-line
manual through a comprehension tool which would aid
and support individual learning and understanding.
Admittedly, in the printed format, it has always
been possible to let readers make their own
decisions about which parts of a text to read, the
sequence in which they are read, and the reference
books or glossaries that are consulted. It is also
true that alternative representations in the form of
graphs, charts or illustrations have commonly been
used by writers and editors to make good the
deficiencies of text (eg. Hartley, 1985). What is
new in an on-line environment, however, is the
notion that the reader can query a text in a way
which will make the text easier to understand. The
emphasis here is on the fact that a *text needs a
reader* in order that the communication process might
be complete; appropriate computerised tools can make
it possible for readers to assume conscious control
of the way they 'complete' (i.e. interpret) texts,
for example by providing improved indexing and
searching aids. This allows us to surmise that
interpretations will become more accurate.

What is of particular import here is the recognition
of the inherent *imperfection of texts*, and at the
same time the realisation that readers are not
powerless when faced with this fact. If, as readers,
we can unravel the obscurities which have been
unwittingly twined with the thread of a text, we are
better equipped to avoid misinterpretation.
Computer-assisted reading can serve both readers
(reading for comprehension) and writers (reading
with a view to editing for a given readership).

The problems associated with texts are many. To
begin with, a text must be seen in the context of
other writings treating the same or similar subject;
for example, a technical manual may be one of a
series of manuals covering related aspects of the
operation of a machine. In this sense, a given text
is incomplete, and even if no other written source
exists, there may be a body of personal knowledge
and experience which is not easily captured by the
written word. Furthermore, practical constraints in
terms of available space may preclude a
comprehensive treatment of a subject. Next, there is
the writer's skill and knowledge, which may be
limited, and the same applies to the reader's prior
knowledge and reading skills. The relationship
between the reader and the text is then a
compromise, and necessarily imperfect. Finally, the
words themselves are not just a means, but also
potentially an obstacle, to textual communication.
The success of communication depends greatly on
writers' and readers' mutual understanding of terms.

Readers have traditionally looked to dictionaries or
glossaries to clarify the meanings of terms
encountered in texts. While recognising the fact

that electronic reference tools can provide an
unprecedented wealth of information in a
computerised reading or writing environment, it is
important to consider how they are best constructed
and employed. Text processing makes it possible for
dictionary information to be complemented or
replaced by data gleaned from texts. For a given
comprehension problem, the most relevant information
might be that relating to previous and subsequent
uses of a term in that text. This is where the
concept of a lexical reading and writing aid becomes
applicable (eg. Kukulska-Hulme, 1990a). Writers will
want to refer to previous uses of a term in their
own document, in related documents or in standards
or guidelines. Readers can build up a picture of a
concept by looking at the various contexts of usage
of a term. A lexical analysis program can in
addition make significant use of information from a
reference tool, and may feed information into it.

Machine-readable dictionaries are now a vital
component of computer-based natural language
processing systems, and their important role is
being appreciated more and more. As part of such
systems, their content may be intended for internal
use, or as part of an interface module in the
dialogue between user and machine. In instances
where they are part of writing or reading systems,
computer-based learning, or computer-aided
translation, the user may have direct access to the
lexicon, where the organisation of entries can be
more or less transparent. This immediately raises
the question of whether meaning representation
should differ substantially for these different
ends.

Dictionaries which have a rich information content

may qualify for the denomination 'knowledge base', in the sense that they provide the knowledge for lexical decisions; this is a fairly loose way of using the term 'knowledge base', which has a more precise definition within computer science. As knowledge based systems come to the forefront of computing, the reexamination of knowledge and its representation gains a significant place. Sustained by advancements in the simulation of intelligence on computers, the philosophical debate concerning knowledge and its relationship to language and logic proceeds apace (Thomason, 1989; Fetzer, 1990). Clearly, there is not a direct link between results in knowledge representation for computers and the representation of knowledge for humans, but the research has opened up a whole new field of exploration which is already contributing to an improved understanding - if only by way of contrast - of how humans process the knowledge which is to be found in texts.

Some form of intelligence can be planted inside systems, and also at the human interface, where 'intelligent' often means 'cooperative', or 'in tune with the human task'. There is now widespread recognition of the fact that it is not enough for a system to produce the right results, nor to give the user choices - it has to present them effectively, too. Thanks to a growing body of research concerned with the human-computer (man-machine) interface, enriched by investigations into the processing of visible language (Kolers et al, 1979), and by the rapid development of alternative interface technologies (icons, touch-screens, windows, mice, voice), computer users can enjoy a feeling of control which they had hitherto been denied.

As the interface between linguistics and computer science grows in scope and more cognisant bonds are formed between the disciplines, there comes the realisation that it is not sufficient to borrow theories or methods outright, but that they must be reviewed in the light of new requirements. Linguistic descriptions devised for linguistic ends are not directly applicable to computer science - and vice versa. What is more, as soon as we include human users in the computer-based language processing environment, the cognitive dimension becomes imperative. Given that each of these disciplines is currently making great headway, the challenge in research which straddles them is to strike a balance between using each discipline's most recent advances and selecting the right 'mix' of approaches for productive results.

In recent years we have seen significant developments also in the field of educational studies, with an emphasis on interactive and learner-centred learning. The learning of foreign languages now takes place in a 'communicative' context, and multimedia environments are making active learning more widespread. As reading is still one of the basic components of learning, it too must become much more interactive. The design of technical documentation will have to take account of these changes; on-line documentation presents new opportunities in this respect.

Texts may not be the most effective way of communicating specialist knowledge, but the fact is, they are still the most common way of reaching a wide and varied audience. The problem of 'knowledge transfer' or the effective communication of specialist knowledge through the medium of texts

depends crucially on the question of whether texts can be made to respond to their readers' needs. It may be a question of manipulating the representation of knowledge, of being able to transfer knowledge from one representation into another in an automatic or semi-automatic way, on the reader's demand, in order to transfer it more effectively to the reader. Language, being conventional, depends upon the interlocutors' mutual understanding of the relationships between meanings and forms; in specialised language, that understanding is more precarious than ever, so that there is a real danger of language obstructing comprehension. To counteract this situation, we can seek out these obstructions, analyse their nature, and give readers a means of seeing beyond the linguistic representation in a text.

Special subject knowledge is not confined to specialists; computer users are keenly aware of this. In order to satisfy the requirements of their job, which itself may be highly specialised, users have to take on board the special knowledge which will enable them to handle computers in an effective way. Often with limited training, they tackle applications as best they know how, turning to colleagues and printed or on-line sources (eg. manuals, help facilities) when problems occur. Many users will - consciously or not - endeavour to keep their computer knowledge to a minimum, to the level required to handle only the task in hand. The consequences are significant: for example, *computer security*, a domain of knowledge which cuts across computer applications, software and hardware systems, is an area of computing which many users, sometimes at their peril, ignore. This important area has been selected for the present study of the

retrieval and comprehension problems arising in
computer manuals. It is intended that a clearer
picture should emerge of how computers can
themselves be used to help solve the problems which
they help to create.

Comprehension in a terminological perspective

A.Theoretical reference points

1. Comprehension as knowledge-building

The research described in this thesis is broadly concerned with *the comprehension of computer manuals,* and more specifically, with *the language of information retrieval* from manuals. It is our view that any research dealing with human comprehension must take into consideration the very broad spectrum and variety of factors which come to bear upon that process. This does not, of course, preclude a narrow focus in the investigation of a particular phenomenon, and the present thesis is characterised by its progression from a wide perspective to a point where a particular aspect of comprehension - terminological choices in information retrieval - is scrutinised with only selected factors in mind. According to one author, comprehension can be defined very simply as "the opposite of confusion" (Smith, 1982;15); it is, however, a complex phenomenon, even "an internal, subjective process that is in general not open to external observation" (Carroll, 1972:5) - a view characteristic of the early days of comprehension research - making it difficult to define. Carroll (1972) states that comprehension contains at least two stages: (a) apprehension of linguistic information and (b) relating that information to wider context, and he makes the point that comprehension may be impeded by

the fact that messages are degraded in various
ways - by transmission failures, by "unclear" or
"poor" writing. Comprehension involves not only
understanding "the words and grammatical structures
of a message as linguistic symbols, but also taking
account of those knowledges, facts, or ideas that
underlie the message but are not explicitly built
into it." (Freedle & Carroll, 1972:360). More than
twenty years on, these remarks still hold true, and
there is an abiding difficulty in trying to define
'comprehension'. From de Beaugrande's perspective
(1988), to understand something "is to situate it in
a network of relations that constrain its properties
and connections" (de Beaugrande, 1988:10). It may be
apposite to say that comprehension covers the total
chain of events which begins with the writer's skill
and decisions having a bearing upon the product of
writing and its chances of being properly
understood, includes the reader's decoding
strategies, and ends with the assimilation of
knowledge, enabling the reader to make decisions
about action, based on new knowledge gleaned from a
text.

2. Terminological processing and knowledge transfer

It is not our intention here to offer a complete
psychological model of the comprehension process
(see eg. Freedle & Carroll, 1972; Kintsch & van
Dijk, 1978; Sanford & Garrod, 1982). There is,
however, a valuable contribution to be made towards
the investigation of linguistic phenomena in that
process. Even more importantly, there is a new line
of inquiry to be pursued which concerns the
processing of *terminological* information by readers.
The label "socioterminology", used in francophone

countries (fr. *socioterminologie*, eg. Gambier, 1987; Delavigne & Guespin, 1992) is useful here in highlighting the "user" aspects of terminology. There are two major reasons why terminology, with a particular orientation towards users, is so important. Firstly, in today's culture of specialisation, the terminological barrier to communication and understanding is a visible hurdle. Secondly, the advent of widely available computerised retrieval systems has brought into sharp focus the question of which terms are best used for information retrieval. There is now a much firmer conviction that "information retrieval systems are fundamentally linguistic" (Blair, 1992:200), prompting a reassessment of language theories and philosophies (eg. those of Austin, Searle and Wittgenstein) in this light. Successful retrieval is one of the keys to comprehension, conceived as knowledge-building, in a specialised domain, but comprehension is also a condition for successful retrieval. Ideally, the person who needs to retrieve information understands perfectly the domain and its terminology. In practice, this understanding may be partial or inaccurate, and the terms used when formulating a query may not belong to the domain. What is true of computerised systems is also true of conventional retrieval by way of indexes and headings, with one significant difference being that computerised retrieval is more 'remote' from the text, so that there is increased scope for manipulating the terms of a query to match those of the text.

As has been pointed out in the introductory chapter, the fields of terminology and knowledge representation are currently flourishing, and fruitful results may be obtained from research which

brings together these seemingly disparate fields. We also see a need to make more widely known the methods and findings of the field of terminology, better appreciated in mainland Europe and in Canada, but not yet sufficiently well implanted on British soil. Moreover, the design of computer-mediated communication and learning is sometimes seen as the prerogative of computer scientists, with linguists playing a supportive rather than a central role, and with apparently little heed paid to the communicative dimension of language. The research presented in this thesis represents a terminologist's view of text comprehension and the textual communication process, and incorporates research findings from a number of disciplines besides terminology: semantics, text linguistics, cognitive psychology, information science, epistemology, the philosophy of science, education, computer science (and specifically knowledge engineering). It incorporates also a description and analysis of the domain of computer security. Our particular orientation brings into focus the contribution that terminological study can make to the understanding of human processing of specialised texts, *making an explicit link between terminology and knowledge* in the context of comprehension. Above all, we are exploring the role of *terminology as a means of information retrieval*, and noting in particular that *domain knowledge can be acquired by repeated information retrieval*, whereby a picture of the domain is built up gradually. Information retrieval is thus part of the process of 'comprehending' the domain. Obviously, comprehension problems are not just terminological, or even linguistic, in nature, but information retrieval is necessarily concerned with terminology.

As has been pointed out by Alberico & Micco (1990), it is extremely difficult to construct models of information searching because "the interpersonal encounters and the information sources which are part of both reference and online searching are unpredictable and dynamic" (Alberico & Micco, 1990:65). They also point out the paradox of the searching process: trying to describe the unknown in terms of what is known. Often, the formalisation of a user's information need leads to a compromise, taking into account the chances that the information system will satisfy the need (Taylor, 1968).

Although the research undertaken here could be classed as belonging to the field of knowledge representation or *knowledge engineering*, it has been our intention to consider the specialised text as a means of knowledge *transfer*. This is to say that we are putting a special emphasis on the dynamic processes of knowledge encoding and decoding which have the text as their focus, not forgetting also that readers may infer from a text as well as decoding it. It is our view that a text does not represent knowledge in a static way, but that there is an interaction between the symbolic representation in the text and the knowledge which is activated when symbols are encoded and decoded.

'Translation of knowledge' could be used as a close alternative to 'knowledge transfer', if it did not inherit the rather narrrow connotations of the term 'translation', when this refers to foreign language translation, or to the conversion of computer programs into a different programming language. Current usage apart, the term's meaning could be extended to cover any change over from one system of representation to another. Amongst the ways that a

26

text might be 'translated' in this broad sense are operations to condense or amplify it, paraphrase it, explain it, interpret it, represent it in symbolic form, or to change it from one language register into another. 'Transposition' is another term sometimes used in these contexts, though this also has a special meaning in encryption (message scrambling to prevent understanding). Generally, such conversions are attempts to keep – to 'carry over' – the text's conceptual framework (as far as possible) intact, while changing its outward manifestation. Sager (1991) uses "text modification" as a global term, and talks about "derived" or "dependent" texts, encompassing foreign language translations, abbreviations, extraction of sections and modifications of form, with the proviso that "no dependent text can introduce anything – information or argument – that is not in the original" (Sager, 1991:252). The rationale behind these transformations is to make a text more accessible to readers; the same applies when the 'reader' is a machine (cf. the notion of an 'interlingua' in machine translation).

'Knowledge transfer' is an expression routinely encountered in the literature on knowledge acquisition for knowledge-based systems (eg. Gaines & Boose, 1988), where it refers specifically to the transfer of expertise from expert to knowledge base via a knowledge engineer. It is also an expression which is currently being used to designate the global transfer of scientific or technical knowledge, often to developing countries (eg. Nedobity, 1990). These are not the specific meanings intended here. We are using the term to denote the transfer of domain knowledge from computer manuals to their readers, whatever their profession,

geographical location or economic standing. It is
worth clarifying, however, why we speak of
'knowledge transfer' rather than 'information
transfer'. "When information is organized into
bodies of meaningfully interconnected facts and
generalizations, it is usually referred to as
knowledge" (Gagné & Briggs, 1979:79). The computer
manual is being considered here as a learning tool,
so that every instance of information retrieval from
the manual contributes to the process of building up
knowledge structures in the mind of the reader:
"Every search for information is a learning process"
(Alberico & Micco, 1990:17). Scardamalia & Bereiter
(1993) have pointed out the advantages of the
constructivist standpoint, favouring knowledge-
building above mere transmission or reproduction.
When we examine the domain of computer security, we
can see that it has two aspects: preventive and
remedial. Prevention involves risk assessment and
planning, and this must be done on the basis of
knowledge of the domain, rather than on the basis of
separate pieces of information. Remedial action
resulting from security breaches means tackling the
problems associated with that breach, and also
implementing corrective measures to ensure future
security. Any information retrieval from a manual
dealing with security issues thus takes place in
relation to knowledge of the domain, and provides
the inputs to a process of reasoning about the
domain.

'Knowledge transfer' is closely related to
'comprehension' and 'learning'. Both these related
terms, however, focus attention on texts and
readers, whereas we would rather highlight the
process of textual communication, with particular
emphasis on the identification of knowledge.

'Knowledge transfer' is, terminologically speaking, our 'preferred term'.

3. The notion of effective transfer

It is interesting, and also necessary, to explore the notion of the *effectiveness* of knowledge transfer because in a business environment - the environment we are considering primarily -, problem solving with reference to manuals has to be productive within time constraints. Hart (1989) defines an expert's effectiveness as using knowledge to solve problems, with an acceptable rate of success. In our context, the 'effect' of knowledge transfer would show itself in the correct implementation and maintenance of computer security as a result of consulting manuals. In the long run, this could be measured in terms of how secure the system turns out to be. In the shorter term, we can consider how knowledge transfer can be made more effective by improving the process of retrieval. Furthermore, effective *knowledge transfer* is not exactly the same as effective *retrieval*. Retrieval performance can be measured in terms of recall and precision - the number and relevance of references resulting from a query (though relevance is being challenged as a chief criterion for information retrieval by other criteria such as value, utility, impact, information quality, source traits, document traits, etc. - see eg. Saracevic, 1992; Barry, 1992, and similarity measures, eg. Wilbur & Sirotkin, 1992). Effective transfer is also in the realm of understanding - the speed, the ease, the quality of understanding (see also Kukulska-Hulme, 1990b). Knowledge transfer is, in this sense, a broader term than knowledge retrieval. The key question is: how

easily can the individual manage to reach a point of resolution, a satisfying response to a query, a state of understanding ? The emphasis is thus firmly on 'linguistic ease' in the process of retrieval, and on the feelings of intellectual satisfaction and growth in knowledge that come from a query being correctly matched to an answer. It is a question of effectiveness over time, since knowledge and understanding are built up over time. The difficulties in wishing to evaluate this kind of effectiveness are explored later.

One possible definition of effectiveness would be a minimalist one: *a manual is proved to be effective in its knowledge transfer function whenever a reader manages to access with ease relevant knowledge within its texts*. It is then sufficient to show that, from a linguistic or terminological standpoint, a clear route exists from a reader's query, via an entry point to the text, to the item which - in some measure - answers the query. The goal then is to ensure that 'clear routes' are the norm rather than the exception, and to eliminate the situation where a terminological mismatch means that a search is futile or frustratingly lengthy. If this criterion is met, then at least we have created the conditions for effective knowledge transfer to take place.

4. The special role of terminology

We take the view that a text does not represent
objective knowledge of a domain, but that it is a
representation of a writer's subjective knowledge of
that domain. This has implications for the concept
of ambiguity in relation to knowledge transfer:
ambiguity can result from a writer's understanding
or expression of a knowledge element of the domain,
or from a reader's interpretation. Furthermore,
writer and reader do not necessarily share the same
specialised language. The key question, then, is how
does the reader elicit knowledge from the writer,
the 'expert' on the subject, when communication is
mediated by a text ? The reader's needs will be
expressed through the rhetoric of questions, which
are then reduced to terms at the 'entry points' to
the text: the table of contents, index, and various
prominent headings in manuals, or the query
interface of a computerised retrieval system. As
questions range from the general to the specific,
their vocabulary will vary accordingly, and will
contain words from the general language, as well as
semi-specialised and highly specialised terms,
perhaps from more than one discipline. The
vocabulary items become the 'fishing tackle' for
pulling out knowledge structures from the textual
knowledge base, and much depends on the correct
choice of tackle. This is where, traditionally, a
thesaurus might be used to aid retrieval, as it
suggests, or indeed prescribes, the use of certain
terms in favour of others. Its retrieval language

"... saves the searcher the intellectual
effort of bringing to mind closely related terms
which might improve the search. A good entry
vocabulary is another thesaurus feature which will
lead the searcher from highly specific natural

language concepts to the terms used in the system to represent those concepts."

(Aitchison & Gilchrist, 1972:5).

This is the point at which formal and informal terminologies collide, and the 'user friendliness' of the knowledge base is tested.

In this light, the role of terminology is decisive; sensitivity to this issue could make the difference between enabling access to knowledge and exclusion. Terms can be viewed as linguistic pointers to knowledge, as having "special reference within a discipline" (Sager, 1990:19), and very importantly, terms are conceptually linked with one another and physically juxtaposed in text: for comprehension, relations between terms - paradigmatic or syntagmatic - might be more important than the terms themselves. If that were so, those terms which were 'better at networking', i.e. had richer, more extensive links to other terms, might be the most productive or effective in the process of knowledge acquisition - a hypothesis which would need to be tested. However, we are not asking "Which terms are the best for knowledge acquisition ?", but rather, "How can we ensure a closer match between the terms actually used by readers for knowledge acquisition (retrieval), and the terms used by writers to express knowledge in a given domain ?".

5. Language and knowledge interrelated

The definition of 'knowledge' is philosophy's central and perhaps most difficult issue, and even the narrower notion of 'special subject' or 'domain'

knowledge presents us with substantial problems.
What is undeniable, nonetheless, is that knowledge
can be expressed at least in part through a
terminology, and we know also that a high proportion
of the words which make up specialised texts are
terms. Indeed, "texts have been identified as
special by statistical methods determining the
frequency of occurrence of terms" (Sager et
al.,1980:233). A text may not explicitly assume any
background knowledge on the part of its readers, its
writer may even insist on the fact that it does not.
It is partly a question of one's awareness of what
is assumed. In particular, a knowledge of the
general language must be supposed, and this cannot
be separated in all confidence from a knowledge of
the special language, since many words of the
language function in the two spheres. In addition,
writers have to gauge continually the technical
level of the vocabulary they are using, and to try
to predict how it will be understood, using their
experience and intuition as a guide. From this we
infer that the reader is faced with certain
linguistically motivated assumptions about
background knowledge which must be decoded in the
comprehension process.

B.Terminological investigation: aims, scope and
methodology

1.Relating terminology and knowledge to readers and
texts

The fundamental problem being addressed is that of
the relationship between a reader's usage of
specialised terminology, and the terms used in a
sample corpus of texts. Our hypothesis is that even

newcomers to a domain such as computer security already possess some knowledge about it because of the inseparability of specific domain knowledge from knowledge of related domains, of science and technology in general, of knowledge of the world, and more particularly - knowledge of language. These spheres of knowledge are also interlocked within texts. We therefore aim to specify the interdependence of the different spheres of knowledge, and to establish a mapping between the concepts and terminologies used by, on the one hand, the readers, and on the other, the writers, of a specific set of security texts. The ultimate objective is to suggest a generalised strategy or methodology for devising improved retrieval tools based on our findings. As the index to a manual is very often the point of entry (this is discussed later), we focus much of our attention on the index, and ask the question: "What (if anything) is wrong with the index ?" If the problems of knowledge transfer can be thoroughly researched and understood, the solutions to those problems will surely follow.

In a wider sense, the research aims to help redefine the relationship between readers and writers, with particular implications for a more flexible environment such as that offered by computers, i.e. on-line documentation and text retrieval systems. The approach taken is to try to define the problem, and suggest a methodology, which might be transferable to a practical information retrieval aid. The results could benefit either reader or writer - if the latter is willing to explore the comprehension problems of intended readers. This approach gives greater power to readers, who have hitherto had to resign themselves to playing a

34

largely passive role.

The motivation for this research stems from the realisation that a number of problems associated with computer manuals remain to be addressed. It is common knowledge that typically, manuals are resorted to when other sources are unavailable or lack the necessary knowledge, rather than being the first port of call. The difficulties in finding the right information are partly terminological in nature. Firstly, there are problems at the level of entry to the text (eg. via the index and table of contents). For example, we can consider some fairly simple words which spring to mind when one considers computer security: *insurance, disaster, piracy, fire protection, legal requirements, power supply, classified information, illegal sign-on, genuine user.* A perusal of an IBM manual on computer security showed that as far as could be seen from a first reading, none of these terms had been used in the manual, which suggested that a casual user might have difficulty in obtaining information from the manual. Subsequently, a small experiment was carried out (Kukulska-Hulme, 1992) to see whether terms and phrases from an introductory paper on data security (Bradburn, 1990) - see Fig. 1 - could be found in the data security section of the manual. The results showed that the degree of correspondence between the two was low - only about one word in five could be found.

Secondly, in manuals generally, there is a problem with the vocabulary of explanation, which may be unfamiliar to the uninitiated reader. Glossaries, when provided, often use and reuse the same inaccessible terminology of the text, leaving readers no wiser as to the meanings of terms. It is

this two-level terminological barrier which we set
out to examine, focusing particularly on the points
of entry to texts. In computerised information
retrieval environments, an intermediary might be
used to transform a user's query into one that can
be processed effectively by the system. But today's
users are increasingly keen to do their own
searching, and in a paper-based environment an
intermediary is not normally available.

corruption of data
deliberate sabotage
fraudulent manipulation of data
loss of data processing facilities
disclosure of sensitive data
potential risks
disruption
financial loss
breaches in data processing security
principal areas of risk
possible sources of protection
protective measures
means of damaging the computer or data
techniques to enhance security
degree of vulnerability
value of data
effects of loss or damage
disastrous effects
to reduce the probability of a fire
restrict the effects of a fire
method of recovering from damage
quick and effective recovery
replacement of equipment
preventative techniques
rigorous procedures
disaster recovery plan
adequate controls
restriction of access
positive identification of the user
backup copies of software
responsibilities of user departments
internal audit
effective day-to-day procedures
management of data security
precautions
excessive dependence on important individuals
fire and flood protection
to contain the hazard
exotic hazards such as tornadoes

Fig. 1 Examples of phrases from Bradburn's paper

Finally, it seems important to say that the whole
question of domain knowledge transfer - and its
vital components: terminology, knowledge,
communication, learning - is so fundamental to human
endeavour that any attempt to understand its
problems and suggest directions for improvement is a
worthwhile undertaking.

2. Computer security as an important domain

The readers under consideration are acquainting
themselves with the domain of computer security.
Computer security is a very important issue for
those whose business depends on the functioning of
their computer systems, and on the security of the
information which they contain. It is an issue which
cuts across different hardware and software systems,
so that in spite of individual peculiarities, basic
concepts prevail. However, as it is often felt that
the function of assuring security does not merit a
special dedicated post for that purpose, the role
may be taken on by people who have little previous
experience of computers or security. Madden (1990)
makes the point that "security is not just the MIS
director's headache; it is the reponsibility of
management, programmers, and end users too" (Madden,
1990:26). According to computer security consultants
Mike Rentell and Peter Jenner,

"... incidents involving computers where
victims have suffered serious, sometimes fatal,
consequences happen surprisingly often...most of
these losses could have been prevented had sensible
computer security precautions been in place."

(Rentell & Jenner, 1991:1)

There is evidence that in the computing world, increasingly more attention is being paid to the question of security. Viruses are a growing concern for all users, and a frequent subject for magazine articles (eg. "Fighting off infection", Lang, 1993). In 1984, the Data Protection Act was introduced, addressing a concern "arising from the threat which mis-use of the power of computing equipment might pose to individuals" (Guidelines to The Data Protection Act 1984). The Computer Misuse Act came into force in August 1990. This provides a means of prosecuting hackers, virus creators and others who deliberately seek to access or modify computer-held data or software without authority. Software copyright is covered by the Copyright, Designs and Patents Act of 1988, and watched over by FAST - the Federation Against Software Theft. The Department of Trade and Industry has developed a special scheme - the *Information Technology Security Evaluation and Certification Scheme (ITSEC)* - providing an independent evaluation of computer-based security products and the certification of products which meet appropriate standards; the United States has provided such a service since the early 1980's ("Orange Book" issued by the Department of Defense), and European standards are gradually being harmonised. There has also been an IT Security Awareness Programme, managed by the National Computing Centre (NCC), comprising information packs, consultancy, management briefings and courses (eg. "Keeping IT safe"), and extensive publicity in the media.

In spite of these initiatives, some attitudes are slow to change. Two comments from users noted on our survey questionnaires (details of the survey are given later) will serve to make the point:

"Senior management do not realise the importance of computer security nor the ramifications of illegal copying of software. The 'it won't happen to us' attitude prevails."

"Problem with bosses who appear to give little credence to the need to adhere to conventions and want everyone to have access to everything, in the name of efficiency !"

It is not only system operators who need to be helped to understand security issues but also their bosses. While computer security is the domain selected for the research presented here, it must be stressed, nonetheless, that the nature of the research and its findings means that the implications go far beyond this particular domain, and indeed beyond computing and computer manuals in general. This is discussed in the conclusion.

3. Sources of information on computer security

NCC Consultancy has published a 'Survey of Security Breaches Report' (NCC, 1992), detailing how organisations are establishing security plans, and dealing with legislation and disciplinary procedures; it also gives an overview of types of breaches encountered. There are numerous books on computer security, including those published by NCC Blackwell, such as Elbra's "Computer Security Handbook" (Elbra, 1992; see also Ellison & Pritchard, 1987; Smith, 1988; Simons, 1989; Roberts, 1990; Hearnden, 1990; and other books mentioned in later section on the domain of computer security). Articles appear regularly in academic and professional journals and the computing press (eg.

Albert et al., 1992; Burns, McDermid & Dobson, 1992), and in specialist periodicals, eg. Elsevier's "Computer Fraud and Security Bulletin". The ACM (Association for Computing Machinery) Special Interest Group on Security, Audit and Control publishes a quarterly "SIG Security Audit and Control Review", and organises conferences on the subject of security (eg. Annual Computer Virus and Security Conference). Computer security standards are published by standards organisations (eg. ISO, BSI, ANSI), and the Institute of Electronic and Radio Engineers in the UK. On-line conferencing systems (BIX, CIX) are another source of information in this field. On the terminological front, Butterworths has published a dictionary of security terms and concepts (Fay, 1987); the Department of the Secretary of State of Canada is a good source of up-to-date translation glossaries, particularly in the many related fields (eg. English-French Security Equipment Glossary, 1993).

Apart from this information of a general nature (i.e. applicable to computers in general), there is information pertaining to specific systems. For example, the magazine "IBM System User" regularly publishes articles on IBM system security. For IBM's AS/400 (Application System/400) mid-range system ('mid-range' being roughly equivalent to a 'mini' computer, in more widely accepted terminology), which has been chosen as the focal system for this project, there are several sources of information on security. The main written source is manuals; occasional articles may be found in the journal "News 3X/400", for IBM System34, System36, System38 and AS/400 users (eg. Conte, 1990); information about security is also available in the "on-line help" facility on the system. On-line help can be

accessed in several ways:-

- "User Support" - offers general information about the system, a database facility for logging questions, and "online education" - a library of interactive tutorial modules tailored to different types of user. The latter includes a module entitled "System Security Concepts"

- commands related to security can be accessed by calling up a list of command groups; 'security commands' form a distinct group, but commands relating to security can also be viewed by selecting a different group type and then making further selections, eg. to find the command for changing passwords:-

verb commands   -> change commands
                        -> change password command CHGPWD
       OR

subject commands -> security commands
                        -> password commands
                        -> change password command CHGPWD

Help panels are available to explain the commands, with 'extended help' if required, hypertext links available from highlighted words, and suggestions of related topics.

- one further way of getting help is through a "search index", which will accept words or phrases (whole sentences can also be typed, but the system clearly works on a word/phrase basis). This can be a good way of getting quickly to a chosen topic. It exhibits, however, the classic linguistic shortcomings of this type of retrieval system,

discussed in Chapter III.

4. Primary data: users and texts

The research described in the present thesis is
based on two main sources of primary data:
linguistic data elicited from a sample group of
computer users (in charge of an IBM AS/400) by means
of a questionnaire in which they were asked to write
down questions about security on their system; and
textual data from a sample manual on computer
security. The manual, "AS/400™ Security Concepts and
Planning - Version 2", 1st edition, April 1991,
document number SC41-8083-00, was published by IBM
Rochester (permission to use this manual for
research in printed and electronic form is
gratefully acknowledged). This manual was originally
only available in printed form, but was subsequently
also issued in CD-ROM disk format, with accompanying
BookManager retrieval software. It is the main
comprehensive manual on AS/400 security; as with all
manuals, updated and improved editions are being
regularly issued as new versions of the operating
system are released, but for this research it was
necessary to refer to a single edition. Security
issues are also dealt with in IBM publications
related to AS/400 Communications, OfficeVision/400
software, operations, the application programming
interface (API), and programming and utility
security. Further details on the user survey are to
be found in Chapter IV.

[™ Trademark of the IBM Corporation]

5. Evaluation criteria for effective transfer

At this point we must come back to an issue raised
earlier about the evaluation of the effectiveness of
knowledge transfer. As noted, knowledge transfer is
to do with understanding, with the emphasis on
'linguistic ease' in the process of retrieval, and
on feelings of intellectual satisfaction and growth
in knowledge - the confidence of knowing. An
evaluation of these largely (though not entirely)
subjective outcomes would be a separate and
substantial undertaking, and although our research
can postulate the likelihood of an improvement in
the effectiveness of knowledge transfer conceived in
this way, it stops short of actually testing that
particular hypothesis. What is more, the evaluation
of long term system security, as a result of
knowledge transfer, might prove a better criterion,
and though the effect would clearly be secondary, it
should show the vital link between knowledge and
action (using knowledge effectively). Awareness and
knowledge can contribute to effective security, as
can adequate technology, but there is one other key
factor, commitment: "Technology can and should be
used to preserve security but it will depend on
human commitment to be effective" (Elbra, 1992:1).
The problem is complex, but by limiting our scope,
we are able to concentrate on the detailed
terminological and knowledge-based investigations of
retrieval problems which are the hallmark of our
research in this sphere. The essential criterion,
then, is successful look-up: finding what one is
looking for thanks to a mapping between user needs
and access devices. However, "the critical issue to
be examined...", as Itoga (1992:330) put it,
advancing an 'alternative framework' for mapping
information needs in communication, "... is not how

to describe human information seeking behaviour, but
how to understand another person's information
needs". The need for our user-oriented, language-
oriented approach is confirmed also by Smeaton
(1992) when he writes about the current limited role
of Natural Language Processing techniques in
information retrieval research:

"... This role does not really address issues
of retrieving information for users based on the
language used in queries or in texts. Fundamental
issues and questions dealing with the notion of a
retrieval model and document relevance will need to
be integrated with what NLP techniques have to offer
if really significant progress in retrieval
effectiveness is to be expected."

(Smeaton, 1992:277)

6. Progression in stages in a multifaceted approach

The research path deemed to be the most appropriate
here is one that takes as its starting point a
review of the fundamental notions of "knowledge" and
"domain knowledge" in relation to computer security,
then provides the opportunity to discuss knowledge
transfer environments, before moving on to an
empirical investigation of users' knowledge needs.

The particular stages of the research may be
outlined as follows:

(1) *Discussion of theories and sources of knowledge*;
presentation of a catalogue of knowledge types, with
particular reference to the relationship of
knowledge and language; overview of approaches to
the organisation and representation of knowledge -
individual units and fields of knowledge, and
especially lexical and terminological

configurations.

(2) *Exploration of the notion of 'domain knowledge':*
what is domain knowledge, how is it organised and
represented, what are the ramifications of its
transfer to a person from a written source ?
Characterisation of the domain of computer security
in terms of knowledge representation and linguistic
expression.

(3) *Specification of the means of access to
knowledge in texts:* how is knowledge encoded by
writers and communicated to readers ? What are the
real needs of readers in particular situations, and
how do they access texts to fulfill those needs ?
What specific access mechanisms are there in texts ?
How is knowledge assimilated ? What specific
communication and access problems are there in the
domain of computer security ?

(4) *Review of existing environments* which aim to
facilitate the transfer of knowledge, accompanied by
an evaluation of computer-assisted retrieval and
learning, and discussion of the value and
effectiveness of self-instruction from manuals.
Specific discussion of retrieval on the IBM AS/400
system.

(5) *Empirical study to establish readers' knowledge
needs,* with reference to knowledge types sought and
the use of terminology to express specific needs,
using a questionnaire. Creation of a database of
queries. Analysis of reader profiles.

(6) *Qualitative analysis of knowledge needs
expressed in the query data:*

45

Lexical and terminological analysis: by grammatical category, with discussion and further semantic and knowledge-based categorisation

Intentional and rhetorical analysis: query clustering by knowledge-seeking purpose and rhetorical form; discussion of rhetorical difficulties and the question of ambiguity

Overview of knowledge types and dominant knowledge needs: summary of needs as revealed by the qualitative analysis

(7) *Quantitative analysis of source preferences and expectations*, with a comparison of the two.

(8) *Characterisation of the computer security manual*, with particular reference to its organisation, readership, and access devices. Special study of its index as a point of entry, in relation to the needs established earlier.

(9) *Summary review of the relationship or mapping between readers' needs and the access facility of the manual*: what are the problems which hinder knowledge transfer ?

In summary, we have set out to examine the relationship between the knowledge needs of readers, as expressed through their language, and the knowledge expressed by writers within the confines of the computer manual. We would like to see terminology as a facilitator of knowledge transfer; it is hoped that our conclusions and proposals will help bring this about. We begin by exploring, in the next chapter, the notions of 'knowledge' and 'domain knowledge', with special reference to the domain of

computer security.

Chapter II

The nature of domain knowledge

A.Theories of knowledge

1.The inference/experience dichotomy

In order to gain a better understanding of the
knowledge representation issues being tackled in
this thesis, we begin with an overview of the basic
distinctions that have been made within the theory
of knowledge by those who at one time or another
have set themselves the task of trying to ascertain
what is to be understood by 'knowledge'. As
epistemology is a general science, it is equally
important to examine the relevance of its
propositions to the specific fields of computing and
computer security, and to determine the distinctive
features of the latter with respect to the substance
and organisation of knowledge.

The essential search for the foundations of
knowledge has over the centuries been embodied in
two contrasting philosophies, one hailing reason
(Descartes), the other experience or perception
(Locke, Ayer, Russell). Both are centred on the
individual person's self-access to knowledge.
Perception may be the ultimate source, but it would
seem that some knowledge can also be gained by way
of inference, and memory has a supportive role in
this respect. The empiricist would not accept that
human reason or thought can itself produce new
knowledge; mathematics and logic are, however,

considered to be exceptions. The a *priori* knowledge possible in these disciplines is in contrast with *a posteriori* knowledge, the more common type. Noninferential (a posteriori) knowledge may be attained with or without observation (as in the case of knowledge of our sensations).

It is partly in response to the skeptic's doubting of the very possibility of knowing that knowledge has been formulated in propositions, the truth or falsity of which could be subjected to verification. A priori knowledge can often be verified on the basis of analytic definitions ('a crayfish is a crustacean') or synthetic a priori truths (eg. 'every event has a cause'), but if something is only contingently true, it should be established by experience. Verification is particularly important in the context of knowledge-based computer systems, where incorrect knowledge will lead to incorrect inferences. Knowledge is bound up with the criteria of truth and meaningfulness; the truth-value of a proposition could change with time; in fuzzy reasoning, there are degrees of truth; in illocutionary acts (Austin, 1975), the true/false distinction does not apply. Knowledge must furthermore be considered in relation to the notions of certainty and confidence: the "first degree of factual knowledge" can be knowing, but not having the confidence of knowing (Pears, 1972). In spite of this element of uncertainty, for most philosophers, knowledge is not to be equated with beliefs or states of mind: states of mind are changeable, beliefs can be false. It is to be noted that the scope of knowledge is wider than that which may be obtained as a result of one's own individual powers (Hamlyn, 1970). Hence the need for both spoken and written communication.

For knowledge engineers, knowledge may be distinguished from expertise on the grounds that: "... experts are experts because of what they are able to do with their acquired knowledge" (Hart, 1989:15), or in a similar vein, expertise is knowledge plus inference, and ultimately, *"knowledge is concerned with action ... effective use of knowledge leads to the formation of plans of action"* (Graham & Jones, 1988:21). However, it is important to add that, as well as having the ability to act, experts have the ability to theorise, and so to further knowledge, to refine it. And since the more we know, the more we realise there is much we do not know, an expert's knowledge develops into *value judgment*: what is worth knowing, its relevance to a given situation, the probability that something will occur, what needs to be discovered. In this sense, if we equate belief with judgment, we can accept an expert's belief as being knowledge, even though belief or judgment can change. However, it is also possible to identify belief with "received knowledge", and judgment with "experiential knowledge."

2.Sources of knowledge

Knowledge "... has two important characteristics:
(a) it is modified and extended by experience, and
(b) it is linked in a number of ways with previously acquired knowledge" (Last, 1989:115). Experience is an important aspect of knowledge acquisition, for even if it is not the primary source of an element of knowledge, it may come in at a later stage, altering and supplementing knowledge in some way. In everyday language, experience amounts to successes and failures - we talk about learning from

experience. Knowledge is thus tested in a situation, and modified as a result - its major characteristic is change.

We also sense the importance of 'hands-on' experience when it comes to learning something we will have to imitate, and observation is often better than explanation. First-hand experience seems superior to second-hand knowledge, though it may be coloured by subjectiveness. To take the point further, we can ask whether there are areas of knowledge which may be inadequately - or not at all - acquired through language. It seems that we are questioning here the adequacy of language as a means of knowledge representation; if the representation is inadequate, knowledge transfer must be imperfect by implication. This is a problem well known to those who build expert systems, trying to capture and represent 'expertise'. Belkin, Brooks & Daniels (1988), who have used the technique of discourse analysis of user-intermediary interactions in a situation of computer-based document retrieval, list the standard knowledge elicitation techniques as being interviews (informal or structured), verbal protocol analysis (recordings of experts thinking aloud), and observational studies. Experience has shown that some aspects of knowledge may not be open to introspection or verbalisation:-

"... people's awareness of their own mental processes is rather limited. The proceduralization of knowledge and automatization of cognitive skills that accompany the development of expertise, serve to make expert thinking even less accessible to introspection."

(Slatter, 1987:33)

Moreover, we have to take into account the

limitations of natural language: "Natural speech is marked by part-sentences, contradictions, omissions and repetitions... The use of words is often inconsistent and imprecise" (Hart, 1986:61). Gaines (1988) makes a related point when he states that expertise "may not be expressible in language. An expert may not be able to transmit the expertise explicitly because he is unable to express it", and "... expertise may not be understandable when expressed in language. An apprentice may not be able to understand the language in which the expertise is expressed" (Gaines, 1988:4). On the other hand, the most characteristic feature of discourse in natural language is "its ability to deal with incompletely and inexactly expressed concepts and to resolve and assimilate contradiction" (Graham & Jones, 1988:77). These linguistic problems have long stood in the way of expert systems becoming widespread in use, since one of their aims is to hand on expertise to the less experienced: it is not enough to represent knowledge, you also have to communicate it well.

Given that in some situations written documents are the only available, though secondary, source of knowledge, we have to accept and bear in mind their shortcomings, including their fragmentary, incomplete nature. Texts represent fragments of knowledge, but they are also its *focus*: they have a recording function, they serve as a canvas for working out ideas. Their organisation and specific function - whether didactic, descriptive, narrative, expository, synoptic, encyclopaedic, enumerative, argumentative, legislative and so on - harnesses elements of knowledge and imposes a structure. To this, the writer adds personal selection and idiosynchratic ordering of knowledge elements. The coherence factor of any piece of discourse can make

it difficult to extract one element without losing
the ties necessary for understanding. Also,
understanding depends on where you are in the text,
and is likely to be compromised if a reader did not
start at the beginning of a sequentially organised
text. For fast retrieval of knowledge, a thematic
organisation is normally preferable to a sequential
one, but even then there is always implicit
reference between themes. Even reference books such
as encyclopaedias and dictionaries cannot delimit
the scope of their entries in any final way. To a
large extent, it is up to the reader to use skill
and judgment to reduce the volume of information on
offer and to select only that which is relevant,
useful, or appropriate in other ways.

3.An overview of knowledge types

A basic typology of knowledge may be derived from
the grammatical structures of the verb 'to know' -
hence *'knowing that'* (*factual, theoretical or
declarative knowledge - current or historical*),
'knowing someone or something' (acquaintance with
particulars or universals, directly - through
experience - or by description), and *'knowing how'*.
However, no clear demarcation line between these
types is implied, and there is a certain mobility:
one type can become another (eg. when theoretical
knowledge is applied, or when a description goes on
to explain how something should be used). Categories
of knowledge are more often than not "hopelessly
entangled one with another", as Last has put it
(Last, 1989:121).

'Knowing how' may imply knowledge of a methodology
or procedure, or the ability to successfully carry

out a procedure oneself. Common alternative labels
include *'practical' or 'applied' knowledge*, in
opposition to *'theoretical'*. Furthermore, since
'knowing how' is usually the result of active
experience rather than passive observation or
cogitation, the successful application of knowledge
is a matter of *cognitive or motor skill*. Whiting
(1975) defines skills as complex, intentional
actions which "through the process of learning have
come to be organised and coordinated in such a way
as to achieve predetermined objectives with maximum
certainty" (Whiting, 1975:6). *'Procedural knowledge'*
is a present-day label used in computer science. It
is helpful in making the connection between
knowledge and learning: procedural knowledge
requires procedural learning, declarative
knowledge - declarative learning (Waern, 1989: 71).
In the field of knowledge engineering, researchers
have found that, "as expertise develops, there is a
shift towards procedural forms of knowledge
representation" (Slatter, 1987:29). Knowledge can be
modelled, but some domains, "... such as child-
rearing, have no definitive models - forcing the
expert to rely heavily on *heuristic knowledge*"
(Slatter, 1987:27) - rules of thumb, or "inspired
guesses" (Hart, 1986:20), based on experience as
well as theoretical knowledge. 'How to' knowledge
can also be described as *'instrumental'* or
'operational'. When knowledge is the result of
having experienced or studied case histories (eg.
the case method in the American legal system),
'casuistic knowledge' (used in a derogatory sense in
Ethics) or *'case knowledge'* might be more
appropriate. The development and refinement of an
expert's knowledge eventually leads to the
possibility of *value judgment*.

Researchers working on knowledge representation systems make use of the notion of *control knowledge*. This is knowledge that eg. a rule based (production) system must have to control the 'firing' of rules: "Metarules are rules which control the use of domain rules" (Williams & Bainbridge, 1988:106). Control, or *strategic knowledge* (Clancey, 1983), specifies the ways in which knowledge elements are used; it describes "when" and "why" rules should be applied. "Control knowledge describes what to do, when and why (knowledge which can be extremely important to a novice)" states Anna Hart (Hart, 1989:59). It seems that the notion of control in relation to knowledge systems could be a very important one, given that "the most essential factor in the dynamic operation of a system is <u>control</u>, definable as a process for managing the relation between determinacy and indeterminacy" (de Beaugrande, 1988:9). Alexander (1992) uses another label, "*conditional knowledge*", to indicate that it is concerned with "when" and "where".

Gaines (1988) has devised a model of knowledge acquisition based on the premise that knowledge is culturally transmitted, i.e. by a process whereby each person shares the results of knowledge creation and development by other members of a culture. According to this model, at the lowest level there is *reflexive knowledge* or *mimicry*, which has no verbal component and comes directly from experience or from watching an expert at work. This informal knowledge is in contrast with *formal knowledge*, rule-based, induced, transmitted verbally or by reinforcement - eg. by working under expert supervision. Next, there is computational or *technical knowledge*, usually transmitted by rational, technical explanation, perhaps through

books, followed by *comparative knowledge*, transmitted by simile or metaphor, by transfer from related worlds or cultures. At the highest levels, we have *abstract knowledge*, induced or transmitted through mathematical representation, encompassing basic laws, and finally, *transcendental knowledge*, which refers to the transfer of general, systemic principles.

What emerges all along here is the transition from discrete items of knowledge to an awareness of systematic, ordered knowledge, whether in the form of relationships, rules, principles, or steps - understanding how things fit together, how one thing affects another.

Another basic distinction can be made between *foreground* and *background* knowledge, useful for processing purposes (computational and human). In computer systems which are programmed to understand natural language, for instance, 'background knowledge' includes hierarchical representations of systems of concepts, knowledge about the basic laws of the universe in question (eg. properties of objects and how they change), knowledge about methods of performing actions, and how a microworld will change as a result of such actions, procedural knowledge in the form of global schemas - scripts (Memory Organisation Packets - MOPs), plans, plot units, and themes (eg. Thematic Abstraction Units - TAUs) (Schank & Abelson, 1977; Galambos et al., 1986). Schema theories of knowledge representation presuppose top-down processing, where larger structures are used to interpret new pieces of information as they are encountered in the comprehension process. This type of knowledge can be said to be *generic* (see also "domain knowledge"

discussed later). *Causal knowledge* can be conceived
in terms of chains of events comprising actions,
states, and the causal relations connecting them
(Schank's conceptual dependency theory - Schank,
1975). The actions in a schema can be prioritized
(central/subsidiary), and are characterised by their
relative distinctiveness and varying frequency of
occurrence. *Temporal knowledge* has posed a
particularly difficult problem for knowledge
representation in artificial intelligence systems,
yet it is necessary for "... sophisticated world
models that can capture change over time within
them" (Kwong, 1988:190). A further category of
knowledge may be proposed: *restrictive knowledge*,
comprising conditions necessary for concepts to
occur, non-equivalence between concept systems in
different languages, or conceptual gaps. Closely
related to this is *negative knowledge*, encompassing
all that we know to be untrue.

Background, or *prior knowledge*, has been examined by
psycholinguists researching the phenomenon of
inferences in text processing (Rickheit & Strohner,
1985). Inferences are said to be generated from
prior *world knowledge*, much of which is *social
knowledge*. This includes, for example, knowledge of
personality traits in general - extroverted,
hostile, crazy, etc. - allowing a reader to predict
how a person with these traits will behave;
stereotype categories - waitresses, policemen, black
people, middle aged, etc.; *prototypic situations* - a
doctor performing surgery, a cowboy riding a horse
(Clark, 1985), and so forth. Closely aligned to this
is *cultural knowledge*: knowledge of the cultural
context with its conventions of communication. But
text comprehension is also influenced by *personal
knowledge* - knowledge of the world and of linguistic

rules and conventions, conditioned by sex, age, education, occupation, personal attitudes and emotions. Next, there is *verbal context* to consider, i.e. knowledge of those parts of the text which have been presented previously. A distinction can then be made between a priori inferences, drawn from the reader's background knowledge, and a posteriori inferences, drawn from the text already processed (Crothers, 1979). Sometimes this distinction is conceived in terms of *current* information and *stored* knowledge. Kakkuri-Knuuttila & Kusch (1991), Finnish researchers representing a philosophical approach to text interpretation theory, quote Hintikka's notions of *active* and *tacit knowledge*, developed to include *potential* and *virtual knowledge*, made apparent through question-posing. Nystrand (1986) writes about *mutual knowledge* as being the knowledge that two or more individuals possess in common, which allows for establishing a mutual frame of reference in communication.

Background knowledge is not only encyclopaedic in nature but also *linguistic*. Our personal knowledge of linguistic rules and conventions covers grammar, phonetics, semantics, pragmatics. It encompasses our personal understanding of the meanings of items in the lexicon, which may include incorrect perceptions of meaning, and the special meanings of words which function as terms in narrow subject domains.

4.Knowledge and language

As we take stock of the many different types of knowledge, we are drawn inexorably to consider the relationship between knowledge and language. In particular, we can ask whether knowledge which has

been gained through direct contact, observation or experience, and then expressed (described) in language, has been altered in some non-trivial way, for example due to the linear, sequential nature of discourse which may be at odds with the concomitant nature of aspects of an event experienced or observed. The crux of the matter is that readers of manuals can only access knowledge which can be mediated by language (graphical representations apart), and language comes charged with prior meanings - that is to say, the reader has previously encountered most of the words - and many of the terms - in other contexts and will have a preconceived idea of their meanings, based on usage in the general language or in other knowledge domains. After all, according to Ludwig Wittgenstein's dictum, *the meaning of a word is its use in the language*, and most words' meanings will have been determined by previous use. As well as having "inherent features", each lexical unit displays "contextual features" (Chafe, 1972), and both can prime the meanings encountered in new contexts. The knowledge of a large collection of rules relating word form and meaning is a part of the language user's 'competence'. The reader, then, has the ability to understand language, and the ability to misunderstand.

When we come to examine and try to classify types of knowledge expressed in discourse we come up against the problem of language ambiguity. Ambiguity is present at the lexical level in the shape of polysemy, synonymy and imprecision, it is present at the syntactic level, it is recognised in the phenomenon of unclear antecedents (anaphora), and it is brought on by grammatical inaccuracy and spelling error. Ambiguity is particularly striking in the

question form, a major means of knowledge
elicitation. A 'simple' question: "Can you swim to
the other side ?" can have three meanings - (1) do
you know how to do it (2) please do it (3) are you
allowed to do it. Similarly, Browning (1984) draws
attention to the ambiguity produced by 'may' ("You
may initialize the equipment by pressing switch 1").
The problem can be resolved, to some extent, by
probability and contextual clues, but this is
nevertheless a substantial difficulty.

The relationship between words and meanings has
absorbed scholars since the very beginnings of
linguistic and philosophical enquiry, and it is
worth pausing to consider the views which have been
put forward. The main challenge has been to capture
the essence of word meaning in abstract terms, but
there has also been the need to make explicit the
meanings of individual items in the lexicon, and to
explore the compelling notion that these are somehow
interlinked. In particular, if word meanings were to
be handled and conveyed - a need which imposed
itself when the first dictionaries were conceived -
they had to be made tangible and explicit.

However, not only do the meanings of words change
over time, but the perception and representation of
meaning is influenced by a number of factors,
notably current technological trends. Technology
provides a conceptual framework through which
meaning may be viewed. In current vogue is the
information processing analogy, and to a certain
degree, this dictates the choice of meaning
representation. The representation may be
implemented on a computer, and since computers are
used to simulate mental and linguistic activity,
this in turn may colour the perception of meaning.

It may further be supposed that the specific needs of individual systems or applications can become the overriding factor where meaning representation is concerned.

One thing can be stated with absolute confidence: meanings are not encapsulated in individual items of the lexicon. The dependence of word meaning on word settings and the interrelated nature of the 'intensions' of words is well documented in modern linguistic theory. Indeed, one of the tensions characterising meaning representation is how to reconcile the practical necessity of representing the meanings of individual lexical items with the desire to provide a full description by reflecting their relationship to other words and meanings in a given sample of language or in language as a whole.

In the central problem of lexical semantics - that of the relationship of words to their meanings - a long-established approach is referential: the essence of meaning is specified by establishing an interdependence between words and the things or concepts they denote. This is typified by the 'meaning triangle' of Ogden and Richards (1923): a word symbolises a concept, and stands for a referent. It may be fair to say that the analysis of the interrelation between the linguistic sign and concept or referent is not strictly speaking the object of linguistic enquiry. The main alternative is then a functional approach to meaning, where the aim is to study the meanings of a lexical unit through its relationship to other units of the language rather than its relation to either concept or referent. It is also possible to view the functional approach as being complementary to the referential one.

Although lexical items express meanings, these meanings are not fixed: they change when viewed from a diachronic perspective, and synchronically lexical meaning is dependent on linguistic and extra-linguistic context. The study of shifts and developments of meaning over time, which includes tracing the passage of technical terms into the general language and vice versa, is complemented by a study of potential meanings: the potential lexicon.

The grammatical interdependence of linguistic forms makes meaning a function of a lexical item's position in the discourse structure in which it occurs, and a product of the interplay of meanings of items which make up the context. Meaning is also dependent on the culture of a particular language community - including professional cultures - as evidenced by the phenomenon of non-equivalence between languages, and the communication problems associated with the use of 'jargon' terms. Neither is meaning tied to form: homonymy and homography are commonplace phenomena.

Lexical meaning can be perceived as a pattern of syntagmatic and paradigmatic relations or affinities between the constituents of a lexicon in a given language (Cruse, 1986). A lexical unit could be a morpheme, a word, a variable word-group, even a phraseological unit (eg. Ginzburg, 1979). John Lyons (1981) uses the term 'lexeme' to designate 'lexically simple expressions', in opposition to lexically composite ones. The question as to what constitutes a unit of lexical meaning has implications for all aspects of the study of lexis, and is especially important for natural language processing (eg. Sparck Jones, 1985).

From a lexicographic perspective, methods of word definition give us further insight into approaches to word meaning: a definition can identify a word's position in relation to other words of the language (synonymous, analytical and synthetic definitions), or it can make reference to extra-linguistic reality (eg. denotative and ostensive methods). A lexical item is therefore most commonly defined either by connotation - a set of criteria for belonging to a class, or denotation - a set of referents. In technical writing, there is sometimes a need for operational definitions (Sides, 1984:30) or for definition by exemplification (Kukulska-Hulme, 1990c:54). A further method of defining words - the provision of a citation - suggests that meaning can be implied through verbal context. In the course of this century, collocations have been added to the lexicographer's arsenal of methods: to list collocations of words is to express their meaning in terms of their collocability (Firth, 1957).

The last twenty years have seen a rise in interest in the relationship of syntax to semantics, and by extension in the relationship between syntax and lexicon. In the early transformational-generative work of Noam Chomsky, grammatical description was non-semantic, but gradually the need to incorporate semantic restrictions was recognised. Consequently, lexical meanings used for generating sentences were decomposed into semantic features. More recent work in generative grammar tends to make the dividing line between grammatical and lexical meaning progressively less sharp. Proponents of Montague grammar have attempted to establish an even closer correspondence between the two. Hudson (1984) has argued that the internal structure of a word can be generalized to act as the basis for generating

syntactic structures, so that there is no
fundamental distinction between 'rules' and 'lexical
entries'. Mellish (1985) advocates 'early semantic
analysis' as an alternative to the traditional
approach of syntactic parsing followed by semantic
interpretation. In 1980, a semantics-oriented 'word
expert' language parser was reported by Small
(Small, 1980); a semantics-driven approach
characterises also the Distributed Language
Translation project (Papegaaij, Sadler & Witkam,
1986). Specific interest in lexical knowledge
representation has led Viehweger (1991) to conclude:
"Meanings are knowledge representations of states-
of-affairs in reality characterized by complex
internal structures that are flexible and dynamic"
(Viehweger, 1991:263).

In information retrieval, we have to consider the
semantics of individual terms, multiword
terminological units, phrases, as well as simple and
complex sentences. Of particular relevance is the
performative or *illocutionary* nature of language;
Blair (1992) has, in fact, proposed a document
indexing structure based on John Austin's taxonomy
of illocutionary acts (Austin, 1975). We must
remember that user queries perform the action of
information retrieval, even though an intermediary
mechanism or algorithm must be used to map a query
effectively onto the knowledge in a database.

5.Domain knowledge and specialisation

We have so far discussed knowledge - and language -
without explicitly relating them to specialisation.
Knowledge of a specific field of learning or
experience - a science, a technical field, an arts

subject, a sporting activity - can usefully be referred to as 'domain knowledge'. It is, however, necessary to consider definitions of this term, given that it is used by different people in different ways. Alexander, who specialises in this question (see eg. Alexander, 1992), defines domain knowledge as "the realm of knowledge that individuals have about a particular field of study" (Alexander, 1992:34). In her view, background or prior knowledge is advanced to the level of a domain when it becomes a focus of study. When knowledge is based around fundamental generalisations and is highly organised, it can be termed *"discipline knowledge"*. This represents a learner's perspective of domain knowledge, where a progression in knowledge is emphasised. Possession of domain knowledge is what is generally said to distinguish *an expert* from *a layman,* quite apart from their different modes of operation and abilities. There are, indeed, degrees of knowing, so that one can, for example, talk about someone having a 'basic' knowledge of a subject, 'advanced' knowledge, and so forth; knowledge can also be 'partial' or 'incomplete', and there have been numerous studies of knowledge acquisition examining the "novice-expert shift" (eg. Anderson, 1985; also Nystrand, 1986, who describes "knowledgeable" and "unknowledgeable" readers). Research into text comprehension has yielded the labels *"high domain knowledge"* and *"low domain knowledge"* to help explain the relationship between expertise and the number of inferences drawn during text comprehension: the greater one's knowledge, the more inferences are drawn (Rickheit, Schnotz & Strohner, 1985). In information retrieval systems, domain knowledge can be conceived as "information about important topics and concepts in a specific domain

and how they relate to each other" (Brückler, Florian & Kalcher, 1988:151). This is a subject-oriented view of domain knowledge. We have to take both perspectives into account, since we are concerned with matching a user's (learner's) developing domain knowledge to the domain knowledge represented in a body of texts.

According to Hart (1989), experts are characterized by the following features: *effectiveness* (using knowledge to solve problems, with an acceptable rate of success), *efficiency* (deducing probable solutions quickly, determining relevant information quickly), an *awareness of the limitations* of their knowledge, and *versatility* in unfamiliar situations. We would hope to be able to use an expert "as we would a text book with a question-answer facility, where the answer meets our particular requirements, and is phrased in terms which we can understand ...", and to be able to question the expert about how he or she reached certain conclusions, but it is also recognised that experts are characterized by an "... inability to explain high-level problem-solving activity without ambiguity" (Hart, 1989:17-18).

The practical need - as well as academic aspiration - to model expertise has created the concept of 'domain knowledge'. This special subject knowledge has a corresponding special subject language - a language for a 'special purpose' (Kittredge & Lehrberger, 1982; Hutchinson & Waters, 1987) - though such language is not a separate, delineated entity in relation to language as a whole. The distinction between 'general knowledge' and 'knowledge of a special subject' is, of course, not new. What is relatively new, however, is the way we model special subject knowledge ("a subspace

of the knowledge space" - Sager, 1990:16), and the need to represent domain knowledge in a symbolism other than natural language in textual form.

B.The organisation of knowledge

1.Units of knowledge

In this section we review several approaches to the identification of knowledge units, in an attempt to shed light on the relationship between terminological units and units of knowledge. From an encyclopaedic perspective, knowledge can be expressed in texts, with referential links between them, but since a 'text' is a hazy concept in terms of its length, content, and even general characterisation, it seems more productive to look at better defined units of knowledge - not necessarily primitive units, but ones that are more distinctly formed.

From a philosophical perspective, elements of knowledge consist of propositions, a proposition being an abstract entity, an object of thought. A proposition can be expressed in a sentence ("a sentence expresses the proposition which is its meaning" - Landesman, 1972:6); a statement is a sentence uttered with the intention of asserting something true. Propositions refer to individuals and universals, and can be existential, attributive, or relational. True propositions are otherwise known as facts; if there is uncertainty, assumptions are made. The philosopher (eg. J.R. Searle, 1979) might make a distinction between *brute facts* and *institutional facts* (flavoured with conventions and activities characteristic of human institutions).

67

Drawing on this philosophical basis, computer
science has provided us with another perspective on
knowledge units, through its need for accurate and
workable representations. In the field of knowledge
engineering, systems are built on the basis of
knowledge elicitation techniques which identify
facts, assumptions, and rules. An expert has a large
number of perceptual patterns, or "chunks", that "
... directly index part of the expert's knowledge
store. A chunk is a familiar configuration of
elements that through repeated exposure comes to be
recognized as a single unit" (Slatter, 1987:28).
Experts may be unable to make these chunks explicit
by formulating rules, but may be able to list
symptoms or characteristics, with possible
decisions, and subsequently match the two. Knowledge
engineers have to deal with the incompleteness and
uncertainty of an expert's knowledge.

In the context of computer database systems, natural
language interfaces can be based on knowledge about
the meaning of words relative to a specific
environment. The environment is modelled, and a
certain portion of the model is assigned to the word
as its meaning. Relations are then established
between model parts and are represented as
hierarchical, network, relational, and binary
relation models. Entities, attributes and
relationships are the basic ingredients of the
conceptual structure, or schema, of a database.
Conceptual design aims to represent these
ingredients in a form that is comprehensible to the
user and independent of any specific system. "The
problem is that representation mechanisms that are
user oriented tend not to be very database oriented,
while representation mechanisms that are design
oriented force the users to make many representation

decisions in order to get the information into a
processible form" (Teorey & Fry, 1982:57). Since
users are interested in the information content of a
database, not in its physical or logical structure,
designers have had to develop high-level information
representation structures for conceptual modelling.
Conceptual design can be seen from two perspectives
(Teorey & Fry, 1982). One of these, 'object
representation', aims to define the relative
structure of the abstract objects or concepts of a
system. In this view, objects are related to other
objects in two ways: as a collection and as a class,
which can be formalised through 'aggregation' and
'generalisation' (Smith & Smith, 1977). Aggregation
forms an object as a relationship between other
objects, while generalisation forms an object from a
class of other objects. The more traditional
perspective for conceptual design is entity
modelling, particularly as elaborated by P. Chen
(1976). In the entity-relationship model,
information is represented through entities,
attributes (properties or characteristics) of
entities, and relationships between entities. A
relationship can be defined as an association
between one or more entity types, reflecting
relationships in the real world. It is worth noting
that the standard ANSI/SPARC framework for database
systems (Tsichritzis & Klug, 1978) is based on three
schemata (conceptual, internal and application
schema) which map directly to the three vertices
(concept, referent, symbol) of the referential
meaning triangle of Ogden and Richards (1923). The
GLOT terminological data bank in Stuttgart (Mayer &
Maier, 1987), implemented using the ORACLE
relational database, has been conceived along these
lines. It caters for generalisation and aggregation
hierarchies, and for associative relations.

69

This brings us to the theory of concepts as encountered in the field of terminology. Terminology work in its best known form is based on Eugen Wüster's General Theory of Terminology (1974). The theory aims to delimit concepts before attempting to assign terms to those concepts. Concepts can be defined by intension (specification of the concept's characteristics) or by extension (enumeration of all species at the same level of abstraction or of all individual objects belonging to the concept). A concept is roughly equivalent to the 'meaning' of a term, but unlike linguistic meaning, a concept exists independently of a term: they are two separate entities united arbitrarily. Terms are linked to concepts through relations of monosemy, polysemy, synonymy, homonymy, and equivalence (Picht and Draskau, 1985).

Concepts as units of knowledge "do not exist without being related to other concepts" (Budin et al, 1988:52). They are said to be directly related if they have the same characteristics in their intensions. They are indirectly related if the individual objects which they represent are contiguous in space or time. According to Felber (1984), three main types of relationship are possible: logical, ontological, and relationships of effect. Logical relationships include intersection, subordination, coordination, and diagonal relationships. Ontological relationships can be partitive, successive, and can relate material to product. Relationships of effect comprise causality, tooling, descent. Picht and Draskau (1985) classify logical relations as consisting of identity, implication, intersection, disjunction, and negation. Other relations cited by them from the German DIN standard 2330/2331 are ontological

70

spatial/temporal relations, cause-effect, producer-product, material-product, sender-receiver, tool-application, argument-function. Sager (1990) lists a selection out of what he emphasises is a very large number of possible relationships, including phenomenon-measurement, object-counteragent, object-container, activity-place, and so on.

Terminography aims to record terminological data which gives a precise description of a concept and of the relationships between a concept and other concepts. This information, along with data such as sources, field of application, grammatical notes, definition, context, etc. constitutes a terminological record. A typical record for a documentation thesaurus stores information about conceptual relationships of three types: hierarchical, associative, and equivalence (preferred term). A complex terminological record can indicate concept coordination and overlapping, or indeed any of the relationships analysed in the course of terminological work. The type of information actually included depends on the orientation of the data collection, typically standardisation, translation, or language planning.

At this point we can make some observations concerning the relationship between terminological units and knowledge units. We have seen that from a terminological perspective, it is concepts, not terms, that constitute units of knowledge. As has been pointed out by Sager (1990), concepts are notoriously difficult to define; it is, however, possible to group them into four basic types: *class concepts* (or entities, generally corresponding to nouns), *property concepts* (or qualities, for the most part corresponding to adjectives), *relation*

71

concepts (realised through different parts of speech), and *function concepts* (or activities, corresponding to nouns and verbs) - [see Kukulska-Hulme (1989) for an account of an analogous grouping in the design of dictionaries for translators]. We can consider systems of concepts and endeavour to specify relationships between concepts, uncovering the knowledge structures which bind them together. But we cannot do the same with terms. Terms are existential in nature, that is to say they signal the existence of an entity, a relationship, an activity, a quality. Considered outside of verbal context, they can express neither facts nor rules, they can say nothing about the manipulation of knowledge. Furthermore, many concepts, particularly of the relational and functional types, are designated by words of general, not special, reference. Here, then, is the crux of the matter: in a situation of knowledge retrieval from a written source, how can a terminological unit represent a specific knowledge need, other than one which seeks only to discover the concept (meaning) of a specific term or to confirm the existence of a concept ? The procedural, functional, relational, control aspects of knowledge are not well served by terms as symbols of knowledge units outside of context. This must be borne in mind when considering the process of retrieval and knowledge transfer.

There is one other aspect to this. As Smeaton (1992) has pointed out,

"It has always been assumed by researchers that in language it is the noun phrases that are the content-bearing units of information. This is not true for a full representation of meaning but noun phrases are good indicators of text content and for traditional information retrieval, that is what is wanted."

(Smeaton, 1992:272)

72

The emphasis on 'noun phrases' and thereby on class concepts in preference to other concept types may indeed be appropriate for traditional information retrieval, retrieval which presents the user with a list of references as its outcome. But in full-text systems, and in relation to natural language user queries, one has to keep an open mind as to which concepts and elements of expression will be the most effective, the most representative of a user's intention.

2.Domain knowledge representations

We have already seen that it is difficult to separate units of knowledge from the structures in which they are embedded. All knowledge structures have a function: organising, guiding, helping understanding, making access to knowledge easier. The basic structure of knowledge can be said to have three components: categories, rules for category membership (distinctive features), and category interrelations.

The domain knowledge representation problems which are still being addressed today were already recognised in the 1960s and 1970s. Early natural language processing systems made extensive use of semantic primitives and networks (eg. Masterman 1961, Ceccato 1964, Quillian 1968, Wilks 1972, 1973). It was partly due to the lack of success of the earliest machine translation systems that researchers turned to the broader issues of how language understanding might be simulated, and how to enable effective human-computer communication to take place. Natural language communication became an important area of research both in artificial

intelligence and in database technology. A
lexicon-driven approach to the analysis of language
characterises much of recent research, eg.
Cullingford's (1986) ERKS system.

Natural language understanding is part of the more
general aim of simulating human cognitive processes
on the computer. Models of linguistic competence and
performance are built in an effort to simulate this
aspect of human cognitive activity. The difficulty
lies in the fact that "characteristic for human
language understanding is the fact that meaning
cannot directly be constructed from some basic units
of meaning. Instead, complex relations between the
model objects enter as well into the meaning of
language units such as words" (Krageloh & Lockemann,
1978:50). Highly complex models of the linguistic
component of cognition are required, since the
semantics of natural language statements must remain
largely unrestricted.

The questions that concern us here are whether
lexical semantics can be separated from world
knowledge or encyclopaedic knowledge, and whether
the representations which might be used for words
can have the same form as representations of other
kinds of knowledge. In 1987, a round table meeting
of experts under the auspices of the Commission of
the European Communities (McNaught, 1987) confirmed
the lack of consensus in this sphere. For instance,
a multilingual lexical knowledge base could be
designed to contain 'language-oriented' or
'real-world oriented' knowledge, or a combination of
the two. Some existing systems make an explicit
distinction: for instance, in their national
electronic dictionary project, the Japanese
(Ishiwata, 1985) have opted for a two-part modular

design (a linguistic knowledge base and a distinct
conceptual taxonomy); an R & D project at
Carnegie-Mellon University (in McNaught, 1987)
separates syntactic (domain independent) knowledge
from domain-specific (i.e. semantic) knowledge, and
uses a lexicon to link the two.

In a typical knowledge base, both structural and
causal relationships between component entities are
recorded. The symbolic representation of these
relationships can be achieved by IF-THEN rules,
semantic nets, frames, or Horn Clauses. Frames can
be used for representing concept prototypes (eg.
Nissan (1987) makes use of static, consulted frames,
implemented as deeply nested relations), since they
can contain default values. Fass (1987) has
implemented a lexicon of 'sense-frames' consisting
of arcs and nodes that correspond to the genus and
differentia of standard dictionary definitions; the
arcs of all the sense-frames comprise a
'sense-network': a structured semantic network of
word-senses. Graham & Jones (1988) make the point
that different knowledge representation formalisms
are appropriate for representing different kinds of
knowledge, eg. frames for object knowledge, semantic
frames for associative knowledge, production rules
for causal knowledge, and so forth.

Systematic enquiry into the nature of the lexicon
relies largely on the application of principles
borrowed from other disciplines, specifically from
mathematics, and more generally from the
philosophical ideas which have underpinned the
development of science. The origins of many lexical
analysis techniques, like componential analysis,
associative pairing, or antonym pairs, can be traced
as far back as the theories of the ancient Greeks,

Indians and Chinese. In the West, Aristotle was the first to distinguish between two kinds of associations of ideas, namely logical relationships (based on similarity) and ontological ones (based on contiguity in space or time). He developed deductive reasoning, thus laying down the first laws of logic. In more recent times (18th Century), the logical relationships of superordination, subordination, and co-ordination were explicitly described by Immanuel Kant.

Symbolic logic is based on an extensional approach that has proved useful in mathematics. Often natural language needs to be represented in fuzzy logic (Zadeh, 1975), rather than discrete logic, since it reflects the continuous, non-discrete nature of the world. And, as has been pointed out by Ilson (1987), natural language itself is probably the best metalanguage: it has "the flexibility - and the fuzziness - to describe the properties and emulate the behaviour of its own lexical units" (Ilson, 1987:71).

The rigorous nature of analysis methods based on mathematics and logic contrasts with the more flexible approach inherent in a cognitive view of lexis. It is true that technical vocabularies reflect the ordered nature of the subject fields they represent; to these collections, strict rules of classification may be applied. The general vocabulary, however, tends to elude such rigour. No single method for specifying word meanings can hold good for all items of the lexicon. Some words are best described by listing their features, some by enumerating their parts. Sometimes meaning can be elucidated through context, or through related words. In other cases one might use an antonym, or

place the word's meaning on a scale, or name an
object's function, or attribute it to a class.
Equally, no single configuration is applicable to
every set of words. In the end, what matters is the
purpose for which meaning is specified.

For automatic processing, a standard representation
is required. Even in cases where a certain liberty
is permitted, for instance in naming relation types,
the format of the representation must be made
uniform, relation types catalogued, access routes
and operations sharply defined. Without this,
processing cannot take place. Yet standardisation
can compromise accuracy of meaning. And so for human
use, even if intermediate storage in a computer
system is envisaged, lexical items and their
meanings need a representation which is primarily
compatible with the cognitive process in which the
lexicon is an aid. The problem is that while a
computerised process can be clearly specified - be
it a conventional algorithm or a set of inference
rules - an intricate human process (for instance,
reading) cannot be captured in quite the same way.
This makes it even more difficult to devise an
appropriate meaning representation for such an end,
since the precise nature and progression of the
process remains largely unknown.

Symbolic logic continues to be applied to the
problem of the representation of meaning, although
it is essentially different from natural language
with respect to both syntax and semantics. In the
field of knowledge engineering, 'conceptual graphs'
are an attempt to address this problem. According to
J.F.Sowa (1984), conceptual graphs "form a knowledge
representation language based on linguistics,
psychology, and philosophy. In the graphs, concept

nodes represent entities, attributes, states, and events, and relation nodes show how the concepts are interconnected" (Sowa, 1984:69). A conceptual graph, which asserts a single proposition, is part of a larger 'semantic network'. Conceptual graphs are used to represent the meaning of propositions as an intermediate stage in language parsing and generation, using relations such as agent, recipient, object, attribute, cause, destination, duration, instrument, material, negation, successor, and so forth. The content of conceptual graphs is determined by conceptual analysis, whose essential goal is to produce a precise, formal catalogue of concepts and relations.

Charles Fillmore's case grammar (1968) has had a strong influence on semantic work in artificial intelligence, providing convenient labels for relations: agent, instrument, object, source, goal, etc. Language parsers have made use of these conceptual relations, but the idea of incorporating them into dictionaries has come more gradually (eg. Somers, 1980).

Nearly all types of conceptual analysis carried out across the disciplines of knowledge representation and lexical semantics share the basic distinction between intension and extension. 'Intension' is a roll-call for concept, connotation, sense, attribute, property, feature, primitive, data description, definition, inference rule, criterion, class...; it is the basis for sense relations which are logical, superordinated, intersected, overlapping, networked within the language; it creates associative, semantic, conceptual fields; it accounts for fuzzy sets and prototypes. 'Extension' evokes referent, denotation, species, object,

physical entity, file, list..., and relationships
which are ontological, partitive, collective,
successive, contiguous in space or time.

3.Lexical and terminological configurations

It is interesting, though not surprising, to note
that the problems of knowledge representation
overlap with those of meaning representation, i.e.
lexical semantics. Lexical semantics is concerned
with the meanings of lexical items but also with the
identification and representation of semantic
relations between lexical items. Lyons (1981) has
called these 'sense relations'; the sense of a
lexical expression is the set of sense-relations,
both combinatorial and substitutional, that hold
between it and other expressions in the language
(eg. synonymy, hyponymy). An earlier examination of
lexical relations by R. A. Waldron suggests that it
is also possible to consider the way in which "the
referential function of one term is linked up with
the referential function of other terms, so that the
vocabulary appears rather as a system of reference
than as a battery of separate referential words"
(Waldron, 1967:95). Saussure discussed different
kinds of associative relations between words,
association being established through a common
root-element or suffix, through similarity of sound,
or meaning (Saussure, 1972:174).

Cruse (1986) has provided us with a survey of
lexical semantics in which he makes reference to
paradigmatic (congruence) relations, comprising the
logical categories of identity, inclusion, overlap
and disjunction, and to syntagmatic relations,
comprising philonymy, tautonymy (pleonasm) and
xenonymy (dissonance). This is not a clear-cut
classification: the existence of partial and pseudo
relations is also acknowledged. These and other
sense relations can be used to construct lexical
configurations, typically in the form of hierarchies

(taxonomies, meronomies, chains, helices),
proportional series, lexical opposites and sets of
synonyms.

Ginzburg et al.(1979), in an earlier account of
semantic classification, distinguish four
approaches: conceptual fields, hyponymic relations,
semantic similarity and contrast, and word families
or clusters. A conceptual field can be seen as a set
of lexical items in which the meaning of each item
is determined by the presence of the other items.
The classification of words into word-families or
clusters means that a group may be composed of words
with semantically and possibly phonemically
identical root-morphemes, or of words with identical
affixational morphemes, where further categories
such as agent or action can be seen to emerge.

Lexical groupings may be based on the notion of
associativity, in which case we are dealing with
unstructured, unsystematic groupings, or 'clusters'.
Associativity has been a dominant concept in
research concerned with the workings of the mental
lexicon. In the 1960's and 70's, the associative
structures of the 'subjective' were investigated,
using methods such as linear graph analysis,
hierarchical clustering, and multidimensional
scaling on carefully selected domains (eg.
Fillenbaum & Rapoport, 1971). The aim of this type
of research has been to find out to what extent
generalised associative structures can be uncovered
for a given domain, how restricted domains are
interlinked in the subjective lexicon, and how
semantic knowledge develops.

Summarising relevant psycholinguistic research in
this field, Aitchison (1987) describes the two major

components of the mental lexicon: the
semantic(-syntactic) and the phonological component.
Her account is concerned with the spoken word; it
has to be stated that the interrelationship between
the spoken and written forms is not well understood
(eg. Allport & Funnell, 1981). It appears that the
semantic component of the mental lexicon is arranged
in a network divided into semantic fields, with
multidimensional links of various strengths between
words. The links can be of several types -
coordination, collocation, superordination, synonymy
- but the connections between coordinates,
especially those of the same syntactic class, and
between collocates, are particularly strong, while
the other links are weaker. The division of the
lexicon into the two components reflects the
different needs of speech production and
recognition: the semantic component is geared
towards production. The method of retrieval of
lexical items can be explained through a 'spreading
activation' model (eg. Stemberger, 1985), whereby a
large number of nodes on the network are activated,
and those that are not required are then gradually
suppressed.

The semantic net was devised explicitly as a
psychological model of human associative memory.
According to A. Narayanan (1986), association is to
be found between items which occur simultaneously or
in close succession, or which are similar or
contrary. Association can be direct (stored in
direct physical or logical contact in memory), or
indirect (occurring at recall time). The most recent
metaphor for memory organisation and function is a
distributed parallel processing system. In this
model, it is possible to say that "each node of a
semantic network corresponds to a particular pattern

82

of activity over a large number of units. A node can then partake in many different patterns of activity (...)" (Narayanan, 1986:249); a node can be interpreted as representing a concept, but it also represents a pattern of activity at a lower microlevel. Links between concepts are generated by many simultaneous interactions at the level of their microstructures.

Another approach, under the banner of 'cognitive grammar', equates meaning with conceptualisation, where the latter "encompasses novel conceptions as well as fixed concepts; sensory, kinesthetic, and emotive experience; recognition of the immediate context (social, physical, and linguistic); and so on" (Langacker, 1986). Lexical items are recognised to have a considerable array of interrelated senses, which may be conveniently represented in network form.

The theory that the mental lexicon is organised along the lines of a network is one of two predominant theories or models of this aspect of the human mind. The other main theory, that of semantic primitives (with its associated technique of componential analysis, or lexical decomposition), was the dominant viewpoint some twenty years ago (eg. Wierzbicka, 1972). The work of Schank (1972) is much quoted in this sphere; some of the 'meaning atoms' or 'primitives' identified by him were later shown to be complex notions, which were further decomposable, and other important criticisms have been levelled at this theory. In 1976, the psychologists Miller and Johnson-Laird tried to elaborate it by linking semantic primitives to perceptual primitives; however, other aspects of the meaning of words (eg. the function of objects)

are not perceptually based, and these aspects had
not been captured. Although there is no substantial
evidence to support the theory that word meanings
are split up into primitives in the mind, semantic
decomposition is still a useful approach for
organising meaning in the 'objective' lexicon (eg.
Cullingford, 1986). For example, a study using
componential analysis carried out by Ortony, Clore
and Foss (1987) has produced a taxonomy of the
affective lexicon. A few typical classification
problems have emerged in the course of this work:
the variety of syntactic forms in the lexicon;
problems of ambiguity where a word has both a
physical and a psychological interpretation; the
elusive nature of antonym pairs, some of which
belong to the same category, while others do not.

Componential analysis, though it has been shown to
be defective both theoretically and empirically (eg.
Lyons, 1981), is still considered to be a useful way
of formalizing the focal, or prototypical meaning of
lexemes. It has served to show that for some
polysemous words, it is difficult to identify a
'semantic core', and continues to reaffirm the view
that in general, word meanings are fluid, or fuzzy.
Only a very small number of words have fixed
meanings, so that a set of necessary and sufficient
conditions can be specified. However, when
categorising objects, people appear to know the
characteristics of an 'ideal exemplar' or
'prototype' of the object in question (Rosch, 1975).

Another approach to the formalization of lexical
structure is the relation of entailment (a relation
between propositions p and q, such that, if the
truth of q necessarily follows from the truth of p,
then p entails q). Although entailment is normally

84

applied to propositions, it can be used in
connection with lexical items, for example:
'dog'=>'animal'. Entailment is at the heart of
'meaning postulate' theories which assume that there
are no semantic representations for words, only
inference rules (eg. Fodor, Fodor & Garrett, 1975).
Meaning postulates were introduced by Carnap (1956)
to state the relationships between the intensions of
words. Some doubt has since been cast on the
validity of such theories, eg. by P.N. Johnson-Laird
(1978, 1987), who has, in turn, advanced his own,
psychological theory of the representation of
lexical meaning. This asserts that there appear to
be comprehensive lexical entries in the mental
lexicon, containing specifications of the senses of
words. Elements of a lexical specification can
consist of "(a) relations to other words, which
could be represented by a mechanism akin to a
semantic network, and (b) ineffable primitives that
are used in constructing and manipulating mental
models of the world" (Johnson-Laird, 1987:208).
Words can enter the lexicon through direct
acquaintance with their denotata, or, if they have a
more complex semantics, they may be acquired from
definitions, or from encountering instances of the
word in use. In the lexicon, most entries for words
are in fact likely to possess elements of both types
of information (i.e. specifications of their truth
conditions obtained through direct acquaintance, as
well as relations to other words obtained through
verbal definition or encounter). The contents of an
entry may be incomplete, and in the case of 'natural
kind' terms, a major component of the representation
of sense will consist of default values.

C.The specialised domain of computer security

1.Introductory remarks and definitions

Having discussed the concepts of special reference, domain knowledge and specialisation, we now turn to look in some detail at the specialised domain of computer security. In Chapter I, we listed a number of sources of information on computer security. There are many books available on this subject, some covering the entire field (eg. Becker, 1977; Hsiao et al., 1979; Hearnden, 1990), others a particular aspect of computer security (eg. fraud - The Audit Commission, 1985, 1987; access control - Wood, 1985; viruses - Highland, 1990), or a given sector (eg. commerce and industry - Oliver & Wilson, 1983). Some list potential problems and report cases of security breaches (eg. Rentell & Jenner, 1991), giving 'facts and figures' (Pritchard, 1979), suggesting practical solutions and security products (Hruska & Jackson, 1990). Recent books tend to include network (communications) security, and some of the issues of computer security are also touched upon in books dealing with 'health and safety' (eg. Broadhurst, 1991).

We do not propose here to give an exhaustive account of the subject, since this has been done very well by the authors mentioned, who are computer security specialists. In particular, for our purposes we are not so much interested in a systematic classification of the field, nor in a list of problems and solutions, but rather in its characterisation from the point of view of knowledge types and terminology. Nevertheless, the starting point and subsequent reference framework must be a brief description of the main constituents of the

domain, which we will base largely on James Cooper's
exhaustive and authoritative volume: "Computer and
Communications Security: Strategies for the 1990s"
(Cooper, 1989), published in the United States, and
supplement it with information from Tony Elbra's
"Computer Security Handbook (NCC Blackwell -
Manchester/Oxford, UK, 1992). Cooper's book also
contains copious references to further reading on
all aspects of the subject, while Elbra's book has
extensive "checklists" and "guidelines" for every
aspect of security. The necessarily brief
description presented in the next section is only a
top-level representation of the subject.

The Shorter Oxford English Dictionary defines
'security' as, on the one hand, 'the *condition* of
being secure, protected from danger, doubt, or care'
(synonymous with safety), and on the other hand, as
'a *means* of being secure' (a protection, guard,
defence, guarantee). Specialists sometimes choose to
make a distinction between security and safety, for
instance on the basis of degree of harm caused
(security problems cause relative harm, eg. gain or
loss in competitive advantage, safety problems -
absolute harm, eg. when a service or resource is
impaired, a company goes out of business: Burns et
al., 1992). Cooper mentions the effect of computer
security on safety (eg. in air traffic control), but
security is the key term in his book, where he
states: "Briefly, security is protecting "assets""
(Cooper, 1989:11).

It is worth bearing in mind that 'computer security'
is a subdomain, or particular concern, of a broader
'security function'. In commerce, industry and
government, 'security' is a function which aims to
protect a particular organisation, installation, its

machinery and data, etc. from known threats. Certain
security procedures are required or regulated by
legislation such as the Health and Safety Act, the
Data Protection Act, the Official Secrets Act, the
Theft Act, and so on. Security and safety are
guaranteed by implementing and maintaining measures
resulting from an assessment of risk. The 'security
function' can encompass all of the following major
issues:

Policy and management
    appraising risks
    formulating a security policy
    disaster planning
    planning for industrial action
    access control
    staffing of security
    allocation of responsibilities for security
    maintenance of security records and reports
    security audit

**Offences against security - prevention and legal
action**

theft	industrial espionage
sabotage	negligence
criminal damage	bribery and corruption
assault	trespass
arson	blackmail
bomb threats	burglary
extortion by kidnapping	
forgery	
fraud, inc. computer fraud	

Environmental hazards - prevention and action

fire	flooding
power failure	explosion

2. Brief description of the domain

Aitchison & Gilchrist (1972) suggest that to define
a subject field, you should establish the boundaries
of the subject and distinguish between the central
area, and marginal or peripheral subjects.
"Boundaries" seems too definitive a word, but it may
be said that from a computing perspective, or in an
organisation or department where computers play a
crucial role, computer security is central within
the security function. It is also worth noting that
computer security is implanted in the wider field of
technology. This dual allegiance is easily explained
by the two components, 'computer' and 'security':
either can be emphasised. Significantly also,
computer security 'rubs shoulders' with the domains
in which computers are applied, eg. business, and
with the legal system.

The three key concepts of computer security are said
to be confidentiality, availability, and integrity.
Computer security has particular concerns which stem
from the fact that information produced by computer
systems is often time-critical, that systems harbour
potentially sensitive and important data, and that
access takes place through communications systems
which can be difficult to control. A distinction is
usually made between *physical security* and *data
security*, and more recently, *network security*.
Standby facilities are all-important, and the main
hazards are environmental disaster, loss of
information (accidental or deliberate) through
computer or human error, negligence, unauthorised
access, sabotage and fraud.

Cooper's book covers risk analysis and other forms
of security analysis, and resource allocation;

security is then broken down into six 'security environments': *physical, personnel, regulatory, hardware, software, and networks*, covering the following topics:-

Physical: intrusion prevention
                      intrusion detection
                      information destruction
                      power protection
                      fire protection
                      water protection
                      contingency planning

Personnel: personnel as assets
                      personnel as threats
                      personnel ingenuity
                      personnel security techniques

Regulatory: national security
                      sensitive unclassified information
                      privacy issues
                      computer and communications security laws
                      international topics

Hardware: hardware integrity
                      hardware access control
                      electrical and electromagnetic
                                  threats
                      information-tapping techniques
                      personal computer security
                      tamper-resistant seals

Software: software threats
                      software access control
                      National Computer Security Center
                                  resources

Network: network architectures
                communications security threats
                dialup security and hackers
                encryption and cryptanalysis
                authentication and digital
                      signatures
                automated network administration

This classification, which has further levels of
subdivision, shows quite clearly how the growth of
the domain (coupled with the growth of computing)
has made it necessary to identify separate areas for
analysis, each one a candidate for being labelled a
specialised domain in its own right. However, the
boundaries are not clear-cut, and the protection of
an asset will typically require several layers of
security, eg. regulatory, physical and software. The
environments are subsequently prioritised depending
on the type of organisation; and so for
organisations concerned with commerce or business,
the order of priority is: personnel, physical,
software, regulatory, network, hardware. Across
different organisations, "... physical protection
has a strong overall role in security, as does UPS
(uninterruptible power system) protection.
Separation of duties, background screening, and
effective password systems also have high overall
importance" (Cooper, 1989:359). Elbra's (1992)
classification gives additional prominence to
management (managing security/ risk management),
internal auditing, back-up and recovery.

In an introductory section, Cooper lists and defines
the six key terms in computer security, the
"entities within the environments": assets, threats,

*vulnerabilities, risk, protective measures, and
responses -*

> "... "assets" may be information, hardware,
> software, peripheral supplies, people, communication
> media, processing capabilities, or money ... the
> "threats" to assets are 1) people who choose to be
> adversaries, and 2) happenstance due to people or
> nature ... Vulnerabilities are features (design,
> configuration, procedure) that allow threats to
> affect assets ... "risk" can be viewed as the
> probability that a given asset will be lost through
> a specific vulnerability due to a particular threat
> ..."Protective measures" are security features that
> are incorporated to minimize vulnerabilities and/or
> risk. "Responses" are security moves made after an
> incident. These may be in the form of corrective
> measures, analysis, or necessary actions
> (prosecution, recovery)"

<div align="right">(Cooper, 1989:12)</div>

He then makes the point that security problems
should be addressed in a sequence of steps - first
identify the assets to be protected, then identify
threats to those assets, examine vulnerabilities,
assess risk, select protective measures to reduce
the risks, monitor events in order to take
responsive action. This approach corresponds closely
to the CRAMM risk analysis methodology developed
under the auspices of H.M. Government's Central
Computer and Telecommunications Agency (CCTA). A
"CRAMM review", usually conducted by a licensed
consultancy (eg. NCC Consultancy - NCC, 1991),
produces an asset valuation, threat and
vulnerability study, risk assessment and recommended
countermeasures. This general procedure is also
advocated by Elbra, who highlights the need to
allocate responsibility for IT security.

The domain knowledge representations which emerge
from this high-level view of computer security
confirm that a 'mix' of representations is necessary

to convey the various types of knowledge in the domain. There is a need for structure, and within the structure a need for listing facts, describing entities with their attributes and states, specifying relationships, outlining procedures, prioritising and evaluating, linking cause to effect. Terminology helps to delimit and relate conceptual knowledge, and in this way assumes a 'high profile' role in representing the domain. The characterisation of the domain of computer security in terms of both terminology and knowledge types, reported below, will contribute to the emergence of a clearer picture of potential knowledge transfer problems. Some of these problems are caused by the complexity of the domain. As has been pointed out by Kettle (1992): "Overall the subject is so complex that it is not easy, even for an insider in computing, to see any order in the pattern of threats and countermeasures" (Kettle, 1992:198).

3. Characterisation in terms of knowledge types

As noted, the domain of computer security is related to a number of other fields of knowledge. It is not possible, and not necessary, to draw clear boundaries between such fields of knowledge. In particular, the narrower a field, the more difficult it is to describe without constant reference to the wider fields in which it is embedded. Besides, one field can be viewed from two or more completely different angles: for example, "computing" can be seen from an electronics/ electrical engineering perspective, or from a business viewpoint, amongst several others. Bearing these points in mind, it may be said that in Cooper's book, the following domains are represented:

science & technology

- a history of technological developments in the introduction; scientific & technological concepts throughout

   computing

- concepts throughout; some explained in glossary

   mathematics

- mathematical formulae, esp. for risk analysis, resource allocation, encryption, password generation, information redundancy

   security

- the key concepts (assets, threats, risk, etc.) are global security concepts

   business

- business environment identified as having special security concerns; business case studies throughout

   law

- identified as a sphere for consideration (regulatory environment); case studies with legal overtones

   world

- world knowledge used to promote understanding of issues

   language

- specific attention drawn to terminology; familiarity with abstract words assumed

The reference point for a more detailed analysis is the 'overview of knowledge types' presented earlier. In what follows, the aim was to find out whether some (or all) of the knowledge types identified could be traced in the book by Cooper. If so, the relevance of these types to the representation of knowledge in the field would be noted; however, no attempt was made to assess the degree of representation. What follows are examples of items of knowledge, with a typological label (and a page

reference). As indicated earlier, categories overlap, and a statement can be found to express more than one knowledge type.

It was found that a very wide range of knowledge types was represented in the book. As the book emphasises strategic planning and awareness, along with expert knowledge of threats, risks and responses, rather than the practical implementation of solutions (cf. Hruska & Jackson, 1990), there is a tendency towards description and enumeration (tools, techniques, options, case histories, typical situations, examples), which means that tactical control knowledge is less well represented. It is, of course, necessary to stress the incidental nature of the examples given: a statement formulated in one way will represent one knowledge type, but reformulated in another way may represent a different type. Among other factors, pragmatic considerations are known to influence an author's choice of words (see eg. Myers, 1989). Still, it is the ability of formulations to signal knowledge which is our prime concern, even as we recognise that there is no perfect relation between formulation and knowledge type.

The following knowledge types - 19 in total - were identified. Examples are given for each type.

declarative *value judgment*
comparative *prior verbal context*
restrictive *non-verbal*
heuristic/experiential *cultural*
prototypical *transcendental*
case *world*
practical *stereotypical*
procedural *linguistic*

causal   *strategic/control*

temporal

DECLARATIVE

(factual)
"Software bugs can also cause disasters." Pg. 91

(historical fact)
"The Los Angeles Department of Water and Power was victimized in early 1985." Pg. 223

(technical fact)
"Various forms of overvoltage protection are available such as arc-breakdown devices and metal-oxide varistors." Pg. 76

COMPARATIVE

(analogy outside of domain)
"Homer's description in The Illiad of a Trojan horse ... has a software analogy. Hence the name." Pg. 222

(similarity, within domain)
"The similarity to "checksums" is also apparent." Pg. 170

(difference, within domain)
"A cipher is distinguished from a code by virtue of the secrecy implications." Pg. 311

RESTRICTIVE

"This appears to be an attractive concept, but there are significant problems." Pg. 235

"This technique is feasible where large controlled fenced areas enclose a facility." Pg. 204

HEURISTIC/ EXPERIENTIAL

"Users in general are incapable of generating (or unwilling to generate) secure passwords." Pg. 235

PROTOTYPICAL

"Overwriting typically involves multiple overwrites of alternating magnetic polarity signals ..." Pg. 66

CASE

"For example, a London chemical company was victimized ... Scotland Yard intercepted and arrested the men. The lesson is ..." Pg. 59

PRACTICAL

"There are several ways in which entry can be protected. One is to try to ensure that software vendors are known ..." Pg. 231

"Some of the tools and techniques are: ..." Pg. 271

PROCEDURAL

"... it was recognized that seven steps would be necessary to verify the implementation specifications: 1. ... 2. ... " Pg. 259

"Once the alignment is found, groups of letters thought to represent a word are picked out of one ciphertext stream and checked ..." Pg. 317

CAUSAL

"This vulnerability can result in the insertion of a
TH (Trojan horse) or a virus, for example, into a
network thought to be immune to outside threats."
Pg. 301

TEMPORAL

"It is important to periodically test processing at
reciprocal sites." Pg. 88

"Then, if the virus is purged from a particular
version of the system (before or after destructive
action), the seeds of further destruction have been
planted, ready for action when the backup copies are
used." Pg. 223

VALUE JUDGMENT

"... no asset approaches the value of people. Some
of the reasons for this value judgment are ..."
Pg. 99

"These solutions are only partially effective."
Pg. 232

"... I believe the benefits outweigh the risks."
 Pg. xviii

PRIOR VERBAL CONTEXT

(explicit reference)
"This was the technique described in the previous
section ... " Pg. 314

"Returning to the case where the available budget
... is less than the expenditures ..." Pg. 33

NON-VERBAL

(mathematical)
Equations for the 'Doppler' effect in intrusion
detection. Pg. 62

(schematic)
Diagrams for the operation of the "Data Encryption
Standard". Pg. 319

(ostensive)
A photograph of a fingerprint verifier. Pg. 186

CULTURAL

(explicit - not prior)
"France has several unusual potitical features."
(Re: French legislation on security) Pg. 148

TRANSCENDENTAL

"We feel a moral and humanitarian obligation, one
that transcends any monetary investment, to protect
human well-being." Pg. 41

WORLD

"Love is one of the most powerful of all human
motivators, so it should be no surprise that love
plays a role in computer crime." Pg. 107

"Nature can also create exposure." Pg.36

STEREOTYPICAL

"... basically honest people who suffer from momentary temptation." Pg. 106

LINGUISTIC

"A new vocabulary has developed in order to describe a series of new attack methods ..." Pg xvi

"Complete agreement on terminology is often not achieved." Pg. 11

"Hot sites are computer facilities designed to be occupied in an emergency." Pg. 88

""E" occurs about 13% of the time ..." (Re: cryptanalysis) Pg. 315

STRATEGIC/CONTROL

'Strategic' is a global term in the context of Copper's book, since the title specifies the coverage of "strategies for the 1990s". Hence, all of the above knowledge types may contribute to 'strategic knowledge'. Sometimes the word 'strategy' appears in the text:

"The final strategy mentioned ... requires that ..." Pg. 89

Although 'control knowledge' can be synonymous with 'strategy', it can also be interpreted as having a more immediate, tactical resonance, catering for the need to know 'what, when, and why' - as well as 'where and how' - in specific circumstances. This kind of knowledge is not in evidence in Cooper's

book.

4.Terminological and lexical characterisation

Computer security terminology is growing and
changing, reflecting the growth of the domain.
Cooper is very much aware of the problems of
terminology. He points out that new vocabulary items
are constantly developing in order to describe new
attack methods, eg. viruses, worms, Trojan horses,
time bombs, logic bombs, trapdoors, salami attacks.
These new techniques, with their imaginative
denominations, testify to the popularisation of the
domain. At the same time, specialists are still
struggling with the core terminology ("Complete
agreement on terminology is often not achieved, even
for those of us who are professionally involved in
computer and communications security", Cooper,
1989:11), and with the terminologies of
technological advancements (eg. local-area networks
(LANs) and wide-area networks (WANs) - "These types
of terminologies are not crisp", Cooper, 1989:283).
Some concepts present veritable "dilemmas", such as
the definition of 'risk' - should the emphasis be on
high probability, or high potential loss ?

Hruska & Jackson's (1990) book is interesting from a
terminological standpoint: the table of contents
features very accessible expressions (eg. "forgotten
or lost passwords", "rubbish disposal", "time-bomb
from ex-employee", "wiping disks securely")
alongside more technical headings ("electromagnetic
radiation", "virus in CMOS RAM"). Hearnden's edited
handbook (1990) has chapter titles completely devoid
of specialised terms. Most books on the subject
published in recent years contain a glossary of

about 100-350 terms at the back of the book. Typically, these glossaries explain computing and communications terms ("EPROM", "broadband"), scientific and technological terms ("Halon", "Faraday shield"), security terms ("sensitive"), and computer security terms ("virus", "scavenging").

An analysis of a wide range of computer security handbooks shows that the following categories of words, terms and expressions may be encountered (some relatively uncontentious examples are given, though polysemy precludes a final categorisation, and it goes without saying that categories overlap). The analysis is lexical as well as terminological, as words of general reference have been included. At this stage, a separation of the two spheres of reference is not required. As with knowledge types, the wide range of reference fields should be noted. 'Value-laden' items (expressing value judgment) occur in several categories, not only in the general language.

GENERAL LANGUAGE

building, windows, completely, similar, different, malicious, accidental, major, damage, attack, sufficient, currently, incident, important, motive, disadvantage

ABSTRACT and SCIENTIFIC LANGUAGE

strategy, plan, approach, analysis, factor, feature, procedure, rule, constraint, effective, problem, solution, detect, identify, facility, conditions, effect, error, failure, probability

TECHNOLOGY

install, machine, voltage, sensor, device, pressure, batteries, ferroresonant, inverter, waveform, circuitry, infrared, ultrasonic, lock, filter, surge protector, transformer

COMPUTING

Formal:
data file, screen, backup, electronic mail, buffer, CPU, file directory, formatting, ROM, LAN, operating system, plotter, modem, sequential access, write protect, byte, ASCII

Informal:
the computer is down, hacker, system crash

MATHEMATICS

Boolean algebra, arithmetic coding, statistics, equation, Euler totient function, operand, sum, integer, exponentiation, prime number, factoring algorithm, parameter, primitive constant

SECURITY

risk, intrusion prevention, fire protection, contingency planning, privacy, classified information, sensitive information, deterrent, surveillance, authorisation

COMPUTER SECURITY

multi-level security, asynchronous attack, privileged user, ciphertext, deciphering, data diddling, digital signature, exhaustive attack, hot

site, identifier

BUSINESS

transfer of funds, personnel, accounting,
manufacturer, responsibility, cost-effective,
overhead, audit, asset, resource, scenario, check
list, policy, manager

LAW

Data Protection Act, legal requirements, register,
illegal, unlawful, national security, apprehend,
license, enforce, penalty, damages, fines,
imprisonment, compensation, provisions

LINGUISTICS

vocabulary, term, basic terminology, fundamental
terms, jargon, definition, dictionary definition,
glossary, variance in common usage, to mean,
phonetic password, letter frequency in English

Without labouring the point, a number of sense
relations can easily be identified, eg.:

> **opposites** (similar - different, problem-
solution, encryption - decryption);
> **cause-effect** (fire - disaster, negligence -
product defects);
> **synonyms** (bulletin board - electronic bulletin
board, personnel screening - background
investigation, MAC - message authentication code);
> **near-synonyms** (encryption - enciphering);
> **taxonomy** (power protection: line monitor,
voltage regulator, uninterruptible power system,

surge protector, filter, grounding);

word family (protect - protector - protection, circuit - circuitry);

semantic field (information-tapping, direct-connection tapping, passive coupling, electromagnetic energy interception, fiber optics tapping);

associative/ collocational relations (malicious attack, salami attack);

associative cluster (intruder - theft - barrier - alarm - password - secure - sensor - disgruntled employee)

Having explored the vital issues of knowledge and language which impact upon our research, we now turn to the question of how knowledge - and knowledge of computer security in particular - is to be communicated and accessed: the knowledge transfer question.

Access to domain knowledge

A. Knowledge transfer

1. Encoding of knowledge in texts

When considering the process of textual knowledge
transfer, we have to start with the writer's choices
and limitations. The process of writing has been
shown to differ from one individual to another
(Hayes & Flower, 1980). It is rarely a linear
process, so that although sub-processes may be
identified, they do not fit into a stage model:
planning, producing, and reviewing take place in a
recursive mode. Nevertheless, it is possible to
distinguish between the process of producing ideas,
and the process of producing text for those ideas
(eg. Collins & Gentner, 1980). According to Sharples
(1985), who refers to the work of Cooper and
Matsuhashi (1983), the writer generates alternative
text forms at the sentence level, and subsequently
selects one form - this is then verified and
optionally transformed. It is at the sentence level
that words and grammatical forms are chosen. The
writer must juggle a very large number of
constraints: applying critical judgement to the
selection of ideas in accordance with a
communicative purpose, adhering to the linguistic
conventions of written discourse, maintaining
connective flow and consistency of style,
structuring at paragraph, sentence, phrase and word
levels, maintaining grammatical and semantic

accuracy, and so on. The experienced, or "advanced" writer, is likely to make subsequent revisions to a text, checking for errors such as structural faults, repetitions, ambiguity, missing context, and inconsistencies, while less experienced writers will tend to check spelling, grammar, punctuation, and make small additions or deletions (Faigley & Witte, 1984).

The choice of words and terms is therefore embedded in a complex process of selection and "juggling". James Hartley (1985), in his book on designing instructional text, makes several references to the conscious selection of words for a target text. Word length is mentioned ("it is easier to understand short familiar words than technical terms which mean the same thing", however "some long words, because of their frequent use, are quite familiar, eg. communication"), word type ("concrete words and phrases are shorter and clearer than abstract ones"), and ambiguity resulting from excessive use of abbreviations and acronyms. He writes about the option of using readability formulae to check the suitability of a text for a given reader age group, and gives examples of "simpler wording" from "The Good Forms Guide" published by the Department of Health and Social Security (eg. "demonstrate" = show; "commence" = begin; "discontinue" = stop, etc. - *it is worth noting that simpler forms can be longer*, eg. "overleaf" = on the other side of this page). Joan van Emden (1990), devotes a section to vocabulary choice (with notions of accuracy, synonymy, precision, confusion, clichés, jargon, "simple language", American English) in her book on writing for engineers, and counsels her writers: "Use words which the reader will understand" (van Emden, 1990:22). She points out an insiduous danger

in the use of technical language: "... the reader
may assume that he understands and the writer may
assume that he is understood. Both may be
understanding different meanings." (van Emden,
1990:18). Charles Sides (1984), cautioning about the
use of jargon in papers and reports on computer
technology, says this: "... the issue of jargon is
audience-dependent. Always use what the audience
will understand" (Sides, 1984:5). This is sound
advice, yet on reflection, it is so cursory that it
is doubtful whether it can genuinely be followed.
The writer can strive to get to know the audience,
and even think about providing definitions or a
glossary, but there is still the matter of knowing
how to select or adapt one's "jargon". Use "fewer
and simpler words", advises Sides, referring to S.T.
Coleridge as an authority on the matter. The problem
is that knowing which words are simpler is not
simple. Some authors giving advice on writing
computer documentation (eg. Browning, 1984) use the
word "jargon" to designate writing which lacks
clarity for reasons other than the use of unfamiliar
terms (eg. poor sentence structure which is only
confusing to a non-specialist); others (eg. Stuart,
1984) make hardly any comment at all on the use of
language. Derek Rowntree, Professor of Educational
Development at The Open University, notes the
different meanings of "jargon" and gives designers
of self-instruction materials detailed practical
advice on how to avoid it and how to use it when
necessary, mentioning the need to cut out "surplus"
words, use short, familiar, precise words, strong,
active verbs, and to use specialist vocabulary "with
care" (Rowntree, 1986:211-232). With all these
authors, the emphasis is on "simplification".

Given the complexity of some forms of technical

discourse, graphic or diagrammatic representations may sometimes be used to complement or replace the written word. Sentential representations, expressed in natural or formal language, are essentially sequential, corresponding to propositions in a text or to a list, whereas diagrammatic representations, organised by location in a plane, naturally express topological or geometric relations. Diagrams also automatically support a large number of perceptual inferences (Larkin & Simon, 1987). Of course, the value of a diagram is dependent on the ability of the individual to interpret the diagram; in certain cases, that ability can be developed by training.

In terminographical analysis, the graphic symbols used to represent various relationships between terms are borrowed from mathematics (eg. > symbolises 'smaller intension') or devised specifically for terminographical work (eg. >- symbolises 'part'). Elaborate systems of concepts can be represented graphically by various tables, charts, and diagrams. Typical representations for logical systems are tree and chain diagrams, rectangular or circular field diagrams, grid tables, and numbered or coded schedules of various kinds. Systems which classify subjects rather than concepts (eg. documentation thesauri) can be based on a terminographical analysis using rectangular or circular "arrowgraphs" - graphs with arrows linking related subjects. While these representation means are "pretextual" in the sense that they are part of a process of analysis which might precede the composition of a text on a given subject, it is worth noting the strong underlying need for graphical representation - it suggests that linguistic symbols are not enough. The challenge, then, is how to "map" diagrammatically analysed

conceptual and terminological data onto a sequential
text describing those concepts; or alternatively,
how to integrate graphical representations with
running text. As this is not strictly within the
scope of our investigations here, the reader is
referred to the body of research which deals with
the comparison of visual and textual means (eg.
Kolers et al., 1979; Bernhardt, 1986).

2.Textual communication of knowledge

If composing a specialised text were merely a matter
of expression and not communication, then it would
be enough for the writer to find the most apt means
of expression at any given time. In reality, the
writer must direct his or her writing towards a
reader. In a situation of dissemination, regard for
the reader may not be as crucial as it is in a
situation of tuition by means of texts. On the one
hand, we have one-way delivery, in articles and
books, with the reader permanently at the receiving
end. On the other, there is actual or simulated two-
way communication, as in distance learning or self-
instruction materials, where the reader is expected
to play a more active part in the communication
process, if only by "regurgitating" material to
prove that it has been read, or by elaborating the
knowledge structures of the text through the
addition of elements of his or her own knowledge and
experience.

At this point we have to confront head-on the
problems of knowledge transfer through texts. The
constraints of such transfer are obvious: limitation
of scope, necessary selection and prioritisation of
ideas, difficulty in tailoring material to an

individual reader's needs and in providing feedback
on progress or understanding. As a learning
situation, textual transfer lacks the personal human
factor, the possibility of thrashing out meanings in
the course of interaction. Well designed self-
instruction materials "must carry out all the
functions a teacher would carry out in the
conventional situation - guiding, motivating,
intriguing, expounding, explaining, provoking,
reminding, asking questions, discussing alternative
answers, appraising each learner's progress, giving
appropriate remedial or enrichment help ... and so
on", writes Rowntree (1986:11), but even with tutor
back-up (eg. by correspondence) it is not possible
to emulate the immediacy of the face-to-face
situation. There is also the loss of control by the
provider of knowledge (the writer): a reader cannot
be compelled to read in sequence and in totality,
from beginning to end. It follows that from the
point of view of language understanding, a term may
be received "out of context" - without a preceding
context which might have served to clarify its
meaning. This is especially true of technical and
user manuals, which are more likely to be read out
of sequence. The effectiveness of knowledge transfer
is then potentially compromised.

As a counterbalance to these limitations and
drawbacks, there are the conventions of technical
discourse which invite, if not actually prescribe,
elucidations in the form of illustrations, examples,
definitions, and glossaries. A further potential
advantage of textual communication over the spoken
medium is the reader's ability to go directly,
without preamble or digression, to relevant
information - providing that effective mechanisms of
access (eg. index, headings) are in place.

Several problems associated with the use of
terminology in the process of communication deserve
to be highlighted here. Problems of inconsistency
and inadequate standardisation are fairly well
known, as are the challenges of choosing an
appropriate technical level. What has not been
explored sufficiently so far is the relationship
between a specialist's discourse, with its
relatively high density of terms, and a layman's
discourse, which must contain fewer, as well as
different, terms. In a situation where the layman
asks questions of the specialist (whether directly
or through text), it is likely that the layman's
discourse will contain whole phrases which map onto
single terms in a specialist's more compact
terminology. If, as we know, nominalisation is a
feature of specialist discourse, we must ask how
that relates to the necessarily more verbal
character of the layman's language. We must also ask
how we should deal with the vagueness, uncertainty,
and fuzziness of general language. Interestingly, an
expert explaining his or her subject will use belief
words like "possible", "probable", "likely",
"certain", etc. and value words like "fatal",
"serious", "dangerous", "undesirable" (Hart, 1989),
but texts are more definitive and more authoritative
in nature - they strive to freeze the results of
experiment or thinking - so words like these are
less likely to figure. But just as they figure in
the mind of the specialist, so they are a feature of
the layman's thinking - and also the layman's
questioning discourse. The disparity between speech
and text in this matter is particularly visible in
texts of an explanatory or instructional nature,
where the specialist (author) must not be seen to
waver.

Solutions to the problems of specialist-layman
communication have been worked out in a variety of
approaches ranging from the prescriptive (eg. Basic
English, controlled English, terminological
standards issued by the International Standards
Organisation and its national equivalents,
authoritative dictionary definitions), to the
descriptive (explanation by example, analogy,
synonymy, paraphrase), from the analytical (breaking
down complex notions and procedures into constituent
elements and steps) to the synthetic (abstracts,
chapter summaries). But even if material is in
itself well explained, how does one ensure that the
reader gains access to it in a lengthy text ?

3.Expression of knowledge needs

It has been said that "as a necessary condition to
satisfying the need to obtain information, we must
be able to formulate our informational needs"
(Wessel, 1975:4). The formulation of needs in the
form of questions is a skill which can be developed
(Kukulska-Hulme, 1988), and it is also the most
widespread, natural way of obtaining information.
The ultimate goal of information retrieval from a
computer manual is not merely to extract
information, but to gain understanding, and to do
something with the knowledge gained - to act in an
informed way. "Comprehension is less a matter of
being able to reproduce the facts in a text than of
what one does or is able to do as a consequence of
interacting with the structure of the text." (Smith,
1982:65). The success of information retrieval in
terms of comprehension and action depends on the
questions that an individual asks. This fits in well
with the 'function-content' approach to cognition

favoured at Yale (Galambos et al., 1986), where the emphasis is on the content of cognitive activities - motivation, plans, goals, and outcomes -, and on function - how content will be used subsequently in tasks. In this light, we can view retrieval as a goal-oriented cognitive activity related to real world tasks.

If we recognise that retrieval is related to subsequent action, this may be sufficient grounds to cast doubt on the prevalent tendency in information retrieval to discount verb forms as descriptors in favour of nouns. The practical advice proffered by information scientists for the control of thesaurus terms has for a long time been:- "Terms should be in noun form, and verbs should be avoided" (Aitchison & Gilchrist, 1972:14). The reason for this advice could be that it is necessary to match retrieval terms to terms in a body of specialised texts, which are seen to be predominantly nominal. But if we shift the focus from text to user, we are faced with the possibility that needs might be centred on verbs if they are concerned with action. The same authors state that "... terms arising from questions likely to be put to the system are as important as those taken from the literature ... The questions should be collected from users or from records of questions already encountered" (Aitchison & Gilchrist, 1972:69). Consequently, we might have to work out the relationship between verbs and nouns in this specific situation and in relation to the specialised domain. As will be seen, this is an issue which was very much at the forefront of the terminological investigations reported in later chapters.

4.Access devices and reading skills

In order to fully appreciate the scope and
limitations of retrieval, we have to consider the
strategies available to readers. In a conventional
manual or book, there are a number of typical entry
points to texts. Rowntree (1986) gives a useful
rundown of these "access devices" in materials
destined for self-instruction: *explanatory title,
contents list, concept map/flow diagram, list of
objectives, introduction/overview, links with other
"lessons", (numbered) headings, instructions, verbal
(rhetorical) and visual (typographic) signposts,
tests, summaries, glossary, index.* To those access
devices may be added *keywords and illustrations*
(charts, tables, graphs, maps, drawings,
photographs, diagrams). However, this list does not
necessarily represent what is normally found in
manuals; in addition, manuals may be destined for
reference rather than tuition or self-instruction.
Browning (1984), writing about software manuals,
points to the other side of the coin: even if you
are writing a manual for reference, "your readers
could be technicians or students trying to learn
from your manual because it is the only
documentation available to them" (Browning,
1984:27). She then has this to say about tables of
contents: "Readers scan the table of contents once
and then promptly forget it exists, turning to more
important things like the index" (Browning,
1984:99). In her view, the table of contents is
mainly a way of helping the writer to organise a
text and to ensure that all information is included.
Indeed, it would seem that in reference works the
index is a more important access device. Peter
Hansjörg, a Swiss barrister concerned with making
legal texts (pertaining to environmental

legislation) more readily accessible to those
outside the legal profession, writes this about one
such document: "Ce sont des non-juristes qui vont le
consulter et qui chercheront dans l'index. Il faut
tenir compte de ce fait et prévoir un grand nombre
de mots-clés qui satisfassent aux attentes des non-
juristes" (Hansjörg, 1992:35). So what items should
go into an index ? According to Browning,
"important" words or phrases, and "not only words
that actually appear in the text but also related
words or synonyms that might be significant to
readers" (Browning, 1984:115). In an otherwise
perspicacious book which rightly emphasises the
importance of indexes, the vagueness of this well-
meant advice is all too apparent.

Knowledge retrieval is also dependent on reading
skills and strategies, including identification cues
to do with text type and language processing
heuristics (word recognition, syntactic/semantic
cues, anticipation), skimming and scanning of larger
chunks of text, and the ability to follow links or
references. It may be supported by note-taking and
shorthand skills, and the use of dictionary
reference tools for decoding. It is related to
motivation, thinking and reasoning processes.
According to Hart (1989), risk is extremely
important in reasoning: "A low probability high risk
situation might warrant investigation before a high
probability low risk one" (Hart, 1989:111). Risk
could be a significant factor in the selection of
knowledge for retrieval; in the domain of security,
a preoccupation with risk is practically self-
evident. Retrieval is thus part of a larger process
of real world problem-solving. Nowadays, problem-
solving strategies are considered to be domain-
specific rather than general, with problem-solvers

116

moving between detail and overview in an arbitrary fashion. It is not possible to predict exactly how a text will be used for problem-solving, but both detail and overview may be required.

5.Knowledge assimilation

We are constantly reminded of the nature of retrieval as a recurring event in a process of knowledge acquisition. On the one hand, "understanding...is necessarily based on what we already know of the world" (Abbott & Black, 1986:123) - prior knowledge being used to connect related elements, create explanations, make predictions, ignore irrelevancies. On the other, the results of retrieval must be assimilated into and alter existing knowledge structures. Factors which come into play in understanding and learning include motivation, learning strategies and learning ability. Gagné & Briggs (1979) point to the different varieties of learning in schools: intellectual skills and strategies, information, attitudes and values, motor skills. The success of learning depends in part on the accurate definition of learning objectives, and the same authors draw attention to the need for precise language in the definition of objectives: for instance, choosing verbs carefully to describe specific intended capabilities (eg. "discriminate", "identify", "classify") and actions (eg. "match", "name", "define"). Readers of manuals define their own objectives (- not always a crisp definition), and their problem-solving situation provides the motivation for learning.

6.Computer security: knowledge transfer problems

Earlier, we introduced the idea of 'knowledge-building', a process of knowledge acquisition and assimilation. Fig. 2 illustrates the development of knowledge; an individual may, however, tread a different path, starting with experience rather than theoretical or formal knowledge, and in some disciplines observation is the starting point.

The essential problem of knowledge transfer through specialised texts is that the reader (novice, learner) is trying to 'tap into' the expert knowledge of the author, not having gone through the process leading up to expertise. In other words, the reader is seeking a shortcut to expert knowledge, but at the same time is experiencing uncertainty and needs opportunities to relate new knowledge to experience and to prior knowledge. A pedagogical or explanatory text will recognise these needs, but a reference manual (or a manual which assumes a certain level of knowledge) will not. This is potentially a problem. A reader's knowledge needs will reflect uncertainty and gaps in knowledge. How is this to be mapped onto the confident expertise reflected in the text ? The answer must surely involve both knowledge and language.

The domain of computer security brings with it certain additional problems, which can be summarised here:-

- problem-solving orientation: knowledge is closely allied to action, so must in turn be 'translated' into action

```
NOVICE      ->  LEARNER                          ->  EXPERT

THEORIES    ->  PRACTICE     ->  Repetition      ->  EXPERTISE
FACTS           APPLICATION      (experiential)      HEURISTICS
                EXPERIENCE                            PROTOTYPICALITY
                OBSERVATION
                                 Comparison
                                 (intellectual)

NEW THEORIES    <-           <-                   <-
NEW FACTS

CERTAINTY   ->  DOUBT        ->  UNCERTAINTY      ->  CONFIDENCE
FACT                             conditions           based on
ASSUMPTION                       restrictions         experience
                                 considerations       and
                                                      knowledge
                                                      of
                                                      probability
```
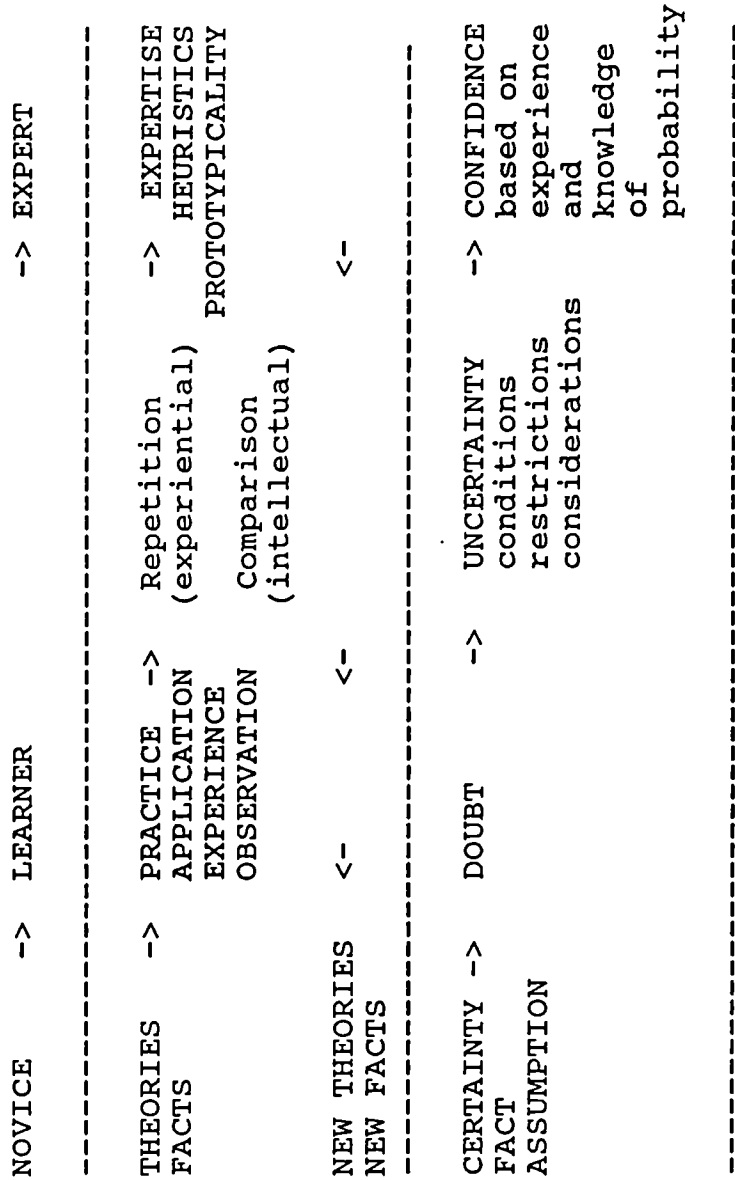
Fig. 2 The development of knowledge

- abstract notions (abstract language) for strategic
planning coupled with practical knowledge (concrete
language) for implementing plans

- relationship to a wider operational context: legal
implications, and people (human resources), combined
with an introspective complexity: technical and
mathematical concepts

- wide range of knowledge types and wide scope of
general and special reference, with ensuing need for
varied methods and language variety in knowledge
transfer

- terminological instability: change and growth, with a
lack of concensus on the meaning of some terms

- terminological obscurity: currency of acronyms
(UPS, EFT, STK, ESD, CSMA, ...) with attendant dangers
of mystification

- general applicability of computer security concepts
(implied by handbooks) versus the need to relate these
to specific computer systems, specific configurations,
specific computer functions

B.Computer environments for knowledge transfer

1.Features of computer-assisted transfer

It is hardly possible to conduct research into the
terminological design of computer manuals without
stressing the point that, with every year that
passes, more and more documentation is accessed on
the screen rather than on paper. According to McGrew
& McDaniel (1989), in their introductory work on on-
line text management, an "on-line text access
system" can be

> "any type of on-line information system. It
> can refer to on-line help for a specific
> application, a computer-aided instruction course, or
> the software used to develop such a course. Or, it
> could be an on-line system for retrieval of the type
> of information traditionally found on paper, such as
> user manuals and reference documents."
> (McGrew & McDaniel, 1989:3)

Access can take place in a "free-standing" text
environment, or one that is "context-sensitive" -
that relates information to the task at hand.
Documents can have a traditional, linear
organisation, or be structured in a hierarchy or a
network, or have associative links; "hypertext"
techniques can use all these options. *Hypertext
systems* embody object-oriented documenting, i.e.
document elements such as words, phrases, sentences,
or paragraphs become objects which can be linked to
other objects, giving the user access to related
topics. *On-line tutorials* are usually "constructed
so that the user is guided from topic to topic in a
directed manner, building on the knowledge
presented" (McGrew & McDaniel, 1989:78); this is a
form of *computer-based instruction*, though the
latter term normally refers to systems which are

121

interactive in nature. There are also *intelligent tutoring systems*, which "adapt to the learning style of the student, providing guidance, instruction, and support in performing complex tasks" (Alberico & Micco, 1990:17). Although the availability of associative navigation can mean that access to information is freer, if document signature (full text) indexing is used, many *noise* references can result.

The nature of human language is such that there are numerous problems associated with keyword-based retrieval and with the automatic processing of text for indexing purposes; Smeaton (1992) has discussed these in an article on the application of natural language processing to information retrieval tasks. In essence, the problems revolve around the basic property of natural language: ambiguity. We made reference to ambiguity earlier, in relation to the identification of knowledge types in text, and now we revisit the problem in relation to computerised retrieval, since automated processing actually amplifies the problem. The culprits are ambiguous words, as well as structures, for instance prepositional phrases, nominal compounds and various forms of conjunction; discourse level ambiguity is also a problem. Nonetheless, progress is being made in improving information retrieval, with the application of techniques such as *conceptual information retrieval* (eg. Wyllie, 1990), *statistical term weighting* (where the likely senses of a word are weighted highly), and *knowledge-based machine indexing* (Genuardi, 1990). The challenge, then, is to design text retrieval systems which genuinely help the user, and do not provide an excessive amount of information, some of which is scarcely relevant. This is where *expert systems for*

information retrieval come in, with their aim of augmenting or emulating the expertise of the intermediary.

Computer-assisted knowledge transfer systems share many of the common linguistic problems which affect all modes of knowledge transfer, but in cases where linguistic sensitivity is all-important - for example, communication between a specialist and a layman - the computer environment is at a definite disadvantage in comparison with person-to-person contact. We must therefore continue to look for ways of describing and emulating human linguistic awareness and skill in responding to the special needs of "special language" (language for special purposes, scientific, technical language) communication. Applications will abound: for instance, in the domain of computer security, Cooper (1989:362) projects a need for "efficient and effective training aids (such as automated interactive systems)".

2.Existing systems for knowledge transfer

It is worth pointing out that, alongside sophisticated, intelligent tutoring and retrieval, knowledge transfer can be enhanced by reading and writing tools which do not necessarily form a complete system, and may have little programmed "intelligence". It may be a question of giving writers feedback about the quality of their technical writing in terms of its "readability", its organisation, grammar, style or vocabulary, based on pre-conceived rules and statistical analysis. In this section, we give a brief description of a selection of different systems and research projects

which have some relevance to the line of enquiry we are pursuing, in order to give a flavour of the variety of possible approaches.

David E. Kieras, at the University of Michigan (Kieras, 1981-87; Mayer & Kieras, 1987; Britton & Glynn, 1989), describes an advanced computerized aid for the writing of comprehensible technical documents resulting from a long-term project supported by the Office of Naval Research and aimed at improving military equipment manuals. The system is designed to help writers in improving the clarity of their writing. It is assumed that both writers and readers have the necessary background knowledge, but both lack "reading skills". The system, based on research in comprehension, aims to help the writer to edit his or her text by detecting problems specified by rules relating to reference, sentence structure, and textual coherence. In a similar vein, J.P. Kincaid (Kincaid et al., 1981) reports a Computerized Readability Editing System (CRES) intended for military (naval) texts. This writer's aid also gives feedback about the quality of writing. Output from the system is an annotated copy of the original document pointing out stylistic problems, non-standard terms, and giving a Kincaid-Flesch readability score.

Bell Laboratories (Macdonald et al., 1982) are where the more generally applicable Writer's Work Bench (WWB) originates. It provides global statistical information about a document, including scores for several readability formulas, part of speech statistics, sentence lengths, and statistical comparison with "model" texts. At the University of California, Morton Friedman's system, WANDAH, (commercially published as 'HBJ Writer'; Friedman,

1987; Britton & Glynn, 1989), is designed to assist student writers in all phases of writing (planning, transcribing ideas, revising). Based on research by Hayes and Flower, the planning aid prompts the writer to reveal purpose, identify audience, and outline ideas. The revising aid addresses problems of grammar, style, and thematic organization; it can highlight specific stylistic features of the text (eg. abstract words, transitional words and phrases, pronouns), and will produce an outline based on the first sentence of each paragraph, or sentences selected by the writer. An innovative project in Computer Assisted Writing Techniques has been reported by Jeannine Beeken (University of Leuven; Beeken et al., 1990). This addresses student and business writing, with two basic components - a questioning procedure (or predefined text-frames for business writing), and a "thesaurus" (system of lexicons) comprising collocations, functional & text-cohesive items (eg. expressions of consequence, concession, comparison, etc.), normative information about lay-out, typography, spelling and grammar, systematic technical terminologies, advice on text structure. The system will include visual tree diagrams of selected text structures, where nodes can be amended through a 'zoom-in' facility. Reading, writing, reference manuals, online help/documentation and distance learning are all applications envisaged by Victoria A. Burrill, University of Reading (Burrill in: Van Vliet, 1986) for her system, VORTEXT, which tries to map the design of a paper book onto the computer screen, with an open page in the centre and closed pages either side (with headings). This gives the reader orientation within the system and within the subject matter. Text can be concorded and headings indexed, and speed of text presentation can be varied.

Moving on to a system concerned with question-answering based on knowledge from texts, Boris Katz (AI lab of MIT; Katz, 1988) has reported START, which analyzes English text and automatically transforms it into a formal representation (the 'knowledge base'), incorporating the information found in the text. The user can query the knowledge base in English, and the system's response is also in English. It has been used in a number of domains, including medicine, politics, space (Mars observer mission), vision, common-sense physics. An example of an intelligent aid to bibliographic information retrieval is the Austrian SAFIR (Smart Assistant for Information Retrieval; Brückler et al., 1988), a system with a user interface which adapts to different user levels and provides "sufficient help" when needed; it has a single command language which translates into different host languages; online information about hosts and databases; and a domain knowledge base with knowledge acquired from the user (including a record of all search histories), used for defining the model of the information need.

Alberico & Micco (1990) describe a system by B.C. Vickery & H.M. Brooks (PLEXUS - A Knowledge-Based Reference System) which represents knowledge about terms and concepts related to gardening; the sources of knowledge are printed reference works, human experts and gardening associations. Using PLEXUS, which has a natural language interface, involves matching terms in queries to terms in its dictionary. Another expert system for bibliographic retrieval has been developed by Gauch & Smith (1993), focussing on (Boolean) "query reformulation" (eg. broadening, narrowing, changing query structure) resulting in improved efficiency. Bordogna & Pasi (1993) propose the use of linguistic

descriptors to specify the degree of importance of terms, in preference to more traditional numeric query weighting.

Ingrid Meyer, Doug Skuce et al. in Ottawa (Research Grant Request, 1989; Skuce, 1993) have developed a knowledge-based approach to conceptual description in terminology, based on Meaning-Text theory, and using CODE, a generic conceptual analysis tool, with graphical representation of conceptual relations. Ivar Utne (Utne, 1987), at the Norwegian Term Bank, describes a terminological databank system which has, apart from a "conventional" thesaurus, conceptual networks of static and procedural information, represented in frames and schema. Frames include the relations 'part of', 'cause', 'sequence', 'connected to', etc. and various attributes (colour, shape, etc.); the procedural schema contain information on events/actions (eg. verb with related roles: agent, object, place, result) and on series of events or actions.

There is a also a rich body of literature reporting research and development work on computer-based learning and intelligent tutoring systems. An interesting experiment has been carried out by Ford & Ford (1992) at Sheffield University, where a simulated expert tutoring system was used to find out about learning strategies; a qualitative analysis of learners' question data relating to the Precis package revealed 11 categories of question, grouped into four general categories, and four levels (general to detailed), which could be used to predict more and less successful learning strategies.

Further information on the fast growing fields of

computer-aided reading and writing, intelligent
computer-aided instruction, intelligent retrieval,
and knowledge representation systems which in one
way or another facilitate knowledge transfer can be
found in a wide range of periodicals, for instance
"Intelligent Tutoring Media", "Online Review", "The
Electronic Library", "Advanced Information Systems",
"Interacting with Computers", "Information and
Software Technology", "Journal of Artificial
Intelligence in Education", "Expert Systems",
"Computers and Education", "Machine Learning",
"Instructional Science", "User Modeling and User-
Adapted Interaction", "British Journal of
Educational Technology", "Information Processing and
Management", and others.

3.Searching on the IBM AS/400 system

The possibility of electronic searching on the
AS/400 has been mentioned earlier; here we must
pause to consider the "search index" and
"BookManager" facilities from a terminological
perspective. The "search index" option on the system
may work well for a very experienced user, but for
someone less experienced or with little knowledge of
security, there are problems:

(a) the system's ignorance of orthographic
variation,
        eg. 'backup' yields a list of topics, the
            first of which is 'Backup guidelines';
            'back up' yields the same list, but
            without the guidelines;
            'backups' yields nothing at all.
            A user who types 'back up' and 'frequency'
            will not get the guidelines, which deal

with frequency of backups.

(b) appropriate terminology expected,
     eg. 'confidential mail' and 'personal mail'
       are acceptable, but 'other peoples' mail'
       is not (as in the query: "Can I look at
       other peoples' electronic mail ?")
     eg. 'hack'/'hacker'/'hacking' are not known
       to the system

(c) accurate terminology expected
     eg. An answer to the query "Is there a way to
     display all commands that a user has *authority*
     to ?" is provided by the 'display user
     permission' command, not the 'display
     authorized users' command, which is potentially
     confusing.

The last example also shows how prior knowledge is
presupposed. Faced with a list of topics as a search
result, the user has to choose from that list, and
so must have an idea of what each topic is about.
The search facility cannot deal with conceptually
complex queries (eg. "Can users be automatically
deleted if they do not sign on for X months ?"; "Can
a user 'passthrough' to another system and gain
greater authority than he has on the original system
?"); it cannot deal either with knowledge not
represented, eg. the AS/400 is apparently immune to
virus attacks - but users, not aware of this, may
still wish to ask about this potential hazard. The
search index appears not to know about viruses.

The BookManager retrieval software for the CD-ROM
version of the security manual, despite many
excellent features - including a sophisticated
ranking facility (by location, frequency, exactness,

uniqueness, sequence similarity) - , shows
weaknesses similar to those of the help index. For a
novice user, or one unfamiliar with security
concepts and terminology, attempts to use 'everyday'
words as search terms will prove futile: items like
'disaster', 'illegal' or 'piracy' result in no
matches at all. A "word check" can be performed
before searching, and will show words with similar
spelling, but offers no thesaural or other
substitution (eg. for 'right' it does not offer
'authority' or 'permission', etc.). So far, the
search facilities available have not addressed, much
less resolved, the terminological knowledge transfer
issues which are the subject of our research.

Chapter IV

Empirical study - the reader's perspective

A.Reader survey

1.Objectives and method

In order to find out more about the relationship
between the natural language used to express
knowledge needs in the area of computer security,
and the language of computer manuals, it was decided
to establish a corpus consisting of questions
formulated by users, which could be analysed
subsequently from a terminological and knowledge
type perspective.

The corpus data was collected by means of a survey.
Gathering truly 'natural' language data is always a
very difficult undertaking. It was important to give
the user the opportunity to express queries
spontaneously, without, as far as possible, having
undue constraints as to content or form. This is
different to taking data from an on-line (or
telephone support) query answering service, where
answers are provided by either people or computers,
in that an initial filtering process would have
already taken place - eg. a user turning to the
support service *after* other avenues (colleagues,
manuals) had been exhausted - whereas a survey could
capture queries as soon as they arose in the mind of
the user. Admittedly, asking a user to write down a
number of queries (for the survey) is in some way
artificial. Firstly, in real life situations,

queries are not always uttered: some remain at the
stage of thought, in which case their form cannot be
described reliably - it may not even be linguistic.
In a linguistically-oriented project, it is not
possible to deal with such data. Secondly, it would
be excellent if queries could be captured as and
when they arose, over a period of time, in various
circumstances; such an ideal, however, cannot be
achieved without considerable input from a truly
dedicated (yet large enough) group of users.

As users of the AS/400 will, in principle, have
access to manuals, we can also refer to them as
"readers", where the term is understood to denote
"actual and potential readers". Clearly, it would be
wrong to suppose that reading a manual could ever
provide the answers to all queries. There are bound
to be questions which will not be answered by the
manual, for reasons of scope as well as the
knowledge representation problems which have already
been discussed. However, a reader cannot be expected
to know or accurately judge whether or not a
particular query can be answered in this way. This
is why it is important to submit all queries to
analysis, without prior filtering. Having said that,
the survey respondents were then asked to indicate
where they would expect to find the answer to each
query - so that a link between query type and manual
consultation could be established and explored.

It was decided that a user's level of knowledge
would be established by asking about professional
experience, knowledge and experience of computers in
general and of the system in particular, in addition
to knowledge of computer security. Users were also
asked about their use of a number of different
information sources on computer security, and in the

main part of the survey (hereafter referred to as Section II), they were instructed to write down, as they would "spontaneously say them", twenty questions they "could ask or could be asked" about the security of their computer system.

The self-completion questionnaire (Appendix I) yielded a corpus of 334 user questions (hereafter referred to as "queries"), representing 3,586 token words. The number of queries supplied per user ranged from 20 (in 6 cases) to 0 (in 1 case), with an average of 9.8 queries. The questionnaire had been distributed to a sample of 76 users known to have some responsibility for security on their AS/400 system. Attempts to obtain further lists of AS/400 users from sources such as IBM, the AS/400 Computer Users Association, several IBM agents, and the NCC in Manchester - despite initial good will - proved futile. The problem appeared to be two-fold: firstly, the desire to protect data held on computer files (not wishing to provide names and addresses where there was uncertainty about users' consent - a case of good security practice !), and secondly, the nature of the domain of computer security - peoples' reluctance to risk revealing gaps in their knowledge of the subject, despite an assurance of confidentiality. However, the main thrust of the research is not quantitative (how many users), but rather qualitative (how are needs expressed).

The response to the survey, after a pilot phase and telephone reminders, was 34 returns, a response rate of 44.73 %. Written and telephone comments from both respondents and non-respondents revealed several reasons for non-completion: (1) time pressures in the business environment, especially the time required to think of 20 queries, (2) a reluctance to

acknowledge a degree of responsibility for security, where it was not unequivocally allocated, (3) a feeling of knowing too little about the subject - inability to think of 20 queries, (4) knowing too much about the subject to think of questions, or already having well-established security procedures, (5) long-term illness (in 2 known cases), (6) user of PC's only (in 1 known case), (7) (speculative) fear of being "shown up" - gaps in knowledge.

2.Reader profiles

The survey was addressed to AS/400 users, with a covering letter explaining that it was aimed "particularly at less experienced computer users", though users with more experience were not excluded. Typically, it is inexperienced users who have the most problems finding information, and the greatest fear of manuals. No age, sex, or professional status criteria were imposed. The users surveyed were based in a number of different business companies; completed returns were from users in professional firms - chartered accountants, and software development (the majority being in these two categories), and in manufacturing/ distribution. In these companies, the AS/400 was known to be used for business, administrative, manufacturing and software development purposes. Geographically, the companies were located throughout England and Scotland.

The results of the survey showed that all respondents considered themselves to have some knowledge and experience of computers. In each of the four categories specified (*i.e. knowledge and experience of computers in general; knowledge of the AS/400 operating system; of computer security in*

general; of security on the system), around half
(47% - 53%) described their knowledge as being
"good". In the majority of cases (85%), respondents'
knowledge of computer security in general was
described as being either good (47%) or limited
(38%), and knowledge of AS/400 security was mostly
good (47%), but also very good (32%), limited (18%),
or very limited (3%). As regards job titles and
professional experience, 18% had a job title not
directly associated with computing (eg. Secretary,
Partner), although in one instance the respondent
had a computing background. Of the remaining 82%,
29% had no previous professional experience in
computing. Currently held computer-related positions
ranged from administrative/ clerical to managerial
(project manager, product manager, technical
director), and included programming, analysis,
operations, and support.

It is interesting to speculate that it may be very
hard indeed to find users with no previous knowledge
of computer security whatsoever, if we consider
that: "Fire and smoke detection alarms are familiar
to most of us because of the availability in recent
years of low-cost home devices. The technology for
computer facilities is similar" (Cooper, 1989:80).

B.Analysis of readers' needs

1.The language of queries

a.Lexical and terminological analysis

A lexical and terminological analysis of the query
data in the survey (Section II of questionnaire) was
undertaken in order to study the word types

occurring in the data in relation to knowledge
needs. The overriding objective was to piece
together, by observing the data from a number of
complementary angles, a picture of the knowledge
needs expressed through different linguistic
elements and devices.

The analysis was begun by segregating words into
grammatical categories. It was decided to leave
aside, initially, some grammatical words, thus
focussing on truly "lexical" items, referring in
this matter to Carter (1987), who makes clear the
distinction between "grammatical" words and
"lexical" (or "content") words. The items set to one
side were articles, prepositions, pronouns,
conjunctions, numerals (other than in names or
codes), and interrogative pronouns. Interrogatives
were subsequently scrutinised in the rhetorical
analysis, and other pronouns and conjunctions were
also identified as having a role in conveying
knowledge needs. Wilbur & Sirotkin (1992) have, in
their own way, challenged the traditional notion of
"stop words" by removing, for the purpose of
improving retrieval, words identified by a vector
method of similarity measure.

For the numerous word forms which could represent
more than one part of speech (eg. 'access',
'control', 'secure', 'change', 'audit', 'damage',
etc.), concordances were run on the computer to
check actual usage in the survey data.

(Note: spelling mistakes occurring in the original
data have been retained, for instance "to setup"
instead of "to set up".)

136

Noun forms

(Note: many nouns are used adjectivally, eg. 'capital' investment, 'user' ids, 'history' log, 'security' measures, etc.)

access, administrator, advantages, air conditioning, alarm, amount, application, application's, areas, arrangements, attack, attempts, audit, auditor, authorisation, authorities, authority, backup, backups, basis, batteries, boot, breaches, building, cabinet, calendar, capability, capital, card, chances, change, changes, chars (=characters), checks, checkers, classifications, clause, colleagues, command, commands, comms, communications, companies, computer, computers, consultation, contracts, control, copies, CPU (=central processing unit), crash, damage, dangers, data, data base, day, days, default, delay, dept, desk, desks, detection, detectors, device, difference, disaster, document, documents, DP(=data processing), EDI(=electronic data interchange), electricity, employment, environment, environments, encryption, event, expansion, expenditure, experience, extent, facility, faults, field, file, files, fire, flexibility, folder, folders, force (in force), freedom, frequency, function, functions, generations, grades, group, groups, hacker, hackers, hardcopy, health, help, history, holders, host, hours, I.D. (=identification), Id's, ids, impact, importance, individual, info, information, infringements, integrity, intervals, job, key, lengths, level, levels, library, libraries, life, lifespan, light, limits, line, link, list, lists, log, logs, loss, m/cs (=machines), machine, mail, management, market, master, measures, media, menu, messages, methods, modem, month, months, need, network, networks, number, object, objects, office, officer, on/off site, options, organisation, output, outside, overtime, owner, pad, pain, panel, parameters, part, partner, parts, password, passwords, payroll, PC, people, peoples', period, person, personnel, place, places, plan, point, police, police station, portables, position, practice, precautions, premises, principles, printers, problem, problems, procedure, procedures, product, products, profile, profiles, program's, programs, proof, protection, quality, queue, rationale, recovery, replacement, reports, resignations, response, restrictions, reuse, rights, rules, safe, safety, saves, screen, screens, security, set, signon, simulation, site, sites, situation, smoke, software, solutions, sort, source, specs (=specifications), spool, staff, standard, status, steps, suite, supply, switch, system,

137

system's, systems, tape, tapes, terminal, terminals, theft, time, times, traffic, trail, updates, UPS (=uninterrupted power supply), useability, user, user's, userid, users, value, virus, viruses, water, way, ways, windows, work, workstations, 3rd party

Proper nouns, function names

AS400, IBM, Lotus, FAST (= Federation Against Software Theft), Office Vision, PC Support, QSECOFR, SECOFR, DSPAUTUSR, SYSOPR, QSECURITY

Verb forms

(NB: some past participles of verbs are used adjectivally, eg. 'shared' folders, 'perceived' delay, etc.; some present participles function as nouns, eg. 'training')

able (to be able), access, accessed, accessing, activated, add, affected, allow, allowed, allowing, am, amend, are, aren't, arise, arrange, audited, authorised, authorized, back up, backed up, backing up, be, becomes, been, being, bolt, book, bother, breached, breaching, break, breaks down, build, can, can't, care, catch, caused, change, changed, check, checked, classed, communicating, conceal, conform, connected, considered, contradict, control, convince, corrupted, could, cover, create, cut off, decrease, decrypted, define, deleted, deny, denied, destroy, detect, determine, dial in, disabled, discover, display, displaying, do, does, doing, done, duplicate, educate, enable, encourage, encouraged, enhance, enroll, ensure, ensures, evaluate, exist, exiting, expand, expect, expired, expires, find, find out, force, forget, forgets, forgotten, found out, found, gain, gained, gave, generate, get, give, given, go, goes, got, grant, guaranteed, guard, hack, hacked, happen, happened, happening, happens, has, have to, have, haven't, having, identify, include, increase, indicates, indicating, informed, installed, invoke, involve, is, keep, kept, know, knowing, lead, leave, leaves, leaving, limited, lock, locked, look, looked, looking, lost, made, make, making sure, manipulated, may, mean, might, monitor, must, necessitate, need, needed, offer, operate, override, owns, pass through, passthrough, passing-through, perceived, persuade, plan, preserve, prevent, prevented, protect, protected, put, raid, read, receive, recover, reduce, reducing, reinstate, related, remain, remove, removed, required, restore, restored, restrict, restricted, return, reviewed,

138

running, saved, secure, secured, see, sees, set up,
setup, setting up, set to, share, shared, should,
show, shut, sign on/off, signon, signed on, signing
on, sit, specify, spread, spreading, start, stop,
stored, storing, suffers, supplied, support,
suppress, take, taken, tell, tendered, terminated,
test, testing, tether, think, trace, training,
travel, tried, try, trying, un-manned, updating,
use, used, using, viewed, viewing, walk, want, was,
will, won't, work, would, wouldn't

Adjectives

adopted, all, any, automatic, available, aware,
best, better, breakable, certain, common, complex,
comprehensive, confidential, detailed, different,
due, easy, efficient, electronic, encrypted, every,
existing, expensive, external, first, foreign, free,
front, full, general, good, greater, illegal,
illicit, inactive, incoming, internal, invalid,
live, local, long, magnetic, main, major, many,
midrange, more, most, necessary, new, normal,
obvious, OK, old, one, organisational, original,
other, outside, own, particular, personal, physical,
possible, potential, practical, present, prior,
public, real, red, regular, remote, resident,
resilient, safe, same, secure, secured, senior,
sensitive, separate, shared, some, specific,
standard, sure, third, total, unauthorised, visible,
vulnerable, wrong

Adverbs

again, always, automatically, away, back, best,
completely, easily, even, ever, far, frequently,
inadvertently, internally, just, last, long, often,
once, only, periodically, physically, quickly,
really, remotely, still, there, twice

One striking feature resulting from this categorisation is the relative abundance and variety of verbs, even allowing for the nature of the language sample (short, simple sentence forms), and the fact that verbs take on more forms than nouns. This suggests that the needs expressed in the questions are associated with _knowledge about actions or events_. The suggestion is further supported by the fact that a substantial number of the nouns in the data are derived from verbs and indicate an action or the result of an action (protection, authorisation, restriction, consultation, detection, expansion, simulation, communication, classification, resignation; replacement, arrangement, infringement, employment). In addition, many of the forms occurring as nouns also occur as verbs, or have the potential to function as verbs (breach, damage, delay, function, access, supply, etc.).

When we examine the nature of the verbs used, we can see that most of them express _active control over actions or events_:

prevent	restrict	secure (against)
preserve	change	monitor
protect	ensure	increase/decrease (access)
conceal	make sure	deny (access)
reinstate	detect	remove
shut	start	stop
cut off	force	suppress
identify	delete	destroy
add	activate	amend
involve	include	limit
check	allow	own
lock	review	arrange
give	keep	put
convince	educate	persuade
take	create	enhance
operate	set up	restore
back up	sign on/off	shut down
evaluate	expand	encourage

bolt	connect	tether
man	audit	book
cause	communicate	consider
exit	guarantee	inform
install	invoke	think
test	train	try
view	store	find out
find	enroll	build
catch	duplicate	grant
guard	manipulate	recover
reduce	save	see
override	disable	define
cover	control	generate
authorise	specify	tell
trace	update	want

A smaller number reflect (in the actual data)
*observed or anticipated events, or accidental
actions*:

hack	break	raid
breach	read	gain (access)
decrypt	access	dial in
contradict	corrupt	travel
tender	pass through	spread
supply	offer	display
show	bother	care
happen	receive	suffer
forget	get	remain
discover	arise	expire
become	exist	terminate
walk	lose	

Some indicate *possibility, advisability, obligation*:

necessitate	be able (can, can't, could ...)
need	have to (must ...)
require	ought to (should ...)

Other verbs are ubiquitous in nature (be, do, have,
leave, go, know, make, use, reuse, relate, work,
support, share, indicate, perceive, mean).

Passive forms occur infrequently in the data. Where
they do, the meaning is often active, for example:

"Can users be prevented from signing on ... ?"

actually means

141

"Can I prevent users from signing on ... ?"

In terms of frequency, the most prominent verb forms
(clustered by meaning) are:

be is are been being am was
can could can't able
do does done doing
have haven't has having
should

Two of these clusters signal POSSIBILITY or
ADVISABILITY (can and should).

Other frequently occurring verbs are:

sign on / sign off
change
access
prevent
will won't wouldn't
use
check
find
happen
know
breach

These reflect the make-up of the total verb list,
with a bias towards *active verbs.*

As the distinction between nouns and verbs can
obscure frequently occurring concepts expressed in
both forms, and as this can be compounded by
variations in spelling, it should be pointed out
that the concept of 'backup' has a particularly high
overall frequency, occurring in the forms:
    back up backup backups backing up back-up
    backed up backed-up

When we move on to examine the nouns occurring in
the data, the following categories can be seen to
emerge:-

142

ABSTRACT NOUNS OF A GENERAL NATURE

difference, importance, extent, part, way, method,
measures, practice, procedure, steps, rules, basis,
rationale, frequency, time, hours, interval, period,
length, month, reuse, area, place, arrangement,
flexibility, freedom, chances, control, principles,
pain, situation, change, need, light, force,
experience, limits, advantages, set, link, point,
sort (of), number, amount, days

NOUNS RELATING TO BUSINESS FUNCTIONS

company, contract, clause, employment, overtime,
payroll, staff grades, job specs, resignations,
office, department, authority, problem, solution,
document, capital, expenditure, information,
management, market, simulation, audit, Total Quality
Management, work, organisation

NOUNS RELATING TO PEOPLE AND ORGANISATIONS

users, staff, people, person, personnel, group,
security officer, master, administrator, auditor,
colleague, key holders, hacker, 3rd party,
individual, partner, police, FAST, IBM

NOUNS RELATING TO COMPUTING AND COMPUTER
APPLICATIONS

system, data, data base, PC, DP, EDI, computer,
machine, portables, workstation, software, programs,
product, environment, backup, command, folder, menu,
communications (comms), device, media, facility,
help, library, network, host, site, option, profile,
screen, character, calendar, capability, command
line, history log, CPU, default, generations,
messages, traffic, saves, tape, object, function,
application, output, queue, spool, hardcopy, file,
field, (illicit) copies, mail, key, update, response
times, lifespan, switch, signon, system value,
AS/400, Lotus, Office Vision, PC Support, QSECOFR,
DSPAUTUSR, SYSOPR, QSECURITY

NOUNS RELATING TO SYSTEM SECURITY

security, password, virus, security level,
authority, authorisation, restriction, UPS, UPS
expansion, batteries, long life, electricity supply,
smoke, water, fire, cabinet, safe, windows, air
conditioning, alarm, computer suite, desk, detector,

premises, site, building, fault, recovery, recovery
plan, check list, integrity, classification,
magnetic card, key pad security, I.D., userid,
status, audit trail, access, external source, crash

NOUNS RELATING TO BUSINESS SECURITY

disaster, health and safety, precaution, protection,
replacement, right, danger, breach, attack, proof,
detection, infringement, attempt, loss, theft,
damage, delay, impact

These nouns, many of which change their meaning
according to context, have been classified with
reference to the contexts in which they actually
appear in the data. Even so, the fluid nature of any
classification must be emphasised: a computer
environment *becomes* a business environment, and
words like 'system', 'file', 'message' or 'mail' can
no longer be assigned to a single domain. Similarly,
a word such as 'signon', though a typical computing
term, has special significance when considered in
the light of security, and so belongs to both
domains. The set of nouns includes computing terms
of various degrees of specialisation (with 'help' at
one end of the scale, 'EDI' at the other), and
security terms which relate to the functioning of
the computer system. It also covers business
concerns in a wider sense, inasmuch as the
functioning of the system affects the functioning of
the business, and has particular implications for
its staff and for its legal ramifications.

Of particular interest are the abstract nouns of a
general nature which provide a clue as to the types
of knowledge represented in the data. Two meanings
dominate:

144

- a concern about HOW actions are to be performed
('way', 'method', 'measures', 'practice',
'procedure', 'steps', 'rules', 'basis', 'rationale',
'principles')

- a preoccupation with timing - WHEN ('frequency',
'time', 'hours', 'interval', 'period', 'month',
'reuse', 'days').

The hazardous nature of security is also in evidence
('chances', 'freedom', 'change', 'control'), and it
is worth noting in the data the presence of "value-
laden" words ('disaster', 'danger', 'pain').

The adjectives present in the data suggest knowledge
needs which seek to establish distinctions by way of
contrast and comparison: the evidence for this may
be found firstly in the presence of OPPOSITE pairs:-

(a) opposites of a general nature
'total', 'comprehensive' vs. 'specific',
'particular'
'same' vs. 'different'
'new' vs. 'old'
(b) opposites concerned with threats to security
'resilient', 'secure' vs. 'vulnerable'
'personal' vs. 'public'
'internal' vs. 'external',
'outside', 'incoming'

Secondly, there are a number of NEGATIONS which
imply an opposite:-
'non standard' => 'standard'
'non breakable' => 'breakable'
'unauthorised' => 'authorised'
'illegal' => 'legal'
'invalid' => 'valid'

And lastly, COMPARATIVE and superlative adjectives
are in evidence:-

     'greater' (greater authority, greater access)

     'better' (better methods)

     'best'   (best virus protection software, best
                methods)

     'more'   (more than one user)

     'most'   (most common ways)

Adjectives may also be classified in the following
categories which mirror the classification of
nouns:-

(1) ADJECTIVES OF A GENERAL NATURE

     general, comprehensive, specific, particular,
     detailed, every, some, many, all

(2) ADJECTIVES RELATING TO BUSINESS FUNCTION

     efficient, easy, practical, good, OK, wrong,
     expensive, long, total, complex

(3) ADJECTIVES RELATING TO PEOPLE AND ORGANISATIONS

     organisational, senior, aware, own, shared

(4) ADJECTIVES RELATING TO COMPUTING

     electronic, magnetic, midrange, main, remote,
     encrypted

(5) ADJECTIVES RELATING TO SYSTEM SECURITY

     confidential, sensitive, personal, public,
     secure, sure, vulnerable, non breakable,
     resilient, unauthorised, virus free, illegal,
     illicit, invalid, physical, full, safe,
     adopted, visible, red

(6) ADJECTIVES RELATING TO BUSINESS SECURITY

     external, internal, outside, incoming, local,
     major, real

The adjectives relating to business function are
largely concerned with the smooth running of the

business, i.e. the best way of implementing security ('Is there an easy way to ... ?', 'What practical steps need to be taken... ?', 'What is the most efficient way ... ?', etc.)

Other adjectives concern

TIMING (including repeating or lasting features)

automatic, regular, prior, resident, common, standard, normal

POSSIBILITY or NECESSITY

possible, potential, available, existing, present, necessary

The <u>adverbs</u> again reflect the types of knowledge needs which can be seen to recur in the data:-

TIMING - frequently, often, periodically, quickly, automatically, twice, long, last, still, ever, after, before, again, always, regular (= regularly), then

MANNER/PURPOSE - completely, best, remotely, easily, internally, inadvertently, back, physically, so (that)

'Inadvertently' reiterates the hazardous element which characterises computer security.

COMPARISON ('very') and RESTRICTION/ NEGATION ('with', 'without', 'no', 'non', 'not', 'only') can also be identified.

Pronouns and conjunctions

Pronouns are an important feature of the user's discourse: they stand for the nouns the user cannot label, or the knowledge which is only vaguely

defined. Many 'people words' occur in the data in the guise of pronouns:-

> everyone, everyone else, anyone, anyone else
> someone, no one, everything

Pronouns can be used in the search for knowledge about people and events ("How resilient is the system to *someone* trying to hack their way in ?"; "How do I control *who* sees *what* ?") or a search for a definition of quantity ("How *much* is there on the system I can't even see ?").

Relative pronouns establish relationships between objects and actions ("objects *that* he/she owns"; "a document folder *that* only I can access"; "user ids *which* a hacker could see") and relate timing to events ("every time (*that*) I leave my computer"). Likewise, conjunctions show relationships between elements of knowledge:-

CONDITION/RESTRICTION   "if", "if so", "unless"

> What do you do if someone's password expires ?
> Can viruses travel through networks - if so,
> are there resident checks to monitor/ ensure
> the health of incoming data/ mail ?
> How can I prevent a user from updating an
> application's set of files unless they are
> using the application ?

OPPOSITION   "yet", "but"

> Why is spool file security so complex ? Want to
> stop people viewing spool yet be able to
> control printers.
> Can users be given access to commands but still
> be 'limited capability' ?

TIMING   "when", "while"

    What should happen to the system when the
    office is un-manned ?
    Can we back up security while people are signed
    on ?

COMPARISON   "than"

    Can a user "passthrough" to another system and
    gain greater authority than he has on the
    original system ?

COORDINATION   "and"

    Can I allow Office users to share a signon I.D.
    and have separate passwords ?

This lexical analysis of the query data shows that
the lexical items used belong to a number of related
domains, and it highlights in particular the fact
that a high proportion of items are general language
words with semi-specialised meanings: the language's
internal loan-words, which are easily assimilated by
users. Any terminological analysis of the data,
therefore, has to be based on a very broad
definition of 'terms', as the low technical or
specialised content of these items is the very
essence of the data.

In trying to establish the range of potential access
paths to knowledge of computer security, we have to
investigate the links formed between the user's
question and the target knowledge base (i.e. sources
of knowledge on computer security, such as manuals)
through the intermediary of terms. It is accepted
that terminological concepts can have a number of

denominations, so that for instance the concept of 'user identification' may be designated by the terms 'I.D.' and 'user id' in their various forms. Similarly, 'back up' and 'save' may be observed being used interchangeably for the same concept. Traditionally, due to the nominal nature of specialised texts, terminologies have been perceived as collections of nouns. It is clear that in the case of users' discourse, other parts of speech play a very important role in conveying knowledge needs.

One could argue, furthermore, that it would be a mistake to try to isolate, at all costs, the lexical and terminological items expressed in users' queries from their connotational and collocational contexts. At the most simple level, this means recognising that, for example, the term 'audit' is connotationally (and paradigmatically) linked to 'auditor', and collocationally to 'audit trail'. It is, after all, very common for terms to constitute "multiword units". But it also means allowing terms to present themselves in their natural *entourage*: "default passwords" will at some stage be "reviewed", "damage" has to be assessed in terms of its "extent", and so on. When trying to establish a valid mapping between questions and answers, we may have to take these larger units into account.

In terms of frequency of occurrence, the top nouns are:-
> security
> system
> user(s)
> password(s)
> access
> data

By the same criterion, two further nouns, "backup(s)" and "authority(ies)", represent very

important concepts, if one takes into account their verb forms (to back up, to authorise); adding synonymous terms ("save", "right") increases their frequency even more.

b.Intentional and rhetorical analysis

An intentional and rhetorical analysis of respondents' queries was carried out in order to discover the knowledge needs of users as revealed by the types of question asked. The approach taken was to examine entire queries (Appendix II), and to create clusters (categories) of queries which appeared to express similar knowledge needs (intentions). Interrogatives such as 'How' or 'What' can suggest corresponding rhetorical questions (who, what, where, how, when, etc.) and knowledge types, but they are only a starting point in the classification process.

The survey invited respondents to write down questions that "you could ask or could be asked about the security of your system". Therefore, although some questions are clearly ones that the respondents themselves would ask (eg. 'Can I change a user's password for them ?'), or would ask themselves (eg. 'Do they really know what a virus is - do they care ?'), there are others that would have been addressed to them by other users, for example: 'My password has expired. Can you reinstate ?', and some that could fit either category: 'Can I have a document folder that only I can access ... ?'.

We can further speculate that some queries may have been rephrased by the respondent, but that they originate from another user's query, for instance: 'Can users' confidential mail be accessed by any

other user ?' may have originated from a question put to the respondent by a less informed user: "Can my confidential mail be accessed by any other user ?' In this instance, the respondent becomes a mediator for the query.

The survey data includes questions formulated as checklist items. These would be questions that a security officer should be asking about the current state of security on the system. In this context, a question such as 'Do you have off site backup ?' does not mean 'Should you have off site backup ?' but 'Check that you do have off site backup'. The checklist style of question is, in fact, a prominent feature of Elbra's security handbook (1992). One respondent has formulated most queries as problem statements (eg. 'The software has terminated due to unauthorised access'), which suggest an implied question ('Faced with this problem, what do I do ?').

Some queries contain several elements which need answering. These range from coordinated questions ('How can I find out how and when ... happened ?'), to appended questions dependent upon the answer to others ('Can viruses travel ... - if so, are there resident checks ... ?'; 'If passwords available, how often are they changed ?'). The 'if' condition, which as noted earlier, is a feature of the sample data, can indicate dependence upon specific circumstances, for example: 'Can users be automatically deleted if they do not sign on for X months ?'

The emerging classification has been outlined below; it reveals the respondent's perspective on the system, which is that of wanting to exercise control

over its various components - the system itself, its
users, and external influences which impact upon the
system and its security arrangements. The
limitations of this classification and observations
resulting from it are discussed at the end of this
section.

The queries fall into 3 categories, which can be
described through generalised questions, as
follows:-

1. User's control over the system

   "In what ways can I control the system ?"
   "In what ways can I control users ?"

2. User's concern over other peoples' control

   *"In what ways can other people - users or
   intruders - exercise control ?"*

3. User's concern over external factors

   *"In what ways does security impact upon
   other aspects of the system and the
   business and vice versa ?"*

1. <u>User's control over the system</u>

The questions which gravitate towards this category
express the following needs:

   "I need to know ..."

a. What to do
b. Whether I can do it
c. Whether I should do it
d. Whether I have done it already, it has happened
e. Whether it will happen
f. How to do it
g. When to do it
h. What it is, how it works and why

i. Why I should or should not do it;
   Why it happens or does not happen

a. Meaning: I need to know WHAT to do in a SPECIFIC
SITUATION or in relation to a specific problem

What happens if I forget my password ?
What happens if the AS/400 ... breaks down ?
What do you do if someone's password expires ?
What do we do if somebody forgets their
password ?

If ..., what do you do in that situation ?
I want to ... What do I do ?

What precautions can be taken to prevent ... ?
What precautions should be taken to preserve
the integrity of the system/ data ?
What regular checks can be made to ensure ... ?
What should happen to the system when the
office is un-manned ?
What is the procedure for setting up the system
security ?

What sort of backups should I be running for my
day to day data ?

Which files should be secured on backups ?
How far back should backups be kept ?

Making sure certain restrictions are made to
access of ...

There are no security arrangements in force -
Where do I start ?

I am not authorised to use the system.

I want authorisation to this system function.

b. *Meaning: I need to know WHETHER the system PERMITS ME to do this, or provides these FACILITIES - eg. for security measures, security checks, recovery procedures (specific and general)*

Can I change ... ?

Can I restrict access to sensitive files ?

Can you restrict access to ... options ?

Can you restore ... ?

(...) Can you reinstate ?

Can I have a ... that only I can access ?

Can we secure the system against virus attack ?

Can we back up security while people are signed on ?

Can the system itself help me to monitor its own health and safety ?

Can the system help me to monitor infringements and identify vulnerable areas ?

Can passwords be changed ?

Am I able to increase/ decrease security access ?

Is there an audit trail to show ... ?

Is there a way to display ... ?

Is there an easy way to check ... ?

Is there anything on the system an administrator can't find out ?

Are there any checks in software/ data to prevent them being encouraged to "work off the premises" ?

Are virus checkers available ?

Is it possible to check for viruses ?

Is it possible to check the system for viruses ?

Is it possible to lock the computer screen ?

Is it possible to be completely locked out of the system ?

If ..., how will I know ?

If ..., how far back does ... remain ... ?

If passwords available, ... ?

What are the rules on ... ?

What security levels are available ?

What flexibility/ options does the system offer on security ?

What authorization levels do you have ... ?

Does my system support different levels of security ?

How many security classifications are there ?

How many levels of password are there ?

(...) if so, are there resident checks to monitor/ ensure the health of incoming data/ mail ?

What can the Security Officer not do ?

How much is there on the system I can't even see ?

What are the limits of what I can do ?

In relation to users:-

Can I allow users to share ... ?

Can I 'force' users to change their passwords ... ?

Can I prevent users from leaving ...
signed on ?
Can I look at other peoples' electronic mail ?
Can users be prevented from exiting ... ?
Can users be prevented from signing on ... ?
Can users be given ..., but still ... ?
Can users be automatically deleted if ... ?
Can more than one user have ... ?
Can files ... be locked to prevent updates from
... users ?

Do you have password and user Id's at sign on ?

c. *Meaning: I need to know WHETHER to do this - I am
looking for confirmation or persuasion that this
action or state is a good idea*

Should we restore ... on a regular basis ... to
ensure ... ?
Should I create access authority levels without
consultation ... ?
Should I invoke the automatic password change
facility on our midrange system ?
Should I be using data encryption for
communications traffic ?

When I leave ..., should I sign off ?

Do I have to sign off every time I leave my
computer ?

Must I have access control on my PC - it's such
a pain ?

What are the advantages of group profiles over
authorisation lists ?

157

In relation to users:-

> Should I be allowing users to ... ?
> Should users be limited to ... ?

d. *Meaning: I need to know WHETHER I have done this (do this) already, WHAT has been done, WHAT has happened*

> Have IBM default passwords been reviewed ?
> Has user access to AS/400 Command line been removed ?
> Have any authorisation lists been set up ?
> What level of security has the machine been set to ?
> What levels of security will I expect to find ... ?
>
> Do security levels for users get checked on a regular basis ?
> Are hardcopy reports secured in a safe place when required ?
> In the event of a machine crash, do you have a detailed disaster recovery plan ... ?
> Do we have physical security ... ?

e. *Meaning: I need to know WHETHER this will happen*

> Will the UPS work when required ?

f. *Meaning: I need to know HOW to do this SPECIFIC ACTION*

> How do I change my password ?
> How do I check my present m/cs are virus free ?
> How do I shut down a UPS if needed ?
> How do I control ... ?

```
How do you change your password ?
How do you set up a password ?
How do we go about setting up the security ?
How can I find out how and when ... happened ?
How can I prevent ... ?
How can I override ... ?
How can I easily back up ... ?
How can one prevent a virus spreading from ...
to ... ?
How can we best secure the system against ... ?
How can groups ... be set up ?


(If ...,) how can you find extent of
damage ... ?


How do you plan for disaster recovery ?


What ... level ensures a secure system without
reducing ... ?
```

In relation to users:-

```
How do I secure ... to specific users ?
How do I restrict user to specific ...  only ?
How do you add a new user to the system ?
```

g. Meaning: I need to know WHEN to do this SPECIFIC
ACTION on the system

```
How frequently should passwords be changed ?
How often should passwords be changed ?
How often should people change their password ?
How often should the security files be backed
up ?
How regular should we be doing our system
backups ?
How many times should a user be allowed ... ?
```

If ..., how often are they changed ?

What should the back-up frequency be ?
Frequency of backing up system ?

When was ... last backed up ?

For how long can back-up tapes be guaranteed
good ?

Can we back up security while people are
signed on ?

h. *Meaning: I need to understand this FEATURE of the
system, WHAT it is, HOW it works and WHY it works
that way*

What are authorisation lists ?
What is level 40 security ?
What is a group profile ?
What does 'password expired' mean ?

What is the system's resident security
rationale, and how do I operate it ?
What is difference in levels or access given by
system to ... ?
What is the key ... used for ?
What does the security key position do ... ?
What does the ... system value do ?
What does this red switch do ?
What authority does a particular security level
have ?
(...) How do these work ?
What lifespan do backup tapes have ?

Why is spool file security so complex ?

i. Meaning: I need to understand WHY I should do this (this happens), WHY I should not do this (this does not happen)

Why do we need password security ?
Why must I always book the portables out ?
Why do you keep everything locked up ?
Why am I classed as a restricted user ?

Why haven't I got access to all the options in the system ?
Why aren't I allowed to a command line on the AS/400 ?

If you gave me greater access, I wouldn't have to bother you so often (= Why won't you give me greater access ?)

So what's wrong with a couple of illicit copies - no one will know. (= Why won't you let me use illicit copies ?)

2. <u>User's concern over 'other peoples'' control</u>

This typically refers to risk assessment - concerns over potential threats from intruders (eg. hackers) and also from users on the system accessing parts of the system which they do not normally access, with either legal or illegal intent.

"I need to know ..."

a. Whether people can do this, whether it can happen
b. How it can happen
c. What to do if/ in order to
d. Whether we should do it

e. Whether we do it

f. What happens if

a. *Meaning: I need to know WHETHER other people or software have access - WHETHER they can do these things, WHETHER the system allows it, WHETHER it can happen*

Can a user have control over ... ?

Can a user have the same password ... ?

Can a user sign on ... ?

Can a user give away objects ... ?

Can a user "passthrough" ... and gain greater authority ... ?

Can an individual change their own security level ?

Can staff inadvertantly destroy ... ?

Can ... be decrypted ... by potential hackers ?

Can passwords be found out ... ?

Can users' ... be accessed by any other user ?

Can ... be viewed by other users ?

Can ... be restricted by another user ?

Can ... be accessed by everyone ?

Can ... be accessed by other people ?

Can any user read ... ?

Can everyone access everyone else's ... ?

Can 2 users ... pass through without ... ?

Can other companies access our network ?

Can other companies use the same network as us ?

Can ... be secure with ... signon facility available ?

Can a PC virus be spread ... to other users ... ?

Can a PC virus destroy ... via ... ?

Can viruses travel through networks ... ?

162

Can you use any parts of the system without
first having a password and signing on ?

Could a user suppress ... to conceal ... ?

Is my system as secure as possible ?

Are there any ... which a hacker could use to
access ... ?

Does anyone have access to all the passwords ?
Do all DP Staff have access to all live
environments (any restrictions ...) ?
Do you allow dial in access to your
system ... ?
Do you have password security when
communicating with 3rd party ?

Is it possible to find out other peoples'
passwords ?
Is it possible to get around security ?
Is it possible to break passwords ?

*b. Meaning: I need to know HOW software, hardware or
people are able to cause intentional or accidental
damage - HOW it can happen*
How can system security be breached ... ?
How is the system vulnerable ?
How secure is my system ?
How resilient is the system to someone trying
to hack their way in ?
How real is the security problem anyway ?

What areas of the system need to be considered
in the "system security" light ?
What are the most common ways of breaching
security ?

What are the chances of being "hacked" ... ?
What are the potential problems which might
arise when new software is put on to old/
existing data ?

c. *Meaning: I need to know WHAT TO DO IF I detect a*
security problem, or IN ORDER TO prevent a problem

What if I receive messages indicating ...
unauthorised access ?

What practical steps need to be taken to
prevent (hacking)... ?
What is the most practical way to protect ... ?

d. *Meaning: I need to know WHETHER we should do this*

Should staff who have tendered their
resignations ... be removed from the
system ... ?

e. *Meaning: I need to know WHETHER we do this*

Are AS/400 History logs being reviewed on a
regular basis to detect unauthorised access
attempts ?

Has any 3rd Party Software ... been reviewed to
check object authorities ... ? - should not be
public access.
Have ... been checked ... ? - should not be
public access.

Do you allow ...; if yes, what security
installed ?

f. *Meaning: I need to know WHAT the CONSEQUENCES*

164

might be of this event (WHAT HAPPENS IF)

> What problem to security if staff gained
> physical access to ... ?

3. User's concern over external factors

These questions are concerned with the interaction
between security and factors 'external' to
security - other parts of the system, the physical
environment, organisational and business issues.

> "I need to know ..."

a. Whether I can do it/ whether it can happen
b. Whether I should do it/ whether it should happen
c. Whether we do it already/ whether it happens
d. How to do it
e. What happens if
f. Where we do it
g. Where I can do it
h. Who can or should do it
i. Why it happens

*a. Meaning: I need to know WHETHER I can do this/
WHETHER it can happen*

> Is there a check list I could periodically go
> through to ensure security and integrity ?

> Do you really think 'FAST' will raid us ?

*b. Meaning: I need to know WHETHER we should do this
or WHETHER it should happen - is this a good idea*

> Should we set up an internal security audit
> function ... ?

165

Should my system's security ever be audited by a third party ?
Should contracts of employment have a data/software theft clause ?

Is disaster simulation ever a practical way of testing a recovery plan ?

Are all functions within a product necessary ?
(= Should we use all functions ?)

c. *Meaning: I need to know WHETHER we do this, WHETHER it happens*

Is there a building alarm ... ?
Is computer dept protected by smoke alarm detectors ... ?
Is on site backup stored in secure cabinet ... ?

Do you have off site backup ... ?
Does the computer suite have its own security i.e. limited access ... ?

If ... has windows, are they non breakable ?
If ... activated, is electricity supply ... cut off ?

(UPS batteries) - are they still full of life without faults ?

Have we got long life batteries ?

Do security measures contradict the principles of Total Quality Management ?

d. Meaning: I need to know HOW to do this

How do I discover what products are on the
market ?
How do I evaluate these products without prior
experience ?
How do I convince colleagues of the need for
security if ... ?
How do I persuade them that the ... delay when
using security products ... is necessary ?
How do I make staff use non standard
passwords ?
How do I stop them storing passwords in obvious
places ?
How do I arrange training for over 100 staff ?
How do you educate other users, so they're
aware of the dangers ?
How do you encourage them to support existing
security practice ?
How do we make sure that everything that should
be backed-up, is ?
How can I best persuade staff of the importance
of security ?

How comprehensive should a disaster recovery
plan be ?

Computer Disaster Recovery - which are the best
methods ?
System saves ... need to be done after hours
... this is expensive ... better methods must
be found.

Do they really know what a virus is - do they
care ? (= How do I explain it to them or
persuade them ?)

Which is the best virus protection software ?
(= How do I choose ?)

What organisational procedures should we have
to enhance security ?
What system of backup ... should be used ?
What is the most efficient way of
backing-up ... ?

e. *Meaning: I need to know WHAT the CONSEQUENCES
might be (WHAT HAPPENS IF)*

What impact does comprehensive security have on
response times ?

How quickly will a replacement machine be
supplied if the existing one suffers a major
disaster ... ?

If we expand ..., will it necessitate ...,
which could mean the replacement of all
batteries ?

f. *Meaning: I need to know WHERE we do this
currently*

Where are backups kept ?

g. *Meaning: I need to know WHERE can do this*

Where do I find info on new viruses ?

h. *Meaning: I need to know WHO can or should do this*

Who is allowed access to ... ?
Who has access to ... ?
Who should have "security officer" status ?

Who owns the 'master' ... password ?

Is the security officer limited to ... ?

i. Meaning: I need to know WHY this happens

Why won't the Senior Personnel Partner amend
employment contracts to include security as
part of job specs ?

Why won't they tether or bolt PC's to the
desks ?

c. Commentary on intentional and rhetorical analysis

Classification systems are by nature not watertight,
so that a number of items will not fit neatly into
one category or other. The phenomenon of overlapping
or straddling is very well known. There are also
specific reasons why certain types of query are
difficult to classify with certainty. For instance,
in a number of formulations, there is an element of
ambiguity, which means that a query taken at face
value can belong to more than one category:-

- the pronoun 'you' can refer to the person asking
the question or to other users on the system,
changing the perspective of control:

'Can *you* restore deleted objects ?'

can mean

'Can *I* restore deleted objects ?'

or

'Can *users/ hackers* restore deleted objects ?'

- the verb 'to have' can mean 'to be given' or 'to
take', again changing the perspective of control:

'Can a user *have* control over his/her own

profile ?'
       can mean
'Can I *give* a user control over his/her own
profile ?'
       or
'Can a user *take* control over his/her own
profile ?'

- the plural noun 'rules' can refer to system rules
or to company rules, as in the example:

'What are *the rules* on password reuse ?'
       can mean
'What are *the AS/400 rules* on password reuse ?'
       or
'What are *our company rules* on password
reuse ?'

- the passive form 'can be found out' can assume
different agents:

'Can passwords *be found out* on the system ?'
       can mean
'Can *I* find out passwords on the system ?'

       or
'Can *users/ hackers* find out passwords ?'

The same is true of the formulation 'Is it possible
to ...', which can mean 'Can I ...' or 'Can anyone
...'. The notion of permission or possibility is in
fact strongly ambiguous, as it carries the
suggestion of hesitation; to take one example:

'*Can I* allow Office users to share a signon
I.D. ?'
       can mean

170

'*Is it possible* for users to share it ?'
        or
'*Should I* allow users to share it ?'

Other impersonal expressions are also ambiguous:

- 'What happens if I forget ...' can have two meanings:

    'What do I do if I forget ... '
        or
    'What will the consequences be if I forget ...'

-'Is there an audit trail ...' can mean:

    'Does the system automatically provide it ?'
        or
    'Have we set it up ?'

Elliptical questions are obviously ambiguous, for example:

    'Password reuse ?' has several interpretations.

Questions relating to the frequency of an action can conceal that its advisability is also in question, eg. 'How many times should a user be allowed to ... ?' invites the possible answer 'zero times/ not at all', which would then undermine the assumption that the action was permissible, and bring it closer to those questions which focus directly on permissibility ('Can I allow users to ...').

A further problem relates to the time orientation of queries. The vast majority of queries relate to the future, even if from a grammatical point of view

typical future forms are not used. For example, a
query which begins with the words "What do we do if
... ?" is almost certainly asking about what action
might be taken in particular future circumstances,
not about established habits. The only clear
exception to this would be the checklist item. This
is possible because in English (as indeed in other
languages) verb tense and time do not always
correspond. The frequent use of the auxiliary "do"
(in preference to "will") is additionally
interesting because it signals an implicit element
of repetition: in the 'microcosm' of computer
security, many actions and events have a recurrent
nature.

It is worth remarking that when individual queries
are listed and analysed, they are taken out of their
context. In real life, a query might result from a
comment or another query, and it could be
interesting and perhaps beneficial to look at larger
samples of language with embedded queries. However,
a list also reveals something about the relationship
of queries to knowledge needs. Two examples from the
survey data may serve to illustrate this. In the
first, a user has started off with general
questions, moved on to highly specific queries, and
finished off again with queries of a general nature.
In the second, queries have been organised such that
they deal firstly with physical security, then data
security, communications, and finally, general
queries. This suggests that in the minds of at least
some users, there are notions of progression and
order which could be important in the organisation
of information in manuals.

An additional point to be made here is that, again
in a real life situation, a query, which typically

springs from doubt or not knowing, may well be
imprecise - it will have to be refined by further
questioning. As has already been pointed out, some
of the queries in the survey data are in fact
formulated in two or even (in one case) three parts
(eg. joined by the conditional "... if so, ..."; "
If ..., and ..., how ... ?").

2.Identification of knowledge types

a.Knowledge need overview

The lexical and terminological analysis allowed us
to discover the knowledge needs expressed through
the choice of words made by respondents to the
survey. Certain concerns surface again and again in
the data; these can be summarised as follows:-
> TIMING
> MANNER
> POSSIBILITY
> ADVISABILITY
> NECESSITY or OBLIGATION

Linked to POSSIBILITY are the concepts of CHANCE or
HAZARD, and the notion of CONTROL. CONTRAST,
COMPARISON and RESTRICTION were also noted as
specific strategies for obtaining knowledge.

Several domains of knowledge were represented in the
data, and it became ever clearer that computer
security in a business environment cannot be
considered in isolation from its closely related -
broader, and overlapping - domains, especially
business and computing. Computer security was shown
to be particularly concerned with people.

The rhetorically-based classification suggested
subsequently also makes it possible to see a number
of knowledge types in the data. These can be put
together to form the overview shown in Fig. 3 (which
represents only the "knowledge space" of the data,
not all possible permutations).

In brief, the queries about actions and events
concern:-

(a) the ESSENCE (or components) of an action/event
(b) the TIMING (or frequency) of an action/event
(c) the MANNER of an action/event
(d) the LOCATION of an action/event
(e) the CAUSE of an action/event
(f) the AGENT/OBJECT of an action/event
(g) the OCCURRENCE of an action/event
(h) the POSSIBILITY of an action/event
(i) the ADVISABILITY of an action/event

Knowledge about ACTIONS and EVENTS - past, present, or future

```
    <---------- A C T I O N S -------------> <---- E V E N T S ----->
```

(a) Knowing WHAT to do .... **WHAT happens**

    What do I/others do ?   *What happens ?*

(b) Knowing WHEN to do sth .... **WHEN sth happens**

    When do I/others do this ?   *When does this happen ?*

(c) Knowing HOW to do sth .... **HOW sth happens**

    How do I/others do this ?   *How does this happen ?*

(d) Knowing WHERE to do sth .... **WHERE sth happens**

    Where do I/others do this ?   *Where does this happen ?*

(e) Knowing WHY to do sth .... **WHY sth happens**

    Why do I/others do this ?   *Why does this happen ?*

(f) Knowing WHO does sth .... **WHO/WHAT is affected**

    Who does this ?   *Who does this happen to ?*

(g) Knowing WHETHER one DOES sth .... **WHETHER sth happens**

    Do I/others do this ?   *Does this happen ?*

(h) Knowing WHETHER one CAN do sth. .... **WHETHER sth CAN happen**

    Can I/others do this ?   *Can this happen ?*

(i) Knowing WHETHER one SHOULD do sth ...... **WHETHER sth SHOULD happen**

    Should I/others do this ?   *Should this happen ?*

Fig. 3   Knowledge need overview

Notes to Fig. 3:

(1) Although the prototypical questions in this scheme imply the prevalent future tense, most questions have the potential to be transformed into the present or past tense, eg. "How do I do this ?" (in future), can become "How am I doing this ?" (currently) or "How have I done this ?" (in the past). The past and present tenses are prevalent in checklist items.

(2) Question types (a)-(f) can each be transformed to add an element of possibility or advisability, eg. "What do I do ?" can become "What CAN I do ?" or "What SHOULD I do ?"

(3) A number of question types can be either reactive or proactive in nature on any given occasion, depending on the structure of the question as a whole, eg. "What do I do if ... ? " (reactive) versus "What do I do to prevent ... ?" (proactive).

b.Dominant question forms and knowledge needs

A frequency count for recurring initial, i.e. capitalised, question forms in the survey data gives us the following picture:

Most frequent verb forms:-

> Can
> Is/Are
> Do/Does
> Should

Most frequent interrogative pronouns:-

> How
> What

The forms "Why", "Where", "When" and "Who" are relatively infrequent. The most frequent forms are used to expresss the following knowledge needs:-

"HOW"

"How" usually introduces questions concerning MANNER ("How do I do this ?", "How can I do this ?", "How does this happen ?"). It can also introduce TIMING ("How often ... ?"), ESSENCE ("How quickly will ... happen ?"), and POSSIBILITY ("How many levels ... are there ?")

"WHAT"

"What" represents the ESSENCE of an action or event ("What do I do ?", "What can I do ?", "What happens ?", "What happens if ... ?"), POSSIBILITY ("What levels are available ?", TIMING ("What should ... frequency be ?"), MANNER ("What is the system's ... rationale ?", "What is ... used for ?", "What are the ways of ... ?")

178

"CAN"

"Can" signals questions concerning the POSSIBILITY
of an action or event

"SHOULD"

"Should" represents ADVISABILITY

"IS / ARE"

"Is" and "Are" introduce questions which focus on
ADVISABILITY, OCCURRENCE, and POSSIBILITY:-

Is disaster simulation ever a practical
........................ADVISABILITY

Is computer dept protected by
........................OCCURRENCE

Is there a Building Alarm
....OCCURRENCE

Is on site backup stored in secure cabinet
........................OCCURRENCE

Is my system as secure as possible
........................POSSIBILITY

Is there an audit trail to show who
........................POSSIBILITY

Is there a way to display all commands
........................POSSIBILITY

Is there an easy way to check the loss
........................POSSIBILITY

Is there a check list I could periodically
........................POSSIBILITY

Is there anything on the system
........................POSSIBILITY

Is it possible to break passwords
........................POSSIBILITY

Is it possible to check for viruses
........................POSSIBILITY

Is it possible to check the system for
........................POSSIBILITY

Is it possible to lock the computer screen
........................POSSIBILITY

Is it possible to find out other peoples'
........................POSSIBILITY

Is it possible to get around security
........................POSSIBILITY

Is it possible to be completely locked out
........................POSSIBILITY

Are all functions within a product
necessary
........................ADVISABILITY

Are AS400 History logs being reviewed
........................OCCURRENCE

Are there any IBM supplied user ids which
........................POSSIBILITY

Are there any checks in software data
........................POSSIBILITY

"DO / DOES"

"Do" and "Does" introduce ADVISABILITY, OCCURRENCE,
POSSIBILITY, and MANNER:-

Do I have to sign off every time I leave
........................ADVISABILITY

Do security measures contradict
........................OCCURRENCE

Do you have off site backup
........................OCCURRENCE

Do you allow dial in access to
........................POSSIBILITY

Do you have password security when
........................POSSIBILITY

Do you have password and user Ids at sign on
........................POSSIBILITY

180

Do all DP Staff have access to all
........................POSSIBILITY

Do you really think FAST will raid us
........................POSSIBILITY

Do they really know what a virus is
........................MANNER if rephrased

Does my system support different levels
........................POSSIBILITY

Does anyone have access to all
........................POSSIBILITY

Does the computer suite have
........................OCCURRENCE

This means that the most common question forms
embrace the following knowledge needs:-

POSSIBILITY	ADVISABILITY	OCCURRENCE
ESSENCE	TIMING	MANNER

A few final remarks about queries. Some queries have
an urgency about them, others are to do with long-
term planning. Some answers are immediately applied,
some are absorbed 'just in case' ("I need to know
about this danger, this possible breech, this
security procedure in case something happens"). But
more fundamentally, queries are "one-off" questions,
i.e. the user cannot normally challenge the
cooperation of the source (a point made by Blair,
1992, in relation to information retrieval systems
in general). Cawsey et al. (1992) also emphasise
belief revision: users (and information scientists
acting as intermediaries) revise their beliefs about
what is wanted. An effective knowledge transfer
system should take that into account.

3.Source preferences and expectations

Respondents to the survey were asked to express preferences for various sources which they would "turn to for getting information on computer security relating to their system". Table 1 illustrates their choices.

Numbers in the "Blank" column are due to one respondent having indicated a preference for only four of the sources (using preference indicators 1, 2, 10 and 11 - the extremes of the spectrum), and to the fact that some respondents left out "other" (K) as a source. Overall, the figures show that colleagues, manuals, and on-line information are the most preferred sources (in that order), and television programmes the least preferred.

In Section II of the questionnaire, an "expected source of answer" was to be indicated by the respondent against each query. The purpose of this was to find out whether the pattern of overall preferences was the same as in the analysis of the expectations which accompanied the process of specific query formulation. In addition, it was important to see whether manuals occupied an important place in the preference hierarchy, and in expectations. Table 2 shows where users expected to find their answers to queries. The pattern is confirmed, and manuals are shown to occupy an important position - 2nd place - in both tables (preferred sources and expectations).

Order of preference indicated (1-11)

Source	1	2	3	4	5	6	7	8	9	10	11	Blank
A	0	3	4	1	7	7	2	3	5	1	0	1
B	19	3	2	5	1	0	3	0	0	0	0	1
C	0	7	5	7	3	0	3	4	4	3	0	1
D	1	0	0	4	5	8	5	5	2	2	0	1
E	0	1	1	1	4	1	10	7	6	1	0	1
F	0	1	2	3	3	3	5	6	9	1	0	1
G	9	7	7	4	1	3	0	0	2	0	0	0
H	2	10	9	5	1	1	1	2	2	21	0	1
I	0	0	0	0	0	0	1	2	1	0	9	0
J	3	2	3	3	7	10	2	3	1	0	0	0
K	0	0	0	0	1	0	1	1	1	5	16	9

A=books
B=colleagues
C=external tel.support
D=IBM representative
E=Information packs
F=magazine articles

G=manuals
H=on-line information
I=TV programmes
J=training materials
K=other

Table 1. Preferred sources

```
-----------------------------------
B - 89      Colleagues
G - 75      Manuals
H - 36      On-line information

K - 35      Other

C - 26      Telephone support
J - 24      Training materials

? - 18      Don't know

A - 16      Books
F -  9      Magazine articles
E -  7      Information packs
D -  5      IBM representative
I -  0      TV programmes
-----------------------------------
```

Table 2.   Expected sources of answer

Notes to Table 2:

(1) Although only one "expected source of answer" per query was
required, a few respondents gave more than one source, and some
gave none.

(2) "Other" sources, where indicated, were "DP manager", "myself",
and "intuition".

Chapter V

Knowledge transfer - establishing a mapping

A.General characterisation of the security manual

1.Organisation and readership

The manual "Security Concepts and Planning" (IBM, 1991), begins with a table of contents (7 pages), followed by some introductory notices and remarks (3 pages), 9 chapters (235 pages), 6 appendices (82 pages), a bibliography (2 pages), and an index (28 pages), making a total of some 357 pages. For comparison, a typical published handbook on computer security (eg. Elbra, 1992; Hearnden, 1990) will contain some 200 pages. A word count, sometimes used to indicate corpus size, would not be a meaningful measure in this instance, because there are numerous diagrams, tables, forms, sample programs, and screen displays, where words play a special, sometimes complementary, role in conveying meaning.

Chapters 1 and 2 are an introduction to security, with an overview of controls and an explanation of the main concepts and considerations. The next section covers user profiles. The manual then moves on to resource security, which deals with questions of authority and ownership. Next, there are "security tips and techniques". Chapters 6, 7 and 8, the three longest chapters, cover auditing (monitoring), security recommendations and planning, and setting up security. The final chapter (9) is in the form of questions and answers. This organisation implies a progression from familiarisation with security concepts and the development of an

understanding of risks and considerations, to the
implementation and monitoring of security.

Chapter 9, entitled "Security Questions and
Answers", deserves special mention here. It contains
21 questions "that are asked most often about
security" (pg. 9-1), with an answer for each
question. The provenance of these questions is not
specified, neither is their sequencing (there may be
none). It is interesting to note, however, that 8 of
the questions, including the first 6, ask "Why...
?"; that there are 5 instances of "How... ?" (incl.
"if... how... ?", "when... how... ?") , and 4
instances of "Is there a way... ?". Almost half the
questions describe a situation of failure or an
unexpected event which need to be corrected –
questions which might typically be put to a support
service, rather than at the stage of planning. Two
other questions ask about the consequences or
"knock-on" effects of an operation, four about
obtaining information of a global nature, while the
rest (6) are essentially "how to" questions.

The manual is "intended for someone who is assigned
the responsibilities of setting up users and
controlling users' authorities on the system" (pg.
xiii). It presupposes familiarity with information
contained in the "System Operator's Guide", and "New
User's Guide". It suggests the use of online help
information "to activate and maintain system
security" (pg. xiii). It "provides information"
about security concepts and planning, making this an
informative and expository text which also has
defining, enumerative, synoptic, recommendatory and
illustrative features. Manuals by their nature are
also authoritative publications, even though
technical inaccuracies or typographical errors are

bound to occur.

2.Points of entry to the text

Apart from the usual major access devices - table of contents, headings and index - there is much to "catch the eye" of the reader flicking through the manual. Words other than those in headings are sometimes printed in bold (eg. **Note:** ; **Method 1**). Command names (eg. DLTUSRPRF - delete user profile), in capital letters, stand out in the text. Information about security risks and considerations has been placed in boxes which serve to highlight it, and attract attention alongside bullet points, numbered lists, tables, diagrams, flowcharts, screens, programs, and forms.

The index to the manual has 602 main entries (in bold). A large proportion of these are command names, and entry keywords are not unique, eg.

> **audit**
> **audit log command, display**
> **auditing security**

Discounting acronym command names, 31% of main entries (189) relate to distinguishable lexemes, like the 'audit' group in the above example. Nouns (singular and plural), verbs (injunctive form, past and present participles, and one verb in the infinitive form), adjectives, and adverbs all appear in main entries, while prepositions ('for', 'in') can sometimes introduce run-ons. There are specialised terms (eg. 'adopt authority'), semi-specialised (eg. 'display', 'value'), and general language words with a high occurrence in this

187

context (eg. 'working with', 'using'). There are
metalinguistic labels ('definition', 'general
description'), and instances of abstract relations
('difference'/'different', 'comparison'). Most items
have one page reference only.

B.Summary of retrieval needs and retrieval problems

1.Evaluating the index

The specification of retrieval problems must begin
with an evaluation of the index as an access device
to knowledge. The relevant British Standard defines
an index as: "A systematic arrangement of entries
designed to enable users to locate information in a
document" (BSI, 1988). The Society of Indexers
defines it as "a detailed guide to the information
and ideas in a document", which "enables enquirers
to find information they need or to recall half-
remembered passages", stating that "all good indexes
need to be clear, concise, comprehensive and
consistent" (Society of Indexers, 1993). According
to the Society, the training of index compilers aims
to impart a number of skills; these could provide
criteria for the evaluation of an index:-

- distinguishing the chief concepts contained in a
  document
- devising the necessary index terms for those
  concepts
- dealing with synonyms, homonyms, and related terms
- assessing any need for multiple indexes to the
  document
- distinguishing between major and minor references
- indicating difference between references to text
  and illustrations

- preparing copy to a high standard of accuracy

However, although these skills would guarantee a given standard of indexing, the starting point is "the chief concepts contained in a document", whereas we must take the users' discourse as the chief reference point for the evaluation. A different approach would be to try to discover whether look-up operations are successful, i.e. do search terms lead to appropriate locations in the manual ? The main difficulty with this would be that the index can work for one search term/ one query, but not for another - even the use of the same term on two separate occasions can produce results of different degrees of satisfaction - so that a very large volume of test results would be necessary. However, we are concerned with the phase which precedes look-up - finding a search term in the index.

In evaluating the quality of the index, the following stepwise assessment provides a framework:-

(1) Are the *knowledge needs* from users' queries represented in the index ?

(2) Are the *lexical items* from users' queries represented in the index ? When a lexical item is not represented, how important is the missing item, in knowledge terms ?

(3) *How easy is it* to map the language of user queries onto index entries ?

Let us note in passing that the functioning of an

index is determined in no mean measure by the skill,
knowledge and intelligence of the user. This is
because the user has the power to choose which items
to look for in an index, making correct and
incorrect choices. A computerised retrieval system
can provide automatic search term expansion, but
this approach has been criticised, eg. by Ruge
(1992) at Siemens Nixdorf, who uses linguistic
knowledge for term expansion in the "hyperterm"
system REALIST (term modifiers being used to
determine degree of semantic similarity between
terms) but insists that the user should be the one
to choose from proposed "interesting" terms.

Using the above framework for evaluation, an
analysis of the index was carried out; this showed
the index to be an inadequate access device in
several respects.

2.Knowledge needs and problems

a.General considerations

There is an evident yet important point to be made
about users' knowledge needs as expressed through
their queries and it is this: what users are asking
is not necessarily what they should or could be
asking ! In other words, there are aspects of
security that are not being dealt with by users
partly because they are not aware of their existence
or importance; conversely, too, queries can be found
to be targeting areas of knowledge with, objectively
speaking, little pertinance to the domain. This
helps to explain why the rich array of knowledge
types emerging from an analysis of the domain itself
is not reflected in the query data.

Based on the query data, a high level summary of the
knowledge needs detailed earlier falls into four
areas:

- need for instructions

Briefly, this encompasses "what to do" and "how to
do it".
Method, frequency/timing, location, and agent/object
may all be required. Specific conditional knowledge
may also be needed (restriction/precision).

- need for understanding

For confident and effective security, users need to
understand - differences, restrictions, implications
or consequences, causes.

- need for advice

This need is focussed on "whether to do". In a
domain which has risk at its centre, expert advice
is needed on the "best way", and on possibility,
necessity, advisability, obligation.

- need for control

Controlling risk means controlling people,
resources,
and external factors. It transpires in long-term
planning, and in more urgent cases - eg. controlling
an incident.

The first two categories of needs are less
distinctive than the second two. In the learning
strategy experiment DIOGENES reported by Ford & Ford

(1992) and referred to in Chapter III, four major categories of question were identified (descriptive, focussing, concrete and analysis); these can all be found in the need for instructions and understanding. The categories of advice and control are a feature of a domain which relates computer resources to the environment; Ford & Ford's experiment concerned the "closed" universe of PRECIS, a document indexing system whose features were being studied in an academic setting.

By contrast with users' global needs, the difficulty of satisfying the needs of individuals, even in a computerised environment, is expressed pointedly by Rex Last:

"...it is relatively easy to design a system which offers a high level of support to the beginner; it is equally straightforward to present a high-powered but more demanding interface to the expert; but what is a far from trivial exercise is to design a system which meets the needs of both, and of others at varying stages of expertise in between, without losing in speed, efficiency, or appropriateness in the level of help and support."

(Last, 1989:82).

According to Ford & Ford, the design of a tutoring system for even one level of user (novice) is a non-trivial undertaking.

b. Content versus knowledge needs

It is important to make clear here by way of example why the widely accepted practice of indexing texts on "content words" fails the user from a knowledge-based perspective. In the first instance, a term like "access" can be described collocationally as a

192

component of a set of actions and events, as having
certain properties, and a comparative dimension
('difference in access', 'greater access') (see Fig.
4). Its meaning is so strongly determined by
context, that on its own, it has little meaning. By
contrast, the manual's index entry for "access" is a
stand-alone term with a mere three sub-entries:
"limit to system unit", "PC Support access
considerations", and "to display station". For
another important term, "password(s)" (the word
occurs in nearly 16% of queries), there are ten main
index entries of varying length and complexity (eg.
"password", "password and user ID journal entries,
format for", "password control", etc.), some with
sub-headings; however, when one comes to examine
users' queries on this topic (see Appendix III for a
list), one can see that they represent a wide
spectrum of knowledge needs, and that there is
currently no mechanism for effecting a mapping from
query to index entry.

3.Language needs and problems

a.Discussion of language needs

It is important to resist the temptation to see the
language of user queries as being not highly
technical and therefore "not special". It is not
specialised in the usual sense; however, it is a
language variety with a distinctive quality, which
can be described by pointing to the preference for
certain categories of words: modal auxiliaries,
action verbs, nouns referring to method, timing and
causation, comparative adjectives and adverbs,
pronouns - words which converge on a specific sphere
of reference. There is a grammatical distinctiveness

193

```
             dial in access
             limited access
            physical access
              public access
            security access
        unauthorised access
                user access
                     access to AS/400 Command
                     access to commands
                     access to confidential files
                     access to live environments
                     access to menu options
                     access to passwords
                     access to sensitive files
                     access attempts
                     access authority levels
                     access control
          difference in access
              greater access
            to create access authority levels
          to decrease access
              to deny access
  to detect unauthorised access attempts
             to gain access
              to have access
          to increase access
  to make restrictions to access
            to remove access
           to restrict access
```

Fig. 4   Collocational pattern for "access" (n.)

194

as well as a lexical one. It is only when that specificity is recognised that the language is elevated to at least the same level of importance as the more obviously special language of the manual, and the needs of users can be brought to the fore.

The application environment - eg. business - is a strong influence on this language, reflected not only in the nomenclature but in the overriding concern of wanting to find a "better way" of doing things. This is signalled by formulations seeking to discover "how" (best method) and "when" (optimal time), and questions which suggest decision-making ("can I ?", "should I ?") with a view to improvement. The emphasis on modifiers and modality is striking, to the point where we have to seriously question the widespread tendency to look to more conventional "content" words (usually nouns) for knowledge representation and indexing. Notably, the role of verbs must be fully explored in a domain where the needs for instructions and active control are so clearly visible.

It is not easy to reconcile the user's potentially wavering or fuzzy language, which reflects the process of grappling with incomplete knowledge, with the definitive nature of statements in a manual. From variant spellings, to value-laden words like 'disaster' or 'danger', the user's linguistic repertoire for this occasion is made up of items which blatantly eschew terminological control. Yet somehow these items must be mapped onto the manual's more formal system of representation, bearing in mind that it, too, may not be in a state of perfect terminological control.

Another feature of the user's language is its

orientation towards 'people concepts', which may be contrasted with the manual's system-centredness. This may be an area where knowledge needs depart from what the manual can be expected to provide; on the other hand, it is a challenging case for exploring access paths to knowledge which is in some way present, but implicit or 'hidden'.

It must be said also that certain aspects of the query data analysed raise potential problems for the retrieval of information from a source such as a manual. In particular, there is evidence of variable, unorthodox, or incorrect spelling (eg. sign-on, signon; back-up, backup, back up; id, I.D.; non standard, non breakable; inadvertantly, etc.), and the use of abbreviations (DP, UPS, EDI, info, m/cs, chars etc.). Furthermore, inverted commas have sometimes been used, suggesting an unusual meaning, an element of doubt concerning the correctness of spelling or meaning, or an awareness of special use ("force", "passthrough", "master", "lock").

There is one other issue which underlies all considerations of linguistic expression and mapping. Is it sufficient to smooth the linguistic path to knowledge, or does successful knowledge transfer require an appropriate formulation of the knowledge at the end of that road ? A retrieval aid such as an index can "rewrite" knowledge in the litteral sense of using an entry vocabulary consisting of substitute words. A complete rewrite of a manual would be a very substantial, very different undertaking, with wide ranging implications. The 'synoptic rewrite' represented by an index is, above all, a more practical solution.

b.Terminological mapping problems - an illustration

We begin by checking whether the lexical items most
frequently occurring in the query data can be found
in the index. We have already seen that "access" and
"password(s)" are present; "security", "system",
"user", "data" and "authority" can be found there,
too. Surprisingly, however, "backup" (all spelling
variants) is absent; as its importance has already
been emphasised, this looks like a major oversight.
Admittedly, "saving" and "securing" do cover this
concept.

As has been demonstrated, however, although queries
contain such terms, much of their knowledge-seeking
intention is expressed through other elements. Below
are expressions highlighted in two sets of users'
queries on the basis that they represent knowledge
needs formulated wholly or partly in non-technical
language (Appendix IV gives a longer list of such
expressions). Against each group of queries is an
indication of the index entry in the manual which
would be the required entry point. Existence of an
index entry does not, of course, guarantee a
satisfactory answer to a query. Finding the
appropriate index entry (shown underlined) in the
first place can be, as illustrated below, a cruelly
difficult or futile undertaking.

Example 1

User queries:-

Can *more than one user have the same* user id ?
Should I *allow users to share* a signon I.D. ?

Relevant index entry:-
Profile
user

Example 2

User queries:-

How can I *prevent someone looking at my document* ?
How can I stop someone accessing *my office documents* ?
How do I give access to *only certain documents* in a folder ?
Can I have *a document folder that only I* can access (for *confidential work*) ?
Can *confidential documents* be viewed by other users in the output queue ?
How can I be sure no one can access *my documents that are confidential* ?

Can I *"lock" my folder* ?
Making sure certain restrictions are made to access of *confidential folders*.

Can I restrict access to *sensitive files* ?
I require a user to access *parts of* a data file via application software, but no access should be allowed via system utilities.

Can users' *confidential mail* be accessed by any other user ?

Can I *look at other peoples' electronic mail* ?

Can *personal mail* be accessed by other people ?

Want to *stop people viewing* spool yet be able to control printers.

For the above group of queries centred around "confidentiality",

confidential	- not in index
sensitive	- not in index
personal	- not in index
electronic	- not in index
mail	- not in index
folder	- not in index
lock	- refers only to keylock switch on control panel

Relevant index entries:-

<u>document password</u>

<u>document user profile</u>

<u>files, logical & physical</u>

<u>spool control special authority</u>

<u>spool job user profile</u>

<u>spool user profile</u>

Conclusions

The philosophical belief that knowledge is allied to
language and the evidence that in texts, knowledge
is inseparable from language, provided us with a
general foundation for the hypothesis that computer
users, through their knowledge of language, have
some prior knowledge of the domain of computer
security and that language can be a facilitating
mechanism, or an obstacle, in the development of
that knowledge. It also lead to the supposition that
users have a conceptual apparatus based on both
theoretical knowledge and experience of the world
and of domains of special reference related to the
environment in which they operate. In this light,
both the language of representation in computer
security texts and the language of retrieval had to
be examined with reference to the representation of
knowledge.

We did not have in mind an abstract representation,
but one that reflected the perceptions of real
people: users, with their needs, and technical
writers, with their writing brief. We have to
understand *both* knowledge needs *and* the constraints
of language on both sides. Then we can start to see
what mechanisms are needed to effect successful
knowledge transfer. At the same time, it was
necessary to draw out the characteristics of the
focal domain - computer security - so as to provide
an explanation for users' abiding concerns. It is
difficult for users to bring to the level of
conscious reasoning all aspects of their knowledge
needs. Some needs will be elucidated by the nature
of the domain (eg. dealing with risk in computer
security), others become apparent in elements of
language used to express those needs.

If there are recurring concepts in users' queries -
which was affirmed by this research with the
emergence of concepts like authority, backup,
access, password and others - then from the users'
point of view, those are the most important. It is
for writers to recognise that manuals do not impose
a view of a domain. Few are the users who read them
from start to finish; and as we have seen, users
already have a view, a conceptual structure, which
determines what more they want to know. And they
already have a language, which determines what they
say. Their knowledge is incomplete; the manual can
complete it, but in certain respects users know
more.

The distinctiveness of our approach resides in the
decision to examine the issue of knowledge transfer
from the perspective of terminology, but with an
equal emphasis on knowledge, and with the user
firmly in view. Another notable feature is that
retrieval is regarded here as a vital aspect of
comprehension, as an important tool in knowledge-
building. Put very simply, we are convinced that
there is no point in having a perfectly written
manual, if the user cannot easily gain access to the
section which corresponds to a need. Access to the
text is part of the decoding strategy - it should
not derange the reader ! The deliberate accent on
linguistic ease of retrieval is a challenge to more
conventional thinking which focuses on retrieval
results. Productivity gains can ensue from the
elimination of fruitless searches, and from long-
term effective security that comes from better
understanding leading to commitment in implementing
and maintaining it. Furthermore, the method of data
collection adopted, designed to avoid the compromise
inherent in queries directed at a specific source,

has allowed us to glimpse an ideal of user need
satisfaction against which actual satisfaction in
future 'systems' (printed manuals, on-line help, or
tutorial modules) can be compared. It has,
incidentally, shown that users have different styles
of questioning, some organising their thoughts by
aspect or level of generality, others defining their
needs in terms of problem statements. The
progression of questions could be explored further.

The potential danger of dealing with users' needs is
that they may be highly diversified. Certainly,
users' personal characteristics differ, as do their
working environments. A knowledge-oriented approach,
however, has allowed us to identify categories of
knowledge needs within which variation can still
take place. The four global categories distilled
from the query data - instructions, understanding,
advice and control - provide a framework for future
understanding of users' needs in this, and perhaps
other domains. The specific knowledge needs
identified, ranging from concrete notions of timing,
manner, essence, occurrence, location, cause, agent
and object, to modal notions of possibility,
advisability, necessity, and obligation, provide a
catalogue of needs which can be used to guide the
creation of appropriate retrieval aids.

In the query data, we have been able to see how
access to knowledge is dependent on knowledge of
terminology. Conceptual knowledge, expressed in
descriptive or 'substitutional' language (pronouns,
relative clauses) cannot be used for direct access.
We have seen users trying to describe the unknown,
succeeding perhaps from their own point of view, but
failing from the point of view of the manual, which
places specific terminological demands. "The set of

significant words in a text is a fuzzy set" says
Charles Meadow (Meadow, 1992:188); users' queries
are also an approximation, and again we are forced
to consider the implications of Zadeh's fuzzy logic
(Zadeh, 1975).

We found that terminological theory was not so well
developed that we could immediately proceed to
empirical research. Therefore, theoretical
deliberations and submissions have constituted a
substantial part of this thesis. In laying the
theoretical foundations for this research, we were
compelled to explore especially the relationship
between terms and units or elements of knowledge.
This proved a worthwhile undertaking, for we were
able to establish that there are many aspects of
knowledge - procedural, functional, relational,
judgmental, managerial - which are not well served
by terms. This conclusion is important, as it leads
to the realisation that significant aspects of
knowledge could be disregarded or underrepresented
in retrieval systems simply because they do not
enjoy the "status" of terms. Terminology has a role
to play in knowledge ordering, but the expression of
knowledge requires concepts corresponding to lexical
and grammatical items beyond the class of nouns, and
beyond the single word. But first of all, on the
simplest level, we must understand and recognise
that in the context of retrieval, a "term" like
"other peoples' mail" is as good as "confidential
mail". With the introduction of the idea of
"knowledge need", we have to consider also the
intellectual processes of knowledge acquisition,
such as comparison. A complete mapping of users'
discourse onto the discourse of the manual, for
which the methodological principles have been
established here, could be the next step, and would

provide further practical pointers for the design of retrieval aids. This could be done on a part of speech basis, eg. starting with mapping users' verbs to nouns.

The characterisation of the important domain of computer security from the two perspectives of knowledge and language, and the description of its specificity, has set a methodological precedent for the analysis of other domains. It was striking and noteworthy to see so many spheres of special knowledge - science and technology, computing, mathematics, security, business, law, language, as well as world knowledge -, and so many knowledge types - 19 in all - represented in what, on the face of it, might appear to be a monolithic domain structure. The fact that users' queries hardly touch some of these fields may be interpreted in several ways: for instance, they may not be aware of these aspects, or they may lack the confidence to probe further, or it could be that the more specific questions which bring in these domains would arise out of actual situations or events, rather than a position of general questioning and planning. On the other hand, the strong presence of some of these domains points to the blurring of boundaries between fields of knowledge. This is important from a theoretical standpoint, and has implications for practice in that technical authors ought to have, and to show in their writing, a greater awareness of 'the world beyond the system', so that the user can be helped to relate special subject knowledge to the operational environment - eg. business.

Not only do domains of knowledge impact upon one another in practice, but their interdependence results from use of a common abstract stratum of

language and from the limitations of word-stock imposing reuse and multiplicity of meanings. Access to knowledge depends on users' linguistic competence: their ability to express their knowledge needs in language, and especially their accuracy (precision) in selecting elements of language and correct graphological form. The linguistic and terminological access problems which have been identified for users 'as a group' can be intensified by individual difficulties. These difficulties will probably remain; we cannot insure either against the inconsistencies and imprecisions of query formulation in natural language as such; however, further work could be done on identifying linguistic differences in users with different levels of experience.

Although we were not seeking to specify in detail what an ideal access mechanism or index might look like, the research carried out allows us to formulate some new and important principles:-

(a) Readers' needs for instructions, understanding, advice and control should be considered prior to the design of appropriate access mechanisms for a given document

(b) Different types of knowledge should be made visible; for instance, where advice is given in the text it should be identified to the reader as being advice (eg. through the use of typographical, graphical or spacial features) both in the text itself and in the index

(c) Abstract notions from the general language (eg. difference, method, rationale, limits, advantages, new/old, comprehensive, unless, while

...) are a significant feature of readers'
discourse, as are high level domain concepts (like
system, backup, authority, access, personal/public,
illegal ...); such items should be included in the
index

(d) Users' concern with timing (eg. frequency,
reuse, again ...), manner (eg. way, steps, remotely
...), possibility (eg. potential, available ...),
advisability (eg. should ...) , and necessity (eg.
necessary ...) should be reflected in the index

(e) All parts of speech - not only nouns - are
candidates for the index, and should be given
consideration during index creation; verbs should be
given particular attention

(f) The identification of other domains which
have an impact on the comprehension of the main
subject field should be carried out to ensure that
concepts from those domains are included (eg.
business concerns relating to best method, optimal
time, decision-making with a view to improvement,
'people concepts' (eg. staff, no one))

(g) The reader will need to be made aware of
these novel entry possibilities, i.e. the new
indexing principles need to be brought to the
reader's attention and explained

(h) The preparation of an index should not be
left to the end of a writing project - its
compilation will help to determine what should be
the content of the document

Seeing that our investigations have put the
spotlight on matters of value judgment (advice) and

control (both strategic and tactical), it would be easy to jump to the conclusion that an expert system was what was required. The problems with developing expert systems are many, not least of which is the investment of a great deal of time and resources. However, there may be another way. The text of the manual already harbours much of this knowledge. It is then a question of identifying where in the manual it may be found, and developing the access mechanisms. We know how difficult it is to elicit and represent expert knowledge, but we are only just realising that the communication of expert knowledge to the learner is as difficult, and is perhaps the reason why expert systems have not 'taken off' in the way that was expected. This is where efforts must now be focussed.

There is, of course, the question of whether users will want to turn to textual information when colleagues were found to be by far the most preferred source of knowledge. This would need to be investigated further, to see whether the main factor in this choice was, for example, ease of access, language, or knowledge. In any case, we have also confirmed that retrieval problems are not magically removed by automation. Manuals are still preferred over on-line information. Besides, language-related problems can be amplified if users are not given a chance to exercise their judgment. This is where the conventional 'back of book' index has an advantage: large chunks of it can be viewed at a time and scanned. One of the inherent drawbacks of the index is that it is poised between detail and overview.

Our research has also contributed to the exploration of the problems of ambiguity, highlighting the various interpretations that some questions invite.

207

We have in some measure managed to 'define the vagueness', capture the uncertainty, of users' discourse, and have grappled with the notion of 'simpler wording', trying to demonstrate that advice on simplification should be more firmly justified and made more specific.

The strategy which was adopted here and which could be generalised for the improvement of technical documentation and its access tools consists of a two-pronged approach: domain and user. If we do not know users' needs, we cannot respond to them; if we do not know the domain, we cannot fully understand users' needs. In the domain of computer security, there is the uncertainty of risk, the repetitiveness of the maintenance of security, the sometimes thorny question of privacy, and so forth. Comparisons between domains must follow, before the specificities of this one can be confirmed, and it would be instructive to see the differences between questions pertaining to the IBM AS/400 system and, say, PC security. Natural language representations of computer security knowledge could also be examined in relation to discourse type, to have a clearer picture of the constraints imposed.

The need for automated training aids is becoming apparent in many domains of knowledge. As has been pointed out all along, documentation is increasingly computer-based, and the possibility of computerised retrieval must always be borne in mind. The problem identification focus adopted in this research does not preclude an electronic implementation of a solution; equally, it has implications for more traditional presentation methods. The issues addressed are of a fundamental nature, spanning epistemology, human communication and learning

theory. The graphic aspect of knowledge transfer systems - incorporating layout and typographic features of texts, and our knowledge of how these features affect learning - is an important one which should not be lost in the race for ever more "glossy" presentation made possible by sophisticated printing and computer display techniques.

The 'back of book' index seems to imply an afterthought. Let us hope that matters of retrieval will be kept at the forefront of documentation designers' minds.

<u>BIBLIOGRAPHY</u>

Aarts, J., Meijs, W. (eds.) (1986) Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora, Costerus, Vol.57, Rodopi, Amsterdam

Abbott, V., Black, J.B. (1986) "Goal-Related Inferences in Comprehension" in: J.A. Galambos et al., 1986

Aitchison, J., Gilchrist, A. (1972) Thesaurus · Construction: A Practical Manual, Aslib, London

Aitchison, J. (1987) Words in the Mind, Basil Blackwell, Oxford

Alberico, R., Micco, M. (1990) Expert Systems for Reference and Information Retrieval, Meckler, London

Albert, S., Ashby, V.A., Hicks, S.E., Marietta, M. (1992) "Reference Model for Data Management Security and Privacy", SIG Security Audit and Control Review, 10, 2, pp. 44-62

Alexander, P.A. (1992) "Domain Knowledge: Evolving Themes and Emerging Concerns", Educational Psychologist, 27, 1, pp. 33-51

Allport, D.A., Funnell, E. (1981) "Components of the mental lexicon", Philosophical Transactions of the Royal Society of London B295, pp. 397-410

Anderson, J.R. (1985) Cognitive Psychology and its Implications, Freeman, New York

Audit Commission, The (1985) Computer Fraud Survey, Her Majesty's Stationery Office, London

Audit Commission, The (1987) Survey of Computer Fraud and Abuse, Her Majesty's Stationery Office, London

Austin, J.L. (1975) How to Do Things with Words, Clarendon Press, Oxford

Barrett, E. (ed.) (1988) Text, Context and Hypertext, MIT Press, Cambridge, MA

Barrett, E. (ed.) (1989) The Society of Text, MIT Press, Cambridge, MA

Barry, C.L. (1992) "The identification of user criteria and document traits: beyond the topical

approach to information retrieval", abstract, Proceedings of ASIS annual meeting, 29, pp. 330-331

Beaugrande, R. de (1980) Text, Discourse, and Process: Toward a Multidisciplinary Science of Texts, Longman, London

Beaugrande, R. de, Dressler, W. (1980) Introduction to Text Linguistics, Longman, London

Beaugrande, R. de, (1988) "Systemic Versus Contextual Aspects of Terminology" in: Czap & Galinski, 1988

Becker, R.S. (1977) The Data Processing Security Game: safeguarding against the real dangers of computer abuse, Pergamon, Oxford

Beeken, J., Geerts, G., Van Belle, W. (1990) The CONST-Project: Computer-Instructed Writing Techniques, Paper presented at Computers & Writing III, Edinburgh, April 1990

Belkin, N.J., Brooks, H.M., Daniels, P.J. (1988) "Knowledge elicitation using discourse analysis" in: Gaines & Boose, 1988

Bereiter, C., Scardamalia, M. (1982) "From Conversation to Composition: the Role of Instruction in a Developmental Process" in: Glaser, R. (ed.) Advances in Instructional Psychology, Vol.2, Lawrence Erlbaum, Hillsdale, N.J.

Bereiter, C., Scardamalia, M. (1987) The Psychology of Written Communication, Lawrence Erlbaum Associates, Hillsdale, NJ

Bernhardt, S. (1986) "Seeing the Text", College Composition and Communication, 37, 1, pp. 66-78

Blair, D.C. (1992) "Information Retrieval and the Philosophy of Language", The Computer Journal, 35, 3, pp. 200-207

Bordogna, G., Pasi, G. (1993) "A Fuzzy Linguistic Approach Generalizing Boolean Information Retrieval", Journal of the American Society for Information Science, 44, 2, pp. 70-82

Borko, H. (1978) Indexing concepts and methods, Academic Press, London

Bradburn, D. (1990) "An Introduction to Data Security" in: Hearnden, 1990

Bridwell, L.S., Nancarrow, P.R., Ross, D. (1984)
"The Writing Process and the Writing Machine:
Current Research on Word Processors Relevant to the
Teaching of Composition" in: R. Beach and L.S.
Bridwell (eds.) New Directions in Composition
Research, The Guilford Press

Britton, B.K., Glynn, S.M. (eds.) (1989) Computer
Writing Environments: Theory, Research, and Design,
Lawrence Erlbaum Associates, Hove and London

Broadhurst, A. (1991) Health and Safety, Pitman

Browning, C. (1984) Guide to Effective Software
Technical Writing, Prentice-Hall, New Jersey

Brückler, H., Florian, D., Kalcher, R. (1988) "SAFIR
- Smart Assistant for Information Retrieval" in:
Czap & Galinski, 1988

BSI (1988) Recommendations for the preparation of
indexes to books, periodicals and other documents,
BS 3700:1988, British Standards Institution, London

Buchanan, R.A. (1990) Textbase Technology - Writing
with Reusable Text, Paper presented at Computers and
Writing III, Heriot-Watt University, Edinburgh,
April 1990

Budin, G., Galinski, C., Nedobity, W., Thaller, R.
(1988) "Terminology and Knowledge Data Processing"
in : Czap & Galinski, 1988

Burns, A., McDermid, J., Dobson, J. (1992) "On the
Meaning of Safety and Security", The Computer
Journal - Special issue: Safety and
Security/Parallel Computation, 35, 1, pp. 3-15

Burrill, V.A. (1986) "VORTEXT: VictORias TEXT
reading and authoring system" in: van Vliet, J.C.
(ed.) Text Processing and Document Manipulation,
Cambridge University Press, Cambridge

Calkins, L. (1980) "Children's Rewriting
Strategies", Research in the Teaching of English,
14, pp. 331-341

Carnap, R. (1956) Meaning and Necessity, University
of Chicago Press, Chicago

Carroll, J.B. (1972) "Defining Language
Comprehension: Some Speculations" in: Freedle &
Carroll, 1972

Carter, R. (1987) Vocabulary: Applied Linguistic

Perspectives, Routledge, London

Cawsey, A., Galliers, J., Reece, S., Sparck Jones, K. (1992) "Automating the librarian: belief revision as a base for system action and communication with the user", The Computer Journal, 35, 3, pp. 221-232

Ceccato, S. (1964) "Automatic translation of languages", Information Storage and Retrieval, 2, 3, pp. 105-158

Chafe, W.L. (1972) "Discourse Structure and Human Knowledge" in: Freedle & Carroll, 1972

Chen, P. (1976) The Entity-Relationship Model - Towards a Unified View of Data, ACM TODS 1, 1, pp. 9-36

Clancey, W.J. (1983) "The Epistemology of a Rule-Based Expert System - A Framework for Explanation", Artificial Intelligence, 20, 3, pp. 215-251

Clark, L.F. (1985) "Social Knowledge and Inference Processing in Text Comprehension" in: Rickheit & Strohner, 1985

Collins, A., Gentner, D. (1980) "A Framework for a Cognitive Theory of Writing" in: Gregg, L.W., Steinberg, E.R. (eds.) Cognitive Processes in Writing, Lawrence Erlbaum, New Jersey

Conte, P. (1990) "A Principled Approach to Computer Security", News 3X/400, January 1990, pp. 34-48

Cooper, C.R., Matsuhashi, A. (1983) "A Theory of the Writing Process" in: M. Martlew (ed.) The Psychology of Written Language: A Developmental Approach, John Wiley & Sons, Chichester

Cooper, J.A. (1989) Computer and Communications Security: Strategies for the 1990s, McGraw-Hill, New York and London

Crombie, W. (1985) Process and Relation in Discourse and Language Learning, Oxford University Press, London

Crothers, E.J. (1979) Paragraph Structure Inference, Ablex, Norwood, N.J.

Cruse, D.A. (1986) Lexical Semantics, Cambridge University Press, Cambridge

Cullingford, R.E. (1986) Natural Language Processing: A Knowledge-Engineering Approach, Rowman

& Littlefield, New Jersey

Czap, H., Galinski, C. (eds.) (1987) Terminology and Knowledge Engineering, INDEKS Verlag, Frankfurt/M.

Czap, H., Galinski, C. (eds.) (1988) Terminology and Knowledge Engineering - Supplement, INDEKS Verlag, Frankfurt/M.

Czap, H., Nedobity, W. (eds.) (1990) TKE'90: Terminology and Knowledge Engineering, Vols.1 & 2, INDEKS Verlag, Frankfurt/M.

Delavigne, V., Guespin, L. (1992) "Nucléaire: risque et sécurité - une recherche en socioterminologie", Terminologies nouvelles, Rint (Réseau international de néologie et de terminologie), 8, pp. 19-25

Dijk, T. van, Petöfi, J. (eds.) (1977) Grammars and Descriptions: Studies in Text Theory and Text Analysis, de Gruyter, Berlin

Dressler, W. (ed.) (1978) Current Trends in Text Linguistics, de Gruyter, Berlin

Duffy, T.M., Mehlenbacher, B., Palmer, J. (1989) "The Evaluation of Online Help Systems: A Conceptual Model" in: Barrett, E. (ed.) The Society of Text, The MIT Press, Cambridge, MA

Elbra, R.A. (1992) Computer Security Handbook, 2nd edition, NCC Blackwell, Oxford

Ellison, J.R., Pritchard, J.A.T. (1987) Security in Office Systems, NCC Blackwell, Oxford

Emden, J. van (1990) A Handbook of Writing for Engineers, Macmillan, London

Faigley, L., Witte, S.P. (1984) "Measuring the Effects of Revisions on Text Structure" in: Beach, R., Bridwell, L.S. (eds.) New Directions in Composition Research, The Guilford Press

Fass, D. (1987) Collative Semantics: An Overview of the Current Meta5 Program, Memorandum MCCS-87-112, Computing Research Laboratory, New Mexico State University, New Mexico

Fay, J.J. (1987) Butterworths Security Dictionary: terms and concepts, Butterworths, London

Felber, H. (1984) Terminology Manual, PGI-84/WS/21, UNESCO/ Infoterm, Paris

Fetzer, J.H. (1990) Artificial Intelligence: Its Scope and Limits, Kluwer

Fillenbaum, S., Rapoport, A. (1971), Structures in the Subjective Lexicon, Academic Press, New York

Fillmore, C.J. (1968) "The case for case" in: Bach, E., Harms, R.T. (eds.) (1968) Universals in Linguistic Theory, Holt, Rinehart and Winston, New York

Firth, J.R. (1957) Papers in Linguistics, London

Fodor, J.D., Fodor, J.A., Garrett, M.F. (1975), "The psychological unreality of semantic representations", Linguistic Inquiry 4, pp. 515-531

Ford, N., Ford, R. (1992) "Learning strategies in an 'ideal' computer-based learning environment", British Journal of Educational Technology, 23, 3, pp. 195-211

Freedle, R.O., Carroll, J.B. (eds.) (1972) Language Comprehension and the Acquisition of Knowledge, V.H. Winston & Sons, Washington, D.C.

Friedman, M. (1987) "WANDAH - A computerized writer's aid" in: D. Berger, Pezdek, and W. Banks (eds.) Applications of cognitive Psychology: Problem solving, education, and computing, Lawrence Erlbaum Associates, Hillsdale, NJ.

Friedman, M., Rand, E. (1989) "A computer-based writing aid for students: present and future" in: Britton & Glynn

Gagné, R.M., Briggs, L.J. (1979) Principles of Instructional Design, 2nd edition, Holt, Rinehart and Winston, New York and London

Gaines, B.R. (1988) "An overview of knowledge-acquisition and transfer" in: Gaines & Boose, 1988

Gaines, B.R., Boose, J.H. (eds.) (1988) Knowledge Acquisition for Knowledge-Based Systems, Vol. 1, Academic Press, London

Galambos, J.A., Abelson, R.P., Black, J.B. (eds.) (1986) Knowledge Structures, Lawrence Erlbaum Associates, London

Gambier, Y. (1987) "Problèmes terminologiques des pluies acides: pour une socioterminologie", Meta, 32, 3, pp. 314-320

Gauch, S., Smith, J.B. (1993) "An Expert System for Automatic Query Reformulation", Journal of the American Society for Information Science, 44, 3, pp. 124-136

Genuardi, M.T. (1990) "Knowledge-based machine indexing from natural language text: knowledge base design, development and maintenance" in: Czap & Nedobity, 1990

Geoffrion, L.D., Geoffrion, D.P. (1983) Computers and Reading Instruction, Addison-Wesley

Ginzburg, R.S., Khidekel, S.S., Knyazeva, G.Y., Sankin, A.A. (1979) A Course in Modern English Lexicology, 2nd. ed., Vyssaja Skola, Moscow

Graham, I., Jones, P.L. (1988) Expert Systems: Knowledge, Uncertainty and Decision, Chapman & Hall, London

Hamlyn, D.W. (1970) The Theory of Knowledge, Macmillan, London and Basingstoke

Hansjörg, P. (1992) "Un index bilingue du droit de l'environnement", Terminologies nouvelles, Rint (Réseau international de néologie et de terminologie), 8, pp. 32-35

Hart, A. (1986) Knowledge Acquisition for Expert Systems, 1st edition, Kogan Page, London

Hart, A. (1989) Knowledge Acquisition for Expert Systems, 2nd edition, Kogan Page, London

Hartley, J. (1985) Designing Instructional Text, 2nd ed., Kogan Page, London

Hayes, J.R., Flower, L.S. (1980) "The Dynamics of Composing" in: L.W. Gregg and E.R. Steinberg (eds.) Cognitive Processes in Writing, Lawrence Erlbaum Associates, New Jersey

Hearnden, K. (ed.) (1990) A Handbook of Computer Security, Kogan Page, London

Highland, H.J. (1990) Computer Virus Handbook, Elsevier Advanced Technology, Oxford

Hockey, S.M. (1980) A Guide to Computer Applications in the Humanities, Johns Hopkins University Press, London

Hörmann, H. (1977) Psychologie der Sprache, Springer Verlag, Berlin

Hruska, J., Jackson, K. (1990) Computer Security Solutions, Blackwell, Oxford

Hsiao, D.K., Kerr, D.S., Madnick, S.E. (1979) Computer Security, Academic Press, London

Hudson, R. (1984) Word Grammar, Basil Blackwell, Oxford

Hutchinson, T., Waters, A. (1987) English for Specific Purposes: a learning-centred approach, Cambridge University Press, Cambridge

IBM (1991) Application System/400: Security Concepts and Planning, Version 2, SC41-8083-00, IBM, Rochester

Ilson, R.F. (1987) "Towards a taxonomy of dictionary definitions" in: R.F. Ilson (ed.) A Spectrum of Lexicography, John Benjamins Publishing Company, Amsterdam

Ishiwata, T. et al. (1985) Basic Specifications of the Machine Readable Dictionary, ICOT Technical Report, TR-100

Itoga, M. (1992) "Seeking Understanding beneath the Unspecifiable: an Alternative Framework for Mapping Information Needs in Communication", Libri, 42, 4, pp. 330-344

Johnson-Laird, P.N. (1987) "The mental representation of the meaning of words", Cognition 25, 1, pp. 189-211

Kakkuri-Knuuttila, M-L., Kusch, M. (1991) "LSP Research, Philosophy of Science, and the Question-Theoretical Approach - some Tentative Suggestions" in: Schröder, 1991

Katz, B. (1988) "Text processing with the START natural language system" in: Barrett, 1988

Kay, M.H. (1985) Textmaster - a document-retrieval system using CAFS-ISP. ICL Technical Journal, 4, 4, pp. 455-467

Kettle, J.D.N. (1992) "Computer Security: mystery or science ?", IEE Review, 38, 5, pp. 197-200

Kieras, D.E. (1981) "Component processes in the comprehension of simple prose", Journal of Verbal Learning and Verbal Behaviour, 20, 1-23

Kieras, D.E. (1982) "A model of reader strategy for

abstracting main ideas from simple technical prose",
Text, Vol.2-1/3, Special issue: New developments in
cognitive models of discourse processing, pp. 47-82

Kieras, D.E. (1983) "A simulation model for the
comprehension of technical prose" in: G.H. Bower
(ed.) The psychology of learning and motivation, 17,
Academic Press, New York, NY

Kieras, D.E. (1985a) "Improving the
comprehensibility of a simulated technical manual",
Technical Report No.20, TR-85/ONR-20, University of
Michigan

Kieras, D.E (1985b) "The potential for advanced
computerized aids for comprehensible writing of
technical documents", Technical Report No.17, TR-
85/ONR 17, University of Michigan

Kieras, D.E. (1985c) "Thematic processes in the
comprehension of technical prose" in: B.Britton &
J.Black (eds.) Understanding expository text,
Lawrence Erlbaum Associates, Hillsdale, NJ.

Kieras, D.E. (1987) "A computerized comprehensive
writing aid: Final Report", Technical Report No.27,
FR-87/ONR-27, University of Michigan

Kieras, D.E. (1989) "An advanced computerized aid
for the writing of comprehensible technical
documents" in: Britton & Glynn, 1989

Kieras, D.E., Dechert, C. (1985) "Rules for
comprehensible technical prose: A survey of the
psycholinguistic literature", Technical report
no.21, TR-85/ONR-21, University of Michigan

Kincaid, J.P, Aagard, J.A., O'Hara, J.W., Cottrell,
L.K. (1981) "Computer readability editing system",
IEEE Transactions on Professional Communications,
24, 38-41

Kintsch, W., van Dijk, T.A. (1978) "Toward a model
of discourse comprehension and production",
Psychological Review, No.85, pp. 363-394

Kittredge, R., Lehrberger, J. (eds.) (1982)
Sublanguage: studies of language in restricted
semantic domains, Walter de Gruyter, Berlin and New
York

Kolers, P.A., Wrolstad, M.E., Bouma, H. (eds.)
(1979) Processing of Visible Language, Plenum Press,
London

Krageloh, K.D., Lockemann, P.C. (1978) "Access to data base systems via natural language" in: L. Bolc (ed.) Natural Language Communication with Computers, Springer-Verlag, Berlin

Kukulska-Hulme, A. (1988) "Computer-assisted reference skill development" in: Ager, D. (1988) Written Skills in the Modern Languages Degree, CILT, London

Kukulska-Hulme, A. (1989) "Dictionnaires informatisés et traduction", Meta, Vol.34, No.3, pp. 533-538

Kukulska-Hulme, A. (1990a) "A Dictionary View of Technical Writing", paper presented at Computers and Writing III, Heriot-Watt University, Edinburgh, April 1990

Kukulska-Hulme, A. (1990b) "Speed understanding of an unfamiliar domain" in: Czap & Nedobity, 1990

Kukulska-Hulme, A. (1990c) "Subject knowledge in computerised dictionaries" in: Thelen, M., Lewandowska-Tomaszczyk, B. (eds.) Translation and Meaning, Part 1 - Proceedings of the Maastricht-Lodz Colloquium on Translation and meaning, Euroterm, Maastricht

Kukulska-Hulme, A. (1991) "La méthodologie en terminologie et la rédaction technique", in: Guespin, L., Pierzo, V. (eds.) Terminologie et Enseignement des Langues, La Tilv, Paris

Kukulska-Hulme, A. (1992) Terminological Cross-mapping for a Sub-domain of Computing, ITL Review of Applied Linguistics, 97-98, pp. 187-204

Kwong, C. (1988) "Representing Time" in: Ringland & Duce, 1988

Landesman, C. (1972) Discourse and its Presuppositions, Yale University Press, New Haven and London

Lang, M. (1993) "Fighting off infection", IBM System User, Vol.14, No.3, p.59

Langacker, R.W. (1986) "An introduction to cognitive grammar", Cognitive Science 10, 1, 1-40

Larkin, J.H., Simon, H.A. (1987) "Why a diagram is (sometimes) worth ten thousand words", Cognitive Science 11, 1, 65-99

Last, R.W. (1989) Artificial Intelligence Techniques in Language Learning, Ellis Horwood, Chichester

Le Ny, J-F., Kintsch, W. (eds.) (1982) Language and Comprehension, North-Holland, Amsterdam

Lyons, J. (1981) Language, Meaning and Context, Fontana, London

Macdonald, N.H., Frase, L.T., Gingrich, P.S., Keenan, S.A. (1982) The writer's workbench: Computer aids for text analysis, IEEE Transactions on Communications, Vol.COM-30 (1), 105-110

Madden, W. (1990) "Plan for Security", News 3X/400, January 1990, pp. 26-33

Masterman, M. (1961) "Semantic message detection for machine translation, using an interlingua", Proc. 1961 International Conference on Machine Translation, 438-475

Mayer, J., Kieras, D.E. (1987) A development system for augmented transition network grammars and a large grammar for technical prose, Technical Report No. 25, TR-87/ONR-25, University of Michigan

Mayer, R., Maier, E. (1987) "Spezifikation eines Konzeptionellen Schemas fur Terminologiedatenbanken" in: Czap & Galinski, 1987

McAleese, R. (ed.) (1989) Hypertext: theory into practice, Intellect, Oxford

McGrew, P.C., McDaniel, W.D. (1989) On-line Text Management: Hypertext and Other Techniques, J.Ranade IBM Series, Intertext, London

McNaught, J. (1987) Round Table on Multilingual Knowledge Bases and Linguistic Applications, Report to the Agence Linguistique Européenne and the Commission of the European Communities, Directorate-General XIII, Manchester

Meadow, C.T. (1992) Text Information Retrieval Systems, Academic Press, London

Mellish, C.S. (1985) Computer Interpretation of Natural Language Descriptions, Ellis Horwood Ltd., Chichester

Miller, G.A., Johnson-Laird, P.N. (1976) Perception and Language, Cambridge University Press, Cambridge

Moulin, B., Rousseau, D. (1990) "Un outil pour

l'acquisition des connaissances à partir de textes prescriptifs," ICO (Revue de liaison de la recherche en informatique cognitive des organisations), Vol. 2, No. 6, pp. 27-42

Myers, G. (1989) "The pragmatics of politeness in scientific articles", Applied Linguistics, Vol. 10, pp. 1-35

Narayanan, A. (1986) "Memory models of man and machine" in: M. Yazdani (ed.) Artificial Intelligence: Principles and Applications, Chapman and Hall, London

NCC (1991) Abbreviated CRAMM Reviews, Version 1.0, National Computing Centre Ltd., Manchester

NCC (1992) Survey of Security Breaches Report, National Computing Centre Ltd., Manchester

Nedobity, W. (1990) "Methodological and practical aspects of computer-assisted knowledge transfer to developing countries" in: Czap & Nedobity, 1990

Nissan, E. (1987) "Exception-admissibility and typicality in proto-representations" in: Czap & Galinski, 1987

Nuttall, C. (1982) Teaching Reading Skills in a Foreign Language, Heinemann

Nystrand, M. (1986) The Structure of Written Communication, Academic Press, London

Ogden, C.K., Richards, I.A. (1923) The Meaning of Meaning, Harcourt, Brace, and World, New York

Oliver, E., Wilson, J. (1983) Practical Security in Commerce and Industry, Gower

Ortony, A., Clore, G.L., Foss, M.A. (1987) "The referential structure of the affective lexicon", Cognitive Science 11, 3, 341-364

Papegaaij, B.C., Sadler, V., Witkam, A.P.M. (1986) Word Expert Semantics: an Interlingual Knowledge-Based Approach, Foris Publications, Dordrecht

Pears, D. (1972) What is Knowledge ?, George Allen & Unwin, London

Picht, H., Draskau, J. (1985) Terminology: An Introduction, University of Surrey

Pritchard, J.A.T. (1979) Computer Security: Facts and Figures, NCC, Manchester

Quillian, R. (1968) "Semantic Memory" in: Minsky, 1968, 227-270

Raymondis, L.M., Le Guern, M. (1976) Le langage de la justice pénale, CNRS, Paris

Reiser, B.J., Black, J.B. (1982) "Processing and structural models of comprehension", Text, Vol.2-1/3, Special issue: New developments in cognitive models of discourse processing, pp. 225-252

Rentell, M., Jenner, P. (1991) Breakdowns in Computer Security, Computer Weekly Publications, Sutton

Rickheit, G., Schnotz, W., Strohner, H. (1985) "The Concept of Inference in Discourse Comprehension" in : Rickheit & Strohner (1985)

Rickheit, G., Strohner, H. (eds.) (1985) Inferences in Text Processing (Advances in Psychology 29), North-Holland, Amsterdam

Ringland, G.A., Duce, D.A. (eds.) (1988) Approaches to Knowledge Representation: An Introduction, Research Studies Press, Letchworth

Roberts, D.W. (1990) Computer Security: Policy Planning and Practice, NCC Blackwell, Oxford

Rosch, E. (1975) "Cognitive representations of semantic categories", Journal of Experimental Psychology: General 104, 192-233

Rowntree, D. (1986) Teaching through Self-instruction, Kogan Page, London

Ruge, G. (1992) "Experiments on linguistically-based term associations", Information Processing & Management, 28, 3, pp. 317-332

Sadler, V. (1989) Working with Analogical Semantics: Disambiguation Techniques in DLT, Foris Publications, Dordrecht

Sager, J.C., Dungworth, D., McDonald, P.F. (1980) English Special Languages, Brandstetter Verlag, Wiesbaden

Sager, J.C. (1990) A Practical Course in Terminology Processing, John Benjamins, Amsterdam

Sager, J.C. (1991) "A Theory of Text Production, Modification, Reception" in: Schröder, H., 1991

Sanford, A.J., Garrod, S.C. (1982) "Towards a Psychological Model of Written Discourse Comprehension" in: Le Ny, J-F, Kintsch, W., 1982

Saracevic, T. (1992) "The remarkable staying power of relevance as basic criterion for information retrieval: why ?", abstract, Proceedings of ASIS annual meeting, 29, pp. 330-331

Saussure, F. de (1972) Cours de linguistique générale, Payot, Paris

Scardamalia, M., Bereiter, C. (1993) "Technologies for knowledge-building discourse", Communications of the ACM, 36, 5. pp. 37-41

Schank, R.C. (1972) "Conceptual dependency: a theory of natural language understanding", Cognitive Psychology 3, 552-631

Schank, R.C., (1975) "The Structure of episodes in memory" in: D.G. Bobrow & A.M. Collins (eds.), Representation and Understanding: Studies in cognitive science, Academic Press, New York

Schank, R.C., Abelson, R.P. (1977) Scripts, Plans, Goals, and Understanding, Lawrence Erlbaum Press, Hillsdale, N.J.

Schank, R.C., Colby, K.M. (eds.) (1973) Computer Models of Thought and Language, Freeman, San Francisco

Schmitz, K-D. (ed.) (1993) TKE'93: Terminology and Knowledge Engineering, Indeks Verlag, Frankfurt/M.

Schröder, H. (ed.) (1991) Subject-oriented Texts: Languages for Special Purposes and Text Theory, de Gruyter, Berlin & New York

Searle, J.R. (1979) Expression and Meaning: studies in the theory of speech acts, Cambridge University Press, Cambridge

Sharp, B. (1989) Elaboration and testing of new methodologies for automatic abstracting, PhD thesis, Aston University, Birmingham

Sharples, M. (1985) Cognition, Computers and Creative Writing, Ellis Horwood, Chichester

Sides, C. (1984) How to Write Papers and Reports about Computer Technology, ISI Press, Philadelphia

Simons, G.L. (1989) Viruses, Bugs and Star Wars: The Hazards of Unsafe Computing, NCC Blackwell, Oxford

Sinclair, J. (1991) Corpus, Concordance, Collocation, Oxford University Press, Oxford

Skuce, D., Meyer, I. (1990) "Concept Analysis and Terminology: A Knowledge-Based Approach to Documentation", Proceedings of the Thirteenth International Conference on Computational Linguistics (COLING 90)

Skuce, D. (1993) "A System for Managing Knowledge and Terminology for Technical Documentation" in: Schmitz

Slatter, P.E. (1987) Building Expert Systems: Cognitive Emulation, Ellis Horwood, Chichester

Small, S. (1980) Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding, University of Maryland, Dept. of Computer Science, Maryland (USA).

Smeaton, A. (1992) "Progress in the application of natural language processing to information retrieval tasks", The Computer Journal, 35, 3, pp. 268-278

Smith, D.C.P., Smith, J.M. (1977) "Database abstraction: aggregation and generalisation", ACM Transactions on Database Systems, 2, 2, 105-133

Smith, F. (1982) Understanding Reading, 3rd ed., Holt, Rinehart and Winston, London

Smith, J. (ed.) (1988) Information Security Yearbook, 1988/89, IBC Technical Services Ltd.

Society of Indexers (1993) Services to Indexers; Services to Publishers; Services to Authors, The Society of Indexers, London

Somers, H.L. (1980) ISSCO*PTOSYS - Brief description and user manual, Working Paper No.41, Université de Genève

Sowa, J.F. (1984) Conceptual Structures, Addison-Wesley, Reading MA

Sparck Jones, K. (1971) Automatic Keyword Classification for Information Retrieval, Butterworths, London

Sparck Jones, K. (1985) "Compound noun interpretation problems" in: F.Fallside, W.A. Woods (eds.) Computer Speech Processing, Prentice-Hall International, London

Stemberger, N. (1985) The Lexicon in a Model of Speech Production, Garland, New York

Stuart, A. (1984) Writing and Analyzing Effective Computer System Documentation, Holt, Rinehart and Winston, New York and London

Swales, J.M. (1990) Genre Analysis: English in academic and research settings, Cambridge University Press, Cambridge

Taylor, R. (1968) "Question-negotiation and information seeking in libraries", College & Research Libraries, May, pp. 178-194

Teorey, T.J., Fry, J.P. (1982) Design of Database Structures, Prentice-Hall, New Jersey

Thomason, R.H. (1989) Philosophical Logic and Artificial Intelligence, Kluwer

Tsichritzis, D.C., Klug, A. (eds.) (1978) "The ANSI/X3/SPARC DBMS framework", Information Systems 3, 173-191

Turney, A. (ed.) (1988) Applied Text Linguistics, Exeter Linguistic Studies, Vol.13, University of Exeter

Utne, I. (1987) "Technical terminology at the Norwegian termbank – a basis for knowledge representation and information retrieval" in: Czap & Galinski, 1987

Viehweger, D. (1991) "Lexical Knowledge and Text Interpretation" in: Schröder, 1991

van Vliet, J.C. (ed.) (1986) Text Processing and Document Manipulation, Cambridge University Press, Cambridge

Waern, Y. (1989) Cognitive Aspects of Computer Supported Tasks, John Wiley & Sons, Chichester

Waldron, R.A. (1967) Sense and Sense Development, André Deutsch

Wessel, A.E. (1975) Computer-Aided Information Retrieval, Melville Publishing Co., Los Angeles

Whiting, H.T.A. (1975) Concepts in Skill Learning, Lepus Books, London

Wierzbicka, A. (1972) Semantic Primitives, Athenaum Verlag, Frankfurt

Wilbur, W.J., Sirotkin, K. (1992) "The automatic identification of stop words", Journal of Information Science, 18, 1, pp. 45-55

Wilensky, R. (1983) Planning and Understanding: A Computational Approach to Human Reasoning, Addison-Wesley, Reading, MA

Wilks, Y.A. (1972) Grammar, Meaning, and the Machine Analysis of Language, Routledge & Kegan Paul, London

Wilks, Y.A. (1973) "An artificial intelligence approach to machine translation" in: Schank & Colby, 1973

Williams, R., Cummings, S. (1993) An Informal Dictionary of Computer Terms, Peachpit Press, Berkeley, CA

Williams, T., Bainbridge, B. (1988) "Rule Based Systems" in: Ringland & Duce, 1988

Wood, M.B. (1985) Computer Access Control, NCC Publications, Manchester

Wüster, E. (1974) "Die Allgemeine Terminologielehre - ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften", Linguistics, 119, pp. 61-106

Wyllie, J. (1990) "Concept Indexing: the world beyond the windows", ASLIB Proceedings, 42, 5, pp. 153-159

Zadeh, L.A. (1975) "Calculus of fuzzy restrictions" in: L.A. Zadeh, K.S. Fu, K. Tanaka, M. Shimura (eds.), Fuzzy sets and their applications to cognitive and decision processes, Academic Press, New York

Questionnaire

Note: Originally, it was thought that some of the
users to whom the questionnaire was being addressed
might be System/36 rather than AS/400 users; in the
event, all respondents used the AS/400.

Ref._____

Aston University Modern Languages Department
Doctoral Research Project

SURVEY OF AS/400 AND S/36 USERS

Questionnaire

This research project is concerned with the way computer users
express their information needs. It aims to find out to what
extent the language used to express these needs corresponds to
the language used in IBM manuals. The results will help to
establish how the information contained in technical
documentation could be made more easily accessible to users.

We would like to find out what are the questions you have asked
in the past, or would now like to ask, about the security of your
computer system.

All answers will remain **absolutely confidential** and will not be
associated with your identity at any stage.

Agnes Kukulska-Hulme
Lecturer in Computational
Linguistics

In this section you are asked for some details of your background and present position.

1. Which system do you use ?

AS/400	
S/36	

2. Your present position: are you a 'computer professional' eg. computer manager, systems manager, systems analyst, etc. ?

Yes		Job title
No		Job title

3. What is your previous professional experience ?

       (a) in computing

```
..................................................................
```

       (b) other professional experience

```
..................................................................
```

4. How would you rate your knowledge and experience of computers in general ?

Please tick the appropriate box:

1. very good	
2. good	
3. limited	
4. very limited	
5. none	

5. How would you rate your knowledge of the IBM AS/400 or S/36 operating system (whichever you use) ?

Please tick the appropriate box:

1. very good	
2. good	
3. limited	
4. very limited	
5. none	

6. How would you rate your overall knowledge of computer security in general ?
(We are taking 'computer security' in a broad sense here, to cover all aspects you might care to include, such as physical security, data security, network security, passwords, viruses, and so on.)

Please tick appropriate box:

1.	very good	
2.	good	
3.	limited	
4. very limited		
5.	none	

7. How would you rate your knowledge of computer security <u>on your system</u> ?

Please tick appropriate box:

1.	very good	
2.	good	
3.	limited	
4. very limited		
5.	none	

8. Which of the following sources would you turn to for getting information on computer security relating to your system ?

Please indicate order of preference by numbering these from 1 to 11, where 1 = most preferred source
11 = least preferred source

A. books on computer security	
B. colleagues	
C. external telephone support	
D. IBM representative	
E. information packs	
F. magazine articles	
G. IBM manuals on computer security	
H. on-line information	
I. television programmes	
J. training courses and materials	
K. other	

Comments: ....

2

SECTION II

In this section, you are asked to write down twenty questions that you could ask or could be asked about the security of your system.

These can be questions you have asked yourself or others in the past, questions that have been put to you, or questions that now occur to you. Try to write them down just as you would spontaneously say them. Assume you can ask anything at all, and do not worry about whether or not the questions can be answered. You can also include questions you already know the answer to, but consider to be relevant.

Your questions can be as general or specific as you like. For instance, you could ask something very general like "Is my system secure ?", "What are the precautions I should be taking ?" , and you can ask very specific questions like "What can I do if someone forgets their password ?", "Can I put a user on more than one group list ?", etc., or more complex technical questions, depending entirely on your knowledge of computer security. Please indicate the main source where you would expect to find the answer to each question, by entering a letter in the 'expected source of answer' column.

A. books on computer security
B. colleagues
C. external telephone support
D. IBM representative
E. information packs
F. magazine articles

G. IBM manuals on computer security
H. on-line information
I. television programmes
J. training courses & materials
K. other (specify)
? don't know

QUESTION	expected source of answer (A-K or ?)
1.	
2.	
3.	
4.	
5.	
6.	

3

QUESTION	expected source of answer (A–K or ?)
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	
20.	

If you have any further comments or information which you think might
be relevant, it will be very welcome:

Would you be willing to accept a telephone call if there are any
further points I would like to discuss with you ?

Yes	
No	

Thank you very much for your help.

Please send the completed questionnaire to:

        Mrs. Agnes Kukulska-Hulme
        Modern Languages Department
        Aston University
        Aston Triangle
        Birmingham B4 7ET

5

The following is a copy of the database of respondents' lists of queries (Section II of questionnaire).

1 Can I change a user's password for them ?
2 What security levels are available on the system ?
3 Can I allow Office users to share a signon I.D. and have separate passwords ?
4 Should I be allowing users to have more freedom on the system ?
5 Can a user have control over his/her own profile ?
6 Should users be limited to signing on at only one device ?
7 How many times should a user be allowed to try and sign on before the device becomes inactive ?
8 Can a user sign on to the system remotely ?
9 What are the rules on password reuse ?
10 Can a user give away objects that he/she owns ?
11 Is there an "audit trail" to show who has made changes to user profiles, etc. ?
12 Can you restore deleted objects ?

1 Can I have a document folder that only I can access (for confidential work) ?
2 Can everyone access everyone else's calendar ?
3 Can I look at other peoples' electronic mail ?
4 Is it possible to check the system for viruses ?
5 Is it possible to lock the computer screen ?
6 Do I have to sign off every time I leave my computer ?
7 Can other companies use the same network as us ?
8 How do you change your password ?
9 What do you do if someone's password expires ?
10 How do you add a new user to the system ?
11 Is it possible to find out other peoples' passwords ?
12 Can you use any parts of the system without first having a password and signing on ?

1 What flexibility/ options does the system offer on security ?
2 How many levels of password are there ?
3 What is the procedure for setting up the system security ?
4 Is it possible to be completely locked out of the system ?
5 If it is, then what do you do in that situation ?
6 What are the most common ways of breaching security ?
7 How can we best secure the system against a virus attack ?
8 What organisational procedures should we have to enhance security ?
9 How often should passwords be changed ?
10 How often should the security files be backed up ?
11 How resilient is the system to someone trying to hack their way in ?
12 What impact does comprehensive security have on response times ?

1   Physical security: Is there a Building Alarm and if so is it
      connected to local Police Station and are key holders
      informed ?

2   Physical security: Is computer dept protected by smoke alarm
      detectors and if so are they connected to Fire Station ?

3   Physical security: Does the computer suite have its own
      security i.e. limited access (key pad security or magnetic
      card) ?

4   Physical security: If computer suite has windows are they
      non breakable ?

5   Physical security: If alarm detection activated is
      electricity supply to computer suite cut off (i.e. air
      conditioning, computers) ?

6   Data security: Frequency of backing up system ?

7   Data security: Is on site backup stored in secure cabinet
      i.e. fire/water proof safe ?

8   Data security: Do you have off site backup; if yes, off site
      frequency and secure cabinet ?

9   Communications: Do you allow dial in access to your system;
      if yes, what security installed ?

10   Communications: Do you have password security when
      communicating with 3rd Party (i.e. EDI) ?

11   General: Do you have password and user Id's at sign on ?

12   General: If passwords available, how often are they changed,
      password reuse, lengths, etc. ?

13   General: What authorisation levels do you have i.e. Library,
      File, Field ?

14   General: Have IBM default passwords been reviewed on AS/400
      and have they been changed ? i.e. QSECOFR

15   General: Has user access to AS/400 Command line been
      removed ?

16   General: Are AS/400 History logs being reviewed on a regular
      basis to detect unauthorised access attempts ?

17   General: Has any 3rd Party Software on AS/400 been reviewed
      to check object authorities of security related commands -
      should not be public access

18   General: Have IBM object authorities been checked on
      security related commands, i.e. DSPAUTUSR - should not be
      public access

19   General: Have authorisation levels been checked on User
      Profiles (i.e. USER rights or SYSOPR rights, etc.)

20   General: Do all DP Staff have access to all live
      environments (any restrictions i.e. Payroll) ?

1   How secure is my system ?

2   Does my system support different levels of security ?

3   What happens if I forget my password ?

4   Can I restrict access to sensitive files ?

5   Can I 'force' users to change their passwords at regular
      intervals ?

6   Can I prevent users from leaving their workstations signed
      on ?

1 Can a user have the same password twice ?
2 Can passwords be found out on the system ?
3 Can PC Support be secure with automatic signon facility available ?
4 Can users be given access to commands but still be 'limited capability' ?
5 Can users' confidential mail be accessed by any other user ?
6 Can confidential documents be viewed by other users in the output queue ?
7 Can QSECOFR authorities be restricted by another user ?
8 Can a user "passthrough" to another system and gain greater authority than he has on the original system ?
9 Can 2 users on different systems with the same userid pass through without knowing the other user's password ?
10 Can a PC virus be spread via PC Support to other users ?
11 Can a PC virus destroy data on the AS400 via PC Support shared folders ?
12 Can files accessed on PCs via shared folders be locked to prevent updates from unauthorised users ?
13 Can users be prevented from signing on at specific screens ?
14 Can users be prevented from exiting Office Vision to the 'main menu' ?
15 Can users be automatically deleted if they do not sign on for X months ?
16 Is there a way to display all commands, etc. that a user has authority to ?
17 Is there an easy way to check the loss for breaches of authority ?
18 Could a user suppress the 'sign on' information to conceal invalid attempts at signing on ?
19 Can password file be decrypted so that it may be read by potential hackers ?
20 Are there any IBM supplied user ids which a hacker could use to access AS400s ?

1 How can system security be breached internally ?
2 How can system security be breached from an external source ?
3 What specific precautions can be taken to prevent (1) and (2) ?
4 What regular checks can be made to ensure security is not being breached ?
5 If security files are corrupted, how will I know ?
6 How can I find out how (5) and when (5) happened, and prevent same happening again ?
7 If history file indicates illegal sign-on by users, how far back does history log remain on the system ?
8 If security is breached, how can you find extent of damage (if any) caused ?
9 Should we restore backups on regular basis to ensure they are OK ?

1 Can you restrict access to particular menu options ?
2 Can any user read confidential documents ?
3 How do I change password ?
4 Is it possible to break passwords ?
5 When was data last backed up ?
6 Where are backups kept ?
7 What system of backup generations should be used ?
8 What problem to security if staff gained physical access
   to CPU ?
9 Can staff inadvertantly destroy data ?
10 How do I secure library, file or folder to specific users ?
11 How do I restrict user to specific menu only ?
12 What is difference in levels or access given by system by
   master, security officer, etc. ?

1 Am I able to increase/ decrease security access ?
2 What if I receive messages indicating outside
   unauthorised access ?
3 Making sure certain restrictions are made to access
   of confidential folders.

1 How do you set up a password ?
2 My password has expired. Can you reinstate ?
3 How do I change my password ?
4 Does anyone have access to all the passwords ?
5 Can folders be accessed by everyone ?
6 Can personal mail be accessed by other people ?
7 Can other companies access our network ?
8 Is it possible to check for viruses ?
9 What is the key on the screen used for ?
10 When I leave my desk should I sign off ?

1 How many security classifications are there ?
2 How do we go about setting up the security ?
3 Is it possible to get around security ?
4 Can we secure the system against virus attack ?
5 If so, how ?
6 How often should people change their password ?
7 What do we do if somebody forgets their password ?
8 Can we back up security while people are signed on ?
9 Can an individual change their own security level ?

1 What are the chances of being "hacked" and what
   practical steps need to be taken to prevent it happening ?
2 How do I convince colleagues of the need for security
   if solutions involve capital expenditure ?
3 How can one prevent a virus spreading from connected PC's
   to the AS/400 host ?
4 What should happen to the system when the office is
   un-manned ?
5 How do we make sure that everything that should be
   backed-up, is ?
6 What is the most efficient way of backing-up data and
   software ?

7   What should the back-up frequency be ?
8   For how long can back-up tapes be guaranteed good ?
9   Who should have "security officer" status ?
10 How comprehensive should a disaster recovery plan be ?
11 Is disaster simulation ever a practical way of testing a recovery plan ?
12 Should my system's security ever be audited by a third party ?
13 How frequently should passwords be changed ?
14 Should we set up an internal security audit function, eg. using our own computer auditor ?
15 Should staff who have tendered their resignations (3 month period) be removed as users on the system whilst still with us ?
16 Should contracts of employment have a data/software theft clause ?
17 How quickly will a replacement machine be supplied if the existing one suffers a major disaster - IBM lead-times are very long.
18 System saves, etc... need to be done after hours when access is denied to normal users. This is expensive (overtime) - better methods must be found.
19 How real is the security problem anyway ?
20 Do security measures contradict the principles of Total Quality Management ?

1   Which is the best virus protection software ?
2   Is my system as secure as possible ?
3   How do I shut down a UPS if needed ?
4   UPS batteries - are they still full of life without faults ?
5   Have we got long life batteries ?
6   If we expand IBM, will it necessitate UPS expansion, which could mean the replacement of all batteries ?
7   Will the UPS work when required ?
8   Computer Disaster Recovery - which are the best methods ?

1   What areas of the system need to be considered in the "system security" light ?
2   How is the system vulnerable ?
3   What precautions should be taken to preserve the integrity of the system/ data ?
4   What is the system's resident security rationale, and how do I operate it ?
5   How do you educate other users, so they're aware of the dangers ?
6   How do you encourage them to support existing security practice ?
7   Can the system itself help me to monitor its own health and safety ?
8   Can the system help me to monitor infringements and identify vulnerable areas ?
9   Where do I find info on new viruses ?
10 How do I check my present m/cs are virus free ?
11 Is there a check list I could periodically go through to ensure security and integrity ?

12 Can viruses travel through networks - if so, are there
    resident checks to monitor/ ensure the health of incoming
    data/ mail ?
13 Is there anything on the system an administrator can't
    find out ?
14 What are the potential problems which might arise when
    new software is put on to old/ existing data ?
15 Are there any checks in software/ data to prevent them
    being encouraged to "work off the premises" ? - Lotus
    I know about. How do these work ?

 1 I require a user to access parts of a data file via
    application software, but no access should be allowed
    via system utilities.
 2 I need to keep a log of all users that access the
    payroll file via any source.
 3 A user is allowed to enter only display commands.
 4 I have security officer authority, but I do not want to
    sign off my screen when I leave it un-attended.
 5 Specific users only should be "signed off" if a key
    depression is not made within 30 seconds.
 6 Remote system access must be restricted to office hours
    only.
 7 System termination can only be performed by operations
    staff.
 8 A user password must be greater than 8 digits and contain
    3 numbers. The system must control this.
 9 I want to know everyones password.
10 I want to limit certain commands to a physical screen
    location.

 1 There are no security arrangements in force. Where
    do I start ?
 2 How do I discover what products are on the market ?
 3 How do I evaluate these products without prior experience ?
 4 Are all functions within a product necessary ?
 5 How do I arrange training for over 100 staff ?
 6 How can I best persuade staff of the importance of
    security ?
 7 Why won't the Senior Personnel Partner amend employment
    contracts to include security as part of job specs ?
 8 Do they really know what a virus is - do they care ?
 9 How do I persuade them that the real and perceived
    delay when using security products during a PC boot
    is necessary ?
10 How do I make staff use non standard passwords ?
11 Should I create access authority levels without
    consultation, on the basis of staff grades ?
12 How do I stop them storing passwords in obvious places ?
13 Should I invoke the automatic password change facility
    on our midrange system ?
14 Why won't they tether or bolt PC's to the desks ?
15 Must I have access control on my PC - it's such a pain ?
16 Why must I always book the portables out ?
17 Why do you keep everything locked up ?

18 If you gave me greater access, I wouldn't have to bother you so often.
19 Do you really think 'FAST' will raid us ?
20 So what's wrong with a couple of illicit copies - no one will know ?

1 Can I change my own password ?
2 What does 'password expired' mean ?
3 Can more than one user have the same password ?
4 Can more than one user have the same user id ?
5 Why can't I see my password as I key it in ?
6 Can you tell me my password ? I have forgotten it.

1 Who should be authorized to QSECOFR ?
2 How do I ensure that users can access data in files only through the application software ?
3 How do I determine which programs use adopted authority (i.e.of the program's owner) ?
4 How do I ensure the security of off-line data ?
5 What does the 'QSECURITY' system value do ?

1 What are the different levels of user security ?
2 What system security level ensures a secure system without reducing useability ?
3 How can I prevent a user from updating an application's set of files unless they are using the application ?
4 Can I put a time-out facility for workstations that haven't been used for a certain amount of time ?
5 Can I give a user a command line and then specify which system commands that user has authority to ?

1 How often should backups be taken ?
2 Which files should be secured on backups ?
3 How far back should backups be kept ?
4 How often should passwords be changed ?
5 Should passwords conform to any standard ?
6 Who should have access to passwords ?
7 How easy would it be for someone to hack into the system ?
8 Do security levels for users get checked on a regular basis ?
9 Are hardcopy reports secured in a safe place when required ?

1 How often must passwords be changed on the system - has the system value been set up ?
2 What level of security has the machine been set to ?
3 Have any authorisation lists been set up - if not - how is object authority looked after ?
4 Who is allowed access to the QSECOFR password - how often is it changed ?
5 Are any remote terminals authorised to the QSECOFR password ?
6 If only one terminal was authorised to use QSECOFR, and that terminal was lost, how would you get over this problem ?

1   What levels of security will I expect to find and which
    ones am I authorised to ?
2   How can groups of security levels be set up ?
3   What authority does a particular security level have ?

1   If the QSECOFR user profile has been disabled is it
    possible to recover or generate *SECOFR authority
    some way ?
2   Software viruses are common on PC's, what about on
    the AS/400 ? Are there virus checkers available ?
3   Regarding remote security, could someone from an outside
    organisation setup a comms link to our system and signon ?
4   Data saved onto media eg. tape, can this be restored onto
    another system and manipulated ?
5   In the event of a machine crash, do you have a detailed
    disaster recovery plan, to cover all your users ?

1   How do I stop users passing-through to our remote sites ?
2   How do I prevent users from displaying other users reports
    while they are on the output queue ?
3   Can I prevent users from signing-on during particular
    times ?
4   How do I give access to only certain documents in a folder ?
5   What sort of backups should I be running for my day to day
    data ?
6   What are authorisation lists ?

1   What can I do if I forget my password ?
2   How secure is the AS400 ?
3   Can different users have different security parameters ?
4   Can specific AS400 functions be security protected ?
5   Is it possible to remove the security completely ?

1   How can I stop someone physically accessing the AS/400 ?
2   Can someone just walk into our environment - sit down &
    use our system ?
3   Why haven't I got access to all the options in the system ?
4   Why aren't I allowed to a command line on the AS/400 ?
5   Why am I classed as a restricted user ?
6   Can I stop someone viewing a specific file on a system ?
7   What happens if the AS/400 physically breaks down ? Can
    I return to the same point with the same level of
    security ?
8   How can I stop someone accessing my office documents ?
9   Why do we need password security ?

1   How can I access other areas of system ?
2   Can I "lock" my folder ?
3   Can I reduce the security level of my own password ?
4   What are the limits of what I can do ?
5   How much is there on the system I can't even see ?

1   How do I control who sees what on the system ?
2   How can I see who is trying to access confidential
    information ?

```
 3   How can I prevent someone looking at my document ?
 4   How can I define security for a group of users ?
 5   How can I override group security for a particular person
     to a particular file ?
 6   How do I stop unauthorised access to a specific
     application ?
 7   What is level 40 security ?
 8   Should I set my system to level 40 security ?
 9   Can people easily hack into my AS/400 from outside ?
10   How can I break the security on the AS/400 ?
11   What happens if the security officer forgets the password ?
12   What does the security key position do on the AS/400
     front panel ?
13   Can IBM break our security arrangements ?
14   Should I be using data encryption for communications
     traffic ?
15   Can my AS/400 catch a virus ?
16   How can I stop the Security Officer or anyone else from
     viewing my confidential work ?
17   Can I see if any external unauthorised users have tried
     to get into our AS/400 system ?
18   How can I easily back up all of my work on the system ?
19   What can the Security Officer not do ?
20   What lifespan do backup tapes have ?

 1   How do I get into this function ?
 2   I am not authorised to use the system.
 3   I want authorisation to this system function.
 4   The software has terminated due to unauthorised access.
 5   How do I enroll a user to use the system ?
 6   I want to grant a user authority to a group of objects.
     What do I do ?
 7   What is an authorisation list ?
 8   How do I build data base access security into my
     application ?
 9   What system functions exist to enable data integrity on
     the AS400's ?
10   What is a group profile ?
11   What are the advantages of group profiles over
     authorisation lists ?

 1   Is the computer environment secure ?
 2   Do we have physical security to computer environment ?
 3   Who owns the 'master'/'security officer' password ?
 4   Who has access to the security officer password ?
 5   Is the security officer limited to sign-on at certain
     terminals ?
 6   Can I restrict users only signing onto the system once ?
 7   Can I control the number of chars in their passwords ?
 8   Are their passwords visible to any other users or are
     they encrypted ?
 9   Can I force users to change their passwords every 'X'
     many days ?
10   How can I find if anyone has signed on as the security
     officer ?
```

11 How can I trace breaches of security ?
12 What do I do if someone forgets their password ?
13 How do I stop remote users having access to the system ?
14 How do I stop users having access to the system libraries ?
15 How do I stop users having access to a command line ?
16 Can I prevent users displaying spool files for other users ?
17 How regular should we be doing our system backups ?
18 Can AS/400 shared folders be affected by PC viruses ?
19 Can I duplicate the system key ?
20 What does this red switch do ?

1 What happens if the system goes down ?
2 How do I take a backup ?
3 Can anyone access my data ?
4 Can passwords be changed ?
5 If someone leaves how do I stop them accessing the system ?
6 How do I guard against unauthorised access via modem ?
7 How do I check for viruses ?

1 How can I test the systems security ?
2 How can I monitor access to secured files ?
3 How can I see what an individual has access to ?
4 How can I monitor the security officer ?
5 How can I prevent the security officer accessing
   confidential files (i.e. payroll) ?
6 How can I deny access to some foreign systems in the
   network ?
7 How can I be sure no one can access my documents that
   are confidential ?
8 What is the most practical way to protect data - by file
   or by library ?
9 Why is spool file security so complex ? Want to stop
   people viewing spool yet be able to control printers.

1 How often should I back up ?
2 How often should I change passwords ?
3 What should be backed up ?
4 What off site arrangements should be made ?
5 How do you plan for disaster recovery ?

End of Appendix II

237

Appendix III

Concordance for "password*"

53 questions

Can I change a user's password for them ?
Can I allow Office users to share a signon I.D. and
have separate passwords ?
What are the rules on password reuse ?
Can a user have the same password twice ?
Can passwords be found out on the system ?
Can 2 users on different systems with the same userid
pass through without knowing the other user's
password ?
Can password file be decrypted so that it may be read
by potential hackers ?
How do I change password ?
Is it possible to break passwords ?
What happens if I forget my password ?
Can I 'force' users to change their passwords at
regular intervals ?
How do you set up a password ?
My password has expired. Can you reinstate ?
How do I change my password ?
Does anyone have access to all the passwords ?
How do you change your password ?
What do you do if someone's password expires ?
Is it possible to find out other peoples' password ?
Can you use any parts of the system without first
having a password and signing on ?
How frequently should passwords be changed ?
How often should people change their password ?
What do we do if somebody forgets their password ?
How many levels of password are there ?
How often should passwords be changed ?
Do you have password security when communicating with
3rd Party (i.e. EDI) ?
Do you have password and user Id's at sign on ?
If passwords available, how often are they changed,
password reuse, lengths, etc. ?
Have IBM default passwords been reviewed on AS/400 and
have they been changed? i.e. QSECOFR
How do I make staff use non standard passwords ?
How do I stop them storing passwords in obvious
places ?
Should I invoke the automatic password change facility
on our midrange system ?
Can I change my own password ?
What does 'password expired' mean ?
Can more than one user have the same password ?

Why can't I see my password as I key it in ?
Can you tell me my password ? I have forgotten it.
How often should passwords be changed ?
Should passwords conform to any standard ?
Who should have access to passwords ?
How often must passwords be changed on the system - has
the system value been set up ?
Who is allowed access to the QSECOFR password - how
often is it changed ?
Are any remote terminals authorised to the QSECOFR
password ?
What can I do if I forget my password ?
Why do we need password security ?
Can I reduce the security level of my own password ?
What happens if the security officer forgets the
password ?
Who owns the 'master'/'security officer' password ?
Who has access to the security officer password ?
Can I control the number of chars in their passwords ?
Are their passwords visible to any other users or are
they encrypted ?
Can I force users to change their passwords every 'X'
many days ?
Can passwords be changed ?
How often should I change passwords ?

END OF APPENDIX III

APPENDIX IV

Examples of everyday formulations in users' queries.

(This list could be used as test data for an intelligent tutoring system on computer security or natural language retrieval aid.)

more freedom on the system
use the same network as us
to find out other peoples' passwords
most common ways of breaching
how resilient is the system
soeone trying to hack their way in
if I forget my password
at regular intervals
have the same password twice
can passwords be found out
viewed by other users
check the loss for breaches
potential hackers
what precautions can be taken
what regular checks can be made
how will I know
how can I find out
find extent of damage caused
to break passwords
where are they kept
go about setting up security
get around security
the chances of being hacked
practical steps taken to prevent
the office is un-manned
a practical way of testing
machine suffers a major disaster
normal users
how real is the problem
as secure as possible
UPS batteries - full of life
will work when required
recovery - best methods
considered in the "system security" light
how is the system vulnerable
security rationale
aware of the dangers
educate users
encourage users
can the system itself help me to
a check list I could go through
travel through networks
the health of incoming mail
an administrator can't find out

problems which might arise
new software is put onto old data
being encouraged to "work off the premises"
I require a user to
when I leave it un-attended
want to know everyone's password
arrange training for staff
persuade staff of the importance of security
security as part of job specs
make staff use passwords
passwords in obvious places
tether or bolt PC's to the desks
more than one user having the same user id
see password as I key it in
how often should backups be taken
how far back should backups be kept
reports secured in a safe place
a machine crash
access to only certain documents
day to day data
remove the security completely
can someone just walk in
return to the same point
other areas of system
the limits of what I can do
I can't even see
who sees what
confidential information
how can I see who is trying to access
prevent someone looking at
what does the security key position do
catch a virus
tried to get into our system
what can security officer not do
get into this function
the advantages of group profiles
visible to any other users
force users to change
how can I find if anyone
stop users having access
doing our backups
how regular
what does this switch do
see what an individual has access to
how can I be sure
most practical way to protect
stop people viewing
how often should I

END OF APPENDIX IV

241