DEREK AUSTIN

Head
Subject Systems Office
The British Library
London, England

The Role of Indexing
in Subject Retrieval

On first reading the list of speakers proposed for this institute, I became aware of being rather the "odd man out" for two reasons. Firstly, I was asked to present a paper on PRECIS—which is very much a verbal indexing system—at a conference dominated by contributions on classification schemes with a natural bias, as the centenary year approaches, toward the Dewey Decimal Classification (DDC). Secondly, I feared (quite wrongly, as it happens) that I might be at variance with one or two of my fellow speakers, who would possibly like to assure us, in an age when we can no longer ignore the computer, that traditional library schemes such as DDC and Library of Congress Classification (LCC) are capable of maintaining their original function of organizing collections of documents, and at the same time are also well suited to the retrieval of relevant citations from machine-held files. In this context, I am reminded of a review of a general collection of essays on classification schemes which appeared in the *Journal of Documentation* in 1972. Norman Roberts, reviewing the papers which dealt specifically with the well established schemes, deduced that "all the writers project their particular schemes into the future with an optimism that springs, perhaps, as much from a sense of emotional involvement as from concrete evidence."[1] Since I do not believe that these general schemes can play any significant part in the retrieval of items from mechanized files, it appeared that I had been cast in the role of devil's advocate.

By tradition, the role of devil's advocate (and we should remember that every conference needs one) has to be defended by logical argument. I would therefore like to begin by stating some of my grounds for believing that a library classification, as this term is usually understood, cannot function equally well for the dual purpose of organizing shelves on the one hand, and searching machine-held files on the other. This will then serve as a useful introduction to the topic on which I was primarily invited to speak: the role of the verbal subject index in document retrieval, using PRECIS as the example with which I am familiar.

STRUCTURE AND FUNCTION IN LIBRARY CLASSIFICATION

The review by Norman Roberts quoted earlier referred to a collection of essays edited by Arthur Maltby, entitled *Classification in the 1970's.*[2] A rather more direct opinion of this work was expressed by an astute American reviewer, Jean Perreault, who regarded these essays as clear evidence that "the two major purposes of documentary classification, namely for shelf organisation and for mechanised retrieval, are *not* well served by a single system unless consciously modified to cater for the two purposes."[3] Perreault does not suggest how this modification might be carried out, though I strongly suspect that any alteration of a scheme to enhance its performance in one of these roles would almost certainly render it less effective in the other. To demonstrate this point, we can consider the relationship between structure and function in a classification scheme, starting with its obvious function of imposing order upon collections of documents. For this purpose, we can stipulate certain desiderata, of which the most important are probably:

1. brevity of notation—this point was heavily stressed by librarians in a survey of classification needs carried out by the (British) Library Association in 1966;[4]
2. reasonable collocation, or the bringing together of like-with-like on the shelves, while bearing in mind the disconcerting fact that no library scheme, however well conceived, can ever bring together all the documents which a given reader would regard as belonging to his special field of interest;
3. hospitality and specificity—with the introduction of these two complementary characteristics we can already begin to detect an element of strain (i.e., How can any scheme offer these two characteristics, and still retain a short notation?);
4. standardization—becoming increasingly important as international data exchange networks continue to develop. The acceptance of a classification scheme as a general standard could eventually mean that the librarian in

Chicago has no need to reclassify any work which has been handled already by his counterpart in London or Paris. Provided that a decision made in the country of origin of the document accords with standard practice, it should be possible to adopt that decision as soon as it becomes available, either in the form of a magnetic tape record, or via a telecommunication link directly into the foreign data base.

I should now like to consider a different set of desiderata: those which apply to a mechanized file intended for tracing relevant documents in response to users' inquiries. In this context, we could stipulate two important characteristics:

1. currency and hospitality—that is, we need the ability to identify quickly works on newly emerging concepts, or on new subjects which consist of familiar concepts combined in unfamiliar and even unexpected ways. A good deal of the literature we handle on a day-to-day basis contains emergent knowledge which belongs to one or the other of these two categories;
2. we need to identify, in the most *economical* way (which in computer terms means as *quickly* as possible) all the works which may have dealt with a specific concept. For this purpose, a given concept should ideally be represented by just one symbol which can then be used as the key to its retrieval from any part of the file.

If we now attempt to compare these two sets of desiderata—that is, those for a shelf-order system, and those for a mechanized file—we can, perhaps, begin to see why these different needs cannot be satisfied entirely by a single system. Let us consider, for example, the librarian's justifiable need for a short notation, and contrast this with the need, in a mechanized file, to identify each separate element in a compound subject by some unique symbol which could serve as the key to its retrieval from any part of the system. An enumerative classification, such as DDC or LCC, obviously serves the librarian very well in terms of notational economy. A great deal of conceptual information can be packed into a fairly simple class mark such as 621.3, which represents *electrical engineering*. However, this number is not particularly helpful if we consider it as an aggregate of concepts from the viewpoint of machine retrieval. The symbol .3, attached to the stem number 621, means *electrical* in this case, but it does not follow (as it should, ideally, in a ׀mechanized system) that the mark .3 continues to express *electrical* throughout the rest of the schedules. In a different class context, such as 914, for example, an additional .3 denotes *Central Europe and Germany*.

In that case, a given symbol does not consistently represent the same concept. The converse is equally true; that is, a given concept is not

represented consistently by the same symbol. In fact, it takes approximately one hundred different symbols to represent the concept *electrical* or *electricity* in the schedules of the eighteenth edition of DDC. Such a wide range of numbers is partly due to the fact that this scheme is generally enumerative, but it does not follow that the problem has been solved by the makers of faceted classifications. For example, a relatively simple concept such as *iron* is expressed by at least six entirely different numbers in the abridged edition of the Universal Decimal Classification, and by several different symbols in the Colon Classification. This does not mean that these schemes have no role to play in library organization, but it does cast at least some doubts on their effectiveness as tools for mechanized searching. I have tried, in a different paper, to set out the case for regarding these faceted schemes as less than satisfactory for present-day purposes on the grounds that, in trying to satisfy both the librarian and the data base manager, they may have attempted the impossible and succeeded in neither.[5]

THE CRG RESEARCH INTO A
NEW GENERAL CLASSIFICATION

I might point out that this opinion represents more than a theoretical viewpoint. It is also based on some personal experience in trying to devise a scheme which could function equally well for both library arrangement and mechanized retrieval. An opportunity to explore this ground arose in connection with the NATO-funded research into a new general classification scheme which was carried out by the Classification Research Group (CRG) in London during the 1960s. Partially for the reasons I have outlined, the CRG decided that an entirely new approach to classification was needed—one which, it was hoped, would lead to a scheme which could function equally well for both library arrangement and the identification of works on specific concepts.

It was assumed from the beginning of this research that any new scheme should be founded upon the basic postulates for an analytico-synthetic classification established by Ranganathan. These postulates are themselves based on two assumptions which together constitute the keystone to modern classification theory: (1) any compound subject is amenable to analysis into discrete conceptual elements, each of which (at least in theory) could be identified by its own unique symbol; and (2) the compound subject, regarded as a whole, could then be reconstituted out of these parts in accordance with a general formula, and the formula itself could be based upon a single set of logical principles which would apply across the whole spectrum of knowledge.

These postulates are all very well in theory, but what about the practice? At the time when the CRG research began, no one had actually attempted to take these ideas to their logical conclusion and construct an entirely analytico-synthetic classification. Even Ranganathan's Colon Classification is firmly based on a set of main classes, and the notation which represents a given concept can vary from one main class to another. Furthermore, Ranganathan's formula for number building, based on the general categories of Personality, Matter, Energy, Space and Time, is not so generalized at it first appears. In particular, the primary facet, Personality, has caused problems for both teachers and practitioners. Since this is the factor which has to be cited first when building a compound subject from its parts, it is therefore the factor which determines where documents on that subject will be shelved.

In practice, however, it has been found that the interpretation of Personality depends upon the main class structure of the scheme in use; even in the Colon Classification system, this can vary from one class to another. Unfortunately, a good deal of modern literature, even at the monograph level, severely strains the concept of main classes. When faced by a subject such as "the use of computers to handle the payroll of teaching staff in American universities," the interpretation of Personality will certainly vary with the frames of reference of the user (as well as the librarian) depending on whether the user is computer-oriented, is an accountant, a personnel manager, or a university administrator.

These were the kinds of challenge, which appear to be endemic in both enumerative and faceted classifications, which stimulated the CRG research. The solutions we explored can best be considered as simultaneous attacks on two different but related fronts. The first might be called the semantic approach, and was concerned with the organization of concepts (individual units of information) into basic categories to which they appeared to belong in a definitional sense, without taking any account of the ways in which these concepts might occur in different compound subjects, in the sense in which *iron*, for example, belongs to a category called *metals,* and *beauty* to a general class of *human subjective judgments*. Once a concept had been assigned to its general class, it would then have been identified by a single notational symbol which would have served two purposes: (1) to label that concept in a once-and-for-all fashion, so that the symbol could be used for locating documents on that concept from any part of a data base; and (2) it would show, through its hierarchically expressive structure, the general class of ideas to which the concept belonged. The other approach we considered is more closely related to syntax than semantics. This was a search for what might be called a set of generalized rules which would constitute a classificatory "grammar," insofar as

they would determine the order in which concepts should be set down when building any compound subject out of its parts.

The first of these tasks—the assignment of concepts to general categories—obviously called for an explicit act of classification, although not in the library sense. We were here concerned with imposing order on a universe of concepts, not on a universe of subjects. For this reason, I would prefer to use the term *categorization* to describe what we attempted, leaving the term *classification* to be used in its familiar or library sense. In terms of methodology, our general approach to this universe of concepts was not radically different from that employed by the maker of a library classification. Each of these tasks calls for a basically similar technique. Certain principles of division have to be established, and these must then be introduced one at a time, each principle being exhausted before a new one is introduced. We first divided concepts into two basic kinds, those which indicate Things, and those which are the Attributes of things. Each of these classes was then further subdivided. The general category of Things, for example, was separated into two new classes, called Naturally Occurring Entities and Artificial Entities; the latter category was again divided into concrete artifacts (such as *chairs* and *aircraft*), and mental constructs (such as *systems of belief* and *theoretical models*). A similar operation was also carried out for the general cateogry of Attributes. If there had been time to complete this work, the final product would have been what might be called a macrothesaurus dealing with the basic concepts, as they occur in modern literature, which form the quanta from which all compound subjects in any field can be constructed.

I should, perhaps, stress that this is not an entirely new approach to the organization of knowledge. Thesauri, as such, have a long and respectable history, with Roget's serving as the obvious model for the kind of macrothesaurus we are now considering. It is also worth noting that several library classification systems, with DDC as the classic example, already operate in this way to some extent. Apart from the fact that compound subjects can be built by using the *add* instruction, certain classes of general concepts, especially those which are likely to be needed at any point in the schedules, have already been assigned to general categories. These form the auxiliary schedules which now occupy a separate volume in the current edition of DDC, and from which the classifier extracts, as he needs them, commonly occurring factors such as bibliographic forms, places, methodologies, and so on. The approach considered by the CRG would simply have taken this idea to its logical conclusion—that is, *all* the concepts in the schedules would, in effect, have been assigned to the auxiliary schedules, then notated on a permanent basis, ready for use in number-building whenever the appropriate literature appeared.

The second problem faced by the CRG was that of devising a general formula, based on teachable and logical principles, for building compound subjects out of their parts. We had, of course, started with the classical PMEST formula of Ranganathan, but this was found to be inadequate in some respects. We therefore extended this model in various ways. Following the work of Vickery,[6] we defined the parts of a subject more explicitly in terms of their grammatical roles or functions. For examples, Wholes were distinguished from Parts, and it was stipulated that the whole must always be cited before the part. We also identified specific elements of subjects, such as the product of an action, the object upon which the action was performed, the action itself, and its agent. In order to achieve a reasonable level of consistency among classifiers, each of these roles was identified by a numerical code (called an operator) which was given a built-in filing value. When building a compound subject out of its parts, each separate piece of notation representing a specific concept would have been prefixed by an appropriate operator, and the filing value attached to the operator would have ensured, for example, that the whole was consistently set down before the part, that the object or recipient of the action was written before the action itself, and the action before the agent. In effect, we were searching for a generalized grammar of classification—one which could be used as a mental model for regulating the order of concepts in any compound subject. In devising this model, we had deliberately disregarded traditional disciplines as these are usually understood. Nevertheless, the order of concepts had still been selected with a view to providing some kind of helpful collocation in a pan-disciplinary library or bibliography.

It would be foolish to claim that anything resembling a new general classification arose from these efforts. Nevertheless, at the end of the research (when we had used up the £5,000 awarded to the CRG by the NATO Science Foundation), I think we had at least demonstrated the feasibility of the approach we had been exploring, both toward the construction of a general thesaurus, and toward the establishment of a generalized grammar for subject building. Near the end of the project, a provisional notation was applied to the outline categories of concepts which had been developed, and the number-building techniques were applied to a sample collection of research reports. The results were then studied by the members of the CRG, who considered them from various viewpoints. From the viewpoint of collocation, the results were surprisingly acceptable. Obviously, the general formula we had developed did not produce groupings of the kind which are usually associated with the traditional disciplines found in schemes such as DDC. Nevertheless, we appeared to have achieved helpful groupings, especially in those emergent fields which tend to cut across the older disciplines. However, there was still one factor we could not ignore. Although this system might have proved well

suited to the searching of machine-held files, the resulting class numbers were completely unsuitable for library purposes, simply because they were far too long and complicated. As Jack Mills pointed out when reviewing this work: "the code system used . . . conveys the structure of the system succinctly for machine manipulation . . . although it is obvious that the system does constitute a general 'library classification' in the accepted sense."[7]

THE DEVELOPMENT OF PRECIS

Fortunately, this was not the end of the story. During the CRG search for a number-building formula, concepts had been organized in various ways to test a range of hypotheses. In a number of cases, we had actually used words rather than notational symbols to represent specific concepts, partly because a word conveys a more obvious and immediate message than a symbol, and also because in many cases we were dealing with concepts which had not yet been admitted into the thesaurus. These experiments in term manipulation became more than a matter of expediency, and assumed the status of a new research project when the decision was made, in 1969, to produce the *British National Bibliography* (BNB) directly from our own MARC tapes.

From its first issues, BNB has appeared as a classified bibliography; that is to say, full catalog entries for all British monograph output have been printed under their DDC class numbers in the "front end" of the bibliography, and this systematic arrangement has been supported by one or more separate indexes giving access under the names of authors, titles, subjects, etc. It is necessary, at this point, to stress a lesson of MARC which has still not been fully appreciated by many librarians: MARC, if applied correctly, should mean the end of the concept of the main entry. Provided that all the essential components of a full catalog entry have been assigned to their correct fields in a record, so that each is uniquely identified, the librarian can, through a simple instruction to the computer, ask for these data to be organized in any way which satisfies his requirements. This kind of provision is endemic in MARC itself; nevertheless, it was some time before we fully realized the potential of the system. When BNB first became involved with MARC, our exchange tapes were made as an extra operation, and the national bibliography itself was still being produced by traditional means. It then became clear that many of the sequences found in BNB could, in fact, be extracted directly from the MARC records. For example, it was a simple matter to print full catalog entries under DDC class marks as the front end to the bibliography, since all the necessary data are uniquely tagged in MARC records. Some of the supplementary indexes could also be produced in this

way, especially those giving the names of authors, titles, etc. However, when BNB first became involved with MARC, no satisfactory means existed for producing a subject index directly from these records. The decision was therefore made in 1969 to set up a special research project to study the machine production of a subject index, as a necessary preparation for the fully automated production of BNB.

Quite naturally, we tried first to automate the production of the chain index which had been a familiar feature of BNB for some twenty years, but for various reasons this proved to be abortive. We also studied a range of alternative indexing techniques which had already been designed for use with computers, but none seemed entirely capable of producing an index to the standards we felt were necessary in a national bibliography. We therefore made what was probably a courageous decision, and set out to explore some new approach to the production of a subject index directly from machine-readable data. This is the research which led to PRECIS.

Certain desiderata for this new index were established as guidelines at the start of the project, and others were added as the work progressed. The principal characteristics for the index can be summarized under five main headings:

1. The computer, not the indexer, should produce all of the index entries, so that a large part of the clerical drudgery of index-making would be handled by the machine. The indexer's task would be limited to preparing an input string of the terms which are the components of index entries, together with instruction codes which indicate to the computer how these terms should be organized into entries; all the entries themselves would be constructed by the machine.

2. Each of the entries constructed in this way should be equally coextensive. In other words, each entry should express in a summary form the full subject of the document as perceived by the indexer. This should be seen in contrast to the chain index, where only the final entry is actually coextensive with the subject of the work in question, and also to a system of subject headings, where a compound subject may have to be expressed by two or more different headings, none of which by itself expresses the whole of the subject.

3. The system should be based on a single set of logical relationships among concepts; these should apply to subjects across the whole spectrum of knowledge. This would mean that terms in input strings, and in the entries produced from these strings, should be organized according to a kind of indexing grammar which would remain valid in fields as diverse as physics and metaphysics. Obviously, this grammar would not necessarily reflect the order in which concepts are introduced into the schedules of any one

classification scheme. However, this notion of classificatory neutrality was also regarded as important in a system carrying a range of different class marks such as MARC, since the same alphabetical index could then be applied to classified sequences organized by any of the schemes in the data base.

4. Index entries should be meaningful according to what might be called the normal frames of reference of the user. In other words, they should not be based on a librarian's conception of grammar, which accepts an inverted heading, such as *bridges, concrete,* as though it were everyday English. Instead, we should try to come closer to natural language so that the uninitiated reader can use the index with a minimum of instruction.

5. To complement the entries produced from input strings, the system should also be provided with means for constructing references among terms such as synonyms and higher generics, which are semantically related to index entry terms. These *see* and *see also* references would be extracted by quoting a suitable code from a machine-held thesaurus.

On the face of it, this may appear to be a complex set of criteria; considered on a very elementary level, however, it can be seen that we were actually concerned with only two different kinds of relationships among concepts. Furthermore, both of these had already been studied during the CRG research into a general classification. The earlier work had dealt with the search for a general formula for regulating the order of concepts in a compound class number; we were now concerned with a general model for regulating the order of terms in input strings and index entries. We might call this the search for a generalized syntax for an indexing language. Also, during the CRG research we had studied the ways in which concepts might be organized into categories within a macrothesaurus; we now had the task of creating a machine-held thesaurus of this kind to serve as the source of *see* and *see also* references in a printed index. This could be termed the semantic approach to an indexing language.

Examples of output from each of these sides of the index system can be seen in the extract from a typical PRECIS index which appears in figure 1. At the top of column 3, the user is redirected by a *see* reference from the term *pelecypoda* to its preferred synonym *bivalves*. This is one kind of semantic relationship. A different kind can be seen at the top of the middle column, where the term *particles* is linked, through a *see also* reference, to the names of various species such as *alpha particles*, *atoms*, and so on. The same term, *particles,* also appears farther down in the middle column, but this time it functions as part of an index entry and is syntactically related to terms such as *beams* and *scattering*. In this particular index, produced by the British Universities Film Council, the user is then referred, through a UDC number,

Figure 1. Extract of PRECIS Index

to a classified sequence where full details of the appropriate films will be found. However, I should point out that this is just one of several options available in PRECIS. In some cases the entry may refer to a separate file of subject headings which have been derived mechanically from the PRECIS input strings. One or two organizations also use PRECIS as a one-stage index—that is, they print the relevant citation directly below each index entry.

I cannot attempt to describe in detail all the stages in the production of a PRECIS index—that would require a series of papers. However, I would like to deal at least briefly with the basic mechanics of the system, partly to demonstrate the extent to which we met the basic requirements for a printed index considered earlier, and also to show how this indexing system relates to a general classification. I shall deal separately with the two aspects of the system, and consider first the syntactical relations between terms in index entries, then briefly touch on the semantics and the making of a machine-held thesaurus. Syntax itself can be considered from two different viewpoints: (1) the format and structure of index entries, and (2) the organization of terms into the strings from which the entries are produced.

THE PRECIS ENTRY FORMAT

When we set out to establish a suitable format for PRECIS, we found that we had to depart, in some respects, from the concept of a single-line entry which is typically found in systems such as the chain index, KWIC, and subject headings. The reasons for this can be illustrated by referring to the string of terms, and some hypothetical entries, which are shown in figure 2. The string:

FRANCE · TEXTILE INDUSTRIES · SKILLED PERSONNEL · TRAINING

represents a typical PRECIS input, and justifies some explanation. In the first place, these terms have been organized deliberately so that they form what we call a context-dependent sequence. This means that each term in the string sets the next term into its obviously wider context, in the sense in which *France,* for example, establishes the environment in which the *textile industries,* and therefore the rest of the subject, were considered by the author. The next term, *textile industries*, identifies the context in which *skilled personnel* were considered, and this new term establishes the class of persons to whom the act of *training* was being applied. It is worth pointing out that no attempt has been made to organize these terms in such a way that their order reflects their relative importance as shelving factors; we are principally concerned with expressing the meaning of the subject, and we leave the job of indicating shelf position to the classification scheme.

(1) FRANCE · TEXTILE INDUSTRIES · SKILLED PERSONNEL · TRAINING

(2) TRAINING · SKILLED PERSONNEL · TEXTILE INDUSTRIES · FRANCE

(3) SKILLED PERSONNEL / FRANCE · TEXTILE INDUSTRIES · TRAINING

(4) (A) — (B) — (C) — (D)

(5) [A. B] ← (C) → [D]

(6) (C) → [B. A]
          ↘ [D]

(7)
```
           Heading
        ⎯⎯⎯⎯⎯⎯⎯
[LEAD]  [  QUALIFIER  ]
   [   DISPLAY   ]
```
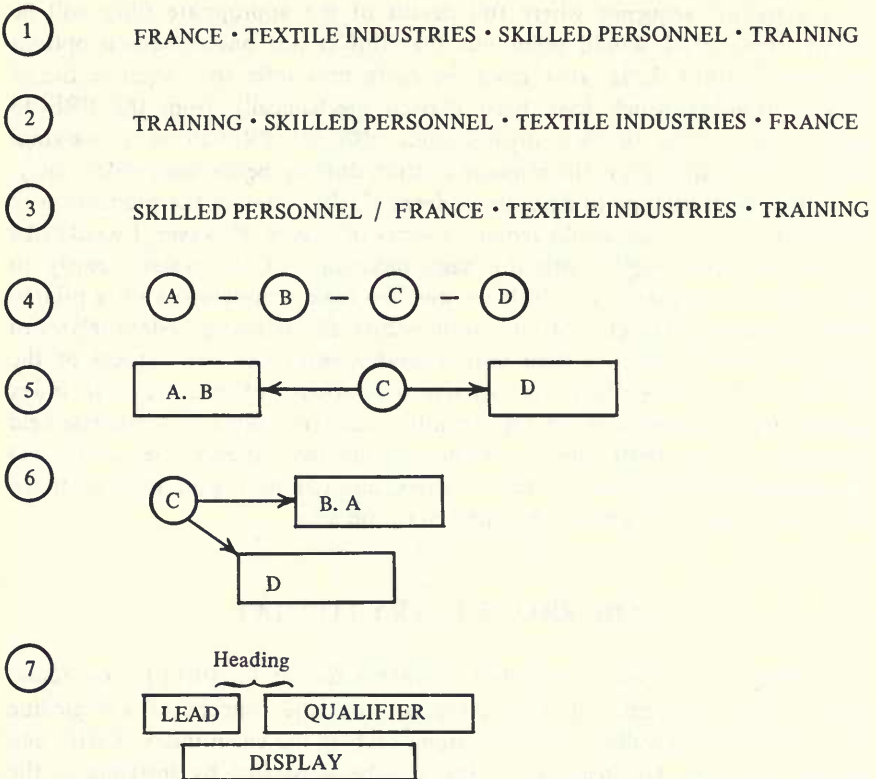
Figure 2. String of Terms and Hypothetical Entries

As a natural consequence of arranging these terms in a context-dependent order, they also form what we call a one-to-one related sequence. This means that each term is directly related to the next term in the string. Both context dependency and one-to-one relations occur in natural language itself, of course, and it may be worth mentioning that the order used in PRECIS was derived from a study of sentence structures. We regard these one-to-one relations as particularly important in conveying the meaning of a subject statement; indeed, in the present example, these relationships are so strong that the meaning of the original string remains unchanged even when the order of terms is reversed, as:

TRAINING · SKILLED PERSONNEL · TEXTILE INDUSTRIES · FRANCE

It therefore follows that either of these strings could function as an index entry which satisfies most of the criteria considered earlier: they are equally

coextensive, and they both convey the same message according to common frames of reference. Both could also be derived, by a very simple algorithm, from a single input string.

However, we start to encounter problems when we consider the production of an entry under one of the middle terms, such as *skilled personnel*. It would be a simple matter to instruct the computer to lift this term out of its place in the string and print it at the start of the entry, as shown at position 3 of figure 2. An element of ambiguity, however, has then been created: when reading this entry, we can no longer tell with certainty how the *skilled personnel* are related to the rest of the terms. Are they being trained, or are they employed in training others? It is not a difficult matter to deduce how this ambiguity arose. When this term was shifted from its original position, the mind automatically closed up the space that was left, and created a new set of one-to-one relationships. In a situation such as this, the problem can be expressed as a question: How can we maintain the original one-to-one relationships in an index entry without distorting the meaning, and without losing any of the terms in the process?

The approach we adopted is shown in the form of a diagram at position 4:

$$A - B - C - D$$

These four letters represent a sequence of four terms organized as a context-dependent and one-to-one related sequence. As we saw, the problem arose when we tried to make an index entry under one of the middle terms, such as C. As shown at position 5, this is due to the fact the term C is related simultaneously to the terms on either side; that is, B (which sets C into its wider context), and D (which is itself context-dependent on C). In order to make these relationships explicit on the printed page, we devised the two-line and three-position format which is shown at position 6. In this case, the term C functions as the user's access point to the index, and this is followed on the same line by those terms which set the lead into its wider contexts. The final term, D, is indented below on a second line, but remains explicitly related to the entry term C.

The layout of terms seen at position 6 shows an obvious two-line and three-position structure, which has now become a typical feature of a PRECIS index. These parts have been separately named, as shown at position 7. The *lead* is the term which functions as the user's access point, and this is automatically printed in roman bold to give it emphasis. The *qualifier* follows on the same line, and contains those terms which set the lead into its context, while the *display* holds the terms which are context-dependent on the lead. Terms in the qualifier and display may be printed in ordinary roman or italic, depending on how they are coded in the input string. The lead and qualifier together constitute what is called the *heading*. If two or more different strings give rise to the same heading,

France — Textile industries — Skilled personnel — Training

FRANCE
   Textile industries. Skilled personnel. Training

TEXTILE INDUSTRIES. France
   Skilled personnel. Training

SKILLED PERSONNEL. Textile industries. France
   Training

TRAINING. Skilled personnel. Textile industries. France
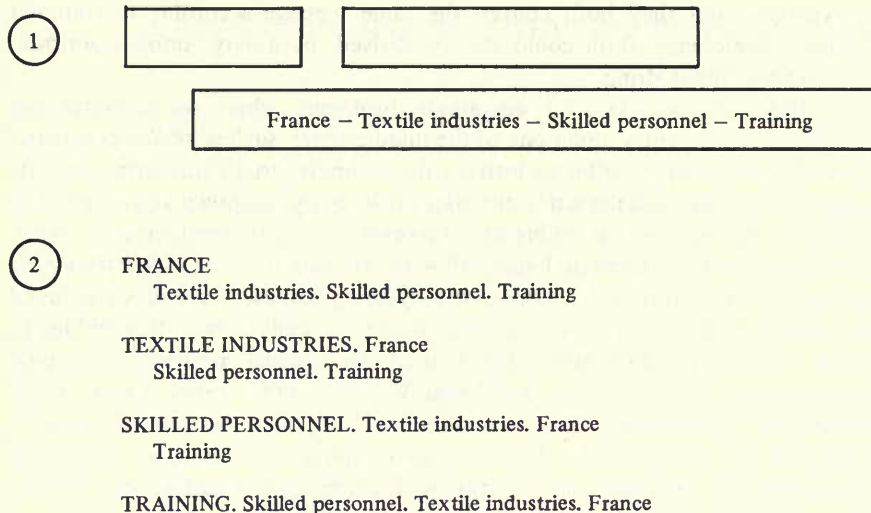
Figure 3. Terms Organized in Standard Format

so that only the displays are different, the computer automatically cancels the
second and subsequent headings, and organizes the displays as an alphabetical
column (see figure 1), where the term *particles* (near the top of the middle
column) has two displays, one starting with the term *beams,* and the other with
the term *counting*.

   The adoption of this format did more than resolve the problem of
maintaining the one-to-one relationships between terms. It also gave us the basis
for a fairly simple program which would allow us to generate mechanically a full
set of entries out of a single input. This is performed by an operation known as
shunting; the choice of this name may become clear if we consider the example
of index entry construction which is shown in figure 3. At the top of this page
we can see the three basic positions in a PRECIS entry, with a set of terms
marshaled in the display position. The lead has not yet been occupied, and no
entry has yet been produced. At the first stage of the operation, the first term in
the display is shunted into the lead, and the remaining terms are then shifted left
to a standard indentation position; this gives us the first entry, under the term
France, which appears in position 2. In generating the next entry, the term
France is shunted across into the qualifier, and its place in the lead is taken by
the next term, *textile industries*; again, the rest of the terms in the display shift
across to the standard indentation level. This operation could be repeated twice

more, to give us the entries under the terms *skilled personnel* and *training.* In this particular case, all of the terms in the input string passed through the lead position to give a total of four entries, but it should be pointed out that the generation of entries is not left entirely to the computer. It is, in fact, always under the control of the indexer, who can stipulate, in the form of instruction codes written as prefixes to each of the terms in the string, which terms should appear in the lead, or in any other part of the entry.

The entries shown in figure 3 represent terms organized in what is called the *standard format.* This format, which is produced by a straightforward application of shunting techniques, accounts for most of the entries in any PRECIS index. Two other formats are also available, but I shall not attempt to describe these here—they are fully explained in some of the technical accounts of the system.[8]

THE TREATMENT OF COMPOUND TERMS

The procedures shown in figure 3 demonstrate the treatment of a typical compound subject—that is, a subject which consists of a string of several terms. We also found that a basically similar technique could be applied to a compound term—that is, a term which has to be expressed in more than one word. The treatment of a term of this type *(fibre reinforced plastics)* is shown in figure 4.

In order to explain this procedure, it is first necessary to make a distinction between the different components of this term. In particular, we need to distinguish clearly between the noun, which is called the *focus,* and the adjectives, which are called *differences.* The focus (the word *plastics* in the present example) identifies the general class of concepts to which the term as a whole belongs. The term *difference* is used in its strictly logical sense to indicate some characteristic which specifies a subclass of the focal concept. In the present example, we have two differences, *fibre* and *reinforced,* each of which has its own logical function. The word *reinforced* functions as what we call a direct difference—it qualifies the focus, and defines a special subclass of the universe of *plastics* called *reinforced plastics.* The word *fibre,* however, has a rather different function, since it does not directly qualify the focus (that is, these are *not* fibre plastics), but instead qualifies the adjective *reinforced* in terms of the material used as reinforcement. It therefore functions as what we call an "indirect difference." This distiction is shown in the diagram at position 2 in figure 4.

Since these logical functions affect the correct form of term in an index entry, they must be indicated clearly to the computer. This is expressed by
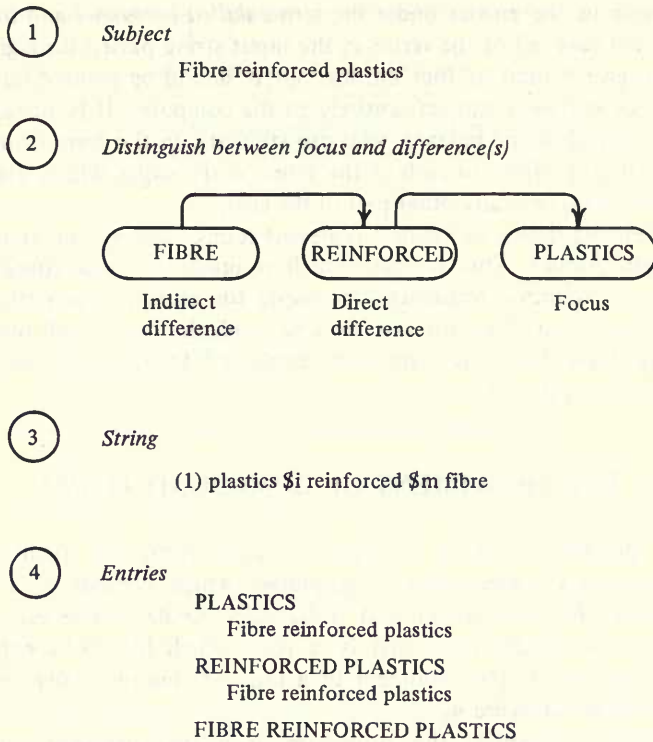
Figure 4. Treatment of Compound Terms

codes which are written as prefixes to each part of a compound term in the input string. An example of an input is shown at position 3 in figure 4:

(1) plástics $i reinforced $m fibre

where the focus, *plastics,* is prefixed by one of the role operators which we shall be considering later, while its status as a lead is indicated by a check. The code $i, which precedes the word *reinforced,* conveys two instructions to the computer: (1) it indicates that this is a direct difference, and (2) it shows that this word should appear in the lead. The code $m, which precedes *fibre,* also indicates a lead, but specifies that this word functions as an indirect difference. The output from this string is shown at position 4, and it can be seen that each of these entries is fully coextensive with the original concept. If the term in the lead is incomplete, as it is in the first and second entries, the whole term is printed, in natural-language order, in the display position.

Using these procedures, it is not possible to produce an inverted heading such as *plastics, reinforced.*

THE SCHEMA OF ROLE OPERATORS

So far we have considered the basic mechanics of entry construction, but we have not yet faced the problem of trying to ensure that a team of indexers will consistently achieve the same order of terms in their input strings. As I mentioned earlier, terms in an input string have to be arranged so that they form a context-dependent sequence. However, this is no more than the statement of a guiding principle. We need something more definite if we are to ensure that a team of different and quite human indexers (including the same indexer on different occasions) will consistently achieve this order.

To this end, the indexers work within the constraints of a kind of grammar. This is represented in the schema of role operators shown in figure 5. In many respects, this schema possesses some of the functions of the system developed during the CRG research. One of these operators has to be written as a prefix to each of the terms in an input string, and the operators then have two functions: (1) the principal codes (that is, the numbered or main-line operators seen at the top of the list) have built-in filing values, and it is these which determine the overall pattern of terms in a string; and (2) the codes act as computer instructions, and determine not only the format of the index entries, but also the typography of each term and its associated punctuation.

It would be quite impossible in the time available to describe in detail the workings of a scheme such as this, which is capable of dealing with compound subjects at any level of complexity. At least I can try to demonstrate how the system operates in practice, using the role operators to carry out an analysis of the subject we considered earlier: the training of skilled personnel in the French textile industries. This analysis is shown in figure 6.

During their initial training, indexers are taught to carry out their analyses in a step-by-step fashion, and are advised first to test each subject for the presence of an action. If an action concept is present, it usually determines how the rest of the subject should be handled, in much the same way that the verb tends to dominate the sentence in traditional grammar. In the present example, it is clear that an action is present in the term *training.* This term could therefore be written first, and prefixed by the operator "2," which represents an action or its effects, as shown at position 2:

Main line operators		
Environment of observed system	0	Location
Observed system (Core operators)	1	Key system: *object of transitive action; agent of intransitive action*
	2	Action/Effect
	3	Agent of transitive action; Aspects; Factors

A ————————————————————————

Data relating to observer	4	Viewpoint-as-form
Selected instance	5	Sample population/Study region
Presentation of data	6	Target/Form

Interposed operators		
Dependent elements	p	Part/Property
	q	Member of quasi-generic group
	r	Aggregate
Concept interlinks	s	Role definer
	t	Author attributed association
Coordinate concepts	g	Coordinate concept

B ————————————————————————

Differencing operators		
(prefixed by $)	h	Non-lead direct difference
	i	Lead direct difference
	j	Salient difference
	k	Non-lead indirect difference
	m	Lead indirect difference
	n	Non-lead parenthetical difference
	o	Lead parenthetical difference
	d	Date as a difference

Connectives		
(Components of linking phrases; prefixed by $)	v	Downward reading component
	w	Upward reading component

C ————————————————————————

Theme interlinks		
	x	First element in coordinate theme
	y	Subsequent element in coordinate theme
	z	Element of common theme

Figure 5. PRECIS—Schema of Role Operators

(1) Subject: Training of skilled personnel in the French textile industries

(2) Check for the presence of an action. Write the appropriate operator
        (2) training

(3) If the action is transitive, and the object is present, code the object as 'key system'
        (1) skilled personnel
        (2) training

(4) If the key system is part of a whole, code the whole as key system; use 'p' to identify the part
        (1) textile industries
        (p) skilled personnel
        (2) training

(5) Establish the environment
        (0) France
        (1) textile industries
        (p) skilled personnel
        (2) training

(6) Entries in 'standard format' (assuming a lead on each term)
        FRANCE
           Textile industries. Skilled personnel. Training

        TEXTILE INDUSTRIES. France
           Skilled personnel. Training

        SKILLED PERSONNEL. Textile industries. France
           Training

        TRAINING. Skilled personnel. Textile industries. France

Figure 6. Analysis of a Compound Subject

(2) training

The indexer next determines whether the action is transitive or intransitive; if (as in this example) the action is transitive, he establishes whether the object is also present. In this case, the act of *training* is being applied to the *skilled personnel*, who therefore represent the object. This concept is frequently coded as the key system, as shown at position 3:

(1) skilled personnel
(2) training.

However, this particular example contains a circumstance which causes a change to be made in this coding. In fact, the *skilled personnel* are part of another system (the *textile industries*), and this whole/part relationship has to be expressed by noting an operator "p" (which introduces a part or property indicator) in front of the name of the part. The numbered operator, "1," is then assigned to the name of the whole, which gives us the revised string seen in position 4:

(1) textile industries
(p) skilled personnel
(2) training

One more concept remains to be coded. This is the term *France,* which establishes the environment in which all the rest of the subject was considered, and therefore should be introduced by the operator "0." The final version of the string is shown at position 5:

(0) France
(1) textile industries
(p) skilled personnel
(2) training

It can now be seen that we have achieved in a fairly mechanical way exactly the same order of terms as in the earlier analysis, when this subject was considered only from the viewpoints of context-dependency and one-to-one relationships. It is only fair to point out that an experienced indexer would *not* go through the stages of this step-by-step analysis. Instead, he would write the correct string in an almost intuitive fashion, without necessarily being aware of the mental processes involved. The entries produced from this input are shown at position 6.

THE SEMANTIC ASPECTS OF PRECIS

I shall turn now to a different side of the system—that is, to the construction of the thesaurus which serves as the source of *see* and *see also* references in the printed index. This introduces a different set of routines and relationships, many of which involve classification in its taxonomic (rather than its library) sense; I will deal with these only briefly. I should start by pointing out that PRECIS works with an open-ended vocabulary; that is, a new term can be admitted into the system at any time, as soon as it has been encountered in a document. Any term marked as a lead in an input string is assigned—as part of a separate operation—to a position in a random-access file.

This position is indicated to the computer by writing a special number as part of each thesaurus input record. This number (the Reference Indicator Number, or RIN) identifies the address where the term will be stored. This address will later be written in a special field on an indexing form, where the presence of this number acts as a machine instruction, and leads automatically to the production of a full set of *see* and *see also* references directing the user toward the term which actually occurred in an index entry.

In figure 7 we can see a batch of thesaurus input records which have been prepared by an indexer, and are ready to be keyboarded and put into the computer. These are the cards which might be written if the indexer encountered the term *penguins* in a string for the first time, assuming that none of the other associated terms had previously been admitted into the system. The input record for *penguins* appears at the bottom of the sequence, and it can be seen that a number of data are recorded on this card. These include special codes, such as $m and $o, which indicate that the term *penguins* is related, in clearly specified ways, to terms which are held at other random file addresses in the computer, such as its synonym, *sphenisciformes*, and the higher generic term *birds*. The codes used for this purpose actually record, in machine-readable form, a range of semantic relationships which has now been established by the International Standards Organisation, and is recorded in a new international standard, IS 2788. These codes and their associated relationships are shown in figure 8; we can see at position 2 that they also determine the kind of reference which should be printed. The code $m produces a *see* reference from a nonpreferred term to its preferred synonym, while $n and $o both generate *see also* references.

Not all the data on these input records have to be keyboarded. It would be pointless, for example, to keyboard the terms from which references have to be made: it is enough to indicate to the computer the addresses at which these terms have been stored. If we turn to the bottom of figure 7, we can see the relatively small amount of data which is actually assigned to the computer when the card containing the term *penguins* is being processed. Once keyboarding is finished, the input cards are returned to the indexer, who then stores them alphabetically to form a term authority file.

Perhaps it is difficult at first to see any coherent pattern in the set of input records shown in figure 7. Nevertheless, this set of cards contains all of the semantic information needed to record within the computer the network of logically related terms shown in figure 9. Once a network of this kind has been established, it can be used in various ways. For example, the address of any term in the network can be quoted as part of an indexing record as soon as an appropriate document is encountered, and the necessary *see* and *see also* references will then be produced automatically. If we had set up this network

SCIENCE
024 430 9

BIOLOGY   SCIENCE
024 212 8 $o 024 430 9

ZOOLOGY   BIOLOGY
024 201 2 $o 024 212 8

ORNITHOLOGY   ZOOLOGY
024 519 4 $o 024 201 2

LIVING SYSTEMS   BIOLOGY
024 018 4 $n 024 212 8

FAUNA
024 571 2

ANIMALS   FAUNA   LIVING SYSTEMS
024 317 5 $m 024 571 2 $o 024 018 4

                 ZOOLOGY
                 $n 024 201 2

VERTEBRATES   ANIMALS
024 526 7 $o 024 317 5

AVES
024 757 X

BIRDS   AVES   VERTEBRATES
024 176 8 $m 024 757 X $o 024 526 7

           ORNITHOLOGY
           $n 024 519 4

SPHENISCIFORMES
024 510 0

PENGUINS   SPHENISCIFORMES   BIRDS
024 784 7 $m 024 510 0 $o 024 176 8

INPUT RECORD FOR 'PENGUINS'

0247847Penguins$ m0245100$ o0241768#

Figure 7. Thesaurus Input Records

① RELATIONAL CODES USED TO LINK TERM ADDRESSES (RINs)
Relationships based on IS 2788

$m = EQUIVALENCE RELATIONSHIP
Synonyms
Quasi-synonyms

$o = HIERARCHICAL RELATIONSHIP
Genus-species
Hierarchical whole-part

$n = ASSOCIATIVE RELATIONSHIP

② MACHINE INSTRUCTIONS BUILT INTO CODES

$m = PRINT *See* REFERENCE, i.e. Aaa *See* Bbb

$n ⎫
$o ⎬ = PRINT *See also* REFERENCE, i.e.

Aaa
See also
Bbb
Ccc

Figure 8. PRECIS Thesaurus

for *penguins,* and we later handled a work on *vertebrates,* we could immediately produce references such as:

Animals   Zoology
  see also   and   *see also*
    Vertebrates   Animals

simply by quoting the RIN 024 526 7 in the appropriate field on the indexing record. We can also quote any of these addresses to link new terms to the names of categories which are already on the file—we need to set up the network only once, then leave the rest to the computer.

THE MANAGEMENT ASPECTS OF PRECIS

Finally, some of the management aspects of PRECIS merit attention. In view of the present conference, particular notice should be taken of its relationship to some of the other subject data appearing on current British MARC records. This immediately introduces the concept of a packet of subject data, which can best be explained through reference to the work-flow diagram which appears in figure 10.
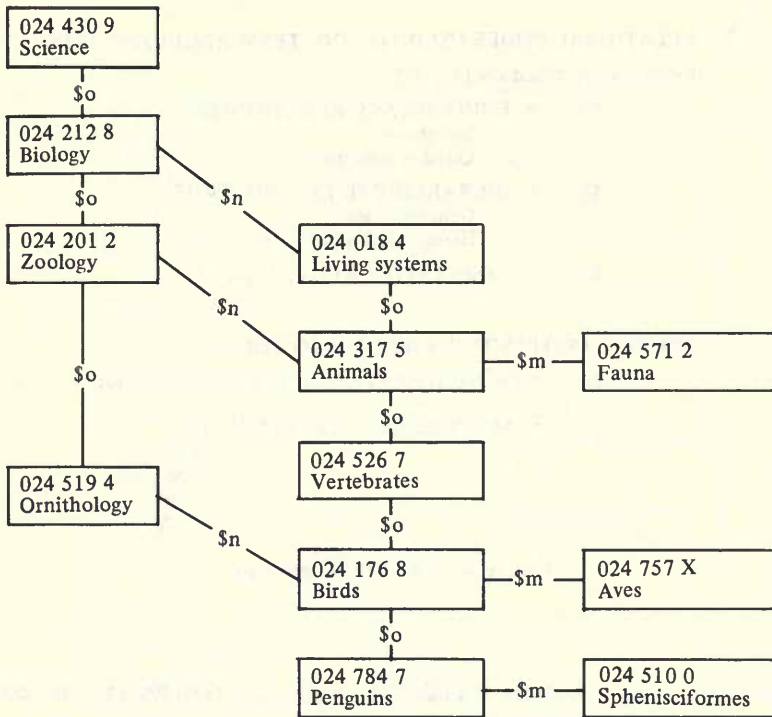
Figure 9. Network of Terms Linked by Thesaural Relationships

When documents are processed in the British Library, descriptive and subject cataloging are handled by separate teams of specialists, and a document enters the subject division after it has been cataloged descriptively. Each document is accompanied by a worksheet containing details of author, title, etc., recorded in the appropriate MARC fields. This worksheet, however, contains no fields for the recognized subject data, such as the PRECIS string or the DDC class mark. It does, however, contain one small box which will later be occupied by a number which will function as the link between the document record and all the appropriate subject data held in a separate file inside the computer.

As soon as a document is received from the catalogers, it is handled first by the PRECIS team. They formulate its subject, then check for a precedent in their master file of all past index entries, using any appropriate term as an access point. Let us assume that an indexer, after examining a document,

```
┌─────────────────────┐
│ DOCUMENT INTO       │
│ SUBJECT DIVISION    │
└─────────────────────┘
          │
┌─────────────────────┐
│ PRECIS TEAM         │
└─────────────────────┘
          │
     ╱────────────╲
    ╱ SUBJECT      ╲
   ╱  ALREADY       ╲────── NO ──────┐
   ╲  ON FILE?      ╱                 │
    ╲────────────╱        ┌─────────────────────────┐
          │                │ WRITE PRECIS STRING     │
         YES               └─────────────────────────┘
          │                           │
          │                ┌─────────────────────────┐
          │                │ CLASS BY DDC            │
          │                └─────────────────────────┘
          │                           │
          │                ┌─────────────────────────┐
          │                │ ASSIGN LCSH             │
          │                └─────────────────────────┘
          │                           │
          │                ┌─────────────────────────┐
          │                │ CLASS BY LC             │
          │                └─────────────────────────┘
          │                           │
          │                ┌─────────────────────────┐
          │                │ CHECK TERMS –           │
          │                │ ASSIGN RIN's            │
          │                └─────────────────────────┘
          │                           │
          │                ┌─────────────────────────┐
          └────────────────│ ASSIGN SUBJECT          │
                           │ INDICATOR NUMBER        │
                           │ (SIN)                   │
                           └─────────────────────────┘
          │                           │
┌─────────────────────┐   ┌─────────────────────────┐
│ QUOTE SIN ON        │   │ KEYBOARD.  READ         │
│ DOCUMENT RECORD     │   │ OUTPUT                  │
└─────────────────────┘   └─────────────────────────┘
          │
┌─────────────────────┐
│ DOCUMENT OUT OF     │
│ SUBJECT DIVISION    │
└─────────────────────┘
```
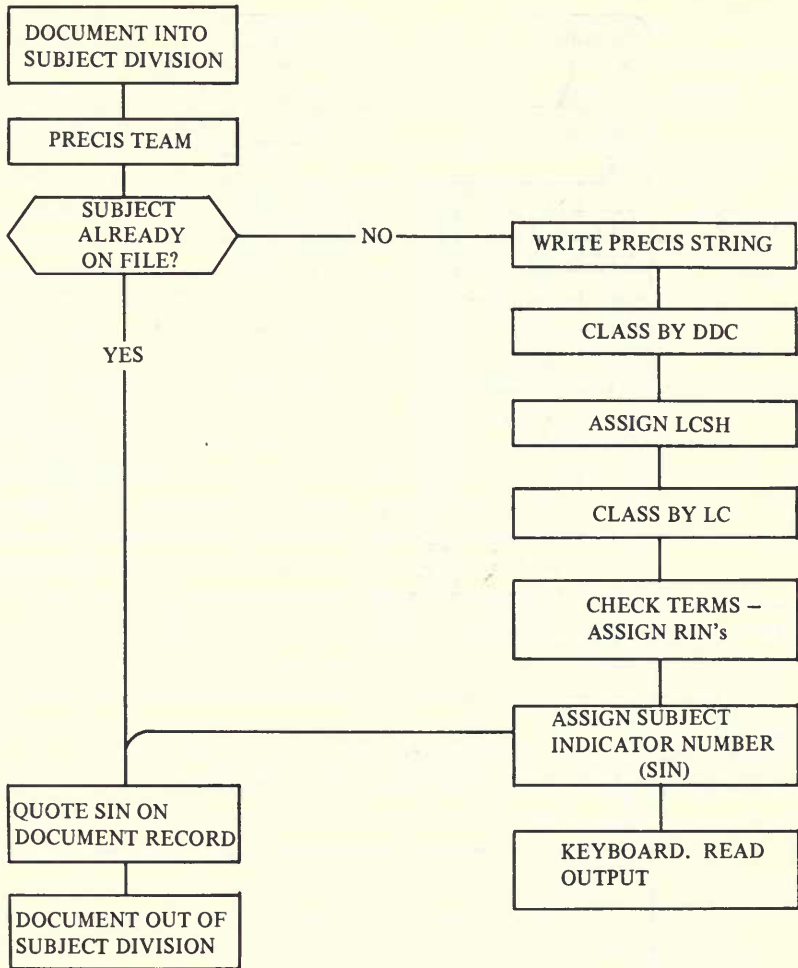
Figure 10. Work-flow through Subject Division of British Library

cannot find an exact match in the file of past decisions. The PRECIS indexer must then move to the right side of the work-flow diagram, to the point where the indexer writes the PRECIS string. This is recorded in a special field (Field 690) on the subject form which is shown in figure 11; the example shows the string for a document about a programming language called BASIC.

The document and the index form then move on together to the DDC team. Working from the string, and if necessary the document, this group

DC class no.   LC class no.   SIN

| 082 001.6'424 | 050 QA 76.73.B3 | 691 004 328 7 |

690

PRECIS string:

$	z	1	1	0	3	0$	a	digital computer systems
$	z	2	1	0	1	0$	a	programming
$	z	3	1	0	3	0$	a	Basic language
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		
$						0$		

RINs:

692	000 7471	692	692
692/1	005 920X	692	692
692		692	692

LCSH:

LCSH
650   Electronic digital computers $x Programming
650/1   Basic (Computer programming language)
6
6

Notes

Figure 11. Sample of Indexing Input Form

assigns the appropriate class number, and records this in Field 082 in the upper left corner of the indexing form. In the majority of cases, they do not expect to find an exact place in the DDC schedules for the subject expressed in the string. A careful analysis of our working methods has shown that only some 15 percent of PRECIS strings can be matched exactly by DDC numbers. We do not regard this in any way as a serious problem. We recognize a clear

distinction between the function of an index and that of a shelf-order scheme. The index entry should tell us what the document is about; the class number then supplements this information by indicating where it appears on the shelves.

The document and form then travel from team to team through the rest of the subject division—i.e., down the right side of the flow diagram in figure 10—and at each step, a different decision is recorded on the worksheet, such as the LC class number, Library of Congress subject headings, and the Reference Indicator Numbers (or RINs) which direct the computer to the machine-held thesaurus. Finally, the form is checked by a junior indexer, who ensures that all the necessary fields are occupied and then strikes out the next available random file address number from a machine-produced list. This is the Subject Indicator Number (inevitably shortened to SIN) which is written in Field 691 on the top line of the indexing form. The SIN identifies the address in a machine-held file at which all these subject data will be stored, as a package, for future use. The same number is also transcribed onto the catalog form, and then becomes the link between the document record and all its appropriate subject data.

The subject form goes on to the keyboarding section, and the data are submitted to the computer, which first carries out a series of validation checks; if these are satisfactory, it then assigns all the data to disk. It also responds by producing a full set of authority file cards of the kind shown in figure 12. These are read for errors such as spelling mistakes, and if they are satisfactory they are filed as a reference tool for the indexers and classifiers. It should be noted that each of these cards contains the whole packet of subject data, including the SIN itself—this is the number 004 3281 which appears on the left side immediately below the index entry.

Let us now return to the work-flow diagram in figure 10, and visualize a different situation. Let us assume this time that after examining the document, the indexer checks the authority file and finds that the subject, as he perceives it, has already been handled in the past. In that case, he can transcribe the SIN directly onto the catalog record, and the process is then completed. The document and its catalog worksheet can now be routed out of the subject division down the left side of the flow diagram. We have kept very careful statistics of the extent to which this left-hand route is used, and calculated that some 55 percent of BNB's total throughput was handled in this way from a three-year-old authority file. Obviously, from a manager's viewpoint, this represents a worthwhile savings of time and intellectual effort.

Figures 13-16 show some figures relating to indexer performance and index evaluation, as well as a survey of the present and potential users of PRECIS. Unfortunately, this list is already outdated. I feel that the most satisfying aspect of this survey of users is its sheer diversity. This applies first

```
Digital computer systems
   Programming. Basic language
0043281                          082010 001.6'424

690000  $z11030$adigital computer systems$z21010$aprogramming
$z31030$aBasic language

692000 0007471 692000 005920X

650000 Electronic digital computers $x Programming
650000 Basic (Computer programming language)
050000 QA76.73.B3
```

```
Programming. Digital computer systems
   Basic language
0043281                          082010 001.6'424

690000  $z11030$adigital computer systems$z21010$aprogramming
$z31030$aBasic language

692000 007471 692000 005920X
650000 Electronic digital computers $x Programming
650000 Basic (Computer programming language)
050000 QA76.73.B3
```

```
Basic language. Digital computer systems
0043281                          082010 001.6'424

690000  $z11030$a digital computer systems$z21010$aprogramming
$z31030$aBasic language

692000 0007471 692000 005920X
650000 Electronic digital computers $x Programming
650000 Basic (Computer programming language)
050000 QA76.73.B3
```

```
001.6'424                          QA76.73.B3
0043281

690000  $z11030$adigital computer systems$z21010$aprogramming
$z31030$aBasic language

692000 0007471 692000 005920X

650000 Electronic digital computers $x Programming
650000 Basic (Computer programming language)
```

Figure 12. Diagnostic Printout on Continuous Card Stationery

1 *Indexing Rates*

Indexing rate (i.e. string writing) is approximately 30 documents per working day of seven and one-quarter hours, i.e. the average time required to string a document on a new theme is about 18 mins. This figure represents 'elapsed working time', as opposed to stop-watch times, which would be 37-50 percent shorter.

Manipulation coding accounts for less than 10 percent of the total string writing time.

2 *Statistical properties of strings*   *Averages (mean)*
Number of strings per document   1
Number of terms per string   2.7
Number of lead terms per string   1.9

3 *Operation of the RIN and SIN systems*
   A. Proportion of documents handled by quoting
      existing SIN's from a three year file (1971-73)   55%

   B. Number of terms in thesaurus after three years   27,000

Figure 13. Indexing Performance Figures Collected at BNB

1 *Test environment*
PRECIS index to 584 journal articles in the field of management.

100 questions; 1 relevant document per question; each question searched once.

28 researchers, mainly students.

2 *Success rate of searches*
No. of successful searches (relevant document retrieved).   83

No. of unsuccessful searches (relevant document missed).   17

Recall ratio   83%

3 *Search times*   *Average per question*
                                     (to nearest 15 secs.)

	Mean		Median	
	Mins.	Secs.	Mins.	Secs.
Successful searches	1	30	1	00
Unsuccessful searches	4	00	3	00
All searches	1	45	1	00

Figure 14. Research at Liverpool Polytechnic

Note on symbols: (a) = form and/or frequency of output
(b) = production: computer or manual
(c) = 1-stage or 2-stage index
(d) = if 2-stage, classification or other address system

A) CATALOGUES OF LIBRARIES OR LIBRARY NETWORKS

1 East Sussex Public Libraries − *(a) COM; (b) computer; (c) 2-stage; (d) DC*

2 London Borough of Hillingdon Libraries − *(a) COM; (b) computer; (c) 2-stage (d) DC*

3 Sheffield College of Education − *(a) card; (b) manual; (c) 1-stage*

4 Polytechnic of Central London − *(a) card, may experiment with COM; (b) manual, but computer planned; (c) 2-stage; (d) DC*

5 Stockwell College of Education (London) − *(a) COM; (b) computer; (c) 2-stage; (d) DC*

6 Media Resources Centre (Inner London Education Authority) − *(a) card; (b) manual, but computer planned; (c) 2-stage; (d) DC*

7 Aurora High School Library *(Ontario, Canada)* − *(a) card; (b) manual; (c) 1-stage*

B) BIBLIOGRAPHIES

1 Australian National Bibliography − *(a) printed & cumulating; (b) computer; (c) 2-stage; (d) DC*

2 British National Bibliography − *(a) printed & cumulating; (b) computer; (c) 2-stage; (d) DC*

3 British National Film Catalogue − *(a) printed & comulating; (b) computer; (c) 2-stage; (d) UDC*

4 A/V Materials for Higher Education *(British Universities Film Council)* − *(a) printed; (b) computer; (c) 2-stage; (d) UDC*

5 HELPIS (A/V materials) − *(a) printed, intermittent; (b) computer; (c) 2-stage; (d) UDC*

6 HELPIS-MEDICAL − *(a) printed, intermittent; (b) computer; (c) 2-stage; (d) UDC*

7 Film Catalogue *(College Bibliocentre, Ontario)* − *(a) printed & cumulating; (b) computer; (c) 2-stage; (d) broad subject headings + serial numbers*

8 British Education Index *(from January 1976)* − *(a) printed & cumulating; (b) computer; (c) 2-stage; (d) PRECIS S/H*

Figure 15. Users of PRECIS

A) CATALOGUES OF LIBRARIES OR LIBRARY NETWORKS

1 Wollongong University *(NSW)* – *(a) card; (b) manual; (c) 1-stage*

2 S.G.M.E. *(Dept. of A/V Materials, Ministry of Education, Quebec)*–*(a) printed, French; (b) computer; (c) 2-stage; (d) Lamy-Rousseau classification of A/V materials*

3 Department of the Environment *(GB)* – *details not settled*

4 British Library Reference Division *(formerly British Museum Library)* – *(a) COM, cumulating; (b) computer; (c) 2-stage; (d) PRECIS S/H*

5 Université de Rouen, Section Sciences – *(a) printed or COM, French; (b) computer; (c) 2-stage; (d) thesis serial number or PRECIS S/H*

B) BIBLIOGRAPHIES

1 British Catalogue of Music – *(a) printed & cumulating; (b) computer; (c) 2-stage; (d) details not settled*

C) BACK-OF-THE-BOOK INDEXES

1 Public Record Office (London) – *indexes to calendars etc*

2 Scottish Record Office (Edinburgh) – *indexes to calendars etc*

AGENCIES PLANNING PILOT PROJECTS, OR
ENQUIRING FOR TRAINING OR PROGRAMS

1 Malaysian National Bibliography

2 South African National Bibliography

3 South African Council for Scientific & Industrial Research

4 Danish Library Centre

5 ONTERIS (Ontario Educational Research Information Service)

6 British Library (Library Association Library)

7 Indian Library Science Abstracts

8 National Film Board of Canada

Figure 16. Pilot Projects

to the size of the organizations involved—they range from a high school in Canada to two national bibliographies. It also applies to the media being indexed, which range from monographs, through audiovisual materials, to archives held in two public record offices. It is, I think, worth recording that none of these factors has affected the use of the system.

REFERENCES

1. Roberts, Norman. "Review of *Classification in the 1970's*," *Journal of Documentation* 28:184, June 1972.

2. Maltby, Arthur, ed. *Classification in the 1970's*. Hamden, Conn., Linnet Books, 1972.

3. Perreault, Jean M. "Review of *Classification in the 1970's*," *International Classification* 1:47-48, 1974.

4. Davison, Keith. *Classification Practice in Britain*. London, Library Association, 1966.

5. Austin, Derek. "Differences between Library Classifications and Machine-based Subject Retrieval Systems." In *Proceedings of the Third International Study Conference on Classification Research, Bombay, 1975*. The Hague, FID. (In press.)

6. Vickery, Brian C. *Classification and Indexing in Science*. 3d ed. London, Butterworths, 1975.

7. Mills, J. "Progress in Documentation: Library Classification," *Journal of Documentation* 26:138, June 1970.

8. Austin, Derek. "The Development of PRECIS," *Journal of Documentation* 30:47-102, March 1974.

9. _____. *PRECIS: A Manual of Concept Analysis and Subject Indexing*. London, BNB, 1974.