A U.S. Indexer attends a PRECIS Indexing Workshop

Barbara M. Preschel

After attending a PRECIS* indexing workshop in the United States, a U.S. indexer and teacher of indexing comments on PRECIS indexing and its future applications.

In October 1976, a three-day PRECIS workshop was held at the University of Maryland. It was the first time the PRECIS system was demonstrated and discussed in depth in the United States, and people from twenty-one states, Canada, Australia, Great Britain, and Denmark were in attendance. Sessions ran from 9 a.m. to 9 p.m. and consisted of lectures by Derek Austin on how to use PRECIS indexing techniques, exercises that allowed participants to try their hands at using the techniques, and speakers who discussed how they had used or tested PRECIS indexing in various working situations and for various types of materials. What follows is a short disquisition on items of interest and observations made at the workshop. It is not definitive since a three-day workshop, no matter how intensive, and no matter how extensive the reading and experience brought to it, is still a relatively superficial acquaintance.

The PRECIS indexing system

PRECIS is a technique of content analysis for bibliographic control that is currently used in the production of the British National Bibliography and has also been used to produce subject indexes to a few other bibliographic tools. PRECIS is based on the following principles.

1. The computer should be used to do the donkey work of alphabetization and manipulation of index entry terms. Human effort should be used only for content analysis of the texts, to write the string of terms used as index terms, and to determine which terms will be used as lead terms.

2. Each string of terms, or index entry, should express the entire theme or topic of the text. Each entry should be co-extensive with the text it indexes and should cover all the indexable concepts in the text.

3. Each index entry should be meaningful when read and similar in format to ordinary language.

4. The system should be based on a set of logical principles that are independent of subject field.

5. The entries should be supported by synectics or cross-references that are semantically related, but not necessarily inherent in the index entries themselves.

6. Each term in the string of terms, or index entry, created by the indexer, should set the next term in the string of terms in its logical context. Each term in the string should be related in a one-to-one relationship with the terms directly ahead and behind it in the string.

The PRECIS indexing system consists of two aspects:

1. The syntax of the system (its structure).
2. The semantics of the system (the synectics).

*As readers of recent issues of The Indexer will know, PRECIS (PREserved Context Index System) is essentially a system for producing printed alphabetical subject indexes.
not be discussed in detail here. However, as an example, one of the simpler exercises used in the workshop is given in Figure 1.

The role operators must be used in a certain order, and the effect of this is that the indexer is forced to think about the subject in the text in a logical manner.

He must first write a paraphrase of the indexable subjects or concepts in the text. This paraphrase should be designed to cover all the indexable concepts in the text. If one paraphrase

FIGURE 1

SUBJECT: Servicing aircraft engines
STRING OF TERMS
PRODUCED BY THE INDEXER: (1) aircraft √
(2) engines √
(2) servicing √

ROLE OPERATORS USED:
(1) Key system
(2) Action/Effect
(p) Part/Property

COMPUTER INSTRUCTIONS Used: √ indicating that this term should be used as a lead term in an entry.

ENTRIES THAT WILL RESULT:
AIRCRAFT Engines. Servicing.
ENGINE. Aircraft.
Servicing.
SERVICING. Engines.
Aircraft.

There are more than twenty-five possible role operators that may be used by the indexer to establish relationships between and among the terms in a string. The numbered operators control the typography and format of the printed entry; the lettered operators regulate meaning. The operator's primary value to the indexer seems to be to force the indexer to fit the subjects in the original paraphrase into a certain logical framework and to provide a step-by-step basis for the identification and labelling of the concepts in the paraphrase. It may even be that the mind of an experienced PRECIS indexer operates on two levels when he is constructing subject paraphrases and the role operators he has internalized interact with the indexable concepts in the document in his mind to produce a subject paraphrase that has been constructed to fit the role operators as well as the indexable concepts in the text.

Much was made at the workshop of the fact that, given the same original subject paraphrase, indexers would almost always create identical strings of terms, and therefore identical entries. This is inherent in the system since the schema of role operators determines the order and the way in which the concepts in the string are identified. This almost complete consistency in the creation of index entries is to be expected from the application of a rigid logical system for determination of entries to a pre-established subject paraphrase. What is not said is, that in most working situations, an indexer is not given a pre-established paraphrase to index, he is given a document and must create his own paraphrase. Past studies that concerned themselves with the consistency of indexer perception of indexable matter in texts seem to indicate that different indexers might create different subject paraphrases for a given document and thereby vitiate the consistency achieved by the application of the PRECIS indexing syntax to pre-established subject paraphrases.

Although some of the literature on PRECIS states that it is a natural language indexing system, it has been pointed out* that this description is imprecise. A PRECIS thesaurus of more than 33,000 terms exists and is consulted by the indexers when they are choosing terms with which to characterize the indexable concepts in the texts being indexed. The thesaurus grew by the accretion of terms used by the indexers, but at present, if a concept has been indexed in the past, an indexer is not free to choose any term he feels is appropriate as an
entry for the concept. He must use the concept as it appears in the thesaurus.

The semantics of PRECIS

The semantics of the PRECIS indexing system are concerned with relationships that are independent of the treatment of a concept in a particular document. These are relationships that are not document-specific, but exist naturally in the world at large, such as the relationship between "birds" and "penguins". There are three general types of relationships that are used as the basis for the syndetics in PRECIS.

1. Equivalence relationships (synonyms and quasi-synonyms).
2. Hierarchical relationships (generic and part-whole).
3. Associative relationships (involving a defining or explaining function).

As each new term is put into the thesaurus, it is examined to see if it has relationships that are not document-specific. If it has, a network of relationships is established for it. This network may include terms that are related by equivalence, hierarchy, or association, and the cross-references generated by the computer for individual entries are based on this pre-established network. Terms in the thesaurus are given RINs or Reference Indicator Numbers and through various relational codes, a kind of classification system has evolved. PRECIS is not based on an existing classification scheme, but a classification scheme has been self-generated. It is a flexible and hospitable classification scheme, since additions, deletions, and corrections can be made relatively easily, and it is possible in this classification scheme for a word to be part of two or more hierarchies, but it is a classification scheme nonetheless.

The PRECIS workshop

The workshop format enabled participants to learn PRECIS indexing heuristically, by doing it. The use of certain role operators was explained in detail and then we were given exercises in which we were asked to analyze subject paraphrases, apply the appropriate role operators, and establish what the resultant index entries would be. We were told about the semantic networks for syndetics and then were asked to construct semantic networks similar to those that appear in the PRECIS thesaurus. The participants in the workshop were, of course, bright and earnest, and primarily librarians. A few were indexers or indexer/librarians who earned their livelihoods indexing, supervising indexers, or teaching indexing. This latter group, from personal observation, seemed to have much less trouble doing the exercises than those people who were not experienced indexers. I believe the explanation for this is quite simple. The syntax of PRECIS forces the indexer to look for specific categories of concepts in the text to be indexed. These categories are analogous to the journalists' "who, what, where, when, why, and how"; or to the West Publishing Company's instructions to searchers of their legal publications' indexes to analyze their query in terms of 1. parties involved (e.g. clergymen, paupers); 2. places in which facts arose (e.g. theatres, restaurants); 3. objects or things involved (e.g. baseballs, hammers), etc. They are analogous to any of a number of other systems' instructions to indexers, abstractors, or searchers on how to choose or retrieve indexable matter. That is, a logic is imposed that forces the indexer to think about the concepts in the text in certain contexts, which concepts in the text are important concepts in these contexts, and which are important only because of their relation to other concepts. People who are indexers have already been trained, or have trained themselves to approach a text in this fashion. Therefore, although one of the most important aspects of PRECIS is that it has established step-by-step procedures for indexing, the importance of this aspect of the system does not lie in the instructions themselves, but in the codification of the instructions and the fact that these instructions are not related to a particular subject discipline, but are based on linguistic categories.

Approximately one-third of the workshop was devoted to the delivery of papers. Some discussed specific instances in which PRECIS was used or evaluated for use. These ranged from a paper on non-mechanized use of PRECIS in a high school library, to PRECIS indexing of films and other non-book media in special libraries and special collections, to evaluation of the use of PRECIS in a university setting. There was a paper by Jutta Sorensen outlining multilingual and translanguaging aspects of PRECIS which seemed to indicate that because there are certain linguistic universals, and because special codes can be introduced into the system for the idiosyncrasies in grammar in each language, it is possible to use PRECIS for languages other than English.
There was a research paper by Phyllis A. Richmond in which Library of Congress subject headings (LCSH), KWIC and PRECIS indexing were compared; and another by Ann Schabas which reported research being done comparing PRECIS indexing, LCSH, and words from titles for purposes of selective dissemination of information to special libraries.

The papers were interesting for many reasons. One reason in particular being that so many of them compared PRECIS indexing to LCSH. In almost all of the examples given, LCSH came off second best. They were less specific and provided fewer access points. In some of the examples, the poor showing of the LCSH was due more to the way in which they were applied than to any shortcoming in the subject headings themselves. If the cataloguer had applied the usual techniques for choosing headings reasonably adequately, the headings chosen would have undoubtedly been better. There were, however, many cases in which the inadequacies were caused by aspects of the LCSH system itself.

The development of Library of Congress subject headings and Library of Congress cards and their international distribution began, of course, prior to the age of the computer. The capabilities of the computer cannot be fully utilized in the present LCSH system. PRECIS is certainly more capable of full utilization of computer capabilities. But more important, LCSH are based, to a large extent, on ideas of information transfer that were formulated prior to the so-called "information explosion" of the post-World War II era. They often result in subject analysis that is not specific enough to enable the user to identify and retrieve, quickly and unambiguously, the one item he wants from the mass of similar items under a given heading.

With this in mind, I had brought to the workshop an example of a PRECIS entry from the British National Bibliography that was so specific that it looked like a detailed multi-level back-of-the-book entry. The degree of specificity it represented could most probably not be achieved with LCSH. I had also chosen the entry with the idea that at least part of it should be so specific that the chance that it would recur, and therefore that it was likely to become part of a thesaurus, was slim. The entry was:


The first four segments of the entry, and the last segment, are all terms that might occur in any co-ordinate index, and might also occur, alone or in combination, in LCSH. The entry as a whole has a degree of specificity that would probably not be attempted by the Library of Congress, since it would prove unwieldy, to say the least, in a card catalogue. However, the interesting segment of the entry is "Effects of installation of pelican crossings." (For the benefit of U.S. readers, I have ascertained that pelican crossings are a type of traffic signal.) This phrase could be handled in a co-ordinate system by indexing for individual concepts, if the terms were acceptable to the system. LCSH could be used, if appropriate terms were available, but it is improbable that a heading of such multi-level complexity and specificity would be assigned. Mr. Austin, after stating that working backwards from the entry would probably result in a different set of indexing instructions from those actually used, kindly attempted to reconstruct the indexing that produced the heading. He arrived at the following:

(z) (0) London
(x) (2) road traffic
(y) (p) accidents
(y) (2) reduction
(y) (2) effects of installation of pelican crossings
(x) (1) pelican crossings
(y) (2) installation
(y) (2) effects on reduction of road accidents

Note that "effects of installation of pelican crossings" is kept together as a single indexing term. I asked if this phrase would be entered in the thesaurus. Mr. Austin replied that it probably would not. There was a relatively small degree of probability that the term as a whole would recur in the system. "Pelican crossings" would be a thesaurus term, however, as would some of the other terms in the example.

The effect of this kind of treatment of subject terms, in that some are standardized and appear in an authorized list of terms and some are created for individual documents—and that these latter would probably not become part of the authorized list of terms, and would probably not be lead terms—is that the PRECIS system has a flexibility that is not possible in a system such as LCSH. Whether or not this spontaneity would have a deleterious effect on
future cumulations is something about which to speculate.

Management aspects of PRECIS indexing

PRECIS indexing at the British National Bibliography is done by teams of personnel. A team consists of six PRECIS indexers and an index supervisor, two people who assign the Dewey Decimal Classification numbers used as locaters in the system and for the classed arrangement of the bibliographic citations in the BNB, three people who assign LC numbers and LCSH, one person who spends half his time on DDC numbers and half on RINs, and one senior supervisor.

The system is cumbersome and expensive. A nine-position code must be assigned to the PRECIS string as instructions to the computer, and there is a different group of codes for proper names, etc. that will be used as index entries. PRECIS strings are, on the average, about 25% longer than LCSH for the same document, 81% longer if the manipulation codes are included. Only 15% of the time is there an exact match between the PRECIS string and the DDC number; the DDC is usually not specific enough and each string may be assigned only one DDC number.

For libraries or publishers of indexes who might like to start a PRECIS indexing system, development costs and training costs are high. Computer programs are available for IBM 360 and Sigma 7 machines, but they are in assembler language. Plans are being made to re-write the programs in COBOL. Use of PRECIS in manual systems is much too difficult. At present MARC tapes and PRECIS tapes are not compatible, but PRECIS indexing appears on all British MARC tapes.

Initially, PRECIS is taught to indexers at the BNB in a ten-day course. As of October 1976, approximately 300 people had taken this course. The course itself is only a beginning, since much practice and continuing interaction with other PRECIS indexers is necessary to perfect PRECIS techniques.

The future of PRECIS

It was obvious at the workshop, that many of the people who use PRECIS envisage it as the LC card system of the computer age. That is, they envisage an international multilingual system of computerized bibliographic control for document retrieval based on PRECIS. They may very well be correct.

The individual aspects of PRECIS are not, taken one by one, things that have never been done before. Many have been used in other computerized information systems or for various subject areas. What makes PRECIS an innovative and powerful system is that these separate aspects have been fused to produce a flexible and integrated system. What makes PRECIS a candidate for a future universal computerized system of bibliographic control is that it exists and is being used by a major international bibliographic reference source. Its elaborate semantic and syntactic input superstructure may be more time consuming and intellectually wearying than seems to be necessary, but it exists, and it works, and a growing body of documents is being indexed using it.

Just as the LC card system, or the Ohio College Library Center network, have grown, not because they were the best system possible, but simply because they were the system that was functioning at the time something like them was needed, PRECIS may grow because the large number of documents currently being published requires a system that will give quicker, more specific access than systems at present in use on an international basis.

What will be needed for PRECIS to develop into a universal international system?

1. Money.
2. Training programmes in schools of library and information science.
3. Adoption by national libraries and information dissemination agencies.
4. The decision by major libraries to close their current subject catalogues and establish new subject catalogues based on PRECIS entries.
5. Universal inexpensive availability of PRECIS indexing products.

Bibliographical References