PROJECT TO AUTOMATE THE CANADIAN EDUCATION INDEX: LOOKING FOR A LANGUAGE (PROJET POUR L'AUTOMATISATION DU REPERTOIRE CANADIEN SUR L'EDUCATION: A LA RECHERCHE D'UN LANGAGE)

Carolynn E. Bett
Canadian Education Association
252 Bloor St. W., Toronto, Ontario M5S 1V5

ABSTRACT

This brief history of the project to automate the Canadian Education Index centres mainly on a consideration of three indexing languages in relation to the objectives of the Canadian Education Index: our own subject headings (L.C. style), ERIC and PRECIS. PRECIS comes closest to meeting our objectives, but as funding has not yet been received, decisions cannot be made. Work continues on the problems of format and cost effectiveness. (Pour le projet d'automatisation du Répertoire canadien sur l'Education, dont on trouvera ici une brève historique, trois langages d'indexation furent étudiés dans l'optique des objectifs du Répertoire: nos propres vedettes-matières (style L.C.), ERIC et PRECIS. PRECIS se révéla le plus approprié à nos objectifs mais, au moment de la rédaction de ce rapport, la question du financement n'étant pas encore réglée, les décisions finales demeuraient encore en suspens. Le travail continue sur les questions de format, de coût et d'efficacité.)

Rather than being a technical discussion, this paper will approach the problem of computerizing a bibliographic system from the point of view of a novice librarian who has no computing science background. The system to be computerized is the Canadian Education Index, an author-subject index, which attempts to be a comprehensive listing of journal articles, books, pamphlets and reports on education in both French and English, published in Canada. The vedettes-matière or French subject-headings given at the back of the Index together with their English translations enable Francophones to use the Index easily, although it is filed in English.

The desire to automate the production of the Canadian Education Index came in response to the tedium of repetitive typing, proof-reading and filing of the cards that are used in the photo-offset printing of the Index. Automation would also combat the error rate inherent in manual production. Moreover, users constantly complain about late delivery of issues that have scarcely gone into production.

CANADIAN EDUCATION INDEX

A cursory glance at our operation showed us the following general benefits to be derived from computerization. Typing and proof-reading time would be cut in half, and the possibility of error would be reduced. Filing mistakes would be eliminated together with the 2 weeks needed to file an issue and the 2 months needed to file the cumulation: the computer would do the filing overnight. One of the most time-consuming jobs is pasting the cards onto sheets of paper to format the Index for photo-offset printing. The computer would give us camera-ready formatted sheets that could be pasted-up in a matter of mere days rather than the month or more each issue now takes. Stripping the paste-up in preparation for interfiling the cumulation would be eliminated.

Finally, the full subject-heading list with vedettes-matière and references would be keyed once only and updated as needed. At present, they are typed and proof-read for every issue. It was the problem of updating the subject-headings that was our first area of concentration in the plans for computerization. But before work could begin, a general background in computer capabilities and techniques had to be established.

Because of the relatively small number of documents handled in each issue of the Index, it was asked if automation would be economical or if the manual production could be streamlined. The number of documents handled has doubled since this first inquiry; however, at that time we were creating the first detailed manuals for indexers, typists and proof-readers. These records of how to handle basic problems were in themselves a streamlining agent, but they are now in need of updating and time is less available for the task. A document describing the present workflow for the Index (Bett 1974) was prepared, which enabled us to see a few minor areas where streamlining was possible. Still, it was apparent that much human effort and time was being expended on work that would be done with greater speed and a lower error rate by the computer.

Having decided to investigate computerization further, the editor of the Canadian Education Index needed to become aware of the capabilities of the various systems available that deal with bibliographic information and needed to find experts who would be willing to give time and aid to the CEI project. Familiarity with the jargon and some idea of computer capabilities were gained from taking a data processing course at Ryerson. A visit to the College Bibliocentre in Toronto brought the PRECIS (PREserved Context Index System) into the stock of known indexing systems. A phone call to the Faculty of Library Science generated contact at the University of Toronto Library/Library Automation Systems (UTL/LAS) and eventually brought a student, with computing science background, to aid with research.

One of the major concerns which arose during the familiarization period was that of compatibility with other systems for education

bibliography in this country. Our concern stems both from the point of view of consolidating education information for wide distribution and also from the point of view of drawing on the experience and work of those who have already built systems. However, it was not until we had decided to pursue our own path that others became distinctly known to us. Since our first inquiry, several of these systems, both automated and manual, have been documented (Lamy-Rousseau, 1974; Olsen, 1974; Beardsley, 1975). We are now in close contact with those who are automated or plan automation as well as with other education bibliographers not mentioned above. Because most of us are now aware of what the others are doing, the possibility of developing a national information system for education seems bright.

Keeping in mind both the broad context of a national network for the collection, storage and retrieval of information in education, as well as the context of production of our particular segment of the information, we turned our attention to problems surrounding the subject-heading list for the Canadian Education Index.

From the outset, the volunteer indexers had been indicating their dissatisfaction with the lack of specificity of the current subject-headings. Because revision of the list would be a mammoth manual task, which would need doing at regular intervals, the library student, Stan Guttman, set out to study computerization of production of the subject-heading list, with the following objectives in mind:

- easy update capability
- quick reprint capability
- a structure that would be useful in a projected automated retrieval system
- compatibility with other subject authority files, if possible.

The last two objectives relate to the wide co-operative context, the first two to our immediate production needs.

Our present subject headings are patterned on Library of Congress but are not L.C. headings; therefore, use of our system strikes out the possibility of compatibility with L.C. users. Although L.C. headings are undergoing pressure to change because of the computer, it is doubtful they will ever reach the efficiency of some other systems for automated retrieval. It soon became evident that a lot of work and expense would be involved to give quick update and reprint capability, leaving the problem of lack of specificity still with us and the objectives of compatibility with other subject systems and efficiency in an automated retrieval system still unmet.

Therefore, Stan's first paper was an investigation of three indexing languages: our own patterned on L.C., ERIC and PRECIS (Guttman, 1974). It should be pointed out that PRECIS is more than an indexing language; it is a "machine-produced alphabetical indexing system, based on syntax derived from a study of natural language" (Austin, 1974).

Therefore, it includes the manipulation programs as well as the instructions for indexing. Considered as a language, PRECIS differs from other indexing languages insofar as it "consists essentially of a set of working precedures, not a prescribed list of terms or phrases" (Austin 1974). However, one of the potential outputs of PRECIS is a printed thesaurus, and it is the efficiency of this potential thesaurus that Stan was considering in relation to ERIC and our own subject headings.

During the writing of the paper, we entertained the possibility of developing from one language to a better one as funding became available. However, it soon became evident that choice of a language meant committment to a system. Change would be a rather complex and costly precedure involving formats and programs.

A very basic comparison of the three systems includes the following observations. The Library of Congress system and our own were developed for manual production and manual search methods. Pre-coordination of terms is a feature of the system that aids enormously in manual retrieval. However, broadness of the terms produce many irrelevant citations during machine retrieval.

The ERIC system was developed for machine retrieval. Unbound terms used post-coordinately make it ideal for machine retrieval but most awkward for manual retrieval. Yet many libraries without machine capability find the content of ERIC essential for their users.

PRECIS was developed to produce, by computer, a printed index for manual retrieval; however, the potential for machine retrieval both by string and term is part of the system design. Pre-coordination of terms is achieved by stringing together natural language terms according to strict rules of syntax laid down in the PRECIS manual. This produces the PRECIS string. A reference file is made for each term using the traditional see and see also language. This is called the RIN (Reference Information). The SIN is the subject information file, used for controlling the subject strings in the system.

PRECIS preserves pre-coordination and facilitates manual search, which seem to be necessary for many small Canadian education libraries at this time. However, should technology develop so that on-line retrieval is available to all our users, PRECIS will be developed to serve in this new situation.

Because PRECIS is a flexible natural language system, the thesaurus can be developed to reflect Canadian usage. The numbering system for strings and reference permit a change of term to reflect a change in usage; for instance, "superintendents" could be substituted for "inspectors" under the same number, as the usage and function changed.

This numbering system can also be used for bilingual indexing by having the indexer put both a French and an English string against the same number. Controversy is arising over the question of automatic switching from one language to another (Lamy-Rousseau 1974), which leads one to think that research done to date (Austin 1974) has been exploratory only.

We considered the International Bureau of Education (IBE) thesaurus in relation to Canadian usage and bilingualism. Essentially the principles behind national development and bilingualism are the same as for PRECIS. However, the IBE thesaurus is patterned after ERIC and, indeed, considers ERIC along with other national education thesauri to be part of its system. Automatic switching is possible in IBE because of its uniconcept structure, but with PRECIS there is the context based on a syntax to complicate the translation. Yet it is precisely this context which gives PRECIS its particular power.

We rejected IBE for immediate use because, like ERIC, it was created for machine retrieval and does not adequately serve present needs for manual retrieval. However, it is quite possible that when machine search by term is developed for PRECIS, the thesaurus we will have built may be linked to IBE. Thus, we are watching with interest the developments of the IBE thesaurus and are maintaining contact with its office.

Some question has arisen over the depth of indexing available with PRECIS compared with ERIC. Studies show that PRECIS gives greater precision than L.C. or DEWEY (Robinson June 1973 & March 1973); however, no study has been published comparing ERIC with PRECIS. The question is important because we wish to use PRECIS for the indexing of education documents and are already well aware of the shortcomings of L.C. and DEWEY. As used in the British National Bibliography, PRECIS strings have been written to index the whole document rather than concepts within the document. An average of two PRECIS strings per document have been written, although I have not read how many terms are used in each string. The approach has been patterned after L.C. rather than ERIC. Nonetheless, funding permitting, there is no reason why more and longer strings could not be written for specific uses. Because of the added feature of context in PRECIS, one is tempted to hypothesize that the number of false drops, on a search due to incorrect linking of terms would be fewer with PRECIS than with ERIC. However, this is yet another area of the system demanding more research.

From the production point of view, PRECIS provides not only easy update and quick reprint of the subject strings but also manipulation of the references, a time-consuming task in the present manual production. The general problems of repetitive typing, proof-reading, filing and paste-up could be solved by any automated system; however, the PRECIS

programs are already written and have been used for education. (Lamy-Rousseau, 1974; Robinson 1973). The particular adaptation of PRECIS to the British Education Index (Shaw 1974) will have specific bearing on the Canadian Education Index.

There are two aspects to this adaptation. One deals with the citation order of documents listed, for which the British National Bibliography uses DEWEY numbers. DEWEY was found not to be specific enough for education; hence, the PRECIS research team has developed a citation order based on broad terms from the PRECIS string. We had considered use of the IBE broad terms for this purpose but will now probably adopt the British solution.

The other adaptation concerns the use of the MARC (Machine Readable Cataloguing) format for periodical articles. Consideration of the MARC format for the Canadian Education Index is the second phase of our investigation. As yet, we have not heard the details of the British solution; moreover, we have the added complication of listing both monographs and periodical articles in the same file. Our student assistant, Stan, has offered a solution to the monograph and periodical problem, which is to be checked by experienced PRECIS users before a final decision is made.

Once format is decided, we will be able to do a fairly detailed costing of the computer operation. This work should be completed by the end of April.

Although funding is not yet established, prospects are promising; thus, we are working as quickly as possible in order to be ready to automate as soon as funding is received. With the generous assistance of our Canadian colleagues and with the brilliant break-through of the British PRECIS research team, we are making significant strides toward this goal.

CANADIAN EDUCATION INDEX

REFERENCES

AUSTIN, Derek 1974 The development of PRECIS: a theoretical and and technical history. Journal of Documentation, 30(1): 47-102.

BETT, Carolynn E. 1974 Work flow system for the Canadian Education Index. Toronto. 31p (unpublished)

[BEARDSLEY, Barbara] 1975 Bibliographic Notes. Canadian Education Association Newsletter, January:4. (short write-up on MERC and ONTERIS)

GUTTMAN, Stan 1974 A report to the Canadian Education Association on the computerization of the Canadian Education Index. Toronto. [30p.] (unpublished)

LAMY-ROUSSEAU, Françoise 1974 Système d'indexation PRECIS II. pp.27-54 in F. Lamy-Rousseau, Traitement automatisé des documents multi-media avec les systèmes ISBD Unifié, Lamy-Rousseau et PRECIS. Propositions S.I.L.P. Montréal, Service général des moyens d'enseignement, Ministère de l'éducation, Québec.

OLSEN, John ed. 1974 RIPE - a report. Vancouver, Educational Research Institute of British Columbia. 35p

ROBINSON, Derek 1973 PRECIS; an introduction. College Bibliocentre Newsletter, no 7 March:10-16.

_____ 1973 Indexing the Film Catalogue: with comparisons between LC subject headings and PRECIS entries. College Bibliocentre Newsletter, no 8 June:7-13.

SHAW, Christine M. 1974 The British Education Index; current practice and future possibilities. Education Libraries Bulletin. 17(3): 1-10.