The Index of the Diefenbaker Speech Collection: An Experiment in Computer-Assisted Indexing of Archives

by **ELIZABETH DIAMOND**

The Diefenbaker Centre in Saskatoon is in the final stages of producing a detailed computer-assisted index and finding aid to its collection of John Diefenbaker's speeches, both manuscript and recorded. The project, begun in 1984, has involved a substantial amount of planning and organization and has been beset by numerous problems. In the relatively new field of computer indexing in archives, every project will proceed by trial and error to some extent, but it is also necessary to learn from the mistakes of others while trying to avoid repeating them. An account of the Centre's experiences with the project may prove useful to other small institutions involved in organizing both audio-visual holdings and computer-assisted indexing projects.

The original deposit of Diefenbaker material at the Centre in 1979 included about 22.9 metres of manuscript speech materials; drafts, notes and final versions of over 1500 speeches given by John Diefenbaker from 1920 until his death in 1979. There were also 96 tape-recordings of his speeches and interviews, in varying states of preservation, some almost unusable. In September 1979 and February 1980, an appeal for further tapes was sent to radio stations across Canada. Copies of Diefenbaker recordings and related material were obtained from the collections in the National Archives and other repositories. John Munro, who assisted Diefenbaker in writing his autobiography *One Canada*[1] and later became director of the Diefenbaker Centre, deposited his tapes and transcripts of interviews used for the book. There are copies of taped interviews made by Peter Stursberg while writing his books *Diefenbaker — Leadership Gained 1956-62* and *Diefenbaker — Leadership Lost 1963-67*.[2] Previously unknown tape-recordings are still being discovered, and other related material is being added to the tape collection from time to time.

The initial reaction of an archivist or historian used to conventional textual records is to relegate speech material to a position of secondary importance. The assumption is made that political speeches are ephemeral, meant only for the time, place and audience to which they were given, the details soon to be forgotten as the course of events moves on. It is assumed that better evidence of a politician's real thoughts can be found in his letters, private journals and permanent writings on policy.

A detailed examination of John Diefenbaker's papers reveals a surprising lack of this conventional kind of evidence. His private letters, of which there are a great many, are

very superficial, giving details of his daily activities but little of his thoughts. There is nothing to equate to the revealing, detailed correspondence of nineteenth-century politicians nor are there private diaries.

On the other hand, the speeches, particularly the drafts and delivery notes, reveal some of Diefenbaker's political thoughts as he edited and changed the material produced for him by his aides into what he regarded as an appropriate expression of his opinion at the time. Basil Robinson, who worked with him as a liaison officer with the Department of External Affairs from 1957 to 1962, has recently written that "A lifetime in opposition politics had taught him to use speeches as a vehicle for formulating policy."[3] Sean O'Sullivan, another aide, recounts how Diefenbaker used to talk of using just eight speeches to change the country's opinion on any given subject.[4] Diefenbaker was one of the great political speakers of this century, perhaps the last whose style was formed before the introduction of television changed the nature of political speaking forever. The collection of his speeches preserved in the Diefenbaker Centre can serve both as a repository of his political thought and an illustration of the techniques of a great political speaker.

In October 1984, a project to create finding aids and a computer-generated index to the speech collection was begun. From the beginning, funding and organization of the project were uncertain. Even before the first two-year Social Sciences and Humanities Research Council of Canada (SSHRCC) grant was received in 1984, the staff member who made the initial application had left. Her successor was unable to consult with her in detail before continuing. The supervision of the project was undertaken by University staff with too little time to spend and too little support from the University administration. The bulk of the data entry was done by the only full-time archivist, who was also responsible for all the other functions of the archives. She was assisted by various temporary employees as funds permitted. As a result of a series of funding crises, staff changed too frequently, without enough overlapping time to ensure that the work was done consistently.

In 1984, when the project was designed, there were neither affordable personal computers nor appropriate database programmes available. It was decided to use a word-processing programme on the University of Saskatchewan's mainframe. Initially, data entry had to be done off-site, due to unexpected delays in installing a terminal in the Centre. Furthermore, the discovery that a large number of the tapes were too fragile to be used, caused an extensive delay while they were remastered.

As a result, even after a second SSHRCC grant had been expended in 1987, the project was not entirely complete. Due to the uncertainty of the Centre's own funding, a gap of nearly a year and another change of staff occurred before the final editing was undertaken. In retrospect, it is obvious that the constraints of insecure and inadequate funding aggravated the problems that this paper seeks to describe. Many of the intellectual problems, however, would exist in the best funded and most stable of projects. Learning from the mistakes of the Diefenbaker Centre may minimize them for others.

When work began in October 1984, the manuscript section of the collection, which was the more complete and homogeneous, was handled first. Each speech was given a distinct file number in a single chronological sequence. Drafts and preparatory materials were placed in files separate from the final drafts and given the same file numbers as the final versions with the suffix 'S' to distinguish supplementary material. All the material for one speech event was filed together.

The next stage in the plan was to complete and standardize the work sheets which had already been begun for each speech. A great deal of thought went into creating these work sheets. The final version was a single xeroxed sheet on which was entered all the relevant information about each speech as it was discovered. In the first section, completed for every speech, there were spaces for speech number, date, place and occasion. To these was added a space for "subject (key words),"[5] in which possible index terms were to be entered during the next stage.

The second section recorded information about the form and versions of the speech; whether there were printed versions in the collection, whether tape or video recordings had been collected, and from where, and whether there were any press versions. The final section included information about advisers and speech writers, press releases and press reaction to the speech. This last section was least often used but provided some very interesting information. The work sheets were filed in numerical order in binders, and were used instead of index cards when the material was being entered on the computer.

All the information was at hand in a standardized form and in categories which corresponded exactly to the fields in the computer programme. The programme as finally written did not allow for the inclusion of all the detailed information in the second and third sections of the work sheet. However, the binders of work sheets were kept and are still a useful reference tool.

Recording the main entries on the computer was completed next. The fields for this first part of the project were volume and speech number, location and event. Working from the work sheets the task was straightforward, and within six months 1563 speeches were entered. The programme sorted them in three sequences: by date, location and event. Perhaps because the speeches were already numbered in an almost exact chronological sequence, it was felt not necessary to generate a list sorted by volume and speech number. This was an unfortunate omission, because such a list would also have been a true shelf list. During the indexing stages the necessity of having a single main entry for each speech was lost sight of, and it was necessary to go back to generate a list on which each speech appeared only once in numerical order. The locations were entered as addresses would be in ordinary language; town or city first, followed by province, state or country. In retrospect, they should have been entered in reverse, so that the computer could have sorted all the places in one province, state or country into a single alphabetical list.

Three lists arranged by date, place and occasion were generated and similar procedures were completed for the tape-recordings. These lists formed the first or finding aid part of the project. The second part required turning the words and phrases entered on the work sheets under the heading "subject (key words)" into a series of uniform index terms to be used in the final references. What was needed was to create a controlled vocabulary from the original entry vocabulary. To do this, one preferred term had to be chosen to represent all the terms in the entry vocabulary with the same or similar meanings. All the other terms could then be listed with a *see* reference to the preferred term.[6]

Thesauri and lists of subject headings are very useful in selecting the best terms for a controlled vocabulary. No existing list seemed to suit the project's very specific needs, yet conformity to a standard of some kind seemed desirable. Eventually the National Archives of Canada Subject Authority List created by Judith Cumming in 1982 was

chosen as the most appropriate. However, as the work progressed it became obvious that even this list would not satisfy all the project's needs. A portion of the terms essential for indexing something as idiosyncratic as Diefenbaker's rhetoric did not and should not appear in any standard subject authority. This difficulty necessitated combining two sorts of indexing.

One of the processes we think of as indexing in archives is devising subject access for all the records in a repository. We also think of indexing on a smaller scale, of providing an index to a small or homogeneous collection, similar to the process of making an index to a single book. The first of these categories is analogous to library subject cataloguing; the second to what librarians call "back of the book indexing." Subject cataloguing is much improved by conformity to standardized subject authorities such as the Library of Congress Subject Headings, National Archives of Canada Subject Authority List or Provincial Archives of Alberta Subject Headings.[7] "Back of the book" indexing on the other hand, does not require such conformity and can be greatly complicated by trying to achieve it. It is generally agreed that this sort of index is too idiosyncratic for using a limited vocabulary, although some authorities recommend beginning a "back of the book" index by consulting a general subject heading authority or thesaurus as an initial framework.[8] Inevitably, however, a certain number of idiosyncratic terms arising from the work's own style and content must be added. No one would attempt to create a list of index terms meant to be used in the indexes of all books. They would certainly not expect such references to be used to the exclusion of all others.

The Diefenbaker speech project proposed detailed indexing which required a combination of "back of the book" indexing and subject cataloguing techniques. The very idiosyncratic nature of the material meant that a standard list of restricted vocabulary could not be used exclusively.[9] The project soon discovered this, and began to add terms to Cumming's list. Terms such as "Free Enterprise," "National Goals" or "Government Expenditures" are basic to Diefenbaker's political rhetoric, but far too vague and unspecified for a subject cataloguing authority. Their very vagueness makes them extremely difficult to fit into a controlled vocabulary and cross-referencing system. In adding these new terms to the list, the project workers lost some control and began to enter too many synonyms for both types of term. Moreover, the appropriate mechanism of references back to general and cumulative headings was not created. The result was that cross-referencing and cumulative grouping was not done consistently. For example, there were entries under "King Gov't Record" which were not included under the more general "Liberal Gov't Record." Many of these mistakes were caught at the editing stage, but others could not be and it would have been better to have instituted a procedure that prevented such omissions in the first place. If the project workers had maintained a separate list of their idiosyncratic terms, they might have been more conscious that they were trying to combine two types of indexing and could have maintained the integrity of Cumming's terms.[10]

The archivist entered the "key words" directly from the work sheets into a word-processing programme on the University's mainframe computer. Within each unit there were fields for date, place and event. In addition, there were fields for as many index terms as were required for each unit. The page number in the manuscripts, or the time position on the tapes of each index term, was entered separately, so that when an index term occurred at more than one place in the unit, it was entered for each occurrence. The completed documents were then extracted from the word processing programme and

used to create an ASCII file. A specially written data conversion programme was used to reorganize the data into a format which could be sorted on the utility sorting programme of the University's PDP 11/70 computer, using the RSTS operating system. The data were sorted four ways: chronologically and alphabetically by place, event and subject. The programme also sorted at a secondary level: chronologically within each alphabetical sequence and conversely, alphabetically within the chronological sequence. Using the computer's utility search programme, the data could also be searched for any given index term. The programme also allows new units to be added to the index, so that material generated by the Centre's Oral History programme and any copies of Diefenbaker's speeches which may turn up from other sources may be added.[11]

Had the project began only a few years later, a database management programme, rather than a word-processing programme, would no doubt have been used. The technology of six years ago is now seriously outdated. Completely automated indexing systems are being created at a very rapid rate. What seemed impossible ten years ago is now routine. It would be theoretically feasible to programme a computer to select the index terms directly from the speech texts.[12] However, both the manuscripts and the tapes would have to be put into machine/readable form, a procedure which cannot yet be completely automated and which, in this case, would have been too expensive.[13] Moreover, the resulting computer-created terms would still have had to be checked and edited carefully. In any case, it had always been the intention to generate the subject entries by hand, and to use the computer to sort and arrange them.

A software package could have been bought or designed to generate cross-references and variants of the main subject entries automatically.[14] Such packages require careful editing and control and cannot generate certain types of reference automatically.[15] For example, the reference "DE GAULLE, CHARLES-PQ-SPEECHES" could be manipulated by a programme to create "PQ-DE GAULLE, CHARLES-SPEECHES" and "SPEECHES-DE GAULLE, CHARLES-PQ." With more sophisticated punctuation and programming, other permutations of the entry could be generated. However, the cross-reference under the term "SEPARATISM" to "See also DE GAULLE, CHARLES-PQ-SPEECHES" could only be created by a personal indexer, and yet would be one of the most important. It was therefore decided to generate variants and cross-references manually as well, which demanded careful checking and editing to ensure that all references would appear under the appropriate headings. Even so, the computer greatly simplified the work, since a word-processor obviated the necessity of retyping the whole entry.

The original intention was to paginate and index the final versions but not the supplementary material. As soon as indexing in detail began, however, it became obvious that this plan would not work. The supplementary material, consisting of background research, drafts, suggestions from aides and other notes, was the most interesting part of any speech. It contained the evidence of how the speech was developed. This evidence is what makes the speech files valuable by giving insight into the thought processes of John Diefenbaker unavailable from other documentation. It is relatively easy to discover what he said; it is equally important to discover what he chose *not* to say. Omitting the supplementary material from the index would have made this discovery much more difficult. Moreover, the background material from his aides and researchers would help to reveal some of the sources of his ideas. Therefore, the plan was revised to include the supplementary files in the final index.

Pagination, which is essential for indexing to the detailed level of this project, presented another problem. The pages of the final versions tended to be numbered already, although rarely was the pagination completely consistent or accurate. The supplementary material was much worse, with page numbers and sequences altered, duplicated or omitted altogether. The decision was made to use the existing pagination in almost every case.

In retrospect this decision proved to be a mistake. Although time was saved at the initial stages, the resulting product was not as accurate as it appeared to be. The final product gives a file number followed by a page number. When the researcher goes to the file, he may find that there are as many as ten pages marked with the number he is searching for. Although his search is thus narrowed considerably in a large file, the experience is still confusing and irritating. Moreover, in most cases, he will want to examine the whole file, not just the page on which the reference occurs. Under these circumstances, it is doubtful whether the time spent meticulously entering imprecise page numbers into the computer programme was justified. Considering the great investment of time in the process of indexing to this degree of detail, it would have been preferable first to create an accurate pagination, or to omit page numbers altogether.

Perhaps the greatest single danger in creating indexes is the problem of the reliability factor.[16] All but the most cynical and sophisticated researchers have a great faith that indexes are comprehensive. There is an assumption that subjects not found in the index are not found in the documents.[17] More subtly, there is the assumption that all references to one subject are listed under one heading. It does not matter so much to researchers that the subject headings conform to those in other systems which they might have used, as that, once a subject heading has been found, all references to that subject will be present. The central purpose of all subject indexing systems is to achieve this apparently simple goal. Perhaps it is even more important in archives than in libraries. Most archival material is unique, and the researcher would not have the opportunity of finding another copy elsewhere as he would with a book.

Some archival institutions have regarded the problems of doing this as to be great that they did not attempt to create subject indexes but only the more easily controlled indexes of personal and place names.[18] No subject index will ever be totally comprehensive, but with careful planning, revision and editing it can aim at a large degree of reliability. The lessons provided by the experience of the Diefenbaker speech indexing project can be of particular help in planning for similar projects.

One concept that presents a problem in this type of indexing is literary warrant: the librarians' principle that the language of the index must be appropriate to or warranted by the language of the work indexed.[19] The concept is particularly important in this case, where we are concerned with the exact words used by Diefenbaker, which are often just as important as the ideas they were expressing. Cumming's list, adapted from the Library of Congress Subject Headings, uses the American term "railroad" as a preferred term to "railway." Yet John Diefenbaker consistently used the more Canadian term "railway." Since both the railways themselves and the notion of a distinct Canadian character were very important to Diefenbaker, we may not be warranted to alter his usage in this case. More generally, it can make the researcher's task more difficult if too much translation between the actual language of the creator of a document and the controlled vocabulary of an indexing language has to be done. In cases of detailed

indexing of idiosyncratic texts, consideration should be given to taking as many of the preferred terms as possible from the text, even when they are at variance with a prescribed list or thesaurus.

Two further library concepts must be considered before starting to index: exhaustivity and specificity. Exhaustive indexing means recording an index term every time it occurs in the text.[20] At first glance this seems both obvious and correct, yet it rapidly becomes clear that following this procedure would create a very large number of references to places in the text where the subject is mentioned only superficially. This is especially a problem where the subject is a very general one. Many researchers will be unwilling to tackle a long list of references, many of which are unimportant. However, as soon as the indexer starts to omit occurrences of a term, the chance of retrieving every relevant document is diminished. With a collection of speeches, such as the Diefenbaker collection, where a half-remembered quotation is sometimes the only clue to searching the index, this problem is significant. In this case, the only practical solution is for the indexer to be conscious of the problem. By avoiding indexing brief references to general concepts, while including a more exhaustive listing of more specific terms, a suitable compromise can be found.

Specificity is a related problem.[21] Indexers must decide how narrowly they will define each concept. The Library of Congress Subject Heading "Canada- Politics and Government, 1945-" might be specific enough in the index of a collection of American or British documents. In the Diefenbaker collection almost all of the documents could be said to be about Canadian politics in one way or another, so the heading would be useless because it was not specific enough. On the other hand, the single reference "Bren Gun Contract," which was initially entered into the Diefenbaker index as a main entry without cross-references, is too specific because it would be lost to a researcher looking for general references to the arms trade or political controversies.

Specific references ought to be linked in most cases to more general terms. In this case, the reference "Bren Gun Contract" was made into a *see* reference to "Arms Trade" and "Liberal Gov't-Scandals." Since neither of these terms occurs more than four times and on each occasion includes a date, this was deemed sufficient. Had the subject of the Bren Gun contract been discussed more often or in greater detail in Diefenbaker's speeches, it might have been better to leave "Bren Gun Contract" as a preferred entry and to add *see also* references under "Arms Trade" and "Liberal Gov't- Scandals."

Large thesauri and subject headings lists try to maintain consistent levels of specificity. Specialist thesauri naturally are more specific about particular subjects than about general topics. An idiosyncratic, "back of the book" index such as that of the Diefenbaker collection must adjust the specificity of its preferred terms to the actual speeches themselves. *See* and *see also* cross-references can alleviate the problem by tying specific to general terms in a continuum. It is here that the guidelines provided by a thesaurus can be most useful, aiding the researcher in the choice of search terms,[22] ensuring that the cross-references are made consistently every time and are not circular, and ensuring that standard as well as idiosyncratic terms are used.[23]

All these problems can arise with manual systems as well as computer systems. They may be made worse in computer systems by a number of factors. The programme designed for the indexing project allowed a field of only thirty-five characters for the subject heading. This automatically limited the use of sub-headings and qualifiers and at

the same time required the use of abbreviations. Since computers sort in a strict and literal alphabetical sequence, certain abbreviations would appear many entries away from the full form of the word. Thus all the general entries under "Commonwealth" had to be entered under CMWLTH, so that longer terms such as "Commonwealth Prime Ministers' Conference" could be abbreviated to "CMWLTH PMs CONF." and sorted in the appropriate position. UNESCO would be separated from UN by UNDERDEVELOPED AREAS and UNEMPLOYMENT with all its modified versions, a space of two whole pages in the print-out. Furthermore, the abbreviations themselves can be ambiguous. CA, initially used for both Canada and Canadian, created CA WEST which could be read as either Canada West or Canadian West, two quite different geographical areas, both of which Diefenbaker might be expected to have mentioned. The task of changing all cases where CA stood for Canadian to CDN was a major one since it involved intellectual decisions that the computer could not be programmed to make. If these problems could have been anticipated in the beginning, a great deal of time would have been saved.

The shortness of the field limited the ability to create more specific headings with the use of modifiers and sub-headings. The separate fields for date, place and occasion serve as modifiers in many cases by restricting the reference to a given period of time. However, there are still places where large numbers of references to unmodified general terms present the researcher with an intimidating task. Perhaps in the end this is the unavoidable result of attempting exhaustively to index rhetoric, which is by nature full of loosely-defined general terms. John Diefenbaker's references to the general subject FREEDOM, which occupy more than a page of the print-out, were exactly that, generalized references to the idea of freedom.

Researchers may want more than access to the subject of the speeches. They may want to examine the relationship of the content to the time and place and to the audience. Knowledge of the history of a given speech or of a series of speeches on a given subject can be obtained from the supplementary files. This information was recorded on the work sheets, but no attempt was made to index it. The names of aides and speech-writers, where they appear, should have been indexed. Perhaps some consideration of the concept of provenance indexing would have enabled access to this information to be given. The addition of provenance information to a content index such as this would have required careful thought and planning, but could have added useful information.

This raises the whole question of what Richard Lytle calls Provenance Method subject access versus Content Information subject access.[24] The first, or finding aid, stage of the Diefenbaker speech project provided access through provenance-related information, that is to say, through date, place and occasion of speech rather than through content or text of the speech. This is provenance access at a more detailed level than archivists normally think of using it, but the same principles apply. The second phase of the project produced a Content Index. Analysis of the way researchers use the end-product in future will provide an opportunity to assess the advantages and disadvantages of both types of access, and to determine whether the use of one type of finding aid is more effective in retrieving the desired information.

From the point of view of the creators of finding aids, the provenance form is less time-consuming and more accurate. It is possible in almost all cases to determine a specific date, location and event for any given speech, enter it in a uniform manner and

use it for straightforward and unequivocal retrieval purposes. Creating a content index type of finding aid is, on the other hand, much more time-consuming. More importantly, content access is *always* subjective and can never be considered complete.

The actual experience of working with the Diefenbaker speech collection has begun to show that both types of access are useful. A researcher looking for a half-remembered quotation or wanting a quick idea of what John Diefenbaker thought about a given subject can use the subject index with a fair chance of success, but can rarely pinpoint the exact quotation. The researcher looking for more general and fundamental statements of policy and attitude, or studying any subject in depth, benefits from a mixture of content indexing and information derived from provenance. Such a researcher will soon realize that it is the provenance-based finding aids which are the most important. It becomes necessary to read widely in the speeches to determine where, when and to whom Diefenbaker made his statements on the subject being searched. It seems clear that while all researchers can begin with the subject index, those wishing to continue their work in depth must have the provenance-derived information. Subject information taken out of context has limited value. The context, discovered from provenance-based information, is essential to any detailed or serious study.

Because the Diefenbaker speech collection is relatively small, the project can serve as a microcosm for larger indexing projects. At present, no repository could even consider indexing large parts of its holdings at the depth of detail of this project, for simple financial reasons. Yet in the next decade computers and optical scanners will have advanced to the stage where such indexing becomes financially possible. It will then become even more necessary to decide how useful detailed subject indexing of large fonds will be. What is clear is that traditional provenance-based access will continue to be essential. Perhaps attempts to combine provenance information with selected subject information will have the most fruitful results.

The creation of a subject index to the Diefenbaker speeches has been a valuable exercise for the staff of the Diefenbaker Centre. However, only further use by researchers will reveal whether the subject index is worth the time and effort it cost. What is already clear is that provenance access is essential to serious study of the speeches. Subject or content access shortens the searching time for each enquiry, but provenance-based finding aids are essential to put the speeches in their context.

In looking back at the Diefenbaker speech indexing project, it becomes clear that many of the problems arose from lack of sufficient planning to guarantee sustained funding and organizational support. Any institution undertaking a similar project should be aware of this pitfall. Meanwhile, the intellectual problems involved in deciding what a particular index is intended to do, and relating indexing theory to specific applications, remain.

Notes

1 John G. Diefenbaker, *One Canada* (Toronto, 1975-77).
2 Peter Stursberg, *Diefenbaker: Leadership Gained 1956-62* (Toronto, 1975) and *Diefenbaker: Leadership Lost 1963-67* (Toronto, 1976).
3 H. Basil Robinson, *Diefenbaker's World; A Populist in Foreign Affairs* (Toronto, 1989), p. 29.
4 Sean O'Sullivan with Rod McQueen, *Both My Houses* (Toronto, 1986), p. 61.
5 The term "key words" was not used in the librarians' strictly defined sense.
6 F. Wilfrid Lancaster, *Information Retrieval Systems: Characteristics, Testing and Evaluation* (New York, 1968), p. 188.

7 Ibid., p. 192.
8 P.F. Booth, "Thesauri — Their Uses for Indexers," *The Indexer* 15:3 (April 1987), p. 143; G.N. Knight, *Indexing; The Art of, A Guide to the Indexing of Books and Periodicals* (London and Boston, 1979), p. 91; Alan Gilchrist, "The Role of Thesauri in Mechanized Systems" in L.H. Harrod ed. *Indexers and Indexing-a Selection of Articles Published in the Indexer* (New York and London, 1978), pp. 381-9.
9 Knight, *loc. cit.*
10 Lancaster, p. 189.
11 The Diefenbaker Centre is indebted to the Computer Services Department of the University of Saskatchewan for its work and advice on the computer programme.
12 Russell Lewis Martin, "Archival Indexing: Problems and Issues," (University of British Columbia thesis, 1987).
13 James D. Anderson and Gary Radford, "Back of the Book Indexing with Nested Phrases Indexing System (NEPHIS)," *The Indexer* 16:2 (October 1988), p. 79.
14 Martin, *op. cit.*, pp. 65-8 recommends the use of the PRECIS system.
15 Brenda Hall, "A Computer Generated Index Technique" in Harrod, *op. cit.*, pp. 353-61.
16 Constance McCarthy, "The Reliability Factor in Subject Access," *College and Research Libraries* 47:1 (January 1986), pp. 48-56.
17 T.D. Regehr, "Do We Need New and Improved Archivists?" *Archivaria* 3 (Winter 1976-1977), p. 117.
18 Marion M. Torchia, "Two Experiments in Automatic Indexing: The Presidential Papers and the Papers of the Continental Congress," *The American Archivist* 39:4 (October 1976), p. 438.
19 Mary Jo Pugh, "The Illusion of Omniscience: Subject Access and the Reference Archivist," *The American Archivist* 45:1 (Winter 1982), p. 43.
20 Lancaster, *op. cit.*, pp. 194-5.
21 Pugh, *loc. cit.*
22 Gilchrist, *op. cit.*, p. 383.
23 Booth, *op. cit.*, p. 143.
24 For a discussion of this topic see Duncan Chalmers, "Computer Indexing in Public Record Office" *Journal of Society of Archivists* 6:7 (April 1981), pp. 399-413; W. Theodore Durr, "Some Thoughts and Designs about Archives and Automation," *The American Archivist* 47:3 (Summer 1984), pp. 271-289; Martin, op. cit. and Richard H. Lytle, "Intellectual Access to Archives I. Provenance and Content Indexing Methods of Subject Retrieval," *The American Archivist* 43:1 (Winter 1980), pp. 64-75 and "II. Report of an Experiment Comparing Provenance and Content Indexing Methods of Subject Retrieval," *The American Archivist* 43:2 (Spring 1980), pp. 191-206.