

Indexing and the Classification Research Group

Ia C. McIlwaine

The article outlines the origins and achievements of the Classification Research Group over the past 50 years in the development of indexing principles and techniques from chain indexing and facet analysis to relational indexing and PRECIS.

Introduction

The year 2002 was not only the year of the Queen's Golden Jubilee; it was also the 50-year anniversary of the foundation of the Classification Research Group (CRG). It therefore seems appropriate to devote a little space to a survey of the contribution of this group of people, the majority of whom were practising librarians, and frequently also teachers who, in each of these roles, made a considerable contribution to developments in the field of indexing. Not least among its members was the late Mary Piggott, president of the Society of Indexers from 1995 to 1997. The CRG exercised a great influence on the rationalization of indexing principles and formulated a grammar that retains its validity today. The Group's work deserves recognition and at the present time its fundamental role is often overlooked. In 2000 a *Festschrift* was published for Pauline Atherton Cochrane, herself a member of the CRG, in which several of the contributors lamented the fact that the Group's pioneering work had, since the mid-1980s, been largely ignored (Wheeler, 2000: 77). In order to redress the balance slightly, a reminder of some of their achievements may not go amiss.

The CRG was set up in 1952 as the result of an expression of concern emanating from the Royal Society's 1948 Conference examining the information needs of the post-war environment in the UK (Royal Society, 1948a: 190). That conference made a number of recommendations relating to indexing, with particular emphasis on mechanized indexing. Also among its recommendations was the following:

12.1 *Committee on classification.* The Royal Society is invited to consider the constitution, through existing organizations or otherwise, of a standing committee on subject indexing in science. It is suggested that alphabetical arrangements should come within the terms of reference. (Royal Society, 1948b: 14)

Thus the CRG came into existence after consultation with the Library Association about the best way to implement the recommendation. Several of its founding members had served in the forces during the second world war, and had met S. R. Ranganathan while serving in India. These encounters were to have a profound influence on the development of classification and indexing in the UK. One result was the first English publication that can properly be described as embodying the basics of all that the CRG was to

achieve in the coming years. This was Palmer and Wells's *Fundamentals of library classification* (1951), a textbook published the year before the CRG came into being, at a time when UK schools of librarianship were beginning to proliferate (there was only one prior to the second world war). Classification and cataloguing were being taught for the examinations necessary to achieve professional status, in many instances by the very people who formed the CRG membership. This factor is directly relevant to the future development of indexing techniques because many students who never became specialists in this technical branch of the profession assimilated the basic principles and retained that knowledge throughout their working lives. The teaching they received was therefore subconsciously reflected in management decisions made by those who were senior members of the profession.

Chain indexing

One of the authors of this textbook, Jack Wells, was the first editor of the *British National Bibliography (BNB)*, which came into existence in 1950 as part of the same post-war development of information services. This set the pattern for the many post-war national bibliographies that proliferated in the 1950s and early 1960s, when the importance of bibliographic control was beginning to be appreciated. It also provided the pattern for the catalogues that were found in libraries throughout the country for the next 30 years, prior to the introduction of the online catalogue, namely a classified catalogue (most often arranged by the Dewey Decimal Classification) accompanied by an author/title and a subject index. That subject index is noteworthy because it was constructed in the form of a chain index, the indexing system promoted by Ranganathan, and the system that underpinned others developed later, in many instances by CRG members. Ranganathan's five fundamental categories – Personality, Matter, Energy, Space and Time – and the concurrent use of a rigid citation order in which these elements were consistently cited, resulted in a predictable arrangement that greatly assisted both the user and the librarian in retrieving information. Chain procedure remains a clear and straightforward way to create an index, once the indexer has grasped the technique. It is based on

the theoretically very simple procedure whereby an entry is made under the most specific element in a compound subject, and then the indexer proceeds upwards through the higher classes or topics, e.g.

House Sparrows – Passerines – Birds – Zoology – Biology – Pure science
 Sparrows – Passerines – Birds – Zoology – Biology – Pure science
 Passerines – Birds – Zoology – Biology – Pure science
 Birds – Zoology – Biology – Pure science
 Zoology – Biology – Pure science
 Biology – Pure science
 Science

This is a grossly over-simplified demonstration of the technique, ignoring the need to include synonyms, hidden concepts and many other aspects, such as the form of a document, but it demonstrates the underlying principles. Although designed to be applied in conjunction with a classification scheme, it is not essential to use it in such a manner.

It is often thought that book indexing is quite different from other forms of indexing, but this is not strictly the case. Chain procedure works equally well for book indexing, if the indexer relies on chapter and section headings and sensible scanning of the text. The terminology of the author will clearly be the starting point, though again, attention may need to be given to such matters as synonyms. Both in the index to a classified sequence and in a book index the use of *see* references should be avoided as far as possible, and direct and systematic entries are the most helpful, though in book indexing the publisher's policy of cost-cutting may militate against this. Chain indexing continued to be the 'standard' practice both in the majority of general libraries that had classified catalogues as their principal means of retrieval and in the *British National Bibliography* until the 1970s.

British (later Current) Technology Index

The strong links with the national bibliography continued throughout the first 30 years of the CRG's existence. In 1957, the same year as the first FID international study conference on classification for information retrieval, commonly referred to by its location as the 'Dorking Conference' (Dorking Conference, 1957), a special classification scheme for music was created for the new offspring of the *BNB*, the *British Catalogue of Music* (Coates, 1960a). This classification was devised by a man who played a major role in the nation's indexing services for over 30 years, and is still an active CRG member, namely Eric Coates. In the 1950s Coates was the editor of the *British Catalogue of Music*. Like many of his fellow CRG members he too wrote a book that remains a staple text, still in print and recommended for use by students today – *Subject catalogues* (Coates, 1960b).

The year 1963 saw the splitting up of the long-standing Library Association's *Subject Index to Periodicals* into a range of different services, each covering a specific field of knowledge. The *BNB*, still at that time an independent commercial undertaking, took responsibility for the technological journal indexing and began publication of the *British Technology Index* (later the *Current Technology Index*).

Coates was appointed editor, a position he held for the rest of his professional career, and he devised the system of indexing that was used for the publication. This, again, has its roots in Ranganathan's chain procedure, and also derived some of its elements from the special expansions developed for use in the *BNB* to supplement the Dewey Decimal Classification, where parts were found unsatisfactory, either because of their strong American bias (e.g. education) or because they were too out of date, especially the 500s and 600s (*British National Bibliography*, 1963). It was experience with these latter classes that gave Coates an awareness of the special features of these disciplines on which he could draw when working out his indexing system for technology. He was greatly influenced by the early-20th-century indexer Kaiser, who separated the elements of items to be indexed into 'Concretes' and 'Processes' (Kaiser, 1911). Coates likewise thought that it was very difficult to visualize an activity without connecting it to something. He therefore considered that the 'Thing' and the 'Action' formed the starting point and then moved on to extend his concept of a 'Thing' to be qualified by the material from which it is made, and the components that together make it up. This gives the order: Thing – Part – Material – Action.

In this way Coates solved the problems of generating specific entries for complex subjects such as are found in scientific and technical periodical literature. He created a system that provided an excellent solution to the problem of generating specific headings for highly complex technical topics, but it was heavily dependent upon a network of cross-references that users found unhelpful. In 1968 the *Index* was computerized and a method for generating these cross-references was devised. Although initially Coates was of the opinion that it was impossible to accomplish this automatically, a solution was found. It was concluded that it is the initial analysis that is all important; the intellectual input lay in the creation of the original string. This conclusion had later implications for the indexing of the national bibliography.

Facet analysis and citation order

An understanding of categorization and the need to think clearly and consistently seem obvious fundamental requirements for the indexer, whatever the type of index being constructed. Ranganathan's five fundamental categories can be seen in the indexing principles developed later by CRG members. Coates's work in the early 1960s was preceded by that of Vickery, who extended the number of fundamental categories from Ranganathan's original 5 to 13, and set out a clear and very digestible set of principles, based on his own practical experience in his *Faceted classification* (Vickery, 1960). The 'standard' citation order, adopted by Mills for the revision of the Bliss Bibliographic Classification that is still in progress (Bliss Bibliographic Classification, 1977–), and expounded in his textbook, *A modern outline of library classification* (Mills, 1960), continues to be that recommended today. It is used, for example, by the editors of the Dewey Decimal Classification when giving the 'preference order' (Dewey Decimal Classification, 1996) in which to apply the different component

elements of a class number and in deciding in which of two or more potential places to classify a document, or how to build up a compound classmark.

Vickery, Coates and Mills are not the only CRG members to hold honoured places in the development of indexing techniques. The 1950s to early 1960s saw the publication of three major works on indexing, which between them span the retrieval problems of the whole spectrum of knowledge. This was the time when Butterworth were publishing a range of classic reference sources for the professional librarian. The first was Vickery's *Classification and indexing in science* (1958), followed by Foskett's *Classification and indexing in the social sciences* (1963) and finally Langridge's *Classification and indexing in the humanities* (1976). These three works, though designed principally as textbooks, expound many universal principles as well as highlighting the specific problems that the various groups of disciplines present and the solutions that have been adopted. All three authors were (and two still are, Langridge having died in 2000) members of the CRG.

Book indexing

Langridge was one of the few members of the CRG to express a view in print about book indexing, which he referred to as 'the Cinderella of library science in this country' (Langridge, 1961). He felt that book indexing was best accomplished by a professional indexer, and not by the author of the work. He recommended the works of Bliss, Ranganathan and Vickery as basic texts that any book indexer should study and maintained that a knowledge of facet analysis was an invaluable tool for the indexer handling a subject on which he or she is not expert, as must inevitably frequently be the case. He compared the role of the indexer to that of the cataloguer making the contents of a library available to the user and maintained that a knowledge of classification assists the book indexer in 'crystallizing' the kind of knowledge needed for efficiently making an index, in systematizing cross-references and in providing guidance on the handling of compound subjects and clarifying the nature of 'unsought' terms.

Relational indexing

In the days when computers were in an embryonic form, indexing methods that were later to be employed by computer scientists in devising programs for retrieval, such as uniterms, the use of punched cards and 'peek-a-boo' cards and other techniques were being developed. Another CRG member, Jason Farradane, was a moving force in identifying the need for a clear distinction of the different kinds of relationship that exist in literature. He developed a system of relational operators that formed the foundation of his system of 'relational indexing' (Farradane, 1961). He was well aware of the problems of over-reliance on the machine to do the work of retrieval, a lesson that still seems to be necessary for the present generation to learn. A machine cannot inject meaning into a group of juxtaposed terms without an element of conceptual analysis first being programmed into it.

He therefore advocated the use of a unique definition for a concept, free from unnecessary assumptions, together with 'semantic factoring' to establish the context. His favourite example was the rabbit: the rabbit can be uniquely defined as a type of mammal and on to that concept the varying contexts such as 'as a pet', 'for cooking', etc. should be appended. He devised a table of nine different types of relations that can be discerned in literature, including the following:

- self activity (e.g. 'Man – walking', 'Bird – migration')
- equivalence (e.g. 'Molasses – fodder')
- physical properties (e.g. 'Lead – melting point')
- parts (e.g. 'Bicycle – brake')
- reaction (e.g. 'Iron ore – smelting')
- causation (e.g. 'Author – book')

He thought language was a poor vehicle for exact expression and therefore a more precise system needed to be devised. Mechanized retrieval systems frequently assume, albeit without their programmers being aware of it, that the machine can inject meaning into a group of juxtaposed terms, although no conceptual analysis or re-synthesis has been programmed. He considered the use of prepositions in the English language to be a particular pitfall.

Although each CRG member worked independently, and total agreement with one another was far from the norm, the effect of regular discussion and interchange of ideas has been reflected in their work. There is a strong resemblance between the categories and relationships identified by Vickery, Coates and Farradane and this became even more pronounced a decade later in the work of Derek Austin, referred to below.

PRECIS and events leading to its development

In many ways, the 1960s might be seen as the Classification Research Group's 'golden age'. A major milestone in its history was the 1963 conference, which resulted in the publication of a pamphlet entitled *Some problems of a general classification scheme* (Classification Research Group, 1964). This conference was the result of discussions arising from a paper on classification problems given by E. J. Coates to the Reference and Special Libraries Section of the Library Association in 1957. At that conference the Section passed a resolution expressing its dissatisfaction with the state of library classification and asking the Library Association to investigate the possibility of making a new general classification scheme. This coincided with the publication of a NATO report on 'Increasing the effectiveness of Western Science', which made a number of recommendations, including one that a new classification for science and technology would be an important factor in achieving the desired end (Foskett, 1964). Negotiations between the Library Association and NATO's science adviser led to the award of a £5000 grant to study the project and produce a pilot scheme. The CRG was asked by the Library Association to undertake this pilot study and the 1963 conference was the genesis of what became known as the 'NATO classification project'; it was

organized to enable CRG members to formulate definite ideas on how best to create such a classification.

The award of this grant enabled them to appoint a research assistant and two people were successively employed in this capacity, the second of whom was Derek Austin. He moved on from the consideration of entities, which had dominated his predecessor's work, to look at the need to distinguish properties and attributes, and to analyse active and passive systems and examine the interactions between systems and subsystems. He produced a system of operators, a system that was considerably influenced by his CRG fellow members, and this eventually formed the basis of the PRECIS system of indexing (Austin and Dykstra, 1984). The CRG reported on its work at the conclusion of the grant in 1968 and a version of that report was published by the Library Association (Classification Research Group, 1969), together with a shorter summary of the work by Douglas Foskett (1970). These two publications are really the final ones that can be described as the corporate work of the CRG.

The CRG's links with the national bibliography continued into the 1970s when, in 1973, the British Library was created and took over responsibility for the national bibliography. The new environment and the establishment of a well-managed editorial procedure for the Dewey Decimal Classification, together with the implementation of MARC cataloguing and the general desire for standardization, led to the decision in 1971 to abandon the special expansions and the chain index from the beginning of January 1972, and to develop a more detailed indexing system to compensate for the shortcomings of the Dewey system in a British environment. By this time the number of items issued in the bibliography was about 39,000 a year and the Dewey classification was not sufficiently detailed to permit anything approaching specificity. The index was therefore seen as the vital tool to provide access to this massive amount of documentation. So it was decided that a new indexing system should be devised, whereby every document entered in the *BNB* was analysed and summarized in a manner that could be expressed in a series of index terms in which each term could, in turn, become the 'lead term' – in other words, a form of 'rotated index' rather than a chain index. All the terms were used in every entry, and permuted. This contrasted with the chain indexing system whereby a chain extending from the most specific to the most general was created, with one term occurring only once as the lead term. Derek Austin was appointed to develop this new system; he drew on the work that he had undertaken in conjunction with the NATO research and developed the PRECIS indexing system.

Mary Piggott (1988: 207) illustrates the contrast between these two systems very clearly:

The chain index would produce these entries:

Mathematical models – circuits – electrical equipment	621.31920184
Circuits – electrical equipment	621.3192
Electrical equipment	621.319

Whereas using the PRECIS system the same example would look like this:

Mathematical models. Circuits. Electrical equipment	621.31920184
---	--------------

Circuits. Electrical equipment	621.31920184
Mathematical models	
Electrical equipment. Circuits.	621.31920184
Mathematical models	

The term shown in bold type is the lead term, while the remainder of the entry is the qualifier, setting the context, hence the name PREserved Context Index System. This system was totally manipulable by machine. It was first implemented in the 1971–3 cumulation of the *BNB*, which revealed the need for some adjustments and led to the production of a revised version of the system in 1974. It gained considerable recognition and was actually considered as a possible alternative to Library of Congress subject headings (LCSH) when the Library of Congress closed its card catalogue in 1981.

For a system such as this to be successful, a careful analysis, based on a clear set of categories, is essential, a conclusion already reached earlier as the result of Coates's work for the *British Technology Index*. Therefore, once again, the first step is to write out a string of terms that comprehend the content of the item being indexed. A 'lead term' has to be selected, together with a qualifier and display terms, each term being tagged according to its role, following the detailed system devised by Austin, not dissimilar to that of Farradane some years before him. The lead term and its qualifier are displayed on one line, with the display terms on the line below, as in the example given above. The terms are then 'shunted' so that each in turn becomes the 'lead'. The complex syntactical system is set out in detail in the *PRECIS manual* and, although it is both complicated and time-consuming to implement, it is interesting in that being based on firm grammatical principles, it is the only indexing system that has been demonstrated to work effectively in several different languages; it was used for some time in Canada for the bilingual *Canadian Film Index* (Dykstra, 1986). However, the time and expertise required to implement it led the *BNB* to abandon its use in the late 1980s, after Austin had retired. After a brief flirtation with a very basic system called COMPASS, the *BNB* has relied upon LCSH as the verbal approach for subject indexing.

Conclusion

The CRG continues to meet regularly and currently has among its members two of the people working on revising the British Standard for thesaurus construction. The main thrust of its interest, however, has for a number of years been upon revising the Bliss Bibliographic Classification, which provides a testing-ground for many of the theories it developed in its earlier years, but has not really broken any new ground with regard to research in indexing. It is frequently assumed that the lively and diverse minds that contributed to its work in the 1950s and 1960s produced a coherent and agreed set of principles. Nothing could, in reality, be further from the truth. The arguments and fierce discussions that resulted in a sound body of indexing principles consumed many of its meetings, and while they were under way it was difficult to discern a coherent theoretical basis emerging. With hindsight it is clear that there was a developmental pattern originating from principles distilled

by Ranganathan, and developed successively, or at times contemporaneously, by Vickery, Coates, Farradane and ultimately Austin, which can contribute greatly to the solution of many of today's problems in the challenge of indexing, at least in part, the content of the World Wide Web.

References

- Austin, D. and Dykstra, M. (1984) *PRECIS: a manual of concept analysis*. London: British Library.
- Bliss Bibliographic Classification (1977–) *Bliss Bibliographic Classification*, 2nd edn ed. by J. Mills [and others]. London: Butterworth.
- British National Bibliography (1963) *Supplementary classification schedules*. London: Council of the British National Bibliography.
- Classification Research Group (1964) *Some problems of a general classification scheme: report of a conference held in London, June 1963*. London: Library Association.
- Classification Research Group (1969) *Classification and information control*. London: Library Association.
- Coates, E. J. (1960a) *The British Catalogue of Music classification*. London: Council of the British National Bibliography.
- Coates, E. J. (1960b) *Subject catalogues: headings and structure*. London: Library Association.
- Dewey Decimal Classification (1996) *Dewey Decimal Classification and relative index*, Edition 21. Albany, NY: Forest Press.
- Dorking Conference (1957) *Proceedings of the International Study Conference on Classification for Information Retrieval held at . . . Dorking . . . 13–17 May 1957*. London: Aslib.
- Dykstra, M. (1986) *PRECIS: recent applications*. Halifax, Nova Scotia: Dalhousie University, School of Library Services.
- Farradane, J. (1961) Fundamental fallacies and new needs in classification. In *The Sayers memorial volume*, ed. D. J. Foskett and B. I. Palmer. London: Library Association, pp. 120–35.
- Foskett, D. J. (1963) *Classification and indexing in the social sciences*. London: Butterworth (2nd edn 1974).
- Foskett, D. J. (1964) Origins of the conference. In *Some problems of a general classification scheme: report of a conference held in*

- London, June 1963*, Classification Research Group. London: Library Association, p. 12.
- Foskett, D. J. (1970) *Classification for a general index language*. London: Library Association.
- Kaiser, J. (1911) *Systematic indexing*. London: Pitman.
- Langridge, D. (1961) Classification and book indexing. In *The Sayers memorial volume*, ed. D. J. Foskett and B. I. Palmer. London: Library Association, pp. 182–95.
- Langridge, D. (1976) *Classification and indexing in the humanities*. London: Butterworth.
- Mills, J. (1960) *A modern outline of library classification*. London: Chapman & Hall.
- Palmer, B. I. and Wells, J. (1951) *The fundamentals of library classification*. London: Allen & Unwin (Practical library handbooks).
- Piggott, M. (1988) *A topography of cataloguing*. London: Library Association.
- Royal Society (1948a) *The Royal Society Scientific Information Conference, 21 June–2 July 1948: report and papers submitted*. London: Royal Society.
- Royal Society (1948b) *The Royal Society Scientific Information Conference: recommendations adopted by the Conference*. London: Royal Society.
- Vickery, B. C. (1958) *Classification and indexing in science*. London: Butterworth (2nd edn 1959, 3rd edn 1975).
- Vickery, B. C. (1960) *Faceted classification: a guide to construction and use of special schemes*. London: Aslib.
- Wheeler, W. J., ed. (2000) *Saving the time of the library user through subject access innovation: papers in honor of Pauline Atherton Cochrane*. Champaign, IL: Univ. of Illinois.

Ia C. Mcllwaine is Professor Emeritus of Library and Information Studies at University College London. Email: i.mcllwaine@ucl.ac.uk

Baffled indexers

No, no – I'm not going to tell you where to look them up . . . ! All right, then, don't whine – under Epistemology. For heaven's sake! Come on . . . !

Where's Epistemology? How should I know? In the back of the book somewhere. In the index. You want me to write the index for you, as well as everything else? Somebody else is doing that! Some specialist index writer . . . ! No I don't look to see what he's put in his indexNo, he doesn't read the text before he writes the index. He's an index writer, not a manual reader . . . What does he put in his index? The same as any other writer puts into what he writes! Whatever comes into his head! Which in his case I should think is more probably words like 'fountain pen' and 'congratulations', because if I know anything about index writers he's as baffled as you are.

Michael Frayn, 'The manual writer's manual', *The additional Michael Frayn* (Methuen, 2000)

indexing specialists

Our team of experts have many years' experience in undertaking and designing effective indexes.

- *Books, journals, manuals* ●
- *Multi-volume works* ●
- *From 100,000 to 100 lines* ●
- *Unfamiliar subjects* ●
- *Design of style and layout* ●
- *Short deadlines met* ●

Phone +44 (0) 1273 738299 Fax +44 (0) 1273 323309
email richardr@indexing.co.uk
website www.indexing.co.uk

INDEXING SPECIALISTS (UK) LTD

202 Church Road, Hove
East Sussex BN3 2DJ, UK