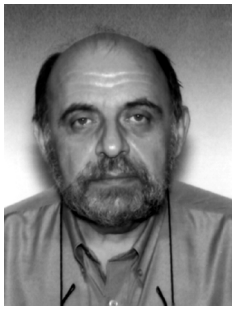
Pre-Coordination and Post-Coordination: Past and Future

Uri Miller*, Ruth Teitelbaum**

*Library of the Wingate Institute for Physical Education and Sports, Netanya, Israel (E-mail: ouri@macam.ac.il)
**Social Sciences Information Center, The Henrietta Szold Institute, Jerusalem, Israel (E-Mail: info-c@szold.org.il)

Dr. Uri Miller is the Senior Indexer at the Library of the Wingate Institute for Physical Education and Sports, Netanya, Israel, and the Thesauri Construction Coordinator at the MOFET Institute – Research, Curriculum & Program Development for Teacher Educators, Tel-Aviv, Israel.



Mrs. Ruth Teitelbaum is the Director of the Social Sciences Information Center at the Henrietta Szold Institute in Jerusalem and teacher at the School of Library, Archive & Information Studies, The Hebrew University of Jerusalem.

Uri Miller, Ruth Teitelbaum. (2002). **Pre-Coordination and Post-Coordination: Past and Future** *Knowledge Organization, 29(2).* 87-93. 31 refs.

ABSTRACT: This article deals with the meaningful processing of information in relation to two systems of information processing: pre-coordination and post-coordination. The different approaches are discussed, with emphasis on the need for a controlled vocabulary in information retrieval. Assigned indexing, which employs a controlled vocabulary, is described in detail. Types of indexing language can be divided into two broad groups – those using pre-coordinated terms and those depending on post-coordination. They represent two different basic approaches in processing and information retrieval. The historical development of these two approaches is described, as well as the two tools that apply to these approaches: thesauri and subject headings.

Introduction

It is only natural that librarianship and information science, similar to any other theoretical foundation of a widespread and important sphere of practical activities, are highly sensitive to the kind of sharp, rapid hi-tech progress and 'technological revolution' that continues to transform our professional environment. Computers and the Internet continue to cause theoretical and practical turmoil even in experts' minds. We professionals find ourselves caught up in a race of sorts, ever hurrying and driven on by the technological onslaught (Miller, 2001).

While these conditions are capable of generating many new decisions and practical 'breaks', they leave too little time to comprehend the theoretical foundations of continuing changes, or, to use military parlance, of 'fortifying the rear'. In addition, there are a number of theoretical principles that, from the outset, appear to be clear and unshakable, that is, *axiomatic*. The concepts of 'pre-coordination' and 'post-coordination', and their interconnections are among these. A striking demonstration of the slump of interest in this problem is evident from a search of these terms in the *Library and Information Science Abstracts* database. Undoubtedly, the high level of their elaboration in the already classical works of Dagobert Soergel (1974),

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -

Antony Charles Foskett (1996), Frederick Wilfrid Lancaster (1986) and others is a very important reason for this established situation. (For interesting historical reviews see: Cleveland & Cleveland, 1990, pp. 60-62; Kilgour, 1997, pp. 340-348).

But nevertheless, now and again, the return to re-analysis of some axiomatic propositions can be very relevant. The need to revisit our foundations is evident from publications that have distorted the real meaning of these concepts. These distortions are represented by two examples written over a decade apart: articles by Ann Schabas (1982) and David Bodoff and Ajit Kambil (1998). Schabas, without ceremony or explanation, tested the Library of Congress Subject Headings (a *pre-coordinated* structure) using *post-coordinated* searching.

Although the recently published investigation by Bodoff and Kambil proposes several new search methods suited to the modern information environment, it nevertheless reveals a certain vagueness of interpretation of some of the basic conceptions underlying subject indexing and searching (both pre- and post-coordinated). First, their definition of "subject queries which return zero hits" as "subject search failures" (Bodoff & Kambil, 1998a p.1254) is *only* admissible in the case of the fixed presence of relevant documents in the information massive, otherwise, "the lack of information is also information". Second, the authors compare pre-coordinated indexing and searching on the ground of *subject headings* (specifically, Library of Congress Subject Headings) to post-coordinated indexing and searching using *full-text free* indexing and searching, where every meaningful word is counted as a keyword. Such contraposition cannot be considered appropriate in the comparison of pre-coordination and post-coordination. Rather, Bodoff and Kambil are comparing the presence or absence of *vocabulary control*. Such comparison can be relevant only to vocabulary-controlled pre-coordinated indexing and searching (based on *subject headings*) as opposed to vocabulary-controlled post-coordinated indexing, and thesaurus-based searching. Ignorance of thesauri as the basic tool of vocabulary-controlled post-coordinated indexing and searching is not only very strange, it also weakens efforts at comparison, due to the inherent impossibility of a different basis for comparison. This methodological confusion allows the authors to reach such questionable and unproven conclusions as the "lower precision" of post-coordination versus pre-coordination and "lack of a theory for selecting post-coordinated keywords" (Bodoff & Kambil, 1998a, p. 1257), the unfounded attack on "narrow keyword terms" (Bodoff & Kambil, 1998a, p. 1257-1258), and so forth.

The widespread evidence of such misapprehensions was vividly illustrated by Catherine Robinson and Janet Knight (1997) in the work especially produced for the elimination of all misunderstandings – the international standard for thesaurus construction ("International Standard ISO-2788. Documentation--Guidelines for the Establishment and Development of Monolingual Thesauri, second edition -- 1986-11-15") and the related discussions.

But perhaps this is useful in prompting us to re-examine our theoretical 'cornerstones'. Without going into details regarding the modern theoretical and methodological principles of thesaurus construction where we can find clear answers to the questions raised (see Miller 1997), we can establish the urgent need for 'stock-taking' of some basic theoretical principles of indexing. In the following discussion, we will review lexical control in the form of pre-coordinate (subject headings) and post-coordinate (thesauri) vocabularies, their application and their philosophical and historical foundations, revealing certain conclusions about contemporary indexing and searching.

The Need for Lexical Control

Once more, the key role played by the retrieval process brings the discussion of the role of lexical control in information processing back on the agenda. Again this raises the 'old' question of the use of free text versus controlled language (lexical control) in information retrieval. While there are many advantages to searching using free text, mostly in terms of availability and rapid information retrieval, the main disadvantages lie in the amount of 'noise' (irrelevant information) accompanying information searching, low precision in search outcomes and a high degree of information loss (low recall). This is the consequence of the lack of control and lack of uniformity inherent in natural language. A free text search means having to cope with a flood of not always relevant information (for a detailed analysis of the central problems of information retrieval, see for instance Blair, 1990, chapter 4). It is perfectly reasonable to search using free text when searching for new terms or concepts which have not yet been expressed in controlled language. A free text search provides a very helpful complement in information extraction, but cannot be the sole search tool (Peters & Kurth, 1991). In some instances, the combined use of a free and a controlled search will produce the best results (Dubois, 1987; Fidel, 1987, 1991, 1992; Rowley, 1990).

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -

The use of free text when searching the Internet is subject to the same constraints. In this case, technology does not entirely solve the 'problem' for us. When searching the Internet we confront a vast array of often irrelevant information because Internet "keywords" are the same elements of free text searching arranged according to the frequency of their occurrence, but not according to all the classical keywords of lexically controlled databases. Neither computers nor the Internet can replace the human mind in processing information meaningfully. Quality data retrieval is impossible without purposeful data processing, particularly since the range of data increases with geometric progression (Weinberg 1995).

Subject Indexing

One form of meaningful data processing is subject indexing. Subject indexing is part of the system known as 'assigned indexing'. Here, a computer program assists the information scientist in processing the data. The program supports the routine work of the indexer with a thesaurus facility, online indexing and indexing quality control. It can look up terms in the thesaurus, proof read and compile indexes.

On the whole, subject indexing uses a controlled indexing language. Indexing languages have two elements: a vocabulary and a syntactical structure. Indexing language, in contrast to natural language, is artificial and is constructed with information processing in mind. The terms used in an artificial language relate to *concepts* as opposed to *words*. Lexical control provides control of synonyms, homographs, alternative spellings and obsolete terminology. Lexical control means that *all material related to a concept will be listed for that concept*, along with pertinent references to other terms.

The syntactical structure consists of rules for terminology construction and for determining the connection between terms. Indexing languages differ in their syntactical structures which become the basis of their designation as either pre-coordination or post-coordination. With pre-coordination, the connection (or coordination) between index terms occurs as the indexer processes the data. With post-coordination, the connection between the index terms occurs during data retrieval. Both approaches employ coordination between the terms, but they differ in timing.

Pre-coordination and post-coordination are two fundamentally different data processing and retrieval philosophies. When pre-coordination is used for indexing, the indexer determines the connection between the index terms in advance. These connections are fixed and rigid. The best examples of this method are subject headings and book subject indexes, which have their own reasons for such methodical approaches (but this is a subject for a separate analysis). With post-coordination, connections between index terms are not decided in advance. The connections are only constructed during the data retrieval process. The thesaurus works on this principle.

The pre-coordination approach used by the indexer during the indexing process only partially resembles the post-coordination indexing process. The stages common to both methods are the essential item analysis and subject selection stages that occur during indexing. An indexer using the pre-coordination method will use a syntactical structure that requires the connection between terms to be decided in advance. The combination order of facets to be predetermined constructs a subject heading that is analogous to a multi-link chain. When using post-coordination, the indexer identifies the appropriate facets, but does not decide on the connection between the index terms. An item will be assigned index terms, while each term retains its independence. Thus, the terms assigned to a particular document are unconnected.

With pre-coordination, there is no clear separation between data processing and its retrieval – they are two stages within a single process. The indexer decides the connections between terms while indexing and creates the subject heading strings. The subject heading string represents the outcome of the search. In other words, the indexer knows the outcome of the search since the outcome is predetermined. When information is extracted using this method, a group of 'ready' subject heading strings is obtained, after which the most appropriate one is selected. This is carried out by browsing through the subject headings which, as we have already noted, are prepared in advance. But such an approach holds the danger of a vocabulary explosion because of the enormous number of subject heading strings required.

With post-coordination, indexing and retrieval are two separate processes, although there is a close relationship between them. In post-coordinate indexing, the indexer tries to express all of the relevant subjects contained in the indexed item according to the facets determined by the indexing policy. The indexer cannot 'predict' the search results since he or she does not make the connection between the terms. Each term stands alone. Each user will construct his or her own search strategy via which the terms are linked, based

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -

on the user's requirements. The indexer using the post-coordination method is required only to carry out the objective indexing. During indexing, the indexer must focus on the indexing process that will influence the retrieval results.

Following this distinction, we can divide databases into the 'traditional', based on the pre-coordinated indexing and retrieval principle (subject headings), and 'non-traditional' databases based on the post-coordinated indexing and retrieval principle (thesauri).

Pre-coordinate terms or complex terms?

Subject headings resemble a chain with multiple links. During pre-coordinate indexing, the indexer connects the various links, that is, he or she expresses pre-coordination via the syntactic structure. Each link in the chain represents a 'subject', for example: "Children and Television". This subject heading has two elements: "Children" and "Television", but is defined as a single 'subject', with two separate distinguishable concepts. The subject heading can have pre-coordination in its lexicon as well; for example: "Physical education for children – Curricula – France". In this heading, we find three different 'subjects' (links in the chain), and four different 'concepts'. The four concepts are: physical education, children, France, curricula.

In contrast, the thesaurus contains descriptors as opposed to subject headings. Descriptors refer to concepts. They may be comprised of one or several words, but will always relate to a single concept. An index term comprising more than one word and which denotes a concept, or 'compound term' (Miller, 1996, 1997; Teitelbaum, 1994), may not be broken down into individual words without there being an essential change of its meanings. Such a break can produce a vast amount of irrelevant information in the process of retrieval, for example: 'high schools', 'latch-key children', 'short working week', 'soap opera' and 'venetian blinds'.

The term 'agricultural schools' is not the equivalent of the sum of 'schools' and 'agriculture'. Similarly, the combination of the terms 'group' and 'discussion' gives us two different terms 'group discussion' and 'discussion group' (Batty, 1989; Milstead, 1984). 'Children and Television' is not a compound term. It is a pre-coordinated term made up of two separable elements. The thesaurus has no place for entries of this kind. The descriptors appear individually, 'children' will be found on its own, as will 'television', and each entry will be accompanied by its own set of references to concepts. The searcher connects them during data retrieval.

It is important to bear the following in mind: a given composite of words may be considered a compound term according to the scope of one thesaurus and a pre-coordinated term according to another. For example, 'sports psychology' is a compound term in sport or psychological thesauri, but it is also a pre-coordinated term in thesauri on more general or different spheres. The conceptual framework in which the term appears is sometimes the decisive factor and the type of interpretation system (pre-coordinated or post-coordinated) in which such a term functions is not important. Information retrieval requirements, which among other things include a high level of precision and which prevent search noise, will determine the use of compound terms in all information retrieval systems (Batty, 1989; Milstead, 1984, Soergel, 1974).

The Development of Pre-coordination and Post-coordination

Pre-coordination and post-coordination are two fundamentally different data processing and retrieval philosophies which were influenced by different general philosophical doctrines. Pre-coordination and post-coordination emerged at different times and represent the moods and fashions then influencing society. When indexing employing subject headings appeared on the American library scene, pragmatism was the prevailing philosophical doctrine in American society (Miller, 1980). The subject heading doctrine posits that the information consumer must trust the indexer, because it is the indexer who determines the connections in advance, and thereby determines the way that information retrieval occurs.

In contrast, post-coordination emerged during the rise of the 'free information' movement. This point of view allowed the consumer a free hand in choosing the information in which he or she was interested and in carrying out combinations of subjects not linked in advance. Here, the indexer's role is to facilitate rapid access to information and ensure its availability to the consumer. The indexer has no control over data retrieval, and all is open before the user. The history of these information processing and retrieval approaches is linked to the different tools employed by each method.

Examples of the tools used for pre-coordination are dictionary catalogues and subject catalogues. The best example of a pre-coordination indexing language is the Library of Congress Subject Headings (LCSH).

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -

Subject headings applied the principles developed by Charles Ammi Cutter who, at the end of the 19th century, laid down the rules for constructing dictionary catalogs, and began to address subject analysis. It is interesting that pre-coordination as a method has a complex character because it was used not only in subject approach, but also in its antipode, a classification approach (as in Cutter's and Dewey's classification systems). LCSH was distributed to and adopted by a large number of libraries throughout the world (see, for instance: Heiner-Freiling, 2000). Its impact has been reinforced by two projects: MARC and the Library of Congress cataloging-in-publication data. These projects used subject headings since they were crucial to the data processing conducted.

With the advent of technological advances, which presented an opportunity for change, subject analysis was transformed into a control-based doctrine. At this point, the crisis worsened since the Library of Congress management was not prepared for a 'technological revolution' and had difficulty in adapting subject headings to the new developments. The well-known researcher Hans Wellisch compared computerized databases that use subject headings to a jet plane rigged with a nineteenth century steam engine. The plane will indeed move, and might even fly, but it will be accident prone, unreliable and its body design will be wasted due to the low speed it moves at (Wellisch, 1972). The Library of Congress Subject Headings have been met with widespread criticism, as have the latest attempts to change the subject headings and make their structure a quasi-thesaurus (Conway, 1993, Dykstra, 1988; Gerhan, 1989; Nuckolls, 1994).

Currently, post-coordination is in the librarianship and information science spotlight. Post-coordination is the outcome of a dialectic process. Although post-coordination was an outgrowth of pre-coordination, it conflicts with its principles and represents an antagonistic approach. Post-coordination emerged with the advent of library computerization and uses computers for information processing and retrieval. Now we see that even the natural adherents of pre-coordination recognize new realities. Thomas Mann emphasized at the "Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web" held at the Library of Congress on November 15-17, 2000: "Neither I nor anyone else is arguing for precoordination *rather than* postcoordination. We need *both* browse displays of precoordinated strings *and* the possibility of postcoordinate combinations of individual elements" (Mann, 2000).

Pre-coordinate information processing and retrieval may be compared to a 'craft', post-coordinate processing to a 'science' and retrieval using this approach to 'an art'. This analogy should not be understood as an attempt to detract from the importance and prestige of pre-coordination; each activity needs to be considered in light of its compatibility with library needs or with the indexed collection. Pre-coordination can be used to provide a high quality, professional standard of indexing. The above analogy relates to the fact that post-coordination information processing activities require associative thinking on the indexer's part, allowing not always visible connections between terms to be exposed. The data retriever using this approach also requires the necessary skills for performing the 'art' of information retrieval.

Subject headings are constructed pragmatically because the subjects are connected to the material contained in a particular library: new subject headings are incorporated only after material relating to a specific concept appears in the library. The thesaurus has no such close link with any concrete database. The presence of *real concepts* is the main reason for building a thesaurus. The existence of *real material* is a secondary issue. Building a thesaurus is like building a model, which is not closely connected with the existence of real material in the database. The thesaurus is a lexical-semantic model of a conceptual reality.

Conclusions

Several conclusions emerge from the discussion on lexical control and developments in the area of subject indexing:

a. Post-coordination is the most sophisticated method of processing data by subject, since it does not impose decisions made by the indexer on the user, and allows flexible data retrieval consistent with diverse and changing data requirements. Post-coordination works efficiently with the employment of syntactical devices that help to increase precision. The quality of the thesaurus as a lexical control tool is affected by numerous factors, among these being: correct choice of terms, building terms according to accepted rules and a consistent, precise method of working. These rules need thesauri with a high degree of order and structure, so that indexers can search for appropriate terms without much effort.
b. At this stage, there is no substitute for lexical control, which offers the most accurate means of data

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -

processing and retrieval. It is worthwhile investing in developing lexical control mechanisms. Obtaining a reasonable amount of highly accurate information justifies the expenditure.

References

Batty, D. (1989). Thesaurus Construction and Maintenance: a Survival Kit. *Database, 12*.13- 20.

Blair, D.C. (1990). *Language and Representation in Information Retrieval*. Amsterdam: Elsevier Science Publishers.

Bodoff, D., & Kambil, A. (1998a). Partial coordination I. The best of pre-coordination and post-coordination. *Journal of the American Society for Information Science, 49 (14).* 1254-1269.

Bodoff, D., & Kambil, A. (1998b). Partial coordination II. A preliminary evaluation and failure analysis. *Journal of the American Society for Information Science, 49 (14).* 1270-1282.

Cleveland, D.B., & Cleveland, A.D. (1990). *Introduction to Indexing and Abstracting*. 2nd ed. Englewood, CO: Libraries Unlimited.

Conway, M.O. (1993). Characteristics of Subject Headings in the Library of Congress BOOKSM Database. *Library Resources & Technical Services, 37*. 47-58.

Dubois, C.P.R. (1987). Free Text vs. Controlled Vocabulary: a Reassessment. *Online Review, 11.* 243-253.

Dykstra, M. (1988). LC Subject Headings Disguised as a Thesaurus. *Library Journal, 113*. 42-46.

Fidel, R. (1987). Controlled Vocabulary and Free-Text Searching: Searchers' Selection of Search Keys. In C. Chen (Ed.), *Proceedings of the 50th ASIS Annual Meeting* (pp. 71-73). Medford, NJ: Information Today.

Fidel, R. (1991). Searchers' Selection of Search Keys 2: Controlled Vocabulary or Free-Text Searching. *Journal of the American Society for Information Science, 42.* 501-514.

Fidel, R. (1992). Who Needs Controlled Vocabulary? *Special Libraries, 83.* 1-9.

Foskett, A.C. (1996). *The Subject Approach to Information* 5th ed. London: Library Association.

Gerhan, D.R. (1989). OCSH in vivo: Subject Searching Performance and Strategy in the OPAC Era. *Journal of Academic Librarianship, 15.* 83-89.

Heiner-Freiling, M. (2000). Survey on Subject Heading Languages used in National Libraries and Bibliographies. *Cataloging and Classification Quarterly*, 29 (1/2), 189-198.

Kilgour, F.G. (1997). Origins of coordinate searching. *Journal of the American Society for Information Science, 48.* 340-348.

Lancaster, F. W. (1986). *Vocabulary Control for Information Retrieval* 2nd ed. Arlington, VA.: Information Resources Press.

Mann, T. Is Precoordination Unnecessary in LCSH? Are Web Sites more Important to Catalog than Books? A Reference Librarian's Thoughts on the Future of Bibliographic control. Available in: http://lcweb.loc.gov/catdir/bibcontrol/mann_paper.html

Miller, U. (1980). *Methodological Challenges of Subject Analysis – A Comparative Analysis of the Principles of Constructing a Subject Heading Language in the US and USSR*. Moscow: Nauka. (In Russian).

Miller, U. (1996). The Sport Database: Some Comments. *Online & CD-ROM Review, 19.* 67-74.

Miller, U. (1997). Thesaurus Construction: Problems and Their Roots. *Information Processing and Management, 33*. 481-493.

Miller, U. (2001) Thesaurus construction and new information environment. In A. Kent (Ed.), *Encyclopedia of Library and Information Science*. vol. 68, suppl. 31, (pp. 319-345). NY: Marcel Dekker.

Milstead, J.R. (1984). *Subject Access Systems: Alternatives in Design*. Orlando, FL Academic Press.

Nuckolls, K.A. (1994). Subject Access to Diversity Materials: The Library of Congress Subject Heading Shortfall. *Reference Librarian, 45-46.* 241-251.

Peters, T.A. & Kurth, M. (1991) Controlled and uncontrolled vocabulary subject searching in an academic library online catalog. *Information Technology and Libraries, 10.* 201-211.

Robinson, C., & Knight, J. (1997). Contemporary Recordkeeping: The Records Management Thesaurus – Response. (Article in response to paper delivered at the 1997 Records Management Association of Australia's National Convention). (Electronic article). Available in: http://www.records.nsw.gov.au/publicsector/rk/aaa/response.htm.

Rowley, J.E. (1990). A comparison between free language and controlled language indexing and searching. *Information Services and Use, 10.* 147-155.

Schabas, A.H. (1982). Postcoordinate retrieval: a comparison of two indexing languages. *Journal of the American Society for Information Science, 33 (1).* 32-37.

Soergel, D. (1974). Indexing Languages & Thesauri: Construction & Maintenance. Los Angeles, CA.: Melville Publishing.

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -

Teitelbaum, R. (1994). The thesaurus in information retrieval systems: Structure and components. *Yad LaKoreh*, *27*. 48-55. (In Hebrew)

Weinberg, B.H. (1995). Why postcoordination fails the searcher. *The Indexer, 19.* 155-159.

Wellisch, H. (1972). Subject retrieval in the Seventies, methods, problems, prospects. In H. Wellisch, & T.D. Wilson (Eds.), *Subject Retrieval in the Seventies: New Directions* (pp. 14-15). Westport, CT: Greenwood Press.

https://doi.org/10.5771/0943-7444-2002-2-87 - am 13.01.2026, 10:31:35. https://www.inlibra.com/de/agb - Open Access -