IMPROVING ACCESS TO MUSIC

A REPORT OF THE
MLA MUSIC THESARUS PROJECT WORKING GROUP

Improving access to music materials has long been a major concern of persons charged with the organization, interpretation, and retrieval of those materials. While the advent of computer technology has provided more effective methods for carrying out these activities, it has also necessitated greater and more efficient bibliographic control. In addition, the tremendous increase in the production of books and articles on music during the past twenty years has created a concomitant need for improved, more thorough indexing, which in turn calls for a controlled vocabulary that is more logically structured and more easily manipulated.

The establishment of any controlled vocabulary, or thesaurus, requires the "vocabulary normalization" that is its main distinction from natural language.1 Such language control has for years been an integral part of the development of online data bases in other fields, particularly in the sciences and social sciences. Medline, for example, the online version of Index Medicus, was constructed to facilitate document retrieval in medical literature. ERIC and PsycINFO offer indexing services with well developed thesauri for education and psychology, disciplines more closely related to the humanities.

Historically the humanities have lagged behind other fields in the "online revolution." Through such recent endeavors as the Art and Architecture Thesaurus Project (AAT), however, this gap is narrowing. In the field of music the Répertoire international de littérature musicale (RILM) has provided much-needed electronic access to music literature.2 RILM is also accompanied by a well constructed thesaurus of descriptors to be used in retrieving abstracts through its online data base. No current list of terms offers comprehensive help to those engaged in providing information services to music researchers, since neither RILM nor Music Index is applicable to the access of music itself.

---


LIBRARY OF CONGRESS SUBJECT HEADINGS

For many years librarians and users alike have expressed dissatisfaction with the Library of Congress Subject Headings (LCSH). The LCSH system has been criticized for the haphazard development of its syntheticonetic structure, for inconsistencies in its system of subject headings, and for its practice of inverting some terms while using others in direct order. A survey conducted by the Art Libraries Society (ARLIS) revealed that a majority of the art librarians polled felt that subject headings and subdivisions in LCSH were often superfluous and ineffective.

The unwieldiness of LCSH continues to present special problems for music librarians and researchers, despite seemingly endless revisions and discussions of LCSH’s content. Designed specifically for a card file that can offer only a minimal number of access points, headings in LCSH often hinder effective subject retrieval online. For example, a user should be able to locate Schubert’s “Der Hirt auf dem Felsen” by searching for any combination of the following: high voice, piano, and clarinet. Under LCSH, the heading assigned is the generic “Songs (high voice) with instrumental ensemble.” Since online systems are able to search on a virtually unlimited number of subject entries, structured subject headings should be broken down into a number of concise descriptors, which may be searched efficiently in any desired combination.

Although LCSH is capable of treating works with multiple topics (e.g., anthologies that include works of different genres), another of its fundamental weaknesses is its failure to allow for the description of works with multiple facets. For example, a work of Christmas music for guitar, under current practice, receives the subject headings “Christmas music” and “Guitar music”; here, a multielement work is treated as a multitopic work. There is no a priori relationship between the terms “Guitar music” and “Christmas music,” and LCSH fails to bring out the true relationship between the two facets of this work. A system that supports both pre-coordinate indexing (in which terms are combined by the indexer) and post-coordinate searching (in which users combine their own desired descriptors from a list of single terms) would more clearly define such multielement works and provide for more accurate retrieval of music materials.

The time has come to redesign and eventually replace LCSH, but a task of such magnitude cannot be accomplished single-handedly. Pauline A. Cochrane, who has studied this problem, has proposed that the

responsibility be divided among a number of groups, each concentrating on a particular subject area.\(^5\)

**MARC FORMATS**

The coded 04x fields in the MARC format may offer a solution to some of the problems raised by LCSH. Since the inception of the MARC music format, librarians have looked forward to the opportunity of conducting Boolean searches on data in the 04x fields, but bibliographic utilities and local online systems have not established comprehensive indexes to these fields and are unlikely to do so in the near future, for good reasons. First, the design of the 04x fields is flawed for the purposes of Boolean searching. The following example makes this clear. A user who wants to find recordings of piccolo concertos might logically search for all the recordings that have a piccolo code in field 048 and the concerto code in field 008 (the "fixed field") or in field 047. This search will result both in relevant hits (desirable retrievals) and in false drops (that is, it will fail to retrieve some items that the searcher would want to find). If a single recording contains both a violin concerto and a suite for piccolo, oboe, and piano, the search would retrieve the piccolo code in an 048 and the concerto code in the 047, even though the recording does not contain a piccolo concerto. There is no way to link 04x codes to a particular title in an analyzed collection, a limitation that is integral to the structure of the MARC Music Format and that probably cannot now be corrected.\(^6\)

Another reason the 04x fields are unlikely to be indexed is that the need for coded fields has declined. When the MARC formats were created, computers were relatively slow, memory was limited, and CPU time was expensive. Coded data seemed economical because it took up little memory and could be searched efficiently. Now that online searching is rapidly replacing the creation of offline products as the primary function of bibliographic utilities and computers are cheaper and faster, coded fields may be viewed from a different perspective. There is a general trend away from coded information in favor of natural vocabulary, whether controlled or uncontrolled, because natural language is easier for both indexers and users.

---


NATURAL LANGUAGE AND CONTROLLED VOCABULARY

Two broad techniques for indexing and retrieval in online systems have been developed over the years, one using controlled vocabulary, the other, natural language. LCSH is an example of a controlled vocabulary, in which standardized descriptors are assigned to an item. In natural-language indexing, any words may be used to describe the item; the words are indexed freely, with no standardization. Each of these techniques offers some advantages over the other, and each presents some obstacles to easy use.

Natural-language indexing is inexpensive, since it often can be completely automated. No skilled personnel are required to develop a vocabulary or to index materials. There are never human indexing errors, either subjective or typographical. New terms can be added to the vocabulary immediately. Finally, searching is simplified because users are not required to become familiar with a complex controlled vocabulary.

The main disadvantage to natural-language retrieval lies in the greater intellectual burden that is placed upon the searcher. All relevant words, and variants of those words, must be covered in the search. Alternative search terms and synonyms are often unknown to the searcher and may be overlooked. Natural-language retrieval of printed music and recordings, the majority of which bear generic titles, poses staggering problems.

Using a controlled vocabulary, on the other hand, solves many of the semantic problems of natural-language indexing and identifies related concepts through a system of cross-references. It guides the user in constructing meaningful searches. Because of the vocabulary's precision, fewer words need to be indexed. If the controlled vocabulary is well constructed and properly applied, the number of desired hits resulting from a search should be high.

There are also disadvantages to using a controlled vocabulary. Creating and maintaining the vocabulary requires the work of a professional staff and other editorial personnel at substantial costs. The vocabulary must be continually monitored for gaps and for terms that have become outdated, such as the LCSH heading "Moving-picture music," which was only recently changed.

In short, controlled vocabulary systems impose their major costs and effort at the time of input, natural-language systems at the time of output. Some information specialists have argued that the capabilities of free-text searching will eliminate the need for controlled vocabularies. Several studies have shown, however, that natural-language searching is not cost-effective, that it places a tremendous burden on the searcher,
and that it creates too many false drops. According to Carol Mandel and Judith Herschman, "comparisons of free-text searching with controlled vocabulary searching . . . have been applied to a variety of databases and systems and invariably lead to the same conclusions: a combination of both is best, with the optimal mix dependent upon the specific features of the database, the system, and the user’s requirements."\(^7\)

**RECOMMENDATIONS**

Limitations in the MARC music format and in online applications of the Library of Congress Subject Headings have demonstrated clearly the need for a standardized vocabulary of music and music literature terms. Such a thesaurus would benefit book and periodical indexers, cataloguers, reference librarians, bibliographers, and most of all, the public. It would cover all musical genres, including Western art, popular, and folk music, as well as non-Western idioms and cross-disciplinary fields.

A primary concern in devising any new controlled-language system should be to address those areas in which conventional indexing and access are weakest in meeting the needs of specialized users. For example, the field of ethnomusicology uses a vocabulary relatively distinct from that of other fields of music. It is also an area that historically has been ignored by those in the musicological mainstream. Pioneering work to rectify this neglect was undertaken by Judith Kaufman and described in two MLA Technical Reports.\(^8\) Kaufman’s studies have demonstrated that different musical genres require different facets of information. Folk music, for example, most often requires geographic and cultural access; the preferred access point for jazz, country, and Latin American music is medium of performance; whereas subject retrieval for rock, rhythm and blues, and other popular music is most commonly by genre. Since many of the Library of Congress’s practices do not transfer very successfully to non-classical idioms, “the needs of the many users who require access to jazz, folk, country and national music . . . are not being filled.”\(^9\)

The Music Thesaurus Project Working Group has concluded that any research undertaken in the creation of a thesaurus of descriptors for music should be directed toward better access to music in all its various manifestations and to information on music in books, articles, disser-


tations, and the like. The thesaurus should be constructed according to accepted standards; it should be capable of accommodating different indexing grammars (e.g., PRECIS or Thesaurofacet, discussed below); it should support both pre- and post-coordinate use; and it should be compatible with LCSH.

Construction might follow the format used by the AAT, which employs the National Library of Medicine's thesaurus, MeSH, as its model. Terms in MeSH are arranged in two sections: an alphabetical list, which gives scope and history notes, applicable subdivisions, and non-preferred terms; and a hierarchical list, which arranges the terms in "tree" structures of broader and narrower terms. Such a two-part arrangement obviates the need for indicating broader and narrower terms in the alphabetical list and creates a more consistent and logical syndetic structure.

Other indexing systems that should be investigated are capable, unlike LCSH, of emphasizing the multielemental aspects of works. PRECIS (Preserved Context Indexing System), for example, creates strings of terms that preserve the context of each facet of a work. Thus Christmas music for guitar receives two subject headings, or strings of separate terms: "Christmas music. Guitar" and "Guitar music. Christmas." Although the PRECIS strings in this case describe the work more accurately than does LCSH, other examples can result in strings that are wordy or cumbersome (imagine a similar set of strings for a Christmas cantata for soprano, flute, guitar, and continuo). In "LCSH and PRECIS in Music: A Comparison," Paula Beversdorf Gabbard presents the results of her study of the strengths and weaknesses of LCSH, and of PRECIS as it is applied by the British Library. Gabbard concludes that while both systems have inherent weaknesses, a composite system incorporating the strengths of each would be "an improvement over each system alone, especially in an online catalog with the capability of single and random word order searches." In a similar study, J. Bradford Young compared printed music retrieval in PRECIS and LCSH and arrived at conclusions similar to Gabbard's: multielement works are more easily and accurately accessed through PRECIS entries than through Library of Congress Subject Headings.

Another subject retrieval system with which a music thesaurus should be compatible and that should be investigated is Thesaurofacet. Originally constructed by Jean Aitchison for use with the English Electric Facet Classification System, Thesaurofacet combines (as its name sug-

gests) a faceted classification schedule with a thesaurus, in two complementary sections. The thesaurus serves as an index to the classed portion of the system. It is constructed in a hierarchical arrangement of broader and narrower terms that are displayed in arrays other than those chosen in the hierarchy for the display of the index terms. Similar terms are given in the thesaurus for concepts related in ways that are not hierarchical, for example, thing/part, thing/property, thing/process, thing/attribute, and thing/application. The two portions of the system are linked by alphabetic classification symbols. An advantage of Thesaurofacet lies in its ability to display concept interrelationships that might not be brought out in a simple hierarchical scheme. Thesaurofacet is constructed so that terms may be combined either by the indexer or by the user (that is, pre- or post-coordinately). It is useful for searching in free-text and natural-language systems because it can easily indicate synonymous terms and can display the interrelationships between precombined concepts and subject fields. The hierarchical arrangement provided by Thesaurofacet is especially well suited to complex forms that are commonly found in music.

The development of a thesaurus for a discipline as broad and complex as music requires much study and planning. Such prescribed standards as the UNISIST thesaurus guidelines or the American National Standards Institute's Guidelines for Thesaurus Structure, Construction and Use (ANSI 239) can be recommended for use in planning any type of thesaurus. ANSI defines "thesaurus" as a compilation of words and phrases showing synonymous, hierarchical, and other relationships and dependencies, the function of which is to provide a standardized vocabulary for information and retrieval systems. ANSI standards include the following:

Identification of source. The project should, as much as possible, draw on existing controlled vocabularies. Preliminary research should be made into the relationships between current lists (LCSH, Music Index, and RILM, as well as the British Catalogue of Music and the British National Bibliography).

Preparation of a draft thesaurus, to be tested in an indexing and retrieval situation. Such a draft thesaurus might take the form of a prototype involving a section of the whole project—for example, non-Western music.

Thesaurus maintenance, requiring continuing interaction with a body of documents and searchers. It should be recognized that any thesaurus

project requires continuous maintenance and revision in order to accommodate changes and new terms. A permanent staff, often under the aegis of a governing board, is commonly needed to administer additions and revisions of the thesaurus.

The Thesaurus Project Working Group proposes that any music thesaurus project be carried out in distinct phases. A preliminary study of existing controlled vocabularies and indexing systems should come first. During this phase, the project directors should contact heads of other, similar projects, seek sponsorship and endorsement from relevant organizations, and request funding for the rest of the project. The study of existing thesauri and indexing systems should result in a method for constructing a prototype thesaurus.

During the second phase, project directors should construct a model or prototype thesaurus. Such a prototype might be built from some subset of music or music literature, such as organ music, Latin American music, or music education. The terms for the prototype could be gathered and verified by a thesaurus staff working in conjunction with subject specialists who would serve as consultants for the project.

In the final phase, the project directors will construct the complete music thesaurus, based on the model developed in phase 2, and establish a permanent maintenance center and staff to carry out the continuing revisions and updating that a thesaurus entails.

The Music Thesaurus Project Working Group has seen a tremendous need, as well as great potential, for a thesaurus of terms related to music, both for online retrieval and for more traditional methods of indexing. That the establishment of any major thesaurus project is an immense undertaking, requiring great resources, both human and financial, is obvious. Such a project should not be planned without the full cooperation of various relevant organizations—for example, the American Musicological Society, the American Library Association, and the American Society of Information Science, in addition to MLA. Such cooperation would necessarily require challenging strategies in planning and coordination, but it would result in a thesaurus that would benefit a large number of music users.