The tasks of creation, management, and maintenance of multilingual subject authority files require significant resources, and are therefore best carried out in cooperation. Four national libraries in Europe (the Swiss National Library, la Bibliotheque de France, Die Deutsche Bibliothek, and the British Library) have worked together on a feasibility study which looked into linking existing subject authority files in three different subject heading languages (SHLs) to offer multilingual subject access to their files. The aim of the study was to establish equivalents (or best matches) among headings in French, German, and English in comparing both headings from selected subject areas and the indexing of publications. Its four key elements were: (1) to establish a methodology for the selection and linking of headings; (2) to link headings and analyze the results in the selected subject areas of sports and theater; (3) to see the practical applications of these linked headings by indexing 40 titles from the field of theater using each language being studied and comparing the results; and (4) to compare the indexing of titles in other subject fields outside the linking study itself for which indexing was available in each of the SHLs. This paper describes the work carried out so far and explains plans for the future. (Contains 11 endnotes.) (MES)
The need for co-operation in creating and maintaining multilingual subject authority files

Genevieve Clavel-Merrin
Swiss National Library
Bern, Switzerland

Abstract

The tasks of creation, management and maintenance of multilingual subject authority files require significant resources, and are therefore best carried out in co-operation. Four national libraries in Europe (the Swiss National Library, la Bibliothèque de France, Die Deutsche Bibliothek, the British Library) have worked together on a feasibility study into linking existing subject authority files in three different languages to offer multilingual subject access to their files.

Paper

Introduction

At a workshop sponsored by the IFLA Section of Cataloguing, Istanbul, Turkey, August 24, 1995 entitled "Multi-script, multilingual, multi-character issues for the online environment", I outlined the discussions at the Swiss National Library concerning a format for multilingual authorities and underlined the fact that the need to offer multilingual access to bibliographic databases affects both institutions in multilingual countries such as Switzerland and those in what might normally considered to be monolingual countries.

The reasons for this are well-known: many of our institutions, national or university, collect material in several languages and make their catalogues available over the Internet to users and scholars around the world. Searching these databases should be possible not only via interfaces
The need for co-operation in creating a conference programme and proceedings in the appropriate languages but also ideally using the user's language to enter search topics (or names of authors, corporate bodies, countries etc.). Currently of course, the majority of subject searches in databases are only possible using the subject entries in the language of the country: in Bibliotheque nationale de France for example, each document, independently of the language in which it has been written, is indexed using a French-language subject heading language. Thus, in order to search by subject headings for documents written in English or German, held in the Bibliotheque nationale de France, the researcher from abroad has to master the French language. The same problems occur to a lesser extent when searching certain author names (especially for transliterated names), corporate bodies or uniform titles, but in this presentation we are going to concentrate on subject access.

As stated in 1995, in theory, the indexer should be able to analyse a document and assign headings in his/her native language, while the user should be able to search in his/her native language. The language of the document itself should have no influence on the language of the subject heading language used for indexing nor on the language used for searching. However, as indicated at the time, we had not chosen a subject heading language for a variety of reasons. The cost of creating and maintaining a multilingual list would have demanded far more resources than the SNL had available, even if we were to pool resources with other institutions in the country. Ideally, we wished to share a list with other libraries in Europe or elsewhere in the world, but such a list did not exist. In the longer term however, we said, only by working on an international level could such multilingual lists be created and maintained.

In the time since those considerations, such international co-operation has begun, and in this paper I want to explain why it was considered necessary by the partners, describe the work carried out so far, and explain the plans for the future.

CENL and CoBRA+

In 1997, The Conference of European National Librarians (CENL) asked CoBRA+ to consider the problem of multilingual subject access to bibliographic databases. The CENL is an independent association of chief executives of those national libraries which belong to member states of the Council of Europe. The conference currently consists of 39 members (1998). In 1998 it was decided to establish the CENL as a private foundation. CoBRA+ (Computerised Bibliographic Record Actions) was a concerted action involving national libraries and other bibliographic agencies across Europe. It was funded by the European Commission from June 1996 until the beginning of January 1999, and it carried on the work initiated by the earlier CoBRA programme in the early 1990s. The programme involved a series of research projects and pilot studies aimed at fostering co-operation between national initiatives in the provision of bibliographic services. Although EC funding has now expired, the partners of CoBRA+ have agreed that the CoBRA Forum should continue to meet and report to CENL.

CoBRA+ is led by a Forum, the members of which are drawn from eight national libraries:

- The British Library (BL), UK (Brian Lang, Chairman)
- Die Deutsche Bibliothek (DDB), Germany (Elisabeth Niggeman)
- Bibliotheque nationale de France (BnF), France (Alix Chevallier)
- Koninklijke Bibliotheek (KB), The Netherlands (Wim van Drimmelen and Johan Steenbakkers)
- Biblioteca Nacional de Lisboa (BNP), Portugal (Fernanda Campos)
- Helsingin Yliopiston Kirjasto (HUL), Finland (Esko Hakli)
- Swiss National Library (SNL), Switzerland (Jean-Frederic Jauslin)
- ICCU, Italy (Giovanna Merola)
- Narodna in Univerzitetna Knjiznica (NUK), Slovenia (Vilenka Jakac-Bizjak)(observer))

In 1997, a working group under the broad remit of CoBRA+ Task Group A was convened to discuss multilingual issues affecting national libraries. The group produced a concept which
The need for co-operation in creating multilingual subject access to materials was presented to the CENL during its meeting in Warsaw on October 2-3, 1997. The CENL agreed that a pilot study on multilingual subject access in French, German and English should be carried out by the group until the beginning of 1999. The information on the study is taken from the subject studies and final report produced by the group. (At the time of writing, these are not yet publicly available, but by August they should be posted on the Gabriel WWW site).

Background to the approach adopted

The problems of creating multilingual thesauri have been widely discussed, though usually confined to a restricted subject field, and a standard exists (ISO 5964) which proposes three approaches to their construction:

- the establishment of a new multilingual vocabulary, without direct reference to the terms or structures of an existing thesaurus;
- the translation of an existing monolingual thesaurus;
- the reconciliation and merging of existing thesauri in two or more working languages.

Looking at these approaches in the context of the study, the following points were taken into consideration:

- libraries covering the different linguistic areas concerned by the study (French, German, English) have already invested considerable time and effort in the creation and maintenance of subject headings languages, which have been used extensively to provide subject access to millions of documents. Creating a new vocabulary, multilingual or otherwise, would be unrealistic and uneconomical;
- the translation and maintenance of a subject heading language into one or more languages and its adoption by other institutions would also require those institutions to 'abandon' their past investments in their own subject heading language and impede access to those documents already indexed;
- it is therefore in the interests of libraries to co-operate in this field and to investigate ways to offer multilingual access to their collections without having to abandon or translate their own subject headings.

The pilot study

As a result, the group moved towards the third approach, in a feasibility study on how to offer multilingual subject access using three different subject heading languages (SHL): RAMEAU, SWD/RSWK and LCSH, by establishing links between the headings in each language. (A similar approach has also been adopted in two major terminology projects, the creation of equivalents between the Art and Architecture Thesaurus and other controlled vocabularies in the same field, and in the Unified Medical Language System (UMLS).) It should be emphasised though that the approach adopted by the group does not correspond exactly to the guidelines in ISO 5964, but that the solutions proposed by the standard have been studied and used in part in the establishment of the group's linking methodology.

The aim of the study was to establish equivalents (or best matches) between RAMEAU, SWD/RSWK and LCSH in comparing both headings from selected subject areas and indexing of publications. Its four key elements were:

- to establish a methodology for the selection and linking of headings;
- to link headings and analyse the results in the selected subject areas of Sports and Theater;
- to see the practical applications of these linked headings by indexing 40 titles from the field of Theater using each SHL being studied and then comparing the results;
- to compare the indexing of titles in other subject fields outside the linking study itself.
for which indexing was available in each of the SHLs (in all 21 titles were found).

It was clear from the start that the linking approach would have consequences for future
cooperation, the most important being that should it be possible to establish equivalents
among the majority of the headings, and that each partner uses these links, fully standardised
subject access would only be guaranteed in the source language of the catalogue in question -
and this access would be different for each institution.

In other words, each document in an institution would not be indexed three times using
SWD/RWSK, RAMEAU and LCSH. Instead, an institution would index documents using its
own SHL, e.g. for DDB this would be SWD/RWSK, then offer an access based on the other
language equivalents from the remaining SHLs i.e. DDB would offer English language and
French language access based on the headings found in LCSH and in RAMEAU. However,
the juxtaposition of those headings in the other languages would not constitute indexing
according to the rules of the other schemes.

While this approach does not correspond to an 'ideal' multilingual access as described in the
introduction, at the same time it has the potential to be more than a simple 'lead-in' term and is
certainly an improvement on the current situation, in which subject access to these databases is
monolingual.

Subject heading languages

It is important to understand something of the nature of subject indexing and also something
about the structure and contents of the three subject heading languages being studied in order
to appreciate the problems faced when attempting to use them as a multilingual access tool, as
well as to be able to judge the benefits of such an approach.

The three subject heading languages contain a list of headings (a controlled vocabulary of
concepts which may be expressed in one or more words), a semantic structure (defined in the
authority records for each heading) and a set of rules by which these headings may be
combined in a string to describe the contents of a document (the syntax of the language).

e.g. Acteurs
Formation
France

are headings in RAMEAU, each with a record in the RAMEAU SHL, while

Acteurs - Formation - France

is a string, constructed according to the RAMEAU syntax rules, that could be assigned to a
document: it has no authority record in the SHL.

The first tasks carried out by the group were situated at the level of the headings, and aimed to
see to what extent links could be made among the three SHLs. Following the creation of
multilingual lists of headings, the group then extended the study to take into account the use of
these headings within indexing strings, and the consequences for access.

Methodology

The selection criteria for these two fields to be studied (Sports and Theater) were based on a
pragmatic approach: Sports was considered to be a relatively uncomplicated field in the sense
that it has no major cultural or national bias, while theatre was considered to be a wider and
more complex semantic field, enabling the group to test the linking methodology defined and
refined in the subject area of Sports. The final methodology (which could be applied to other
The need for co-operation in comprehensive monolingual lists of headings are provided by the libraries responsible for the management of a SHL, in a suitable (homogeneity and size) subject field defined by some agreed main headings, from which all relevant headings are selected according to a hierarchical approach (NT) completed by an alphabetical one, and respecting precise definitions of scope limits (included / excluded types of headings). The three alphabetical lists of selected headings are ready to be compared, after a last consistency check (additions / removals).

From these three lists, a trilingual list containing all the selected headings is established, first by comparing authority headings (terminological level) to find exact one-to-one equivalents, then by working at the authority records level by adding one-to-two equivalents using semantic methods (especially "upward" for references). However this kind of NT/BT equivalence is only applied if it is clearly indicated by an authority record. A further improvement is sought at the indexing strings level to find some heading-to-string equivalents.

However, one-to-one linking is always preferable and therefore some new headings may have to be created in an SHL to correspond to an equivalent in another.

Summary of results for headings in Sports and Theater

- Comparison of terms in the area of Sports:
  - 86% of headings match across all three lists (trilingual equivalents)
  - 8% of headings match in two lists (bilingual equivalents)
  - 6% of headings remain monolingual

- Comparison of terms in the area of Theater:
  - 60% of headings match across all three lists (trilingual equivalents)
  - 18% of headings match in two lists (bilingual equivalents)
  - 22% of headings remain monolingual.

These initial results are very promising, especially given the following:

- it is clear that some headings are assigned more frequently than others e.g. the term 'théâtre' is assigned more often than 'Nô'. When the 50 most frequently used RAMEAU headings in the BN-OPALE catalogue from the field of Theater were identified, it was found that over 86% of these 50 headings matched across all three lists. It is therefore important in a future strategy to give priority to the identification and treatment of the most frequently used headings;

- each SHL has been constructed based on the concept of 'literary warrant' i.e. a heading is created only when it is needed to index a document received by the institution. Thus some perfectly valid heading(s) may be missing from a list simply because no documents on the topic(s) in question have yet been published in the country concerned or acquired by the institution. Each SHL is therefore representative of the culture of the country for which it has been developed, and there are headings that reflect those cultural differences (i.e. the topics studied in publications), that do not have an equivalent or a translation in another language, and are therefore not present in another subject heading language. e.g. "Kasperletheater" exists only in the SWD, and "Living newspaper" does not exist in RAMEAU. It is significant that one of the results of this study has been the enrichment of the different vocabularies through the comparison of their contents (see methodology). This enrichment will also help increase the rate of matches in the future;
However,

- the initial comparison was centred on one-to-one matches between headings. There were cases in which this kind of match was not possible i.e. the existence of multiple matches in one of the lists, or cases in which a heading in one list matched a heading + subdivision in another list. As indicated above in the explanation of the methodology applied, the group defined procedures to cover the questions of multiple matches, and also matches between a heading on one side and a string on the other e.g.

  Enfants acteurs = Child actors = Schauspieler + Kind (string composed of two authorities)

Since at present none of the partners systematically create strings in their authority lists (for reasons of list management and also general philosophy), if the heading / string matches are removed, the results are less encouraging. Either a solution must be found in a structure outside the different systems, or a lower level of equivalence must be accepted, along with the consequences for searching. It is important to consider linking possibilities not only at the theoretical level (headings) but also at the application level (strings) since the ultimate goal is not just to link headings, but to allow users to access documents, typically via subject strings. Indeed, given the nature of heading creation in each SHL (literary warrant) the headings cannot be divorced from their application in indexing. Throughout the work on linking headings in Sports and Theater, it was always possible, and sometimes necessary to refer back to real indexing in bibliographic records in order to check on the meaning of headings, to resolve terminological ambiguities and to confirm some equivalents.

**Comparative analysis of titles indexed using LCSH, RAMEAU and SWD/RSWK**

After the creation of links between headings, the consequences at the level of indexing strings were studied. Initially, it was planned to compare the subject indexing of about 500 recent publications with international imprints. However, given the difficulty of locating publications indexed in all three languages, the group modified the approach:

A comparative analysis was made of the indexing of those titles (in any field) in which original indexing was available in all three SHLs. In addition, a more focused comparison of titles in the subject area of Theater was carried out. This was conducted by asking each library to supply 10 examples of indexing from their current cataloguing in this area which the other libraries could then index using their own system.

There were some inherent difficulties in both approaches. In the first approach, since the group had only carried out a detailed selection and comparison of headings in the fields of Theater and Sports, there were no general multilingual lists available for the comparison. It was therefore necessary to start by attempting to create links 'across the board' for the titles found, without being able to apply the methodology defined in the other subject areas. As a result, some false links may have been made, and others not discovered.

The study did confirm the following points: although the SHLs being studied adhere to the same basic principles, the number of headings and subdivisions which may be combined and the complexity of the strings which may result varies from language to language. The number of strings that may be applied to a document also varies according to the different rules applied. We may say in general that the level of co-ordination and application of rules are closer between RAMEAU and LCSH than between SWD/RSWK and the other two schemes, but not to the extent that they are interchangeable. In addition the number of strings applied to a document may also vary as a result of indexer subjectivity and rule interpretation. Finally, the indexing resulting from the application of the rules within different linguistic and cultural contexts can give rise to variations in the strings assigned to documents.
The need for co-operation in creativity: Programme and Proceedings

http://www.ifla.org/IV/ifla65/papers/080-155e.htm

The group was aware of these different structures and syntaxes at the start of the study, hence the focus of the comparative study, which was to see if equivalents were possible at the SHL level, and then check if those links were reflected usefully in the comparative indexing of actual documents. It was estimated that the overlap in authorities used to index the documents under consideration varied from between 29% to 56%.

The second case was concerned only with titles taken from the field of Theater, and so should have presented fewer difficulties. However, the paucity of documents available in the field resulted in the selection of documents which, although in the field of Theater, were outside the scope of the study in several respects: the trilingual selection and comparison of headings was restricted to common nouns, and several of the documents selected required the attribution of personal names, time subdivisions and form subdivisions, none of which were available in a trilingual form. If we restrict the analysis to the 27 titles without these 'extraneous' elements, we see that in 23 titles there is at least one equivalence established at the authority level. Trilingual access through linked subject headings is however available in many cases.

In general, the indexing comparisons showed that while there may be convergence in the headings used (i.e. at the authority level) there is considerably less overlap in the combination of those headings to create strings. This is in part a reflection of indexer subjectivity, but once again underlines the differences in approach defined in the rules and structures of the SHLs themselves. This is an important point for the application of a trilingual list in user searching, and will be studied further in the prototype phase.

Prototype

The feasibility study aimed at testing the intellectual links among SHLs, and did not plan to investigate the technical aspects surrounding linking and subsequent access to linked headings, or the indexed documents themselves. Nevertheless, management aspect were also discussed.

It was felt initially that if links were maintained in each system, the need for a further layer of management might be eliminated. However, if each partner were to maintain links within each SHL, there would be a duplication of effort, and probably complications in trying to agree each link with each partner. In addition, each extra partner introduced would increase the load of creating and maintaining links. Finally, problems were foreseen in the case of 'one to many' links among SHLs, and subsequent creation of new headings and their corresponding links.

The group saw greater advantages in the creation of an external system or 'metathesaurus' which would contain for each equivalent a record (perhaps using an international identifier), giving the identifier of the heading in the different authority files, and maintaining a link with these. (In the diagram below, Mt = metathesaurus record). A metathesaurus record would be created for each entry, whether it be monolingual (i.e. a heading without an equivalent in another language), bilingual (a heading with an equivalent in one other SHL), trilingual etc.:

Diagram is unavailable. Please contact author.

This metathesaurus would enable a flexible management structure: specialists in each institution could have the right to establish equivalents (and the resulting links) at the metathesaurus level, subject to review. The experience of both BnF and DDB in the co-operative management of RAMEAU and SWD in their own countries will be beneficial here.

Questions remain concerning the management of see-references and related headings. If they are to take place at a local level, they will remain monolingual, and this will have an impact on searching: the 'transparent' parallel approach will be restricted to preferred terms. If the authority structure is to be duplicated at the metathesaurus level, this has implications for duplicated work and future maintenance.
In addition, the impact of searching headings in the metathesaurus as opposed to searching through strings in the databases needs to be evaluated. It may be that only a Boolean search on headings in databases would be possible, but this needs to be tested in a prototype. It should be noted that none of the automated systems used by the partners currently supports multilingual subject indexing, multilingual thesaurus management or multilingual searching. If 'transparent' multilingual string searching is required, the following points will also need to be studied: presentation of headings in subject indexes (separate indexes, or inter-filed); record display (which headings should be displayed - one language or all); default languages; user interface (visible / invisible switch from one language to another). These questions are especially relevant to SNL since ideally the library wishes to offer users as transparent an approach as possible to multilingual searching.

Whichever data management option is adopted, questions of data format and exchange will need to be clarified. A UNIMARC authority structure for the metathesaurus could be envisaged, since it is currently the only authority structure known to the group to accommodate multilingual links. Conversion programmes to and from UNIMARC from the partner authority formats would be necessary since each partner currently uses a different authority format: BnF uses INTERMARC, DDB uses MAB, BL uses UKMARC and SNL uses USMARC. A prototype will need to be tested from the point of view of multilingual searching by the user for bibliographic records within one or more systems, either starting from the metathesaurus and targeting one or more systems, or extending a search in one system through to another in another language via the metathesaurus.

Since users carry out subject searches not only with common nouns, but also with geographic and personal names, the treatment of these and their possible integration needs to be taken into consideration, in co-operation with other work in the same field (e.g. AUTHOR).

In addition, indexer use of the metathesaurus should be investigated to see how it may be used as an aid to indexing, and as an aid to SHL enrichment.

A further point to be considered could be the use of a classification scheme as a mechanism to organise the linking structure.

It is clear from the initial tests that linking equivalents is a labour intensive area and that it will take time to cover all fields. It would be useful to identify the 'most-used' headings, which if linked would cover a high percentage of items, indexed, and work on such sets of headings. The structure of the RAMEAU list and the information system of the Bibliotheque nationale de France enable such information to be established. It would also be productive to study to what degree selection of headings in a field could be automated e.g. using the hierarchical relationships (narrower terms), or by using classification numbers already assigned to headings, according to each SHL classification scheme (for example DDC numbers in the LCSH file used at the BL).

The future

At their meeting of March 18th 1999, and during subsequent discussions at Leipzig on March 24th 1999, the directors of the BL, BnF, DDB and SNL agreed to support the creation of a prototype, according to the following timetable:

- April 1st - May 31st 1999: project description and timetable
- June 1st - June 20th 1999: discussion with partner libraries
- June 21st 1999: validation by partner libraries (CoBRA+ Forum)
- July 1st - December 31st 1999: draw up detailed specifications CENL submits specifications to potential partners. Evaluate replies, and select prototype partner.
- January 1st - June 30th 2000: development and testing of prototype
The benefits of co-operation

Although the feasibility study raises many questions that can only be solved by taking the project a stage further i.e. the creation and testing of a prototype, the work carried out so far has shown the benefits of co-operation on several levels:

- the establishment and use of equivalents in other languages will facilitate access for the user;
- it will also have advantages for the indexer: the existence of indexing in another language for a document is an aid to indexing for all partners;
- the creation of links between headings in different languages will increase the ability to make use of work already carried out by a sister institution;
- furthermore, the comparison of the lists concerning Sports and Theater resulted in the creation of some headings missing in the lists, and has thus enabled each institution to enrich and improve its own vocabulary, which in turn enriches access for the user and the indexer;
- in the medium term, the comparison of the lists will improve and enrich each list, and encourage better convergence;
- although this study was confined to three SHLs, its potential goes well beyond the partner institutions. The three SHLs are applied not only in the partner libraries but are in extensive use in other libraries in France, Germany and Great Britain, as well as in many other English-speaking, French-speaking and German-speaking countries; linking the three SHLs could provide access to millions of documents. In addition, the methodology studied and the approach used was designed from the start with the goal to extend them to other subject heading languages if they proved valid in the restricted context being studied.

We hope that in 2000 one of the members of the working group will be able to report on the outcome of the second stage of this co-operative approach to the creation and maintenance of multilingual subject authority files.

This presentation draws on the results provided by the working group, to whom special thanks is given for their co-operation and efforts.

Endnotes:


2. For more information on CENL see: http://www.konbib.nl/gabriel/en/cenl-general.html

3. For more information on CoBRA see: http://www.bl.uk/information/cobra.html

4. Members of the pilot study group: Genevieve Clavel (SNL), Magda Heiner-Freiling (DDB), Martin Kunz (DDB) Patrice Landry (SNL), Andrew MacEwan (BL), Max Naudi (BnF), Pat Oddy (BL) (until April 1998), Angélique Saget (BnF), Ross Bourne (BL), CoBRA+ Secretary as rapporteur until April 1998, then Peter Dale (BL)

5. http://www.bl.uk/gabriel/

6. International Standards Organisation (ISO). Guidelines for the establishment and


9. as was determined in the study 'Principles underlying subject heading languages (SHLs) / IFLA Section on classification and indexing, Working group on principles underlying subject heading languages. 2nd Final Draft, August 1998 (In press).

10. A proposal to code other language headings in USMARC authority records was discussed by the MARBI group at the 1998 ALA meeting in Washington, but no decision was taken, and the topic does not seem to be a priority for the US.


Copyright © 1995-1999
International Federation of Library Associations and Institutions
www.ifla.org
NOTICE

REPRODUCTION BASIS

This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").