



World Library and Information Congress: 70th IFLA General Conference and Council

22-27 August 2004
Buenos Aires, Argentina

Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>

Code Number: 038-E
Meeting: 145. Classification and Indexing
Simultaneous Interpretation: -

Subject Indexing between International Standards and Local Contexts: The Italian Case

Antonia Ida Fontana

(Director of the National Central Library of Florence, Italy)

Abstract

The Biblioteca nazionale centrale di Firenze (National Library of Florence) has started a renewal of the Soggettario, the subject heading list used by the «Bibliografia Nazionale Italiana» and by most of the Italian libraries. A Feasibility Study on the Revision of the Soggettario, in which the most important languages and tools for indexing at present in use are reviewed and analysed, shows the features of this new tool. Its basic inspiring principles are the respect of the international standards on the vocabulary control and the analysis of syntactic roles in order to build subject strings. The system integrates various features: Rules, stating the principles and guidelines of the indexing language, a Thesaurus and an Application manual.

This tool is run, implemented and updated on line. It is a flexible, modular, interoperational system, that can be used in different documentary contexts (such as archives and museums) and to tap informational resources of a different nature. It is conceived as a tool to be used on a national basis for general libraries, but can be adopted also by local and specialised institutions. The setting out of the language in two different and complementary components, vocabulary and syntax, can foster projects for international cooperation, helping the building of multilingual subject authority archives. The Thesaurus envisages connections with the corresponding DDC numbers, thus creating an articulated joint between the national language setting of the Thesaurus itself and the international, global dimension of the DDC.

The project

The National Central Library of Florence has always been a lab where to test new methods and procedures, carrying on the institutional task of reviewing and updating catalographic tools. Playing this rôle, the Library is taking care of the renewal of the

Soggettario per i cataloghi delle biblioteche italiane, the tool for subject indexing used in Italy since 1956.

Our project is also a peculiar approach to the theme of this workshop, that is implementation and adaptation of global tools to local needs. A real adaptation of a global tool for semantic access to our local needs, made by the Italian national bibliography (BNI), which operates within the National Central Library of Florence, is the Italian translation of *Decimal Dewey Classification*, 21. We hope we'll get resources to translate and adapt also the 22nd edition. On the contrary, the new *Soggettario* will be neither an adaptation of a tool used somewhere else, nor the Italian translation of a foreign thesaurus.

Before starting the *Feasibility study* on the renewal of our tool for subject indexing, we asked ourselves some basic questions:

- What was the spirit of the project supposed to be? To maintain the present system, with some updatings? Or abandon it and chose a new one? To renew it deeply, maintaining the effective components?
- Which foreign experiences had we refer to? How did we want to relate to the tradition of *Library of Congress Subject headings*, largely spread over the world? Were we supposed to translate and adapt the *Subject headings*? And which other indexing languages had we to refer to, also thinking of future projects of international cooperation?

We made the choice of a “renewal” in order to reach a more advanced model of indexing. To this purpose, we studied other indexing systems, looking for methods and principles establishing common denominators, even if operating in different linguistic environments, according to IFLA inquiry on *Principles underlying subject heading languages*¹.

So how is the new *Soggettario* involved in the global-local dynamics?

The main reasons are the following:

1. it uses Italian language, but it is based on international ISO standards; we can envisage a relationship with other indexing systems, especially those in Romance languages;
2. it was born to be a national tool, dedicated to general libraries, but also for local and specialized environments;
3. thanks to its flexibility and modularity, it is possible to use it in various documentary contexts (such as archives, museums, media libraries, documentation centres);
4. it makes the compilation of authority files easier, even multilingual ones;
5. the vocabulary offers links between indexing terms and corresponding DDC numbers so connecting the national linguistic elements of the *Soggettario* to an international and global tool as the DDC.

The global perspective of our new system is represented by the reference to the international standards, the search for principles common to the different systems; the attention to the needs of cooperation, its ability to contribute to multilingual authority files and its connection to DDC. Adhering to the linguistic use and its development, being used

¹ *Principles Underlying Subject Heading Languages (SHLs)*, ed. by Maria Ines Lopes and Julianne Beall, München, Saur, 1999.

in different documentary institutions (such as libraries, archives, museums) and for general and special collections as well as for information sources of different nature, it satisfies local needs too.

Our goals were:

- not to make a simple “restyling”, but an innovative work;
- to preserve the best components of our catalographic tradition;
- to envisage an indexing language based on international standards;
- to establish clear rules and coherent structures, in order to follow the evolution of the language;
- to create a flexible and modular tool to be used both in general and specialized information environments, able to represent subjects of different levels of complexity;
- to get an indexing language easy to be taught, supported by manuals and guides that make it easier to learn;
- to have an indexing language to be shared inside an integrated and cooperative system.

Before describing the new system, it can be useful to explain the main features of the tool we are going to renew, our old *Soggettario*:

- it collected the headings of the subject catalogue of the National Central Library of Florence formulated before the fifties of the past century. Its editors followed the example of the *Library of Congress Subject headings*. It was published in 1956;
- it provides a pre-coordinated indexing, based on main headings and subdivisions. Terms are listed in alphabetical order and semantically connected with others, by cross references such as “see” and “see also”. The use of the cross references is not as consistent as requested, several years later, by ISO 2788;
- the *Soggettario* is essentially a structured list of main heading and subdivisions but it also allows a synthetic use of its terms. In fact, usually main headings and subdivisions are already connected, but sometimes a cataloguer can extract terms from different lists and put them together to build new subject strings;
- the *Soggettario* doesn't provide rules about syntactic or semantic relationships among terms: it supplies only examples of suitable combinations of terms. Citation order is established according to the semantic category the terms is part of. Specificity and exhaustivity are not always respected;
- throughout the years, the *Soggettario*'s terminology was increased by *Liste di aggiornamento*: lists of new terms used by the *Bibliografia nazionale italiana* (Italian national bibliography). In the last updating, BNI partly followed the ISO 2788 principles: in fact, new terms are connected to their corresponding broader terms (BT), already included in the *Soggettario* or in its previous updatings.

Over the years, the structure and the terminology of the 1956 *Soggettario* got more and more outdated. Different libraries (general, special, university) noticed its shortcomings.

The *Feasibility study* was started in 2000. We had to profit from our delay, using both national and international indexing experiences of the last few years.

Since the beginning of the nineties, Italian Library Association (AIB) has been promoting and supporting the Gruppo di ricerca sull'indicizzazione per soggetto (Research group on subject indexing - GRIS), which in 1996 published a *Guida all'indicizzazione per soggetto*, revised in 2001². The *Guida* explains:

- indexing general principles;
- rules for content analysis, referring to UNI ISO 5963;
- the building of subject strings;
- vocabulary control, referring to UNI ISO 2788.

In the last decade, GRIS has been promoting studies and researches in this field, undertaking an intensive training of Italian librarians.

The analytic/synthetic model of language, described in the GRIS guide, has its roots in the modern theory of subject indexing, developed especially through the researches of the Classification Research Group (CRG), the faceted classifications, PRECIS and the international standards on thesauri.

The National Central Library, Florence (BNCF) committed its Feasibility study to a group of experts involved in the GRIS experience. Anna Lucarelli and Marta Ricci of the BNCF worked with them.³

The Feasibility study

The Feasibility study lasted two years and was finished in 2002. It allowed:

- 1) how to reach feasible targets and economically sustainable solutions;
- 2) to prepare a middle-term plan, in order to proceed in steps;
- 3) to reach a preliminary project including:
 - the features of the new indexing language
 - a budget stating the requested human and technological resources.

We pointed out the main steps of the Study several times⁴.

The Feasibility study produced different *Proposte sulle caratteristiche del linguaggio* (Proposals for the features of the new language). These proposals were discussed by

² ASSOCIAZIONE ITALIANA BIBLIOTECHE, GRIS GRUPPO DI RICERCA SULL'INDICIZZAZIONE PER SOGGETTO, *Guida all'indicizzazione per soggetto*. Roma: AIB, 1996, <http://www.aib.it/aib/gris/gris.htm>.

³ Members of the team, lead by Luigi Crocetti, were Alberto Cheti, Daniele Danesi, Massimo Rolle, Stefano Tartaglia; they were assisted by Carlo Revelli and Diego Maltese for scientific consulence. Maria Lucia Di Geso, of the Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, attended the meetings.

⁴ For more details, see: A. LUCARELLI, *La Biblioteca nazionale centrale di Firenze e il rinnovamento dell'indicizzazione per soggetto*, «Bollettino AIB», 41 (2001) n. 3, p. 307-313; M. ROLLE, *Per un nuovo modello di linguaggio documentario*, «Bollettino AIB», 41 (2001) n. 3, p. 315-326; L. BULTRINI, *Subject indexing and Classification in Italy, 2001-2002*, «Newsletter. Classification and Indexing Section. Division of Bibliografic Control. IFLA», November 2002, nr. 26, p. 16-17, <http://www.ifla.org/VII/s29/pubs/ci26.pdf>; ALBERTO CHETI, *Il punto sul GRIS e gli sviluppi attuali*, «Bibliotime», 6 (2003), n.1, <http://www.spbo.unibo.it/bibliotime/num-vi-1/cheti.htm>; A. LUCARELLI, *La revisione del Soggettario*, «Bibliotime», 6 (2003), n.1, <http://www.spbo.unibo.it/bibliotime/num-vi-1/lucarell.htm>.

international experts during a seminar held in Florence in April 2001.⁵ The final results were exposed to the Italian professional world in June 2002.⁶

For a detailed presentation of the project, see the technical document on the Web site of BNCF:

The methodological choices

We can sum up the new *Soggettario*'s basic requirements:

1. to distinguish semantic and syntactic relationships;
2. to comply with ISO standards on the vocabulary structure and the control of semantic relationships;
3. to use syntactic role analysis for the building of subject strings;
4. an indexing oriented to a singular specific coextensive string;
5. to retain *Soggettario*'s terminology, on the basis of new strict norms;
6. to reorganize and implement the terms of *Soggettario*, according to the three-relationship model of thesauri.

As a consequence, we wanted:

- to overpass the approach "main heading-subdivision" focus on a thesaurus of equally important terms, usable in any combination;
- to give certain rules for citation order to build subject strings;
- to integrate the traditional criteria of specificity with the criteria of coextensivity.

The project contains both traditional and innovative elements. Continuity can be identified with the respect of some basic principles of *Soggettario* (pre-coordination and retaining of terms already used in the past); innovation consists of establishing unequivocal rules and distinguishing terminological and syntactic components.

The structures of the new *Soggettario*

The new system has been conceived as a complex of distinct subsystems. The system's architecture is made of four components, which are connected and interact thanks to a proper software:

1. the rules on vocabulary and syntax;
2. an apparatus of instructions, consisting of syntactic notes connected to some terms and a manual for cataloguers;

⁵ Mary Dykstra, Lynn El Hoshi from the Library of Congress, Max Naudi from the Bibliothèque nationale de France, Marion Sedelmayer from Bayerische Staatsbibliothek, Maria Ines Cordeiro from the Fundação Calouste Gulbenkian, Lisboa, Patrice Landry from the Schweizerische Landesbibliothek were present in 2001 at the Seminar held in Florence, as well as many Italian specialists.

⁶ Accounts and documents on the Feasibility Study were published in *Per un nuovo Soggettario. Studio di fattibilità sul rinnovamento del Soggettario per i cataloghi delle biblioteche italiane*, Milano, Editrice Bibliografica, 2002.

3. the thesaurus available on-line;
4. the subject string file

Norms and Thesaurus constitute the effective documentary language. Therefore the new language is based on norms and structures. Syntax and terminology are different domains.⁷ On the other hand, syntactical notes and manual entries should make the application of general rules to both individual terms and full categories of terms, easier. Subject strings expressed by BNI will offer a pattern, a reliable example of the new language. The National bibliographic agency will select and validate subject strings implementing the Library National Service (SBN) data bases.

Figures 1, 2, 3 show the system components, the structure of vocabulary entries and the format of the syntactic note.

The structure of the new Soggettario and the methodological choices on which it is based define the flexibility and modularity features on which depends the interoperability of the system. First of all, the combinatorial character of the syntax allows to combine simple terms to build “ad hoc strings” so that the language can express more or less complex subjects. Therefore, it’s possible to satisfy needs of different kinds of libraries.

Second, the distinction between the semantic and the syntactic components of the system, allows, when needed, to separate them. In fact one can use the semantic structure (the Thesaurus), with all its general and specific levels, independently from the syntactic one. Thanks to this property the system can be used in different documentary environments, leaving the uniformity of the vocabulary untouched.

In the end, the separation between vocabulary and subject strings gives us more opportunities in the international projects of building equivalents in different languages.

⁷ The mentioned ISO 2788 standard provides a clear distinction between semantic and syntactic relationships in controlled languages, confirmed also by a IFLA document, *Principles Underlying Subject Heading Languages (SHLs)*, cit.

Figure 1
Components of the New Subject Heading List system

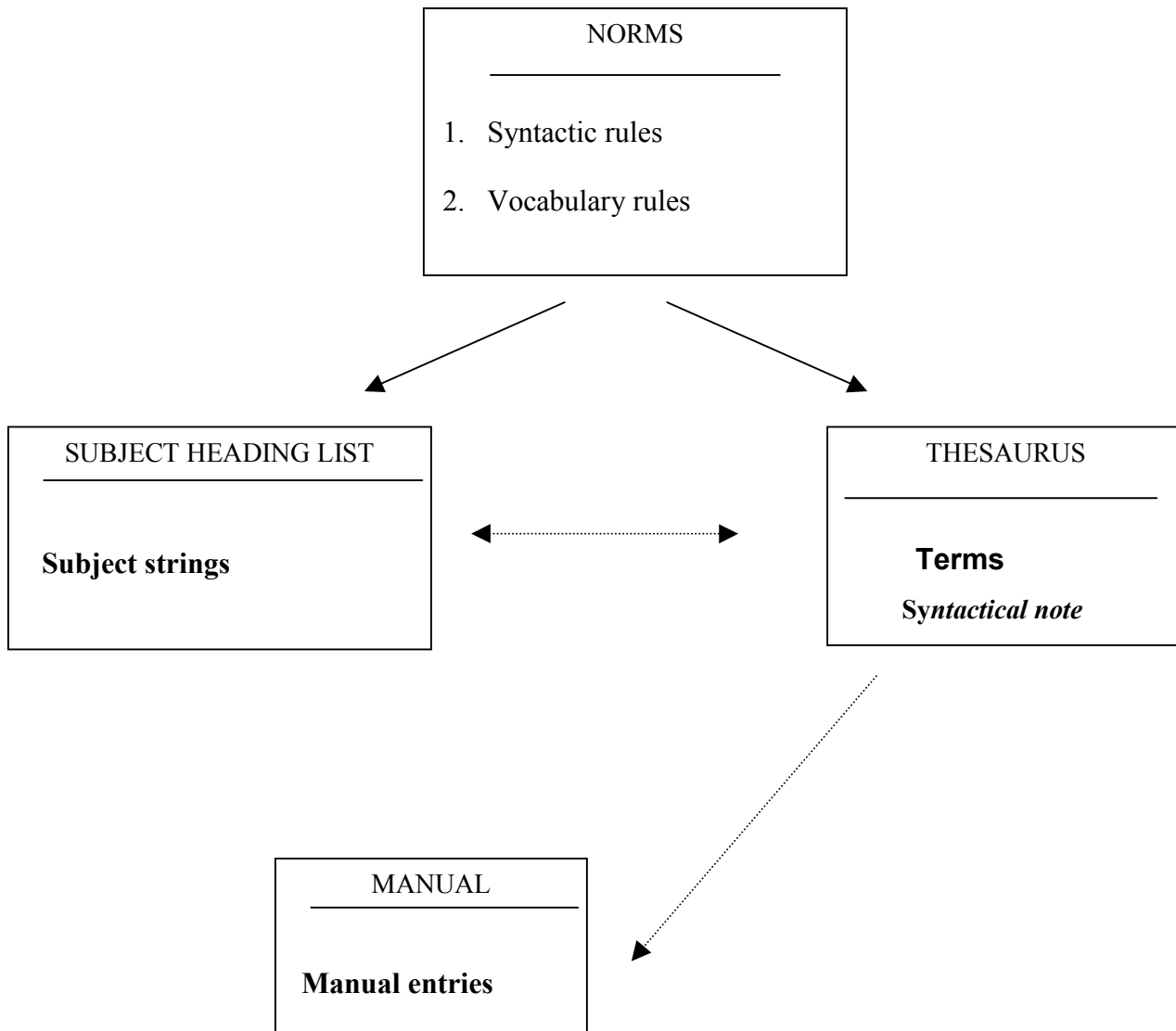


Figure 3
Examples of structured terms

Example 1

Hats

BT Headdresses

NT < *by material* >

Straw hats

Wool hats

NT < *by users* >

Men's hats

Women's hats

RT Caps

Category: Objects

Classific.: 391.43 (DDC21)

Source: *Soggettario 1956*

Example 2

Diseases

BT Pathological processes

NT *<by agents>*
Parasitic diseases

NT *<by way of transmission>*
Hereditary diseases
Infectious diseases

NT *<by organs and parts>*
Cerebrovascular diseases
Muscular dystrophy

NT *<by patients >*
Childhood diseases

RT Sick people

RT Pathology

Syntactic note: Part/property. It follows the term representing its owner (single individuals, groups of people, organisms and their parts), e.g. Leopardi Giacomo – Diseases; Teenagers – Diseases; Cats – Diseases; Digestive tract – Diseases; Children – Nervous system – Diseases [*Before*: Nervous system – Diseases – Childhood]; Larynx – Blood vessels – Diseases

Category: Processes

Classific.: 616 (DDC21)

Source: *Soggettario 1956*

In Example 2 the syntactic note gives instructions about the use of the term in building subject strings. It therefore creates a connection between the terminological and the syntactical level.

State of the art and new perspectives

What we did

In the last year we have been working on the project developing many different directions:

- Testing carried out by the Italian national bibliography. The BNI will use the new indexing language in obedience of the Recommendations of the Copenhagen Conference on national bibliographies (ICNBS,1998). It will offer a prototype and

exemplify the method to other libraries sharing resources in our Library National Service (SBN). It will collect proposals for updating and manage the on line implementation of the tool, in cooperation with the Central Institute for the Unified Catalogue and bibliographic information (ICCU).

In the last few months some experiments of the new subject heading list have been carried out:

- a) on the semantic categories of terms;
- b) on the syntactic aspects.

The National Bibliography has also established a collaboration with groups interested in specialized and general terminology.

- Testing implemented in a special library. This stage of the project (funded by the Regione Veneto) enabled us to verify the principles of the new Soggettario in an Italian library specialized in enology and viticulture (the Biblioteca internazionale La Vigna, Vicenza). The National Library of Florence was responsible for scientific coordination. Thanks to this testing, interesting issues emerged, notably problems relating to the integration of already structured technical and sectional terminology in a general Thesaurus.

What we are doing

We are working with great care on:

- the software of the system;
- the optimization of the new language in OPAC;
- the employment of MARC for structuring data;
- the display of the Thesaurus apparatuses and the strategies of internal browsing;
- the cooperation condition. We'll have to identify the right strategies to accept and validate subject headings coming from other cataloguing centres. We are interested in other projects of cooperation such as SACO (the subject searching component of the Program for cooperative cataloguing of the Library of Congress);
- the testing of the new subject heading list for indexing of non bibliographical material (for example audiovisual aids, photographs, archive material, museum items, e-publications);
- the development of the language towards multilingualism. We followed with great interest the MACS project.

It's important to envisage a connection between the new heading list and a future authority control system, unfortunately not yet issued in our reality.⁸

⁸ A. LUCARELLI, *Authority control semantico e nuovo Soggettario*, in *Authority control. Definizione ed esperienze internazionali: Atti del convegno internazionale, Firenze 10-12 febbraio 2003*, a cura di Mauro Guerrini e Barbara B. Tillet, con la collaborazione di Lucia Sardo, Firenze University Press-Associazione italiana biblioteche, Firenze Roma 2003, pp. 265-75
http://www.unifi.it/universita/biblioteche/ac/relazioni/lucarelli_eng.pdf

The Thesaurus' software will require great attention and it will have to offer advanced and efficient plain solutions. We will have to increase the technical interoperability of our new tool, considering new non book domains in order to allow the possible future integration of different archives, in a web-oriented communication system. The tool will have to answer to standard protocols, to be a technical service, and be supplemented with other applications based on different kinds of cataloguing software. This way its role could also grow relating to the Italian Digital Library, the project our Ministry is promoting studies on and investing resources and competencies. At last, we should not underestimate the economic aspects of the system. The new heading list should have limited operating and maintenance costs. The software will have to be competitive, compared to automatic and post-coordinated search systems, inefficient from the semantic point of view, as we well know, but economically very convenient.

Conclusions

As we have explained in this presentation, our new heading list system represents a flexible, modular tool, applicable in various and different documentary contexts. Moreover, being organized in two distinct and complementary components, vocabulary and syntax ought to foster cooperation projects at an international level. The Italian Thesaurus, if correctly and strictly structured, could communicate or interact with tools in other languages based on the same standards.

The new *Soggettario* system, based on consistent conceptual structures and conceived so that it satisfies the information needs of the final users, is totally attuned to the FRBR model.⁹

The National Library of Florence, putting the project into effect, focused on the current discussion about indexing questions. From the very beginning we have been testing the project, collaborating with other institutions, also on an international level.

It's in our wishes to have the possibility to maintain the method of confrontation and collaboration to ensure the best outcome to our project.

 **Back to the Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>**

⁹ P. BUIZZA, M. GUERRINI, *A Conceptual Model for the New Soggettario: Subject Indexing in the Light of FRBR*, «Cataloging & classification quarterly», 2002, 34 (4); p. 31-46; Federica Paradisi and Raffaella Gaddoni presented a paper about *GRIS e modello FRBR* at the Conference Bibliocom 2003 (Rome, 10-30-2003).