Semantic authority control and the new *Soggettario*

Anna Lucarelli

Biblioteca Nazionale Centrale, Firenze

I was entrusted with the coordination of the project, financed by the National Central Library in Florence (BNCF) for the Feasibility study on the new *Soggettario per i cataloghi delle biblioteche italiane* (Subject headings for Italian library catalogues), the indexing tool also applied in the Italian National Bibliography. In this capacity I cooperated with the Group of experts from IFNET who worked out the Study and developed an Application Project[i]. These experts brought into their task the wealth of experience and elaboration mastered within AIB (Italian Library Association) Research group on subject indexing.[ii] The National Library's enterprise was carried out in observance of the international standards, of IFLA guidelines, and relying on the contribution of traditions and indexing experiences even different from ours. In 2001 we organized an international seminar to receive opinions and advice coming from Italian and foreign colleagues, we also got in touch with other national libraries, institutions, documentation centres and corporate bodies involved in cataloguing not exclusively bibliographic materials.

The work was completed last year in May, then the project proposals were presented in depth in further professional meetings, in order to elucidate the new system and propagate data about its practical implementation from an organization and management viewpoint. Now I'll describe its features and potential in relation to what its actual use may imply in the field of authority control.

As Gloria Cerbai told you yesterday, until now we have not been able to create national author, title, and subject authority files. Anyway, the Italian National Bibliography (BNI) ensured a form of control with its subject headings. We are going to see how the shift to a new indexing language might facilitate the start of a real system for authority control and how BNI might confirm its role as reference frame for Italian bibliographic services.

The *Soggettario*, published in 1956 with BNCF as its editor, is a vocabulary of controlled terms linked by a network of references for various semantic relationships. The language of the *Soggettario* is mostly enumerative (with adaptations of a synthetic nature), pre-coordinated and based on a main heading/subheading structure. It does not create subject strings but it offers examples for possible combinations of terms. In this sense, the choice of a lead term in the string does not depend on the logic relationship it has with other concepts. The *Soggettario*, obviously a child of the cultural time in which it was devised, has been repeatedly analysed. In the past years, the librarians' interest was centred on its outdated terminology because so long a lapse of time that saw new disciplines and research fields, made these gaps evident. Only recently there has been an awareness of the need for revision taking into account its deficiencies in syntax and structure.

The national agency constantly applied the *Soggettario* and updated the vocabulary editing separate lists of new terms used in BNI. In these lists, published in 1977, 1982, 1987 [iii], the terms showed no semantic link among them or with the ones in the *Soggettario,* no syndetic structure except a few references from refused forms. More recently, in 1997 and 1999 the Italian National Bibliography published other updatings, with different criteria [iv]. In fact the terms are linked to other ones already included in the *Soggettario* or in the lists already published according to ISO 2788/1986. Since this was a firs experimental phase, the hierarchical relationship linking the new term to the more general one has been favoured. Only in a few, rare cases, when it was impossible to state the broader term, an association relationship has been made. Besides, each new term has been completed with a link

to the bibliographic record in which it was first used. This experimental work was a first step towards an overall project for the revision of the *Soggettario* that actually started in 2000.

In the long 1959-2003 period the Italian National Bibliography constantly devoted itself to enriching the vocabulary, applying control strategies and activities much more substantial than they might appear given the simplicity of the published lists. The files for use by staff prove the work behind the introduction of a new term: the cards in the old catalogue for use by library staff, as well as in the current one, continuously refer to used sources and reference works; to references from alternative forms and to relationships with associate terms, to references to the linked bibliographic record and, sometimes, to the relative Dewey number. Reading these "paratexts" we find interesting data that shed light on the type of issues the cataloguers faced to provide formal consistency of access points.

Observations on the form and language of terms, particularly the ones related to personal names, come to the surface, issues on uniformity, standardization, as well as problems dependent upon the evolution of the language, for example, related to the semantic change some terms may have undergone in time. In short, information proving a well established methodology in the choice of the form of a subject, information however retrievable for a future effective authority control.

The fact that BNI has always carried out this type of activity does not mean that contradictory, sometimes not uniform, forms cannot be found in the BNI indexes and in the BNCF subject catalogue. This lack of uniformity is due to various causes. The very language of the *Soggettario*, an exemplifying tool, not based on explicit rules, with a partly enumerative partly synthetic structure, leads to incoherent applications both at a syntax and a terminology level. The instance of changes brought by the revision of cataloguing codes and the adoption of new standards: the 1981 *Regole italiane di catalogazione per autori* (Italian rules for author cataloguing) (RICA), the 1984 ISBD, did not occur in semantic cataloguing. Yet, some changes brought by RICA involved the form of subject too.

The general criteria followed by BNI for subjects that are personal names, was to conform to the form stated in the codes for entries in the author catalogue even though with a few well known exceptions, as in the case of names of classical Greek and Latin authors and of Medieval authors, popes etc. Thus, since 1981 – as a consequence of the adoption of the new Rules – there was a formal change in some subject headings. During the years, further changes were made to meet the requirements of the Servizio bibliotecario nazionale (SBN) software. Some other incongruence in the subject headings is due to a change of strategy and renewed choices in the indexing policy.

Corrections to the *Soggettario* and to its updating lists aroused from the control and revision activity carried on by BNI, for example "Giuoco" becomes "Gioco", "Diritto sulla propria immagine" becomes "Diritto all'immagine", "Pianura Padana" becomes " Val Padana". In these innovations there have not always been retrospective action to amend and renew the file of strings in which, nevertheless, lack of uniformity is sometimes caused by simple errors. A relevant action for partial renewal took place in 1997, when the bibliographic records created by BNI from 1958 to 1984 were loaded into SBN.

Thanks to the project called Maintenance of semantic archives (MARS), headings for personal and corporate bodies names, for titles, geographical names and some other categories of subjeccts were unified. Up to now, this enormous work has benefited only SBN. The corrections were not merged into the BNI CD-ROM where we can find lack of uniform headings, as Gloria Cerbai explained; these problems will be solved with new software.

Presently there is a form of control on indexing terms only at local, decentralized level, a prerogative of specific SBN nodes that have organized their own subject managing software. In the BNCF node BNI is responsible for an archive of descriptors where terms, personal names included, can be structured according to semantic relationships. beside the possibility of keeping information in a Notes field, equivalence, association, hierarchy and syntactic links can be stated.

Fig. 1

```
┌─────────────────────────────────────────────────┐
│        S B N - Bibl. Nazionale Centrale di Firenze │
│     GESTIONE SOGGETTI - GESTIONE DESCRITTORE      │
│                                                   │
│                                                   │
│        DESCRITTORE DI PARTENZA:                   │
│        Telefoni cellulari                         │
│                                                   │
│        LEGATO CON:                                │
│        Telefoni portatili                         │
│                                                   │
│   Usato per:        [ X         ]                 │
│                                                   │
│   T.piu'generale:   [           ]                 │
│                                                   │
│   T.piu'specifico:  [           ]                 │
│                                                   │
│   T.correlato:      [           ]                 │
│                                                   │
│   Ha per sudd. :    [           ]                 │
│                                                   │
│   E' sudd. di:      [           ]                 │
│                                                   │
│                                                   │
│       ALTRI LEGAMI (S/N)?  [ N       ]            │
│                        X. USCITA:                 │
│   [          ]                                    │
└─────────────────────────────────────────────────┘
```

The equivalence links and, sometimes, the association ones are activated, while as a rule, the hierarchy and syntactic ones are not. BNI creates links to some variant forms, for example, to the form of a name appearing on the title of the document, if it differs from the preferred form for the subject heading. But these references, that both SBN users and BNI users can see, are managed in the archive for descriptors, not, as they should, in an authority file [v].

Lacking a real national authority file, long wished for by many people, this is the complex, stratified situation of semantic control by BNI, that anyway has been a reference frame for Italian libraries that adhere or don't to SBN. Lucia Di Geso will relate on the cooperative features in SBN also in relation to the tools implemented.

It is interesting, in this context, to point how much the new Soggettario, beside renewing the indexing language, increasing the system efficiency and effectiveness, will enhance, on the one hand, a real authority control, on the other hand, the creation of a national authority file.
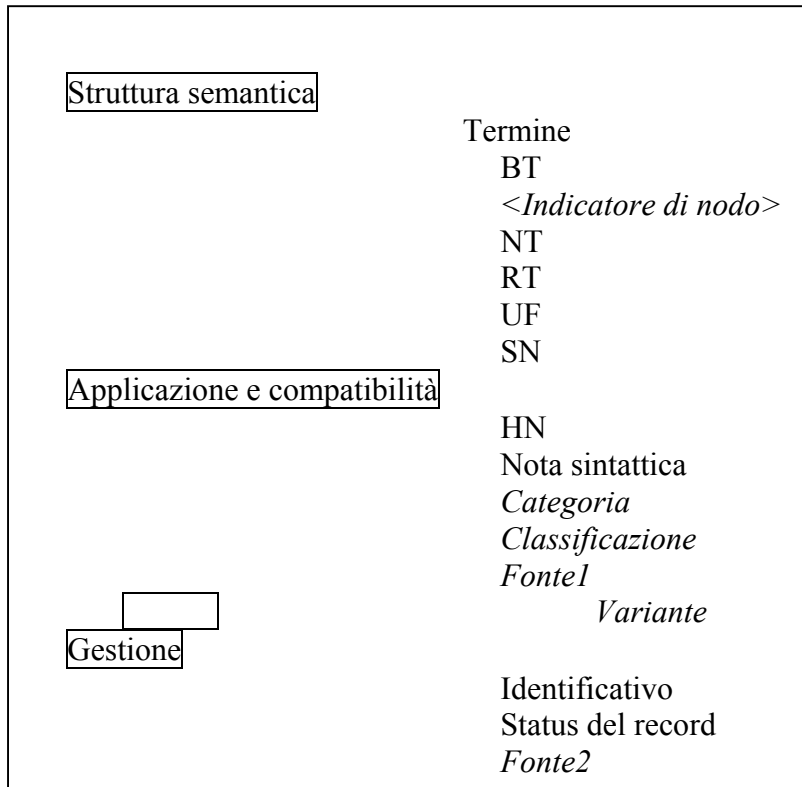
Its features are described in the volume *Per un nuovo Soggettario* in which the documents related to the above mentioned feasibility project have been published [vi]. The proposed language is precoordinated, analitic/syntetic, based on a strict distinction between semantic relationships and syntactic ones, complying with international rules on vocabulary control and structure. It bases its citation order of subject strings on the analysis model for logic roles and it aims, besides specificity, at coextensivness (a single, coextensive string).

The system of the new Soggettario is organized in four parts: rules, vocabulary, syntax-pragmatics, archive of subject strings. In the new language, the rules will be explicit, organic and will have to convey both syntactic and semantic principles. The vocabulary, in a thesaurus form, based on the known principles of international standards, will be the backbone of the new system, i.e. a controlled, structured list of terms, that can each be used combined with any other term. The terms will be given, when necessary, a note to explain how to use them according to the logical roles that

the concept conveyed by the term itself may play in the description of the intellectual content of the work. The same term may have a different position in the string.

Fig. 2
    La struttura della voce (termine preferito**)**

```
┌──────────────────────────────────────────────────┐
│                                                    │
│   ┌─────────────────────┐                         │
│   │ Struttura semantica │                         │
│   └─────────────────────┘                         │
│                          Termine                   │
│                              BT                    │
│                              <Indicatore di nodo>  │
│                              NT                    │
│                              RT                    │
│                              UF                    │
│                              SN                    │
│   ┌───────────────────────────────┐               │
│   │ Applicazione e compatibilità  │               │
│   └───────────────────────────────┘               │
│                              HN                    │
│                              Nota sintattica       │
│                              Categoria             │
│                              Classificazione       │
│                              Fonte1                │
│                                  Variante          │
│       ┌───────────────┐                           │
│       │               │                           │
│   ┌───┴──────┐        │                           │
│   │ Gestione │────────┘                           │
│   └──────────┘                                     │
│                              Identificativo        │
│                              Status del record     │
│                              Fonte2                │
│                                                    │
└──────────────────────────────────────────────────┘
```

As shown in the scheme, already presented in the above mentioned volume *Per un nuovo Soggettario* (p. 244), this architecture meets the requirements for semantic control and, by stating the category the term belongs to, and eventually the classification, it allows vocabulary structuring. The historical note, the citation of sources and variant forms, the thesaurus link it up to the old *Soggettario*. In fact these fields clearly show how the new context of rules will not include only innovative elements but it will also preserve elements of our cataloguing tradition. The new system will allow the retrival of terms and relationships from the *Soggettario* and its updating lists, restructuring their trminology according to a better developed model, through a process guided by rules.

As we did in the volume with the results of the Feasibility study (p. 366), we show an example of heading in which the syntactic note explains possible applications of the term.

fig. 3

Voce corredata di nota sintattica
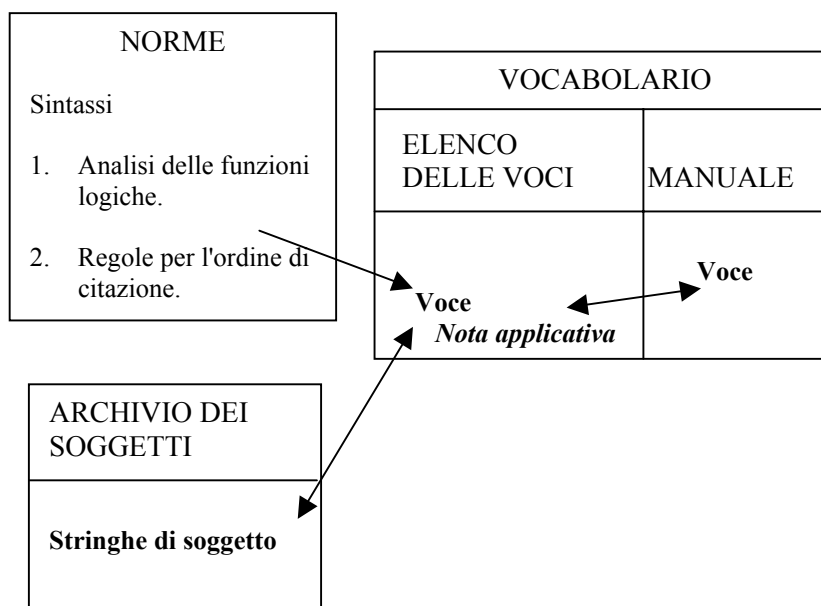
Malattie

BT Processi patologici
NT [Malattie secondo gli organi e parti]
NT [Malattie secondo il modo di
        trasmissione]
NT [Malattie secondo il paziente]
NT [Malattie secondo l'agente]

Nota sintattica: ***Parte/Prop.*** *Segue il termine*
che rappresenta il possessore (singoli individui,
gruppi di persone, organismi e loro parti),
p.e., Leopardi, Giacomo - Malattie;
Adolescenti - Malattie; Gatti - Malattie;
Apparato digerente - Malattie;
Bambini - Sistema nervoso - Malattie
[precedent. Sistema nervoso - Malattie - Infanzia];
Laringe - Vasi sanguigni - Malattie

Faccetta: Processi
Classificazione: 616 (DDC21)
Fonte: *Soggettario 1956*

Besides the syntactic note, it will be possible to find other accompanying instructions in a manual that will integrate the general rules and help the cataloguer to apply them. The archive of subject strings, created according to the rules and the thesaurus, is the final component part of the system.

Fig. 4

NORME	VOCABOLARIO	
Sintassi	ELENCO DELLE VOCI	MANUALE
1. Analisi delle funzioni logiche. 2. Regole per l'ordine di citazione.	**Voce** *Nota applicativa*	**Voce**

ARCHIVIO DEI SOGGETTI

Stringhe di soggetto

A system designed in this way is not only compatible and consistent with semantic authority control, it is even potentially functional to the creation of an authority file. As pointed also by Stefano Tartaglia, the structure of the new language, based on category analysis of semantic relationships, agrees both with the logic of *Guidelines for subject authority and reference entries* and of *UNIMARC. Authorities*, and with FRBR, they too mostly based on general semantic categories [vii].

Furthermore the separation of the syntax and terminology fields at the base of the new Soggettario (a separation that, anyway, finds a junction and a reassemblage point through the syntactic note), is in itself an assumption in favour of authority control since this concerns, *in primis* uniform and consistent terms that can be access points to semantic information.

Of course terminology control cannot be identified with authority control, even less, the terminology control accomplished in a specific language in a specific context with what should be done at a general authority control level. There is no doubt that a tool for a terminology control consistent with international rules and standards, based on category analysis, and that fixes relationships among terms on this base, will not only prove apt to be used at various levels in a documentary context but it will also allow sharing already established data in other lists or authority files.

As the IFLA Working group on Guidelines for subject authority files stated, moving from a subject heading to an authority asks for elaborating a set of relationships and data that partly correspond to what standards require for thesauri, but that, in addition, must guarantee functionality at a wider level than the one required by the specific language used.

The new Soggettario was designed exactly to enhance interoperability and use in quite diverse contexts, proper library contexts, and different ones: special collections for media or photos, museums, archives, etc. The relationships among terms in the controlled vocabulary do not necessarily correspond to the relationships stated in the IFLA standard among uniform headings, parallel headings, correlated (i.e. uniform headings bibliographically related to each other) and variant headings. But undoubtedly the thesaurus includes elements and fields that might be used and thus inherited by a structure for creating authority records. Besides equivalence, hierarchy and association relationships, all the fields related to application and compatibility of the term are significant for authority control, except the syntactic note that might eventually be relevant to guarantee the uniformity of strings and that must not be mistaken for the notes in an authority record.

The historical note "provides indexers with basic data on the use of a term when, during the vocabulary's life span, its role, meaning, structure or form undergo relevant changes, for whatever reason; for example when a term replaces, fully or in part, another term, or when the meaning of a term becomes wider or more restricted"[viii].

The Category (or Facet) brings into the vocabulary a classifying principle similar to the one envisaged by authority control. The same can be said about the classification number that is an element of the information notes area in the *Guidelines*.

The information to be given in Source 1 that, in the thesaurus of the new Soggettario, denotes the reference works the term was taken from or checked, is also mandatory for authority control.

The Variant form in which we should find "all the variants of a term that are not considered equivalence relationships and that must not necessarily follow the bijective term-meaning relationship""[ix] (like singular/plural), allows the creation of a retrospective link with terms in the old Soggettario not migrated into the new one because, for instance, considered linguistically obsolete. The variant form is an atypical component part of semantic control but – in this phase where we move from one indexing language to another – it offers the advantage of flexibility and sharing, since single libraries might retrieve, in this way, a variant, not accepted form considering it a preferred term for their own requirements.

To end, Source 2, where the library or cataloguing agency that presented the new term is cited, is similar to the area in the *Guidelines* where the cataloguing agency responsible for the authority heading, the date of recording, etc. are given (area 1.6).

Various issues are still being studied related to the vocabulary of the new Soggettario. If common names will certainly be the object of semantic control, it is still debated whether proper names (biographical, geographical, of corporate bodies, etc.) will be included in the thesaurus and, in case they will, if their structure will be the same as for any other heading. This issue will be joined by other ones: will the morphology of proper names in the subject catalogue have to be the same as the one in the author catalogue? How must we eventually use qualifiers to disambiguate or identify?

And in case we could actually implement an authority control file, complete with all possible links among variant forms of names, would it still be needed for uniform headings to be the same in the subject catalogue and in the author catalogue? We all know that the various national bibliographic agencies adopt different solutions on this matter. On the other hand the problem could be solved with the help of software providing automatic links among all possible access points.

These and other choices will be made in the near future, keeping in mind that in many of them it will not be possible to leave out of consideration the politics and resources at disposal. A careful assessment of the cost/benefit ratio will have to produce, if not a real national authority file, at least some lists that will be the result of the new language and, at the same time, useful to those who will adapt the new Soggettario to their own requirements. The national bibliographic agency should also be able to validate the subject strings created according to the rules envisaged in the new tool in addition to its control of terms, to which diverse libraries, and the ones specialized in various fields, might cooperate.

A precoordinate language, based on consistent but simultaneously flexible rules, is already shareable and a guarantee of a quality catalogue. Authority control, as Tartaglia is going to explain better, pertains to the vocabulary component, the more so in an analytic-synthetic language like the one of the new Soggettario[x]. But what can we say about the syntax part since it will be characterized by a revised, standardized syntax? The more the language will be used respecting the rule indications the more uniform and coherent will the created strings be. The BNI task will consist in validating them and making them controlled access points in this sense. Of course all this will imply rethinking the type of cooperation among libraries and institutions. The latter will not be carried on only for the production of new terms for the thesaurus but also on the best and shared use of the new method in all its component parts.

The National Bibliography, directly involved in the Project for a new Soggettario, will obviously have to play a leading role on the control of formal coherence of semantic access points. The activity for authority control and implementation of authority files needs resources and investment. New perspectives and new functions at BNI will depend on a real acknowledgment of its role, and a strengthening of its scanty resources. This is the only way for it to confirm itself as a structure for choices, study and in depth investigation of issues related to bibliographic control.

The plan we are dealing with wants to conform the *Soggettario* to present international standards and to redefine choices and solutions within an original consideration. BNI will undertake a double function: its headings will be authoritative subjects in themselves, at the same time the bibliographic agency will validate the strings created by other libraries, even if it does not use them, and control their coherence according to the syntax rules envisaged in the new vocabulary.

BNI will provide the example of a new method, it will keep in touch with the outside world to make known but also to learn new suggestions and proposals. It will have to try and find out new forms of scientific cooperation, of sharing experimental applications for semantic bibliographic control.

For BNI this will be the occasion to try out new strategies as well as to build a bridge to get to environments with which it will learn to dialogue enlarging the typology of its traditional public.

NOTES

i The Group of experts, led by Luigi Crocetti, included Alberto Cheti, Daniele Danesi, Massimo Rolle, Stefano Tartaglia; it profited from the collaboration of Carlo Revelli, beside myself, Marta Ricci from the National Library in Florence participated in the work, while Lucia Di Geso represented ICCU. Diego Maltese brought to the study his scientific contribution.

iii We cite the 1987 edition, cumulating the previous ones: *Bibliografia nazionale italiana* : s*oggetti* : *liste di aggiornamento 1956-1985. -* Firenze : Biblioteca nazionale centrale, 1987.

iv *Voci di soggetto* : *aggiornamento 1986-1998 : Bibliografia nazionale italiana*. – Milano : Editrice Bibliografica, [1999].

v As Paul Weston asserted in *Catalogazione bibliografica : dal formato MARC a FRBR*, "the very exhistence of links between rejected forms and accepted terms, does not guarantee in itself coherence in the choice of headings, it does not allow automatic control when data is transferred nor does it provide operators with the set of rbibliographic eferences that are sometimesessential to solve disputed issues and to place in its bibliographic context the access term chosen" (see: http://www.aib.it/aib/boll/2001/01-3-267.htm. Last contact 30th December 2002).

vi *Per un nuovo Soggettario*: *studio di fattibilità sul rinnovamento del Soggettario per i cataloghi delle biblioteche italiane* / commissionato dalla BNCF alla Ifnet, Firenze ; realizzato dal Gruppo di progetto per il rinnovamento del Soggettario. – Milano : Editrice Bibliografica, 2002.

vii *Guidelines for subject authority and reference entries* / Working Group on Guidelines for Subject Authority Files of the Section on Classification and Indexing of the IFLA Division of Bibliographic Control. - München [etc] : Saur, 1993. - (UBCIM publications; n.s., 12).
Unimarc manual. Authorities format. – 2nd rev. and enlarged ed. - München : Saur, 2001. – (UBCIM publications ; n.s., 22).
Functional requirements for bibliographic records / IFLA Study Group on the Functional Requirements for Bibliographic Records. – München : Saur, 1998. – (UBCIM publications ; n.s., 19).

viii *Per un nuovo Soggettario,* cit., p. 245, note 8.

ix Ibid., p. 246, note 13.

x In other national systems for control of terms, both single terms and, sometimes, combinations of them in strings are included in subject authority files, e.g. in: *Notices d'autorité* di RAMEAU (http://noticesautorites.bnf.fr. Last contact: 30th December 2002).