

FROM INDEXING TO SPECIAL THESAURUSES

Monica LAZĂR

*Librarian, Cataloguing-Classification Department
"Lucian Blaga" Central University Library*

In January 1996, Romanian university libraries launched a program of introducing indexing and then standardising it by the creation of a controlled vocabulary of keywords. Although conceptual, methodological and organisational difficulties were great, this collaboration led to the first level of indexing, based on a Romanian mini-vocabulary of keywords.

The increasing number of records in the database of these libraries and the inconsequence in using the principles of indexing called the attention of the librarians to the need for a control of this vocabulary, moreover, the need for structuring it and developing a relational semantics which would improve the "bringing back" in information retrieval ("bringing back" is a technical term in indexing, denoting the ratio or percentage of all documents linked to a search and those which were actually retrieved.)

In this context, the task of librarians who are collaborating in this project is to adapt the terms of the vocabulary by translating them into Romanian, and to indicate hierarchical relations, references of the type "see also" and to identify excluded subject-headings (synonymous terms).

All these elements prefigure a controlled vocabulary. The quality of validating the terms and the computerised processing of this information could lead to the creation of a modern polythematic dictionary in Romanian.

However, this process is difficult and demanding. It is a great responsibility to translate and choose subject headings and select their synonyms (excluded subject-headings) for various scientific fields, even if the collaborators are specialists in these fields and have been cataloguing scientific material for a long time.

There are many difficulties: linguistic and cultural limits, limits imposed by the evolution of research and the application of sciences in Romania, the politicization of certain terms, the complete lack of terms denoting modern concepts, mostly in the social sciences and philosophy. Other difficulties are the lack of a unique concept regarding the very process of indexing and the low compatibility of UDC (universal decimal classification) used in Romanian university libraries for more than 60 years, with indexing. These two systems of classification are rather complementary, regarding the principles of their

construction and the results of their use, in terms of efficiency, advantages and disadvantages.

Therefore the project of the coordinators of the national program of indexing to develop a vocabulary of keywords starting from the structure and terms of UDC did not reduce the difficulties of indexing, but increased them.

The experience of European university libraries shows two possibilities:

- 1) the translation, adaption and processing of LCSH (Library of Congress Subject Headings)
- 2) the use of LCSH as a reference point in the creation of Romanian special thesauruses for different scientific fields or highly important fields.

The first variant is costly and time-consuming; it presupposes a team of trained librarians or documentarists from a national centre oriented towards carrying out such projects. The second variant seems to be more appropriate for the conditions and possibilities of Romanian university libraries and can use the professional knowledge of librarians as well as the effort invested so far in the project of indexing.

The national project of indexing could be easily oriented towards the creation of special thesauruses, because collaborators are grouped according to scientific fields. Thus it would better fulfill the needs for indexing scientific publications which are priorities for university libraries. In addition, special thesauruses would lead, in a relatively short time, to a better use of scientific information contained by collections grouped now in the libraries of certain faculties.

The compiling of special thesauruses could be planned as the second level of indexing, following the creation of a controlled vocabulary of keywords and preceding the creation of a modern polythematic vocabulary in Romanian.

In certain European countries the idea of a special thesaurus is not linked to scientific fields only, but to highly important or modern fields (economy, finances, law, business, stock exchange or sociology, education, administration), for which the national language does not have an established terminology, or a terminology compatible with international terminology.

The process of compiling a special thesaurus requires the extension of collaboration between librarians, in a certain field, and the acquisition of a software adequate for the exigencies and structure of a special thesaurus.

The methodology of creating a special thesaurus will be presented by the project compiling a *business thesaurus*. Why business? It is a highly important field which combines economic, financial, juridic, political, management, marketing and publicity terminology.

The specific keywords which are being translated and adopted, will be the starting point; other terms will be added from economic journals and the report of activity of different national and international economic organisations.

On the level of methodology the selection of terms (keywords) raises the first problem of the thesaurus: the construction of the *vocabulary*.

The *literary warrant principle* is the first criterion of selection: a term must have a certain frequency in the literature of a field. The *principle of usage*, reflecting the current usage of a term is the second criterion.

In the field of business it is not easy to observe these two principles. Business is a relatively new field in Romania; although we have a European cultural tradition, the majority of the specific terms have been taken over from American special literature, because the US is the home of the great theorists and practitioners of management and marketing.

Romanian theorists were the first to take over technical terms and adapt them to the economic and political reality of socialist Romania. Many concepts, methods and techniques have remained undefined in Romanian language.

Political opening and the changing conditions for economic activities allowed practitioners to adopt new terms which enriched the vocabulary of business, but which do not have yet a literary justification.

The role of the librarian is to follow the evolution of these terms until their recognition, and decide on their introduction in or exclusion from the vocabulary.

The role of the thesaurus is the efficient and neutral combination of the terms of the vocabulary, association, hierarchic rendering and the creation of links between classes of equivalence and the re-grouping of terms in semantic fields.

The *semantics* of the business thesaurus has in view the principles common for all special thesauruses. The keywords of the vocabulary become subject headings, terms selected from a group of equivalent terms, which represent unambiguously a notion or concept which appears in a document or the form of inquiry of a database.

We start from two findings in the re-grouping of subject-headings in fields:

1. business is the centre of interest, whereas economy, finances, law, accounting and taxation are the fields;
2. users' searches/forms of inquiry are of two types:
 - a. descriptive
 - b. instrumental

Let us group subject-headings in the decreasing order of their semantic priority:

1. business
2. finances
3. economy
4. law
5. accounting
6. taxation

The inquiry of the database can be directed

1. towards the descriptive (concourse of events, the event)
2. towards the instrument of analysis (the normative instrument)

This possible structuring of users' inquiry introduces the concept of the facet of the thesaurus. The field of business and related fields can be intersected by two facets:

1. activity → descriptive
2. techniques → the instrument of analysis (normative instrument)

The grouping of subject-headings in the forementioned fields raises a problem: on the one hand, accounting and taxation are empty of content without the financial and economic concepts; on the other hand, law remains substantial even without chapters on business, finance and economy.

The crossing of the following three notions would be a possible solution:

1. the semantic priorities of business, finance and economy
2. the "activity" and "techniques" facets
3. the semantics of each of the six forementioned fields

This results in the final grouping of subject-headings in 9 fields, the so-called semantic fields. (See fig.1.)

The relations between subject-headings are also a semantic problem of the thesaurus; they define *relational semantics*.

Between the subject-headings of the thesaurus there are *hierarchical*, *generic*, *specific*, *associative* and *synonymous* relations. A *generic subject-heading* is a subject-heading which denotes a concept which includes other, more "refined" and detailed notions which are the *specific* terms of the thesaurus. Reciprocally, a *specific term* is a subject-heading which designates a notion included in a larger notion, called a generic term.

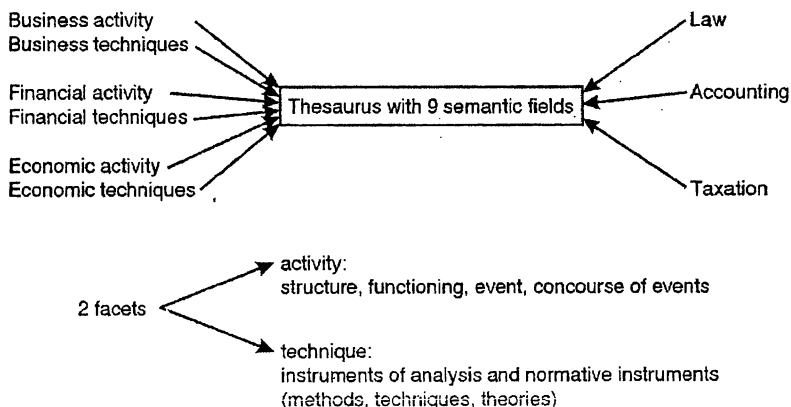


fig.1.

Hierarchical relations between generic and specific terms allow the placement of subject-headings in a certain semantic field and also remove ambiguities.

At the top of the hierarchy is the term which defines the semantic field; it is followed by the generic terms. The subject-headings which form the thesaurus have only one generic term and one or more specific terms.

The hierarchy of such a thesaurus descends to five or six levels at the most, very rarely to ten or more levels.

Hierarchical relations between generic and specific terms are usually represented by a figure with arrows.

Associative relations are not necessarily reciprocal. The relation is not reciprocal when inquiry after an associated term leads to a specific term. It is almost inevitable that a subject-heading appears in two semantic fields. In this case it will be completed by a qualifier which designates the semantic field to which the subject-heading in question belongs:

Commercial societies (Law)

Commercial societies (Business activities)

Similarly, homonyms will be completed by qualifiers which designate the field of the subject-heading:

Chamber of Commerce and Industry (Cluj)

Chamber of Commerce and Industry (Paris)

Synonymic relations or *relations of equivalence* are as important as hierarchical relations. Two or more terms are in relations of equivalence if they refer to the same concept. These terms form a class of equivalence. Usually, one of the terms is preferred and becomes subject-heading; there are references to the other terms called excluded subject-headings.

subject-heading: Stock Exchange

excluded subject-heading: Exchange Market, Stock Market, Financial Market

The special thesaurus has a specific form. The first part presents the semantic fields and the facets, the second is an alphabetic list of subject-headings with the system of references, explanatory notes, hierarchical relations and facets followed, in certain cases by a list of subject-headings grouped by facets.

The quality of a thesaurus can be evaluated by the way in which different trends, professional jargons, political-geographical particularities in a certain socio-cultural context are presented.

The user must be able to make an inquiry to the database using any word from the thesaurus, and this must be structured in a way to be able to lead him from excluded subject-headings to the equivalent subject-heading.

Explanatory notes are very important in a special thesaurus: they specify the date of the recognition of a subject-heading, its evolution and the field, language and country to which it belongs. Therefore a thesaurus is almost a special dictionary.

The functions of a special thesaurus are the following:

1. the control of the use of subject-headings in the indexing of documents
2. the translation of the natural language of the database inquiry into a controlled language
3. facilitating “navigation” within a search, by specific or generic questions
4. facilitating the completion of a search by the investigation of associated terms
5. ensuring a good understanding of the terms by presenting their semantic field, accompanied by explanatory notes where necessary

Special thesauruses are not mere intellectual products; they reflect the stage of development and the level of application of a field at a given moment and help the process of indexing and of efficient information retrieval.

If information is the most precious commodity today, we can assert that special thesauruses are invaluable in the process of information processing and retrieval.