

Note

"The Termi-thesaurus"

Nelida Chan

Meta : journal des traducteurs / Meta: Translators' Journal, vol. 27, n° 2, 1982, p. 231-233.

Pour citer cette note, utiliser l'adresse suivante :

<http://id.erudit.org/iderudit/002212ar>

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <http://www.erudit.org/apropos/utilisation.html>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : erudit@umontreal.ca

THE TERMI THESAURUS*

Today the computer is revolutionizing almost every aspect of our modern lives. It is therefore not surprising to find it penetrating the fields of translation and terminology. A multitude of new computerized equipment often referred to as translation aids is entering the market. A terminology data bank is one such translation aid which is becoming increasingly popular around the world.

Storing thousands and thousands of terms in a computer is obviously useful. The terms must, however, be grouped in classes if this terminological stock is to be used to advantage. Even if a user queried the computer for individual terms only, a classification system would still be indispensable. The meaning of a term often depends on the subject-field in which it is used (for example, doctor in Medicine is not the same as a doctor in the field of Papermaking). Moreover, without a classification system, a terminology data bank could not offer specialized glossaries, nor exchange any of its data with other terminology banks.

Although everyone agrees on the need for a classification system, there is little agreement on the particular system that should be used. For three years, a group of terminologists with the government of Canada terminology bank has been studying and attempting to solve this problem. This project, known as the TERMI Thesaurus, aims to provide a classification system that answers the particular needs of terminology.

The first step was to become familiar with the theory and practice of classification and to examine the existing systems. This study led to two conclusions: that a new classification system had to be developed to meet the needs of terminology and that the thesaurus would be the most effective means of achieving this end.

Let me now explain the basis for these conclusions. The first conclusion rests on the following arguments. In their present form, none of the documentary classification systems, would accommodate terminology. Traditional documentary systems are themselves becoming inadequate and documentalists are turning increasingly to computerized documentary banks which use thesauri as a means of retrieval. None of the current thesauri could be adopted as they now exist; to be useful to terminology a large number of entries would have to be eliminated and others added. Furthermore, indexing documents and indexing terms are two essentially different activities. Documents, which deal with numerous aspects of a field or topic, are now indexed or classed with a number

of headings or descriptors which reflect the content. Terms, on the other hand, designate individual concepts which can be indexed to broader or generic concepts within a network of concepts. Because of this fundamental difference, headings in a classification system used to index documents are often of little use for indexing terms. Many headings or descriptors are interdisciplinary or too specific. For example, there are numerous documents on mathematical geology, econometrics and psychometrics, but are there terms which arise from these fields? Only in-depth research on the vocabularies used in these disciplines will provide an answer. Generally speaking, the terms used in these fields come from two separate fields such as, for example, geology and mathematics; thus, mathematical geology, although necessary to the classification of documents would be of no use for indexing terms. Similarly, headings or descriptors used to index documents dealing with very specific objects such as manometers or rain-gauges would be of no use in terminology, since the number of terms that would be indexed to these concepts would be too small.

Now let me explain why the thesaurus format was chosen. Until now, our terminology bank has used a subject classification, that is to say a breakdown of various disciplines or fields of activity into subfields or sub-sections. Experience has shown that dividing subjects in this way, no matter how carefully done, does not yield a reliable means of indexing terms, because the resultant subject headings are liable to be interpreted by users in a variety of ways. To avoid linguistic ambiguity, the tool used must have mechanisms capable of guaranteeing that all users will interpret subject heading or descriptors in the same way. This is precisely the aim of a thesaurus, which is made up of an alphabetically structured list that specifies the semantic field of each descriptor by indicating its hierarchical, associative and synonymous relations. A scope note, which may be added to any descriptor, gives specific instructions as to the use or the scope of the descriptor in question. If one adds a formal definition for each descriptor, it can be hoped that users will interpret the thesaurus descriptors in a similar, if not identical, manner.

The second step in developing the TERMI Thesaurus was to select and define, at least provisionally, the subject-fields for indexing those terms either already in the terminology bank or likely to be entered in the near future. Because the main function of the bank is to serve the 1 200 translators of the Canadian Government, its terminology data base covers a large number of subject-fields. Consequently, since the bank's contents provided the basis for the research, the thesaurus also covers a large number of fields, although it does not claim to

have exhausted all possible subject-fields. A year and a half was spent identifying the subject-fields which generate terminology. The results of this research were published in a working paper which describes the scope of the fifty-five main subject-fields retained for indexing the present and future contents of the terminology bank. Some subject-fields, such as urban planning and audiovisual media were excluded from the list, because study of these fields showed that the terminology was largely borrowed from other fields. The vocabulary of urban planning, for example, is made up of terms borrowed from sociology, ecology, transportation, economics, administration, law, etc. The working paper was not intended to be used as a classification tool as it lacks the mechanisms of a controlled vocabulary and the descriptions of the subject-fields remain provisional.

Rather, the document was intended to be used as a means of planning the third stage of the project: namely, the construction of the TERMI Thesaurus, to our knowledge one of the first macrothesauri designed for terminology. This working paper permits efficient co-ordination of the research carried out by the various members of the classification group, since it specifies more or less the limits of each subject-fields and, in some cases, resolves the problem of overlapping areas, inevitable within such a wide range of fields. Furthermore, any subfield not yet included can be easily introduced during the third stage of the project since the fifty-five subject-fields listed in the document are not definitive. Nevertheless, these fifty-five fields form the core around which the TERMI Thesaurus is being built.

The third stage of the project began in April 1978, when one of the team members was given the task of developing the Geology thesaurus. In addition to being revised by a professional geologist, this first thesaurus was tested by two terminologists who were assigned the job of indexing, independently, some one hundred terms. I cannot present all the results of the test here; I can, however, say that they were most encouraging. Subsequent to the development and testing of the Geology thesaurus, work methods for building the thesauri for the other fifty-four subject-fields were drawn up in a document.

The Geology thesaurus consists essentially of a list of sixty-five descriptors under which approximately 40 000 terms can be indexed. As indicated by the bibliography, the author carefully gleaned most of the thesauri and classifications in this field, thereby collecting more than 650 potential descriptors. However, only those descriptors for which the concept made it possible to draw up a list of at least one hundred terms were kept for the terminology thesaurus.

In other words, it must be proven that the concept represented by a descriptor is found in at least one hundred terms. This brings us to the basic principle of the TERMI Thesaurus: the decision to index terms according to the conceptual link they have with a generic concept, rather than according to the field of activity in which they are used. For example, not all the terms used by geologists in performing their professional tasks are indexed to the subject-field geology, for this would make it necessary to include terms which come from several other fields such as statistics, computer science, economics, etc. Only those terms which belong to a conceptual network specific to geology will be indexed to descriptors in the Geology thesaurus. Sometimes, a block of terms can well be associated with more than one conceptual network. For example, crystallography can be considered part of mineralogy, chemistry or physics.

In such a case, the descriptor becomes what we call a **free-floating descriptor**, that is, a descriptor that can index a specific vocabulary without being included in a hierarchical system. Some users of the terminology bank may be interested in vocabularies for which a hierarchy has not been established, for example urban planning. Provision has been made for this in the TERMI Thesaurus in the form of interest profiles based on networks of associative relations recorded in the alphabetically structured list of descriptors.

Since the completion of the Geology thesaurus, nine other thesauri have been developed and work is currently under way in thirteen other fields. We anticipate that at the present rate of progress the project will be completed in two and a half to three years.

In closing, I would like to say that, once completed, the terminology thesaurus will be a useful tool which will facilitate the use of the government of Canada terminological data bank.

N.B. Although the author is one of the original designers of the TERMI Thesaurus, this article is based on a French article written by the author's colleague, Noël Lazure, and which appeared in *Actualité terminologique* in January, 1980.

Nelida Chan