Water-related language analysis: The need for a thesaurus of Mexican terminology

Verónica Vargas & Catalina Naumis

Introduction

This research developed out of a series of reflections on the need for a terminological tool that would lend coherence to the indexing and retrieval processes of information related to the resource of water. Experience has shown that the lack of a controlled, holistically treated, water-related vocabulary in Spanish has meant that library scientists in this field do not have a tool that enables greater precision and the necessary rigor to guarantee the quality of information indexing and retrieval processes. In other words, critical inconsistency exists among the terms used in the literature in Spanish and, as we all know, terminology inconsistencies can lead to dispersion of documents on the same subject.

The main objective of this research was to show the necessity of having a terminological tool related to the subject of water by determining whether any existing thesaurus may be adopted and/or adapted to the Mexican reality. To accomplish this, we proposed a descriptive analysis of six thesauri and general examination of term management, as well as the topics encompassed.

Methodology

Taking into account all the human and financial resources involved in constructing a thesaurus, which is why academics recommend exhausting all possibilities of adopting and/or adapting a pre-existing one before considering creating a new one, we undertook the task of descriptively analyzing six thesauri: *Agrovoc*, *General Multilingual Environmental Thesaurus*, *OECD Macrothesaurus*, *Tesauro de Ingeniería Hidráulica*, *Tesauro de Ingeniería Sanitaria y Ambiental* and *Thésaurus Eau*. While the number of thesauri chosen in no way claimed to be exhaustive, the selection was guided by the features we deem relevant in judging the possibility of adopting and/or adapting one that already exists to the Mexican reality.

To carry out the task at hand, inclusion of the subject of water has been considered from all possible perspectives; the institution sponsoring the thesaurus, taking into account its experience and prestige within the scientific and academic communities; and thesaurus availability and accessibility, essential conditions for being able to carry out the respective study. Without a systematic and proven methodology for the analysis of this type of methodological tools in the field, the proposal made by Naumis (2002) was accepted with some added elements.

Three major factors were considered for the analysis: 1) content, which involves thematic coverage, level of structuring and maintenance and updating; 2) presentation, which refers to each thesaurus' distinctive features and how they are treated; and 3) consultation, dealing with thesaurus ease of handling, in other words, the kind of

support, as well as its accessibility and availability and the clarity with which the user guide is presented.

Results

The descriptive analysis led to the conclusion that it is not viable to adopt and/or adapt any particular thesaurus, though it is feasible to adapt the most relevant aspects of each of these thesauri to the Mexican reality. It is possible to make use of the organization of subject matter in the thesauri consulted, as it can be a guideline for developing a structure, as a first step in creating a thesaurus on water for Mexico, without forgetting that our proposal intends to holistically encompass the hydric theme.

While carrying out our analysis, we realized that the significant increase in non-descriptors from one edition to another in some thesauri reflects the need for bringing the natural and controlled vocabularies closer together. One reason might be that most users these days carry out their information searches directly in the databases, though on the other hand, it could be due to the prevalence of idiomatic expressions, which are sometimes closely related to change in political discourse.

Similarly, this research enabled us to observe that water as a resource is basically viewed from the standpoint of the exact sciences, lending little importance to the social and economic aspects. In this regard, it should be underscored that although the solutions to the problems of water must include technological aspects, clearly water, a problem with a social origin and solution, may only be effectively faced using social engineering, from the political science point of view, as well.

We also observed that the appeal to users to propose new terms reflects the importance of users recognizing their information needs. We must not forget that the design of a documentary language that does not take the user into account promises to be incomplete and will lead to a language deficient in its features for information retrieval systems.

Furthermore, we noted that the lack of an introduction and a user guide clearly hinders and limits thesaurus use. The introduction is a substantive part of any thesaurus, since it informs us of its application scope and provides an explanation of its distinctive features.

The same thing happens with the use of foreign terms or deficient translations; without risking exaggeration, we can state that they can cause confusions in the use of the terms, which surely impoverishes the quality of the thesaurus and weakens its credibility, especially if we consider that a thesaurus may also be used as a translation tool. As Domènec Turuguet (1994) mentions, “Translations create additional problems: often a term from the original thesaurus leads to two or more translated ones, it is difficult to have perfect equivalencies and some terms may not be translatable. A temporary solution may be to use the original thesaurus, especially when dealing with scientific and technical subjects, but the difficulty when dealing with humanistic topics is evident. Nevertheless, having a thesaurus thematically akin to the one being planned is always a big help”.

Of the six thesauri analyzed, four discuss maintenance and updating, which speaks of the importance of taking into account the evolution of scientific and technical vocabulary. It is worth mentioning that thesaurus updating is necessary both to incorporate terminology derived from the development of the particular science or discipline and to cover incomplete areas and correct errors detected through its use, as well as to adapt it to the retrieval needs expressed by users through their searches.

We further perceived that the growing consolidation and expansion of the digital environment proves ever more the need for greater availability and accessibility to online thesauri; it is now, in fact, difficult to conceive of a thesaurus in a solely printed version. We realized, however, that the lack of standards for electronically publishing thesauri clearly hampers thesaurus use, interoperability and way of sharing. That is why we consider it fundamental to review the formats and standards used in online thesaurus construction, in order to find a way to incorporate and balance all these aspects.

Conclusions

This research let us see that the thesaurus allows for greater specificity of terms relative to particular fields of study, which implies more precise document analysis. Likewise, it enables the incorporation of associative relationships that favor versatility and help create a real and interdisciplinary universe such as scientific language.

Similarly, the growing consolidation and expansion of the digital environment increasingly proves the interconnections between different spheres of knowledge. In the case of library science, and specifically in the field of information representation and retrieval, the importance of the contributions of linguistics, terminology and automatic computing is undeniable. We believe that such linking makes it possible to alter the dimension of the thesaurus as a terminological tool for information indexing and retrieval, with emphasis on the coherence and relational wealth of conceptual structuring.

Bearing in mind that the thesaurus is, by nature, a dynamic tool, we observe that in an automatized context, the use of a controlled vocabulary turns out to be suitable, due to the literality with which automatic computer programs store, organize and recover information and because the thesaurus enables combining recovery terms. Furthermore, we have noticed that in a global information society, where communication is ever more interculturally expressed, the multilingual thesaurus becomes a basic tool for information indexing and retrieval.

Our thoughts evolve from the idea of resource reutilization, where efforts to compile, organize and systematize huge volumes of information regarding water should not be limited to the construction of one isolated resource, the thesaurus. This terminological tool should, instead, be structured in such a way that it may be used to generate or enrich other resources.