USING ON-LINE THESAURUS IN MACHINE-AIDED TRANSLATION SYSTEMS

Sylvie REGNIER¹, Frédérique SEGOND, Shirley THOMAS²
(position paper)

The scope of this paper is machine-aided translation (MAT) as opposed to fully automatic translation systems. We point out the importance of thesaurus in the translation process and suggest incorporating an on-line thesaurus in MAT systems.

As we are addressing the question of translation we consider it essential to consult professional translators and interpreters. When translating, they do not only rely on bilingual lexicons, although bilingual dictionaries are needed to provide them with vocabulary. Another tool many translators find very helpful, and whose organisation reflects that of the translation process, is the thesaurus.

The translation process in fact involves understanding perfectly the meaning of the sentence in the language A, if necessary by using a mono-lingual dictionary containing semantic definitions and only then finding the best equivalence in language B. And this is the point: to translate is not to find direct correspondences between two texts (or words, or phrases or sentences) but to find equivalences³. In other words, to find the right words or combination of words which will be able to produce the same effect on the reader as they do in language A. From this angle it appears obvious why a thesaurus is needed. It enables an idea-based rather than word-based search. In a thesaurus information is organised under one general heading, which is then subdivided into different grammatical categories which all provide different variants of the original entry.

¹ Aérospatiale
Louis-Bleriot joint research center
12, rue Pasteur BP 76
92152 Suresnes cedex
France

² Institut National des Télécommunications
9, rue Charles Fourier
91011 Evry cedex
France
email: segond at vaxu.int-evry.fr

³ This is the approach adopted at ESIT (Ecole Supérieure d'Interprètes et de Traducteurs), Paris.

Therefore once the general meaning is understood, the thesaurus can help the user to formulate the best equivalence.

Which problems are raised when only bilingual dictionaries are used? Let us consider the following example⁴: "... sélectionner et mettre en œuvre des projets visant à valoriser le patrimoine audiovisuel européen ..." Let us imagine that we only have a bilingual lexicon and we want to translate "à valoriser". The dictionary proposes "to valorise", which with its associations of finance and pricing is not at all suitable in this context. If however we consult a monolingual dictionary in language A, we see that "valoriser" is defined as "adding value". Consulting a thesaurus in the target language under the heading "value" we come to the idea of "usefulness" and among the equivalences proposed we find "to make effective use of" which conveys far better the idea expressed in the source language. A possible translation could be: "... select and implement projects intended to make effective use of European broadcasting archives...".

At present most of existing on-line thesauri are designed for specific technical sub-language and have two main purposes: document indexing and information retrieval. They consist of a list of controlled keywords, selected for representativity and precision purposes, integrated in semantic networks but their general organisation does not reflect that of the general purpose paper-based thesaurus (such as Roget). They do not take into account the textual and translational dimensions which play a crucial role in translation activities. This paper forms part of a research project aimed at developing a MAT system, called "TRUCHEMENT"⁵. The main idea of this project is to provide tools which enable a user who is a native speaker of language B and having some knowledge of language A, to translate a written text in a specific technical field with which he/she is familiar (for instance course material), from A to B.

⁴ INA. Plaquette publicitaire.
⁵ Truchement stands for Dragoman, which is the role of a translator in the translation process, of an interpreter in that of interpretation. This project is conducted at the Institut National des Télécommunications (INT), France.
Bibliographie


MONTEMAGNI Simonetta, VANDERWENDE Lucy, "Structural patterns vs. string patterns for extracting semantic information from dictionaries", Proceedings of COLING 92, pp. 546-552.


