



Solutions in language processing

## LINGWAY Terminology Builder for Systran: a new solution for developing multilingual applications

Paris, May 31, 2005 - LINGWAY, a software company specialized in "Text Mining" announced the launch of **Lingway Terminology Builder for Systran** at iExpo 2005. The solution enables Systran, a world leader in translation technologies, to rapidly build terminologies for its automatic translation software and to develop multilingual applications.

The solution, developed within the framework of the Lingway KM multilingual semantic search, categorization, extraction and reading assistance engine, automatically extracts bilingual terminology, and then imports it to Systran's automatic translation system.

The platform allows users to quickly build terminologies in several languages for multilingual applications (Portals, Intranets, Websites, Knowledge Management applications, E-catalogues, etc.), as well as in research and multilingual "text mining."

*"The building of bilingual terminologies has been the real roadblock for multilingual application development,"* says Bernard Normier, Lingway's CEO. *"Our technology will ease and speed up the process, dividing the time required by a factor of 3 to 5."*

### A 4-step procedure:

1- Two document corpus involving the same technical domain, one in the source language and the other in the target language, are built by using one or more of the following methods: by integrating relevant websites, for example, or conference reports, specialized press articles, patents, and corporate product descriptions.

2- The terms are automatically extracted from the corpus in the source language according to linguistic and statistical criteria. This process is carried out by the Lingway KM terminology extraction tool, which is similar to the one used for automatic text indexing but adapted to this specific task.

3- The Lingway KM search engine sends each term like a "cross-language" query to the corpus in the target language. The corresponding terms in the target language corpus are all potential translation options. Those that appear the most frequently are selected to translate the terms of the original language.

4- The corpus-validated bilingual terminology is then imported to the Systran automatic translation system.

The approach is highly cost-effective. All you have to pay for is the constitution of the corpus that serves as the extraction base, most of which is available free-of-charge on the Web or already in existing corporate sources, and that of a manual revision if required. In addition, the software enables the user to easily consult both source and target texts, and rapidly restore a term to its original context, a useful function for unusual or infrequently used terms.

*"Without a first-rate dictionary, system performance will never be as good as it should be. However, the compiling of these dictionaries is a tedious and time-consuming task: it is usually done manually by specialized translators and at a cost of that ranges from 2 to 5 euros per term,"* explains Normier. *"Now the entire process can be carried out automatically."*