

## Corpus representativeness in the selection of medical terms to be used in translation memory tools

Paula Paiva  
UNESP / FCLAr, Brazil

Scientific journals have become one of the most important sources of information in the medical area as they bring the most recent studies and discoveries carried out in different countries. In Brazil, editors from different medical areas have articles published in two, sometimes, three languages, that is, Portuguese, English and Spanish. This way, study results from Brazilian researchers have become known worldwide. One important aspect from these studies is the common use of new terminology which is connected to recent discoveries. For this reason, a study was carried out in order to observe the use of frequent terms found on journals of anesthesiology, cardiology and orthopedics (Paiva, 2006, 2009).

We have analyzed the most frequent terms used in articles of cardiology and cardiac surgery, based on two parallel and four comparable corpora. The parallel corpora were compiled with thirty articles, originally written in Portuguese as well as their translations in English, which were done by professionals and published on bilingual Brazilian medical journals between 2003 and 2006. The four comparable corpora were compiled with articles from the same medical areas and were originally written in Portuguese and English, however, they were published on international journals by native speakers in both languages. For the selection of keywords to analyze the terms connected to them, we used the program WordSmith Tools 3.0 (Scott, 1999) with the aid of its following tools: WordList, KeyWords and Concordance.

We have already collected the most frequent terms from the corpora of cardiology based on their keyness and representativeness, and compiled a glossary to be used by (future) translators of medical articles. The use of this glossary has been tested on a translation memory tool called Wordfast (Champollion, 1999), which has been the most common program used by professional translators in Brazil (Nogueira & Nogueira, 2004). Since the glossary proposed in this study is based on terms which were chosen considering their representativeness in the corpus, it has proved to be a very helpful aid for translators who (want to) work with medical articles, more specifically, in the area of cardiology. The theoretical basis followed was that of Corpus-based Translation Studies (Baker, 1993, 1996; Camargo, 2005); Corpus Linguistics (Berber Sardinha, 2004; Tognini-Bonelli, 2001) and Terminology (Barros, 2004). We hope this study may help scholars and translators showing how research and practice are straightly connected to each other.