

Designing, Compiling and Sampling Specialised Corpora for Researching Metaphor in Translation: The Case of SCoPE

Jesús Meiriño-Gómez (Universidade de Vigo, Spain)

This poster details the design criteria and compilation protocol, together with the procedures used in the sampling of excerpts, for the Specialised Corpus of Popular Economics (SCoPE). SCoPE includes two different sub-corpora, amounting to around 500,000 words each. One corpus is an English-Spanish parallel corpus, which consists of five American-English books and their Peninsular-Spanish versions. The other is a Spanish monolingual corpus, made up of seven Peninsular-Spanish works. The books covered by both collections address, to a greater or lesser extent, issues related to the economic shifts during the Global Systemic Crisis. They were written respectively by American and Spanish leading experts in the field, and published between 2010 and 2014. SCoPE was originally created within the framework of an on-going PhD research study, which aims to investigate how conceptual metaphors and their linguistic realisations are handled in translation, and describe the similarities and differences that might exist between translated and original texts in the target language.

A brief contextualisation to the research project is provided, followed by a detailed description of the different steps taken and challenges faced throughout the design, compilation and sampling process. The first stage of the project consisted of setting up the design criteria based on the purposes of the study, which allowed the definition of the target population and the sample frame. After specific books were selected for the corpus, they were converted into machine-readable format to allow the subsequent use of corpus analysis tools in their treatment and analysis. This conversion process yielded some changes to the texts in terms of format, content and structure, something which required a subsequent manual proofread. Given the goals of the study, certain elements such as editorial information, page numbers, figures and charts were removed as they were considered irrelevant. In addition and for the same reason, all of the texts were manually segmented at sentence level to make the subsequent consultation and analysis using AntPConc and AntConc easier, while the English and Spanish text pairs of the parallel corpus were also aligned at sentence level by hand. Due to the considerable size of the corpus and its sub-corpora, manual analysis of the complete texts was not possible. For this reason, excerpts taken from the original books were selected at random and then manually analysed, which allowed for a subsequent study by concordancing the previously obtained data. Details on the innovative methodology followed in the selection of excerpts, which involved statistical calculation methods, will be provided. The methodology described above, despite being framed in a particular research study, can be easily applied in other corpus-based studies of a similar nature.