

The Web for Corpus and the Web as Corpus in Translator Training¹

Miriam Buendía-Castro, Clara Inés López-Rodríguez
University of Granada, SPAIN

ABSTRACT

Corpora are rich information sources that can provide the translator with both linguistic and conceptual knowledge that is not found in dictionaries. The question that arises within this context is whether the web can be considered as a corpus. Following the distinction made by De Schryver (2002), there are two corpus-based approaches to the web: (i) web for corpus (WfC), in which the web is used as a source of texts in digital format for the subsequent implementation of an offline corpus; (ii) web as corpus (WaC), which uses the web directly as a corpus. In this paper, we compare and evaluate as translation aid tools, “automatically built” corpora (both general, i.e. WaCky corpora, and specialized, i.e. corpora created by the students themselves through WebBootCat, which are accessible through Sketch Engine), as opposed to the manual building of corpora. To that end, we asked two groups of students in the Translation and Interpreting Degree Program at the University of Granada (Spain) to carry out a technical translation assignment. One of the groups used automatically built corpora, whereas the other group used the web to manually extract texts for the later compilation of corpora. The results obtained showed that these two methods are complementary, and that students should decide for one or the other depending on their needs (i.e. translation assignment, novelty of the translation, directionality and specificity of the translation, time allotted, or level of analysis required).

KEYWORDS: corpora, corpus query system, specialized translation, Web as Corpus, Web for Corpus.

¹ A Spanish short version of this study was published in *TransKom 4 (1)* (López and Buendía 2011). This English version explains the procedure in more detail, and discusses the findings more fully.
Miriam Buendía-Castro and Clara Inés López-Rodríguez, The Web for Corpus and the Web as Corpus in Translator Training, Abstract.