A Cost-Benefit Analysis of Using Online Corpora for Translation Research

James St. André
The Chinese University of Hong Kong

Abstract

While there are an increasing number of electronic databases that contain large corpora potentially useful to researchers in translation studies, they come in a bewildering variety of size, coverage, reliability, and accessibility. For these and other reasons, researchers in translation studies have been slow to make use of such databases, despite the fact that corpus-based translation studies are now well established. This article seeks to explore a set of three interlocking questions regarding the usefulness of large, preexisting online corpora for research in translation studies. First, how useful are such databases for full-text searches and the compilation of corpora? Second, to what extent are privately compiled corpora now reinventing the wheel? Finally, what are the costs and benefits of using online databases versus privately compiled corpora? Through the use of a pilot study, this investigation seeks to make researchers aware of some of the major structural issues involved while working with large online databases. In particular, it is demonstrated that the purpose for which such databases were built means that, when adapting them for use in translation studies, considerable care must be taken to ensure that the researcher is aware of a wide variety of factors influencing the structure, accuracy, and usefulness of the material.

Keywords

English-language corpora, corpus studies, OCR conversion, corpus compilation, online databases