

Arabic-English Parallel Corpus: A New Resource for Translation Training and Language Teaching

Hind M. Alotaibi

Collage of Languages & Translation
King Saud University, Saudi Arabia

Abstract

Parallel corpora can be defined as collections of aligned, translated texts of two or more languages. They play a major role in translation and contrastive studies, and are also becoming popular in translation training and language teaching, with the advent of the data-driven learning (DDL) approach. Despite their significance, however, Arabic seems to lack a satisfactory general-use parallel corpus resource. The literature describes few Arabic–English parallel corpora, and these few are usually inaccurate and/or expensive. Some are small in size, while others are restricted in terms of genre, failing to meet the requirements of many academics and researchers. This paper describes an ongoing project at the College of Languages and Translation, King Saud University, to compile a 10-million-word Arabic–English parallel corpus to be used as a resource for translation training and language teaching. The bidirectional corpus can be used to compare translated and source language and identify differences. The corpus has been manually verified at different stages, including translation, text segmentation, alignment, and file preparation; it is available as full-text in XML format and through a user-friendly web interface that provides a concordancer to support bilingual search queries and several filtering options.

Keywords: Arabic, data-driven learning, English, language teaching, parallel corpus, translation training

Cite as: Alotaibi, H. M. (2017). Arabic-English Parallel Corpus: A New Resource for Translation Training and Language Teaching. *Arab World English Journal*, 8 (3). DOI: <https://dx.doi.org/10.24093/awej/vol8no3.21>