Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327
مجلة اللغة الانكليزية في العالم العربي

Arab World English Journal (AWEJ) Special Issue on Translation No.5 May, 2016

Pp. 317- 336

Bilingual Lexicon Extraction from Arabic-English Parallel Corpora with a View to Machine Translation

Yasser Muhammad Naguib Sabtan
Department of Languages and Translation,
Dhofar University, Oman
&
Faculty of Languages and Translation,
Al-Azhar University, Egypt

Abstract

Automatic extraction of bilingual lexicons from parallel corpora has been recently exploited to overcome the knowledge acquisition bottleneck in a number of research areas in natural language processing, such as machine translation (MT) and cross-language information retrieval. In this paper the author presents a method for automatic extraction of bilingual lexicons from a parallel Arabic-English corpus annotated with part-of-speech tags for potential use in a full-scale MT system. The extraction method, which does not make use of an initial bilingual dictionary, is based on the statistical technique of co-occurrence frequency in the parallel corpus. In addition, dependency relations for some parallel syntactic constructions are made use of to automatically extract head-dependent translation pairs which are then filtered to obtain one-word translation seeds. These seeds are used as anchor points to resegment the parallel corpus in order to bootstrap the lexicon extraction process. Experimental results show that the accuracy of the extracted lexicons was improved after applying the bootstrapping techniques.

Keywords: Arabic machine translation, bilingual lexicon extraction, computational linguistics, natural language processing, parallel corpora