

Validating Terminologies and Phraseological Units Retrieved from Specialized Comparable Corpora in Lexical Semantics: An Interactive Method

The methods of extracting terminologies and phraseological units from comparable corpora in previous studies were inclined to employ statistical machine translation or computational analysis. However, it was apparently insufficient for ensuring semantic level of obtained terms. Accordingly, the present study will apply an interactive method for cross validation of the quality of retrieved terms and phraseological units from specialized comparable corpora. The specialized comparable corpora consist of a Buddhist English Corpus and a Buddhist Chinese Corpus. The proposed interactive method includes filtering the terms and phraseological units with criteria; validating terms and phraseological units with different references sources including Google search engine, English dictionaries, Chinese dictionaries, and Pali dictionaries; implementing various statistical measures such as absolute frequency, log-likelihood (LL), odds ratio (OR) for ensuring the distinctness of obtained terms, and machine translation for comparing terminologies and phraseological units. Additionally, the distinct terminologies and phraseological units extracted from specialized Buddhist comparable corpora are examined in lexical semantics. Furthermore, the change of trend in Eastern and Western Buddhist literature will be explored through comparing the extracted terms and phraseological units from the Buddhist comparable corpora. Based on the results, the study will provide some suggestions for future studies.

Keywords: phraseological units, comparable corpora, lexical semantics