Statistical Comparison of Machine and Manual Translation: A Corpus-Based Study of Gone with the Wind

Authors : Yanmeng Liu

Abstract : This article analyzes and compares the linguistic differences between machine translation and manual translation, through a case study of the book Gone with the Wind. As an important carrier of human feeling and thinking, the literature translation poses a huge difficulty for machine translation, and it is supposed to expose distinct translation features apart from manual translation. In order to display linguistic features objectively, tentative uses of computerized and statistical evidence to the systematic investigation of large scale translation corpora by using quantitative methods have been deployed. This study compiles bilingual corpus with four versions of Chinese translations of the book Gone with the Wind, namely, Piao by Chunhai Fan, Piao by Huairen Huang, translations by Google Translation and Baidu Translation. After processing the corpus with the software of Stanford Segmenter, Stanford Postagger, and AntConc, etc., the study analyzes linguistic data and answers the following questions: 1. How does the machine translation differ from manual translation linguistically? 2. Why do these deviances happen? This paper combines translation study with the knowledge of corpus linguistics, and concretes divergent linguistic dimensions in translated text analysis, in order to present linguistic deviances in manual and machine translation. Consequently, this study provides a more accurate and more fine-grained understanding of machine translation products, and it also proposes several suggestions for machine translation development in the future.

Keywords : corpus-based analysis, linguistic deviances, machine translation, statistical evidence

Conference Title : ICIT 2020 : International Conference on Interpreting and Translation

Conference Location : Copenhagen, Denmark

Conference Dates : June 11-12, 2020