Zhang Tong, Xu Hong✉
Harbin Institute of Technology
Harbin, China
yuqilihit@126.com

Designing and applying language corpus for annotating translation

Abstract
Modern computer technology of corpus linguistics enables to perform deeper analysis of annotating translation techniques while translating from Russian into Chinese. Language corpus can be used to describe and analyze a considerable amount of translations. Using corpus statictics on various types of annotated information, we can determine the frequency of occurrence for each category of information about an original literary work reflected by a translator in comments. We designed our own parallel corpus composed of (i) three iconic novels by three famous Russian classical writers – "Crime and punishment" by F. M. Dostoevski, "Resorection" by L. N. Tolstoy, "Master and Margarita" by M. Bulgakov, (ii) three translations to each original novel, (iii) annotations to these translations. Statistical analysis enabled to identify common and varying patterns in the annotating techniques. The common feature was the highest frequency of comments providing cultural information included in the original text, however the number of actualized categories of comments as well as the number of comments in each category turned out variable. We also discovered three basic challenges of the existing annotations: excessive details, inaccuracy of information, free approach to translation.

Keywords: annotating translation technique, language corpus, information category, statistic analysis