Corpora for Legal Translation: Compilation and Analysis

DOI 10.6342/NTU201603145

parallel and monolingual corpora
corpus analysis
tools
translation reference tool
legal translation
legal English

Corpora, the well-organized bodies of "naturally occurring language data" sampled to represent a variety of language (McEnery, 2003, p. 449), have been making a growing impact in the field of translation (Bernardini, Stewart, & Zanettin, 2003). Scholars have asserted the immense value of corpora as reference tool for translation practice in specialized subject domains (Bowker & Pearson, 2002; Varantola, 2003), where intrinsic features of the language may cause difficulties for the translator. To address the potential lacking in reference tools for specialized translation assignments, this study explores a number of methods and computerized tools in compiling and analyzing a parallel and monolingual corpus of Chinese and English legislation. Incorporating semi-automated tools for text processing, part-of-speech tagging, sentence alignment, and phrasal alignment, this study utilizes keyword analysis, n-gram and n-gram part-of-speech sequence, as well as bilingual and monolingual concordance search to address identification of terminology equivalents, stylistic features, usage patterns, and translation strategies for legal contexts. Findings suggest that with the proposed methods, the corpus compiled in this study could effectively provide a number of information to aid the work of legal translators. The information identified can also be applied to compiling other forms of translation resources. It is hoped that in future research, the corpus tools and approaches employed in this study can be applied to facilitating other specialized fields of translation, and that preliminary findings observed here could be further explored to benefit future work in this discipline.


