

Human or machine translationese? A corpus-based comparative study of cohesion and coherence in Chinese-English translations

Ms. Yanfang SU, Ms. Ho Ling KWOK, and Dr. Kanglong LIU

Abstract

Previous studies have identified the existence of unique linguistic traits, often referred to as "translationese," in translated texts, setting them apart from non-translated works. Among these, learner and machine translations have shown a higher occurrence of translationese compared to professional translations (Kunilovskaya, Morgoun, & Pariy, 2018; Siu, 2023). In particular, machine translations have faced challenges in handling text cohesion and coherence (Niu, Jiang & Zhou, 2020; Wong & Kit, 2012). However, the recent advent of Large Language Models (LLMs) like GPT-4, with their enhanced abilities of contextual understanding and generation of fluent texts, has piqued significant interest across various research fields. Despite this, the potential of LLMs in the domain of translation remains largely underexplored.

To shed light on this potential, the primary objective of this study is to investigate the extent to which LLMs can mitigate translationese concerning text cohesion and coherence, offering insights into their potential role in translation education. The study conducts a comparative analysis of human and machine "translationese" in three different types of translations: learner translations, machine translations, and GPT-polished machine translations. Non-translated texts were used as a reference for comparison. To evaluate coherence and cohesion, various metrics, including type-token ratio indices, adjacent overlap indices, and connectives indices, are measured using the Textual Analysis and Annotation of Cohesion (TAACO) tool. This analysis reveals the unique challenges faced by human and machine translations in maintaining text cohesion and coherence. The findings provide valuable insights for translation practitioners to adapt to the growing influence of LLMs and shed light on how educators can leverage LLMs to empower translation teaching and learning.

Biographies

Ms. Yanfang SU is a PhD candidate at the Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University. Her research interests include corpus linguistics, corpus-based translation studies, and computer-assisted language learning. She has previously published in linguistic journals such as *Assessing Writing*, *Language Learning & Technology*, *System* and contributed book chapters on corpus-based translation studies.

Ms. Ho Ling KWOK is a PhD candidate at the Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University. Her research interests include psychological approaches to translation, corpus-based translation studies, and the application of technology in translation pedagogy.

Dr. Kanglong LIU is Assistant Professor at the Department of Chinese and Bilingual Studies of the Hong Kong Polytechnic University. He specialises in corpus-based translation studies and his main interests include empirical approaches to translation studies, translation pedagogy and corpus-based translation research. He is currently Associate Editor of *Translation Quarterly*, the official publication of the Hong Kong Translation Society.