

Multi-word Expressions and Multiple Renditions in Translation – Old Challenges and New Solutions through Big Data

Benjamin K. Tsou

City University of Hong Kong, Hong Kong, China

The Hong Kong University of Science and Technology, Hong Kong, China

Recent advances in Artificial Intelligence and technology have irreversibly transformed the landscape in professional translation such that manual translations are increasingly proceeded by Machine Translation (MT), and the translator's role is often changed into one of quality controller who does varying amounts of post-editing. This change has been necessitated by not only the need to obtain relief from repetitive manual work but to gain speed and consistency in quality. In the end, the final quality of translation and thornier issues in cross-linguistic intermediation remain very much in human hands. These developments also have implications for translation education.

The affordable cost and increasingly acceptable quality of MT systems have enabled MT to take over many routine and common tasks of simple prose translation. Reactions to these developments have been mixed, ranging from the lamentations of seasoned professionals to the overrated joy of the uninitiated. There is a timely need to reassess what challenges and opportunities exist for translators in the Age of Big Data and AI, especially in Asia.

We shall trace the basic workflow of the translator from pre-editing to post-editing, and explore how Big Data and technology could render assistance. Assuming the translator is already well versed in the structures of the source and target languages, she/he is well equipped to explore new areas of translation except for the important gaps in specialized lexical knowledge inevitably required in many new areas. This is especially the case in the translation of Multi-Word Expressions (MWE's) involving scientific and technical subjects because of non-uniqueness of multiple renditions in the cross-lingual terms. The absence of up-to-date lexical references has added complications, and the time perennial problem has resulted from technological advancements having always outpaced lexicography.

We shall outline the rationale behind an innovative lexical platform, PATENTLEX which could expand the productivity of the translator and her/his range of subjects. A description will be made of the successful construction of PATENTLEX, to show how it draws on Big Data and AI. This will be made on the basis of 300+K parallel Chinese-English patents and more than 30 million bilingually aligned sentence pairs extracted from them. We shall also report on the efficacy of the prototype system and offer trial runs.

References:

- Lu, B., Tsou, B. K., Jiang, T., Zhu, J., & Kwong, O. Y. (2011). Mining parallel knowledge from comparable patents. In W. Wong, W. Liu & M. Bannamoun (Eds.), *Ontology learning and knowledge discovery using the web: Challenges and recent advances* (pp. 247-271). IGI Global.
- Shen, G. W. (2001). The creation of technical terms in English Chinese dictionaries from the nineteenth century. In M. Lackner, I. Amelung & J. Kurtz (Eds.), *New Terms for New Ideas: Western Knowledge and Lexical Change in Late Imperial China* (pp. 287-304). Brill.
- Tsou, B., Chow, K.P., Lee, J., Yip, K. F., Ji, Y. X., & Wu, K. (2020). Bilingual Multi-word Expressions, Multiple-correspondence, and their Cultivation from Parallel Patents: The Chinese-English Case. In *PACLIC 34, Workshop on MWE's*, Hanoi, Vietnam.
- Tsou, B. K. & Kwong, O. Y. (2015). LIVAC as a monitoring corpus for tracking trends beyond linguistics. In B. K. Tsou & O. Y. Kwong (Eds.), *Linguistic Corpus and Corpus Linguistics in the Chinese Context (Journal of Chinese Linguistics Monograph Series 25)* (pp. 447-471). The Chinese University of Hong Kong Press.
- Tsou, B. K. (2001). Language Contact and Lexical Innovation. In M. Lackner, I. Amelung & J. Kurtz (Eds.), *New Terms for New Ideas: Western Knowledge and Lexical Change in Late Imperial China* (pp. 35-56). Brill.