Pre-editing plus neural machine translation for subtitling of TED Talks: an investigation into pre-editing quality

Masaru Yamada
Yusuke Hiraoka
Kansai University

In the advent of neural machine translation engines which are available online such as Google Translate, monolingual pre-editing of subtitling is deemed more realistic than ever. Also increased demand for video hosting services such as YouTube and TED Talks require faster dissemination of video contents in multiple languages with subtitling. In order to meet these requirements, monolingual pre-editing of the source contents may be better suited in combined use with the neural machine translation or NMT. Since NMT is not yet capable of producing high-quality translation or subtitling for certain language combinations such as English and Japanese, it is worthwhile focusing particularly on this language combination to carry out an investigation into pre-editing plus NMT.
In this study, the authorsThis study developed and evaluated a set of pre-editing rules for TED Talks subtitling to translate Japanese transcript subtitle into English. The simplified rules optimized for NMT (@TexTra® Minnano jido hon'yaku) are simplified as they are specifically intended for use by a monolingual pre-editor of original content to be disseminated in English. The rules, which are are 1a) insert punctuation 2b) make implied subjects/s and objects explicit, and 3c) write proper nouns in English, , which were built based on our previous research on bilingual pre-editing of YouTubers' videoonline contents (Hiraoka, 2018).
The effectiveness of the rules was investigated in terms of both translation quality and readability, e.g., 21-character limit specified in the TED subtitling guidelines. As a result, quality improvement was confirmed significant, e.g., good enough quality MT outputs increased from 12 to 40%. It is also confirmed that pre-editing with the set of rules did not affect the CPS requirement. However, in some cases, no changes or even degrade in quality were observed, which means the pre-edited MT output contains low quality translationsAlthough some improvements in quality were observed, however, there was not much difference in the amount of low quality translations, which are is needed to be solved for overall 'good enough quality' translation. TFor the next step, therefore, the present research aims at developing specific pre-editing rules to improve these low quality further analysis of the low quality MT outputs. will be necessary, looking for specified pre-editing rules and definition of 'good enough quality translation' for subtitling.

References

Hiraoka, Yusuke. 2018. The Application Possibility of Neural Machine Translation for Subtitling. In Proceedings of the 19th conference of the Japan Association for Interpreting and Translation Studies, page 22.