How might machine translation influence the development of a language?

Proposer(s) / Proposatzailea(k): Nora Aranberri

Contact / Kontaktua: nora.aranberri@ehu.eus

Description / Deskribapena

Machine translation (MT) quality is better and better. The translation industry and professional translators are starting to adopt the technology for certain domains, contexts and text types. There are also more and more freely available MT systems and as a result, regular users have also started to take advantage of them more and more often. However, we still do not know what influence MT output has in all these users' writing: how does MT shape the final version of a text? How does this final version of a text differ from what the users would have written without it? What is the consequence for users of differing levels of competence in the source and target languages? Should there be any recurrent characteristic in the writings produced using MT, will this have any effect on the development of the target language itself? This project aims to contribute to the study of how MT might shape the language we use and read by analyzing the differences between texts produced with and without MT.

Goals / Helburuak

To study the difference in texts produced with and without the use of MT

Requirements / Betebeharrak

Linguistic background, basic programming skills, proficiency in at least two languages

Framework / Esparrua

Machine translation for users

Tasks and plan / Atazak eta plana

- Analyse literature on language features, translation and post-editing effects, translationese and post-editese

- Identify texts (1) originally written in a language, (2) machine translated into that language (we can do this ourselves), (3) manually translated into that language, and (4) post-edited into that language (essential 1 and 2)

- Compare the linguistic and textual features of the different texts, in particular 1 and 2 above

- Analyse the results: what differences can we identify? Are there any differences in lexical and structural variation and diversity? Are the same linguistic resources used in the set-ups?

- Write up the report

References

Bawden, R., Sennrich, R., Birch, A., & Haddow, B. (2017). Evaluating discourse phenomena in neural machine translation. arXiv preprint arXiv:1711.00513.

Farrell, Michael. (2018). Machine Translation Markers in Post-Edited Machine Translation Output. InProceedings of the 40th Conference Translating and theComputer, pages 50–59.

Green, S., Heer, J. and Manning, C. (2013). The efficacy of human post-editing for language translation.Chi 2013, pages 439–448.

Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2017). Fine-grained human evaluation of neural versus phrase-based machine translation. The Prague Bulletin of Mathematical Linguistics, 108(1), 121-132.

Otero, P. G., Loinaz, I. A., & Campos, J. R. P. (2019). Distancia diacrónica interlingüística: aplicación al portugués y el castellano. Procesamiento del lenguaje natural, (63), 77-84.

Toral, A. (2019). Post-editese: an Exacerbated Translationese. In Proceedings ofthe Machine Translation Summit XVII, Dublin, Ireland, 20-23 August, pages 273-281.

Toury, Gideon. (2012). Descriptive translation studies and beyond: Revised edition, volume 100. JohnBenjamins Publishing