Informatics, 2021, vol.8, N1

Assessing human post-editing efforts to compare the performance of three machine translation engines for english to Russian translation of cochrane plain language health information: Results of a randomised comparison

Ziganshina L.E., Yudina E.V., Gabdrakhmanov A.I., Ried J.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

Cochrane produces independent research to improve healthcare decisions. It translates its research summaries into different languages to enable wider access, relying largely on volunteers. Machine translation (MT) could facilitate efficiency in Cochrane's low-resource environment. We compared three off-the-shelf machine translation engines (MTEs)-DeepL, Google Translate and Microsoft Translator-for Russian translations of Cochrane plain language summaries (PLSs) by assessing the quantitative human post-editing effort within an established translation workflow and quality assurance process. 30 PLSs each were pre-translated with one of the three MTEs. Ten volunteer translators post-edited nine randomly assigned PLSs each-three per MTE-in their usual translation system, Memsource. Two editors performed a second editing step. Memsource's Machine Translation Quality Estimation (MTQE) feature provided an artificial intelligence (AI)-powered estimate of how much editing would be required for each PLS, and the analysis feature calculated the amount of human editing after each editing step. Google Translate performed the best with highest average quality estimates for its initial MT output, and the lowest amount of human post-editing. DeepL performed slightly worse, and Microsoft Translator worst. Future developments in MT research and the associated industry may change our results.

http://dx.doi.org/10.3390/informatics8010009

Keywords

Cochrane plain language summaries, Cochrane russia, Deepl, Google translate, Health domain, Language translation, Machine translation, Machine translation quality, Microsoft translator, Post-editing, Russian language, Volunteer translation

References

[1] Useem, J.; Brennan, A.; LaValley, M.; Vickery, M.; Ameli, O.; Reinen, N.; Gill, C.J. Systematic Differences between Cochrane and Non-Cochrane Meta-Analyses on the Same Topic: A Matched Pair Analysis. PLoS ONE 2015, 10, e0144980.

[2] Koletsi, D.; Fleming, P.S.; Michelaki, I.; Pandis, N. Heterogeneity in Cochrane and non-Cochrane meta-analyses in orthodontics. J. Dent. 2018, 74, 90-94.

[3] von Elm, E.; Ravaud, P.; Maclehose, H.; Mbuagbaw, L.; Garner, P.; Ried, J.; Bonfill, X. Translating Cochrane reviews to ensure that healthcare decision-making is informed by high-quality research evidence. PLoS Med. 2013, 10, e1001516.

[4] Ziganshina, L.E.; Jørgensen, K.J.; Kazan, C. The first ever Cochrane event in Russia and Russian speaking countries-Cochrane Russia Launch-Evidence-based medicine: Achievements and barriers (QiQUM 2015) International Conference, Kazan, 7-8 December 2015. Int. J. Risk Saf. Med. 2016, 28, 221-226.

[5] Stix, G. The elusive goal of machine translation. Sci. Am. 2006, 294, 92-95.

[6] Zulfiqar, S.; Wahab, M.F.; Sarwar, M.I.; Lieberwirth, I. Is Machine Translation a Reliable Tool for Reading German Scientific Databases and Research Articles? J. Chem. Inf. Model. 2018, 58, 2214-2223.

[7] Wołk, K.; Marasek, K.; Glinkowski, W. Telemedicine as a special case of machine translation. Comput. Med. Imaging Graph. 2015, 46 Pt 2, 249-256.

[8] Dew, K.N.; Turner, A.M.; Choi, Y.K.; Bosold, A.; Kirchhoff, K. Development of machine translation technology for assisting health communication: A systematic review. J. Biomed. Inform. 2018, 85, 56-67.

[9] Dew, K.; Turner, A.M.; Desai, L.; Martin, N.; Laurenzi, A.; Kirchhoff, K. PHAST: A Collaborative Machine Translation and Post-Editing Tool for Public Health. AMIA Annu. Symp. Proc. 2015, 2015, 492-501.

[10] Soto, X.; Perez-de-Viñaspre, O.; Labaka, G.; Oronoz, M. Neural machine translation of clinical texts between long distance languages. J. Am. Med. Inform. Assoc. 2019, 26, 1478-1487.

[11] Wu, C.; Xia, F.; Deleger, L.; Solti, I. Statistical machine translation for biomedical text: Are we there yet? AMIA Annu. Symp. Proc. 2011, 2011, 1290-1299.

[12] Health in My Language. 2020. Available online: http://www.himl.eu/ (accessed on 31 January 2021).

[13] Martikainen, H. Post-Editing Neural MT in Medical LSP: Lexico-Grammatical Patterns and Distortion in the Communication of Specialized Knowledge. Informatics 2019, 6, 26.

[14] Das, P.; Kuznetsova, A.; Zhu, M.; Milanaik, R. Dangers of Machine Translation: The Need for Professionally Translated Anticipatory Guidance Resources for Limited English Proficiency Caregivers. Clin. Pediatr. 2019, 58, 247-249.

[15] Lund, H.; Juhl, C.B.; Nørgaard, B.; Draborg, E.; Henriksen, M.; Andreasen, J.; Christensen, R.; Nasser, M.; Ciliska, D.; Tugwell, P.; et al. Using an evidence-based research approach to place your results into context after the study is performed to ensure usefulness of the conclusion. J. Clin. Epidemiol. 2020, 129, 167-171.

[16] Lund, H.; Juhl, C.B.; Nørgaard, B.; Draborg, E.; Henriksen, M.; Andreasen, J.; Christensen, R.; Nasser, M.; Ciliska, D.; Clarke, M.; et al. Using an evidence-based research approach before a new study is conducted to ensure value. J. Clin. Epidemiol. 2020, 129, 158-166.

[17] Lund, H.; Juhl, C. Doing meaningful systematic reviews is no gravy train. Lancet 2020, 395, 1905.

[18] Lund, H.; Brunnhuber, K.; Juhl, C.; Robinson, K.; Leenaars, M.; Dorch, B.F.; Jamtvedt, G.; Nortvedt, M.W.; Christensen, R.; Chalmers, I. Towards evidence based research. BMJ 2016, 355, i5440.

[19] Memsource. Machine Translation Quality Estimation. Memsource Help Center. 2020. Available online: https://help.memsource. com/hc/en-us/articles/360012527380 (accessed on 31 January 2021).

[20] Memsource. Analysis Overview. Memsource Help Center. 2020. Available online: https://help.memsource.com/hc/en-us/ articles/360013675760 (accessed on 31 January 2021).

[21] Wikipedia. "DeepL Translator". 2021. Available online: https://en.wikipedia.org/wiki/DeepL_Translator (accessed on 31 January 2021).

[22] Wikipedia. Google Neural Machine Translation. 2021. Available online: https://en.wikipedia.org/wiki/Google_Neural_ Machine_Translation (accessed on 31 January 2021).

[23] Wikipedia. Microsoft Translator. 2021. Available online: https://en.wikipedia.org/wiki/Microsoft_Translator (accessed on 31 January 2021).

[24] TAUS. Automated MT Evaluation Metrics. 2021. Available online: https://blog.taus.net/knowledgehub/automated-mtevaluation- metrics (accessed on 31 January 2021).

[25] Jelicic Kadic, A.; Fidahic, M.; Vujcic, M.; Saric, F.; Propadalo, I.; Marelja, I.; Dosenovic, S.; Puljak, L. Cochrane plain language summaries are highly heterogeneous with low adherence to the standards. BMC Med. Res. Methodol. 2016, 16.

[26] DeepL. DeepL Pro-Terms and Conditions. 2021. Available online: https://www.deepl.com/en/pro-license/ (accessed on 31 January 2021).