Open-Source LLMs vs. NMT Systems: Translating Spatial Language in EN-PT-br Subtitles

Rafael Fernandes rafael.macario@usp.br
Marcos Lopes marcoslopes@usp.br
Department of Linguistics, University of São Paulo, São Paulo, 05508-220, BR

Abstract

This work, originating as part of a master's thesis, investigates the challenges of translating spatial language using open-source Large Language Models (LLMs) compared to traditional Neural Machine Translation (NMT) systems. It focuses on the accurate translation of two preposition pair – ACROSS and THROUGH, and INTO and ONTO – which present overlapping meanings when translating from English to Brazilian Portuguese (EN-PT-br). Correctly translating these prepositions is crucial for maintaining the source text's semantic integrity while ensuring fluency and adherence to the target language's lexicalization patterns (House, 2014, 2018; Talmy, 2000a,b; Slobin, 2005). The research contextualizes the challenges of spatial language translation, highlighting NMT limitations and potential LLM advantages. A comprehensive literature review traces the evolution of translation theories, NMT development, and the rise of LLMs, while also discussing the limitations of these approaches. The methodology involves a corpus-based analysis using a bilingual dataset centered on spatial prepositions from TED Talks subtitles sourced from the OPUS platform. This dataset was meticulously pre-processed for automated metrics calculation and manual error analysis. The evaluation metrics used include BLEU, METEOR, BERTScore, COMET, and TER, while the manual analysis identifies and categorizes specific types of mistranslation errors. The findings reveal that moderate-sized LLMs, such as LLaMa-3-8B and Mixtral-8x7B, achieve accuracy comparable to NMT systems like DeepL. However, this relationship between architecture and performance might not always linear; for instance, Gemma-7B, despite being heavily penalized by automatic metrics, performed similarly to more robust models in human reviews. LLMs generally exhibited serious translation issues, including interlanguage/code-switching (in) and anglicism (an), often failing to convey fluency in the target language. DeepL, on the other hand, demonstrated better accuracy and precision in this domain. Nevertheless, manual error analysis highlights ongoing challenges in translating spatial language, with both LLMs and NMT systems consistently making errors related to polysemy (po) and syntactic projection (sp), where they either fail to translate a preposition's meaning accurately or replicate the source language's lexicalization patterns (Fernandes et al., 2024; Oliveira and Fernandes, 2022), accounting for 27.84% of preposition-related errors. The study concludes that despite advancements, significant challenges remain in translating spatial language for this language pair. It suggests that future research should focus on enhancing and curating training datasets, refining model architectures, and developing more sophisticated evaluation metrics that better capture the subtleties of spatial language. This study contributes to the field by providing a detailed comparison of model performance in spatial language translation from EN-PT-br and proposing directions for future improvements.

Keywords

Natural Language Processing (NLP), Open-source Large Language Models (LLMs), Neural Machine Translation (NMT), Machine Translation (MT) Evaluation, Spatial Semantics, Polysemy, Language Typology

References

Fernandes, R., Souza, R., Lopes, M., Santos, P., and Finbow, T. (2024). Spatial information challenges in English to Portuguese machine translation. In Gamallo, P., Claro, D., Teixeira, A., Real, L., Garcia, M., Oliveira, H. G., and Amaro, R., editors, *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, pages 620–626, Santiago de Compostela, Galicia/Spain. Association for Computational Linguistics.

House, J. (2014). Translation quality assessment: Past and present. In *Translation: A multidisciplinary approach*, pages 241–264. Springer.

House, J. (2018). *Translation: The Basics*. Routledge.

Oliveira, A. and Fernandes, R. (2022). *Expressing complex paths of motion in Brazilian Portuguese: a closer look at frog stories*, pages 21–35. In: Chiappara, Juan Pablo and Siqueira, Joelma Santana and Oliveira, Aparecida de Araújo and Gediel, Ana Luisa.

Slobin, D. (2005). Relating narrative events in translation. In Ravid, D. and Shyldkrot, H., editors, *Perspectives on language and language development: Essays in honor of Ruth A. Berman*, pages 115–129. Springer, Dordrecht.

Talmy, L. (2000a). *Toward a cognitive semantics: Concept structuring systems*, volume 1. MIT Press.

Talmy, L. (2000b). *Toward a cognitive semantics: Typology and process in concept structuring*, volume 2. MIT Press.