

Cross-Dialect Sentence Transformation: A Comparative Analysis of Language Models for Adapting Sentences to British English

Authors : Shashwat Mookherjee, Shruti Dutta

Abstract : This study explores linguistic distinctions among American, Indian, and Irish English dialects and assesses various Language Models (LLMs) in their ability to generate British English translations from these dialects. Using cosine similarity analysis, the study measures the linguistic proximity between original British English translations and those produced by LLMs for each dialect. The findings reveal that Indian and Irish English translations maintain notably high similarity scores, suggesting strong linguistic alignment with British English. In contrast, American English exhibits slightly lower similarity, reflecting its distinct linguistic traits. Additionally, the choice of LLM significantly impacts translation quality, with Llama-2-70b consistently demonstrating superior performance. The study underscores the importance of selecting the right model for dialect translation, emphasizing the role of linguistic expertise and contextual understanding in achieving accurate translations.

Keywords : cross-dialect translation, language models, linguistic similarity, multilingual NLP

Conference Title : ICELL 2024 : International Conference on English Language and Linguistics

Conference Location : Rome, Italy

Conference Dates : April 04-05, 2024