Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

EIFFEL: Un corpus d'expressions idiomatiques françaises pour évaluer les biais anglocentriques des GMLs

Charlotte Noel[1] Nicholas Asher[2] Olivier Gouvert[1]
Farah Benamara[3,4] Julie Hunter[1]
(1) LINAGORA, Toulouse, France,
(2) IRIT-CNRS, ANITI Cluster, Toulouse, France,
(3) IRIT, Université de Toulouse, CNRS, Toulouse INP, Toulouse, France,
(4) IPAL, CNRS-NUS-A*STAR, Singapore
Correspondance: jhunter@linagora.com

RÉSUMÉ

Les GMLs multilingues populaires sont généralement entraînés sur de plus grande proportions de données anglaises que de données multilingues, ce qui soulève des questions quant à leur capacité à saisir les particularités linguistiques propres à ces autres langues ainsi qu'à saisir leurs informations culturelles spécifiques. Nous contribuons ainsi à un effort visant à accroître la sensibilité multilingue des GMLs en développant un *benchmark*, EIFFEL, qui teste la maîtrise des expressions idiomatiques françaises en contexte. Nous détaillons la méthodologie employée, incluant la participation de locuteurs natifs français, afin de la rendre reproductible dans d'autres langues. Nous comparons les GMLs multilingues populaires aux GMLs axés sur le français à la fois sur des *benchmarks* standards et sur EIFFEL. EIFFEL met en évidence les avantages d'une proportion plus élevée de données en français et montre les limites des *benchmarks* standards pour mesurer le multilinguisme.

ABSTRACT

EIFFEL : a novel benchmark to measure bias of English heavy training on French idiomatic expressions

Mainstream multilingual LLMs are generally trained on a much higher proportion of English than multilingual data, raising questions about their ability to capture linguistic features particular to non-English languages or to capture information important to non-anglophone cultures. We add to a growing effort to increase multilingual sensitivity in LLMs by developing a benchmark, EIFFEL, testing mastery of French idiomatic expressions in context. We fully explain the methodology, which exploits input from native French speakers, to make it reproducible for other languages. We compare mainstream multilingual LLMs with French-focused LLMs both on standard LLM benchmarks and EIFFEL ; EIFFEL brings out the benefits of higher proportions of French data and shows limitations of standard benchmarks for measuring multilingual competence.

MOTS-CLÉS : multilingualité, évaluation, GMLs, benchmarks, expressions idiomatiques.
KEYWORDS: multilinguality, evaluation, LLMs, benchmarks, idiomatic expressions.

ARTICLE ACCEPTÉ À : The 63rd Annual Meeting of the Association for Computational Linguistics.
URL : https://hal.science/hal-05619209
VERSION FRANÇAISE (ABRÉGÉE) : https://hal.science/hal-05618822

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 14–14.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.