

Exploration de stratégies de prédiction de la complexité lexicale en contexte multilingue à l'aide de modèles de langage génératifs et d'approches supervisées.

Abdelhak Kelious¹, Mathieu Constant¹, Christophe Coeur²

¹University of Lorraine and CNRS/ATILF, ²Consultant

abdelhak.kelious@univ-lorraine.fr

mathieu.constant@univ-lorraine.fr

christophe.coeur@gmail.com

RÉSUMÉ

Cet article explore des méthodes permettant de prédire automatiquement la complexité lexicale dans un contexte multilingue à l'aide de modèles avancés de traitement automatique du langage naturel. Plus précisément, il étudie l'utilisation de l'apprentissage par transfert et des techniques d'augmentation de données dans un cadre d'apprentissage supervisé, mettant en lumière l'intérêt notable des approches multilingues. Nous évaluons également le potentiel des grands modèles de langage génératifs pour la prédiction de la complexité lexicale. À travers différentes stratégies de requête (zero-shot, one-shot et prompts avec raisonnement en chaîne), nous analysons les performances des modèles dans plusieurs langues. Nos résultats montrent que, bien que les modèles génératifs obtiennent des performances prometteuses, leur qualité prédictive reste variable, et les modèles optimisés pour une tâche spécifique continuent de les surpasser lorsqu'ils disposent de données d'entraînement suffisantes.

ABSTRACT

Investigating strategies for lexical complexity prediction in a multilingual setting using generative language models and supervised approaches [Kelious et al. \(2024\)](#)¹

This paper explores methods to automatically predict lexical complexity in a multilingual setting using advanced natural language processing models. More precisely, it investigates the use of transfer learning and data augmentation techniques in the context of supervised learning, showing the great interest of multilingual approaches. We also assess the potential of generative large language models for predicting lexical complexity. Through different prompting strategies (zero-shot, one-shot, and chain-of-thought prompts), we analyze model performance in diverse languages. Our findings reveal that while generative models achieve promising performances, their predictive quality varies and optimized task-specific models still outperform them when they benefit from sufficient training data.

MOTS-CLÉS : Complexité lexicale, multilingue, modèles de langage génératifs, apprentissage supervisé, prompt engineering.

KEYWORDS: Lexical complexity, multilingual, generative language models, supervised learning, prompt engineering.

ARTICLE : **Accepté à** Swedish Language Technology Conference and NLP4CALL.

1. Ce document est un résumé de l'article [Kelious et al. \(2024\)](#)

Références

KELIOUS A., CONSTANT M. & COEUR C. (2024). Investigating strategies for lexical complexity prediction in a multilingual setting using generative language models and supervised approaches. In *Swedish Language Technology Conference and NLP4CALL*, p. 96–114.