

Représenter le style au-delà des thématiques : une étude d’impact sur la dispersion vectorielle de différents modèles de langage

Benjamin Icard¹ Evangelia Zve^{1, 2}

Lila Sainero¹ Alice Breton¹ Jean-Gabriel Ganascia¹

(1) LIP6, Sorbonne Université, CNRS, 4 Place Jussieu, 75005 Paris, France

(2) Infopro Digital, 20 Rue des Aqueducs, 94250 Gentilly, France

{prenom.nom}@lip6.fr

RÉSUMÉ

Cet article vise à étudier comment le style d’écriture influence la dispersion des plongements vectoriels de divers grands modèles de langage. Alors que les premiers modèles de type transformeur étaient principalement axés sur la modélisation thématique, cette étude examine le rôle du style d’écriture dans la configuration de l’espace vectoriel. À partir d’un corpus littéraire faisant varier thématiques et styles, nous comparons la sensibilité des modèles de langage en français et en anglais. En analysant ainsi l’impact spécifique du style sur la dispersion vectorielle, nous cherchons à mieux comprendre comment les modèles de langage traitent l’information stylistique, contribuant ainsi à leur interprétabilité globale. Ceci est un résumé de l’article “*Embedding Style Beyond Topics: Analyzing Dispersion Effects Across Different Language Models*” publié dans les actes de la conférence COLING 2025 (Icard *et al.*, 2025) et accessible à l’URL : <https://aclanthology.org/2025.coling-main.236/>.

ABSTRACT

Embedding Style Beyond Topics : Analyzing Dispersion Effects Across Different Language Models

This paper analyzes how writing style affects the dispersion of embedding vectors across multiple, state-of-the-art language models. While early transformer models primarily aligned with topic modeling, this study examines the role of writing style in shaping embedding spaces. Using a literary corpus that alternates between topics and styles, we compare the sensitivity of language models across French and English. By analyzing the particular impact of style on embedding dispersion, we aim to better understand how language models process stylistic information, contributing to their overall interpretability.

MOTS-CLÉS : modèles de langage, plongements vectoriels, modélisation thématique, style, dispersion spatiale, explicabilité, analyse stylistique.

KEYWORDS: language models, vector embeddings, topic modeling, writing style, spatial dispersion, explainability, stylistic analysis.

ARTICLE : **Accepté à COLING 2025.**

Références

ICARD B., ZVE E., SAINERO L., BRETON A. & GANASCIA J.-G. (2025). Embedding style beyond topics : Analyzing dispersion effects across different language models. In *Proceedings of the 31st International Conference on Computational Linguistics*, p. 3511–3522.