Measuring Embedding Sensitivity to Authorial Style in French: Comparing Literary Texts with Language Model Rewritings

Benjamin Icard[1] **Lila Sainero**[1] **Alice Breton**[1]
Evangelia Zve[1,2] **Jean-Gabriel Ganascia**[1]

[1] LIP6, Sorbonne University, CNRS, France
[2] Infopro Digital, France

Abstract

Large language models (LLMs) can convincingly imitate human writing styles, yet it remains unclear how much stylistic information is encoded in embeddings from any language model and retained after LLM rewriting. We investigate these questions in French, using a controlled literary dataset to quantify the effect of stylistic variation via changes in embedding dispersion. We observe that embeddings reliably capture authorial stylistic features and that these signals persist after rewriting, while also exhibiting LLM-specific patterns. These analytical results offer promising directions for authorship imitation detection in the era of language models.

1 Introduction

Computational stylometry has long been central to digital humanities, using natural language processing (NLP) for authorship attribution and stylistic comparison (Stamatatos, 2009; Koppel et al., 2011). This has involved quantifying inter-textual distance and developing feature-based methods to support author verification (Savoy, 2012a,b; Cafiero and Camps, 2019).

In recent years, embedding methods, particularly Transformer-based contextual models (Vaswani et al., 2017; Devlin et al., 2019), have improved literary authorship attribution through richer textual representations of authorial profiles (Terreau et al., 2021; Kim et al., 2025). Yet embedding-based analysis often centers on semantic tasks, especially topic modeling (Bianchi et al., 2021) and content similarity (Rockmore et al., 2025), rather than style in embeddings, aside from a few studies (Wegmann et al., 2022; Icard et al., 2025). Assessing more systematically how embeddings capture authorial style would improve the explainability of embedding representations and strengthen authorship characterization and attribution in computational stylometry.

The increasing capabilities of large language models (LLMs) to imitate human-authored styles make this issue more urgent. Recent work on style transfer shows that LLMs can reproduce salient stylistic features in literary settings (Mikros, 2025), but also that they still struggle to imitate more implicit writing styles in extra-literary settings (Wang et al., 2025). Because LLMs are language models, this reinforces the need to assess the sensitivity of embedding vectors to style, and the extent to which LLMs preserve or alter that sensitivity during style transfer (Huang et al., 2025).

This paper reports a controlled experiment on French literary texts evaluating embedding sensitivity to target stylistic features in human-authored texts and their LLM imitations. We compile a dataset of French excerpts from Tufféry (Tufféry, 2000), Proust (Proust, 1913), Céline (Céline, 1932), and Yourcenar (Yourcenar, 1951), together with GPT-, Mistral-, and Gemini-based stylistic imitations under fixed-topic conditions. We encode all texts with thirteen embedding models and quantify aggregated stylistic effects on embedding dispersion across human authors and imitated texts.

Section 2 reviews prior work on embedding sensitivity to writing style, with particular attention to the human–LLM distinction. Section 3 introduces our French literary dataset and the evaluation of style transfer under LLM generation. Section 4 presents the vectorization of the dataset with thirteen embedding models, a structural evaluation with clustering, and results on how writing style affect embedding dispersion in human-authored texts and LLM imitations. Section 5 examines LLM-specific effects on embedding sensitivity, bringing explainability to the style transfer evaluation used to validate the dataset. Finally, Section 6 concludes and outlines directions for future work.

All reproducibility materials and results are available at: `https://github.com/sma-libra/style-embedding-sensitivity`

Proceedings of the 6th International Conference on Natural Language Processing for the Digital Humanities, pages 69–82

2 Related Work

Feature-Based Stylometry. Computational stylometry classically relies on surface features such as word frequencies, character n-grams (Cavnar and Trenkle, 1994; Ríos-Toledo et al., 2022) and punctuation patterns (Faye et al., 2024), often represented with TF-IDF term weighting (Salton and Buckley, 1988; Bui et al., 2011), to capture authorial voice (Verma and Srinivasan, 2019; Herrmann et al., 2021; Mani, 2022).

Style Embeddings. More recent work examines whether embedding-based representations also encode feature-based stylistic information (Liu et al., 2024). Terreau et al. (2021) propose an author-verification framework that tests whether embedding spaces encode stylistic features rather than mainly semantic content. Using this framework, they show that specialized Doc2Vec-based author embeddings (Le and Mikolov, 2014; Ganesh et al., 2016; Maharjan et al., 2019) are often more semantically driven, while simpler pretrained sentence encoders such as USE-DAN (Cer et al., 2018) and SBERT (Reimers and Gurevych, 2019) can perform better on stylistic feature families.

Measuring Sensitivity. A related line of work addresses style-semantics conflation more directly through topic-controlled or style-sensitive embedding representations. More specifically, Wegmann et al. (2022) show that content-controlled training improves style-topic separation in BERT-based representations. Adjacent work by Chen et al. (2023) shows that style-sensitive encoders can support efficient detection of stylistic shifts in multi-author documents. In parallel, Patel et al. (2023) introduce the style embedding model LISA whose dimensions correspond to stylistic features. Most directly related to our approach, Icard et al. (2025) use Queneau's fixed-topic variations to measure embedding sensitivity to style under rewriting by a single LLM, rather than focusing specifically on author-defining features and their preservation.

LLM Style Transfer. Recent research on LLMs' representation and high-fidelity transfer of literary style is motivated by their demonstrated capacity to imitate authorial writing. In literary contexts, Mikros (2025) show that GPT-4o reproduces salient authorial stylistic features under thematic control, while Sarfati et al. (2025) and Hicke and Mimno (2025) show that literary style is reflected in LLM internal representations. In parallel, Huang et al. (2024) study LLMs for authorship analysis, while Horvitz et al. (2024) propose TinyStyler, a lightweight approach to few-shot style transfer conditioned on author embeddings.

Taken together, these works have examined embedding sensitivity to style in smaller language models and, more recently, in LLMs. However, little work has investigated whether embeddings are *reliably* sensitive to the defining stylistic features of specific authors, and to what extent this sensitivity remains measurable *after* LLM rewriting. We investigate this question in French using stylistically marked and diverse literary texts.

3 Dataset

3.1 Text Materials

We assembled a dataset of 1,248 French literary texts that combines human-authored originals with LLM-generated imitations that reproduce the human authors' styles under fixed-topic conditions.

3.1.1 Reference Corpus

We first compiled a *reference corpus* of 384 French original literary texts, separated in two groups according to topic and style variation.

The first group, TUFFERY_REF, consists of 96 texts extracted from Stéphane Tufféry's *Le style mode d'emploi* (Tufféry, 2000).[1] We use this collection because, like Raymond Queneau's *Exercices de style* (Queneau, 1947) used in Icard et al. (2025), it keeps a single topic across all texts: the story of a bus journey in Paris. But unlike Queneau's variations, however, Tufféry structures stylistic variation more systematically, including explicit pastiches of major French authors such as Balzac, Flaubert, and Hugo. TUFFERY_REF therefore provides a fixed-topic, style-diverse baseline for stylometric comparison.

The second group, STYLE_REF, comprises 288 texts of length comparable to TUFFERY_REF, evenly distributed across three subclasses:

- PROUST_REF: 96 texts from Proust's *Du côté de chez Swann*, the first volume of *À la recherche du temps perdu* (Proust, 1913);
- CELINE_REF: 96 texts from Céline's *Voyage au bout de la nuit* (Céline, 1932);
- YOURCENAR_REF: 96 texts from Yourcenar's *Mémoires d'Hadrien* (Yourcenar, 1951).

[1] We excluded 3 of the 99 texts from Tufféry's *Le style mode d'emploi* when constructing TUFFERY_REF because they departed from the bus-journey topic.

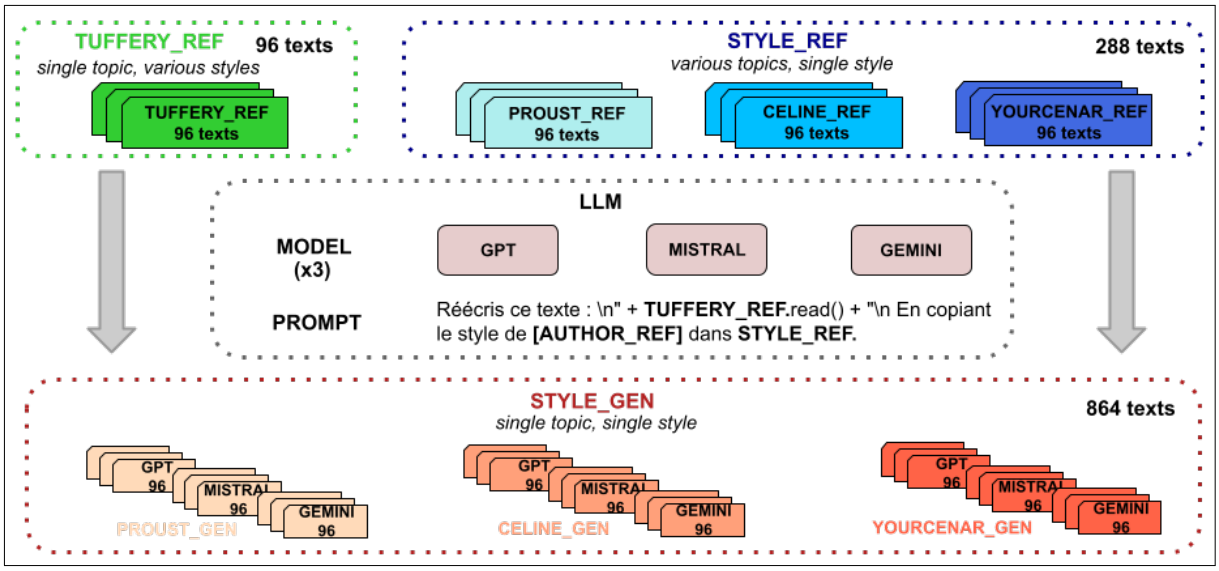


Figure 1: Textual corpora and generation scheme used to construct the STYLOGEN embedding dataset. The reference corpora are TUFFERY_REF (single topic, various styles) and STYLE_REF (various topics, single style per author). Three LLMs generate STYLE_GEN by rewriting Tufféry texts in the styles of the authors in STYLE_REF, using the French prompt shown.

The authors were selected to maximize stylistic contrast across four central dimensions in French literature (Magri-Mourgues, 2006; Brunet, 2014; Hough, 2016): structural complexity, morphosyntactic richness, lexical diversity, and referential anchoring. Proust exemplifies structural complexity (long sentences, high average word length) and morphosyntactic richness (diverse POS patterns), with recurrent proper-name allusions (named entities) (Brunet, 1982). Céline foregrounds a spoken register (short sentences and words, high pronoun/verb share) alongside lexical diversity and nonstandard forms (higher entropy, distinctive character distributions) (Zábojniková, 2006; Styhre, 2011; de Sacy, 2025). Yourcenar sustains balanced prose densely anchored in historical and geographic reference (named-entity density) (Colvin, 2005; Broche, 2022).

3.2 Generated Corpus

We constructed the generated corpus, STYLE_GEN, by prompting LLMs to rewrite the 96 texts from TUFFERY_REF in the writing styles of PROUST_REF, CELINE_REF, and YOURCENAR_REF, respectively. To ensure both diversity of behavior and architectural variety, we employed three LLMs for each target style: `GPT-4o`, `mistral-large-2411`, and `gemini-1.5-flash`. For simplicity's sake, we now refer to these LLM versions as GPT, MISTRAL, and GEMINI.

The group STYLE_GEN comprises 864 texts evenly distributed across three subclasses:

- PROUST_GEN: 288 texts produced by prompting each of the three generative models to rewrite Tufféry's texts in the style of Proust's *Du côté de chez Swann*;
- CELINE_GEN: 288 texts produced by prompting each of the three generative models to rewrite Tufféry's texts in the style of Céline's *Voyage au bout de la nuit*;
- YOURCENAR_GEN: 288 texts produced by prompting each of the three generative models to rewrite Tufféry's texts in the style of Yourcenar's *Mémoires d'Hadrien*.

Figure 1 presents an overview of the pipeline followed to assemble our textual corpora, including the French prompt used. The English translation of the prompt is: *"Rewrite this text: \n"* + **TUFFERY_REF**.*read()* + *"\n By copying the style of* [**AUTHOR_REF**] *into* **STYLE_REF**."

3.3 Style Transfer Evaluation

We assessed style transfer from STYLE_REF to STYLE_GEN as a function of the LLM used as the generator. Specifically, we evaluated the extent to which texts in PROUST_GEN, CELINE_GEN, and YOURCENAR_GEN were

Validation Set		PROUST_REF	CELINE_REF	YOURCENAR_REF	STYLE_REF (held-out 20%)
Corpus-Level	Macro-F1	—	—	—	0.965
	Accuracy	0.947	0.947	1.000	**0.966**

Test Set		PROUST_GEN	CELINE_GEN	YOURCENAR_GEN	STYLE_GEN (100%)
Corpus-Level	Macro-F1	—	—	—	0.669
	Transfer Accuracy	0.601	0.722	0.670	**0.664**
Per-Class	*when* GPT *is used*	0.490	0.542	0.646	0.559
	when MISTRAL *is used*	0.844	0.812	0.625	0.760
	when GEMINI *is used*	0.469	0.812	0.740	0.674

Table 1: Style transfer results with TF-IDF character 3-5-grams + `LinearSVC`. We report corpus-level performances on STYLE_REF (held-out 20%) and STYLE_GEN (100%), and per-class transfer accuracy for each target author label. We indicate corpus-level accuracy in bold, and the highest per-class transfer accuracy across LLMs in blue.

classified as PROUST_REF, CELINE_REF, and YOURCENAR_REF, respectively.

Character n-Gram Validator. For this task, we used a validator that represents each document with TF-IDF-weighted character 3-5-grams and classifies it with a linear support vector classifier (Cortes and Vapnik, 1995), implemented with `LinearSVC` from scikit-learn, into one of three author-style labels: PROUST_REF, CELINE_REF, YOURCENAR_REF. We chose this validator because character n-grams are strong surface-level features of authorial style (Kešelj et al., 2003; Stamatatos, 2009), and because it is embedding-independent, avoiding confounds in our later sensitivity analyses.[2]

Methodologically, we split STYLE_REF into train/validation sets and use STYLE_GEN as the test set:

— **Train/Validation:** using STYLE_REF only (i.e., PROUST_REF, CELINE_REF, and YOURCENAR_REF) split into 80% for train and 20% for validation (stratified by author, with seed = 42);

— **Test:** using 100% of STYLE_GEN (i.e., PROUST_GEN, CELINE_GEN, and YOURCENAR_GEN), evaluated without adaptation to measure style transfer from human to LLM-generated text.

Table 1 reports the validator's performances on the held-out STYLE_REF corpus (20%) and its *transfer* accuracy on the STYLE_GEN corpus (100%), broken down by imitated author class and by LLM used as a generator.

[2]Appendix A.1 (Table 3) reports additional style transfer results on the corpus using another linear `LinearSVC` classifier, this time with function-word frequencies as a classical stylometric baseline for authorial style (Burrows, 2002).

At the corpus level, the validator achieves near-ceiling accuracy on held-out human texts from STYLE_REF (Accuracy: 0.947 to 1.00). When applied to STYLE_GEN, transfer accuracy drops substantially, but remains well above the three-way chance level (1/3) (Accuracy: 0.664).

Transfer accuracy on generated texts varies with the LLM used for imitation. PROUST_GEN is best recognized when generated by MISTRAL (0.844), CELINE_GEN is best recognized when generated by MISTRAL or GEMINI (0.812 in both cases), and YOURCENAR_GEN is best recognized when generated by GEMINI (0.740). By contrast, GPT yields the lowest accuracies across the three generated corpora. The confusion matrix complementing these results is provided in Appendix A.2 (Figure 5).

Beyond style transfer validation from STYLE_REF to STYLE_GEN across LLM generators, we now analyze embedding representations of the same full textual corpora to compare embedding sensitivity to stylistic features before and after LLM rewriting.

4 Embedding Sensitivity to Writing Style

4.1 Dataset Vectorization

Embedding Model Selection. The reference corpora TUFFERY_REF and STYLE_REF, along with the generated corpus STYLE_GEN, were embedded using thirteen models (Table 2), with their embedding dimensionalities reported. Selection criteria were architectural and dimensional diversity, computational efficiency, and strong performance on the Massive Text Embedding Benchmark (MTEB; November 2025 version)[3] (Muennighoff et al., 2022). The resulting embedding dataset for all thirteen model is available in our data repository.

[3]https://huggingface.co/spaces/mteb/leaderboard

Clustering Validation. To assess embedding space separability across the three corpora (TUFFERY_REF, STYLE_REF, STYLE_GEN), we run k-means clustering (Hartigan and Wong, 1979) on the full-dimensional embeddings of the thirteen models, setting $k = 3$.[4] Clustering quality is evaluated using purity (Manning, 2008), an external score in $[0, 1]$ computed by comparing the induced clusters to the three ground-truth corpus labels (Soni and Dwivedi, 2024). Table 2 reports, for each embedding model, the purity score computed in the model's full-dimensional space (FullD).

Embedding Model	FullD	Purity
`xlm-roberta-large`	1024	0.7654
`multilingual-e5-large`	1024	0.6836
`mistral-embed`	1024	0.6809
`e5-base-v2`	768	0.6802
`distilbert-base-uncased`	768	0.6701
`text-embedding-3-small`	1536	0.6663
`text-embedding-004`	768	0.6539
`solon-embeddings-large-0.1`	1024	0.6524
`voyage-2`	1024	0.6493
`sentence-camembert-base`	768	0.6408
`all-roberta-large-v1`	1024	0.6242
`paraphrase-multilingual-mpnet-base-v2`	768	0.6080
`all-MiniLM-L12-v2`	384	0.5721

Table 2: Embedding dimensionality and k-means cluster purity ($k = 3$) for each model in full-dimensional space (FullD), computed with respect to the three corpus labels (TUFFERY_REF, STYLE_REF, STYLE_GEN). Models are ordered by purity scores.

Recovery of the three corpus groups is reasonable across the thirteen embedding models, with mean purity (0.6575) and a close median (0.6539). At the level of individual models, `xlm-roberta-large` scores highest (0.7654) and `all-MiniLM-L12-v2` lowest (0.5721), with intermediate models spread smoothly across this range.

UMAP Reduction. Beyond the consistent purity scores across models, Table 2 shows substantial variation in embedding dimensionality (from 384 to 1536). To enable comparisons across model-specific full-dimensional spaces (FullD), we apply Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018). For each target dimensionality, we repeat the projection thirty times with different random seeds to account for UMAP's stochasticity.

Focus on 2D UMAP. We tested UMAP reductions, specifically 2D, 3D, and 10D, to find the optimal sufficient dimension with respect to numerical fidelity to the FullD purity scores. To quantify alignment, we computed the mean absolute error (MAE) to FullD, as well as the maximum absolute error (MaxAE).

Among the UMAP reductions, we observed that 2D performed best, with the lowest MAE (0.025 for 2D, vs 0.036 for 3D and 0.029 for 10D) but also the smallest MaxAE (0.056 for 2D, vs 0.090 for 3D and 0.066 for 10D), making it the reduction best aligned with FullD purity. To visualize corpus-group structure, Figure 2 shows the 2D UMAP projection of the `xlm-roberta-large` embedding model, which, as in FullD, obtains the best clustering purity score (0.7627).

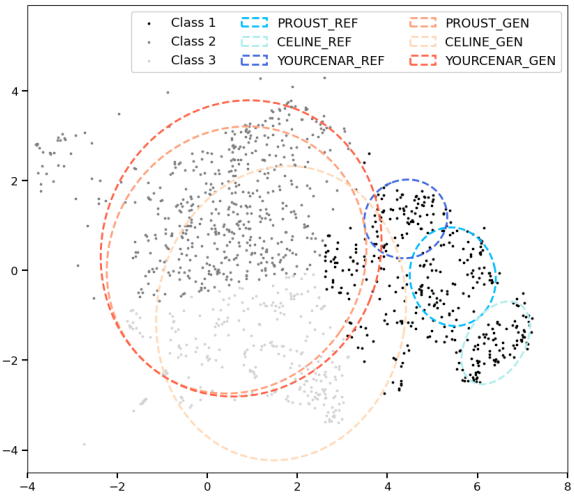


Figure 2: 2D UMAP projection of `xlm-roberta-large` embeddings on the dataset. Points indicate the three k-means clusters (Class 1-3), while dashed ellipses (visual guides, not k-means clusters) indicate label-based coverage regions for the human and generated corpora, drawn so that exactly 80% of the corpus lie inside the ellipse zone.

The dashed ellipses in Figure 2 suggest possible stylistic effects on the embedding representations. For PROUST_REF, CELINE_REF, and YOURCENAR_REF (right side of the figure), the clear separation between ellipses is consistent with differences in both topic and style across the human corpora, whereas the strong within-ellipse cohesion cannot be explained by topic, since topic varies within each corpus, and may instead reflect the homogenizing effect of shared authorial style. For PROUST_GEN, CELINE_GEN, and YOURCENAR_GEN (left side of the figure), the substantial overlap is consistent with intended topic alignment on TUFFERY_REF, but the overlap remains only partial, especially for CELINE_GEN,

[4] Using *scikit-learn* `KMeans` with default parameters and `random_state=0`.

suggesting that embeddings may retain residual stylistic differences after LLM rewriting.

These observations motivate measuring whether embeddings capture stylistic features beyond topic under fixed-topic rewriting. We now examine embedding sensitivity to author-characteristic features in French, both before and after LLM imitation.

4.2 Sensitivity Evaluation Metrics

Following Icard et al. (2025), we compute dispersion-based metrics on an aggregated 2D UMAP reduction derived from the thirteen embedding models listed in Table 2.

Embedding Dispersion. For the j-th UMAP iteration, we define $d_X^{(i,j)}$ as the Euclidean distance of the i-th embedding vector from the centroid $c_X^{(j)}$ of class X as follows:

$$d_X^{(i,j)} = \|v_X^{(i,j)} - c_X^{(j)}\| \quad (1)$$

where $v_X^{(i,j)}$ is the i-th embedding vector of class X in the j-th iteration and $\|\cdot\|$ is the Euclidean norm.

To capture the spatial dispersion of embeddings in the UMAP target space, we calculate the mean Euclidean distance from the centroid of each class X across all 30 UMAP iterations and all embeddings N in X, written $\bar{d}_X$:

$$\bar{d}_X = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{30}\sum_{j=1}^{30} d_X^{(i,j)}\right) \quad (2)$$

Target Stylistic Features. To analyze embedding sensitivity to stylistic variation in human texts and LLM rewritings, we must retain features that reflect the author-characteristic dimensions associated with Proust, Céline, and Yourcenar, as described in Section 3.1.1.

To this end, we drew on the stylometric framework proposed by Terreau et al. (2021), which is based on a comprehensive inventory of eight stylistic features families,[5] from which we retained a subset of five (*structural features*, *part-of-speech tags*, *indexes of lexical diversity*, *letter frequencies*, and *named entities*) most relevant to authorial style along the core dimensions previously considered:

- **Structural:** mean values of selected *structural features*, specifically word length and sentence length, normalized by text length;
- **Morphosyntax:** frequency of *part-of-speech tags* (nouns, verbs, adjectives) to capture grammatical diversity patterns;
- **Lexicon:** *lexical diversity* (using Shannon Entropy, see Shannon 1948) and *letter*-level patterns (character unigram and capitalization frequencies);
- **Referentiality:** density of *named entities* (e.g., persons, locations, organizations) per sentence, as a proxy for referential content.

For convenience, we write $f_X^s(i)$ to denote the average frequency value f of the stylistic feature s measured on the i-th document of class X.

Dispersion-Style Correlations. In the following analyzes, TUFFERY_REF serves as the reference class to study interactions between embedding dispersion and writing style on the corpus.

We denote by $\Delta d(\text{TUFFERY\_REF}, Y)$ the difference in embedding dispersion between TUFFERY_REF and another class Y (human or generated, here), and by $\Delta f^s(\text{TUFFERY\_REF}, Y)$ the difference in frequency of stylistic feature s between the two classes. More formally:

$$\Delta d(\text{TUFFERY\_REF}, Y) = d_{\text{TUFFERY\_REF}}(i) - d_Y(j) \quad (3)$$

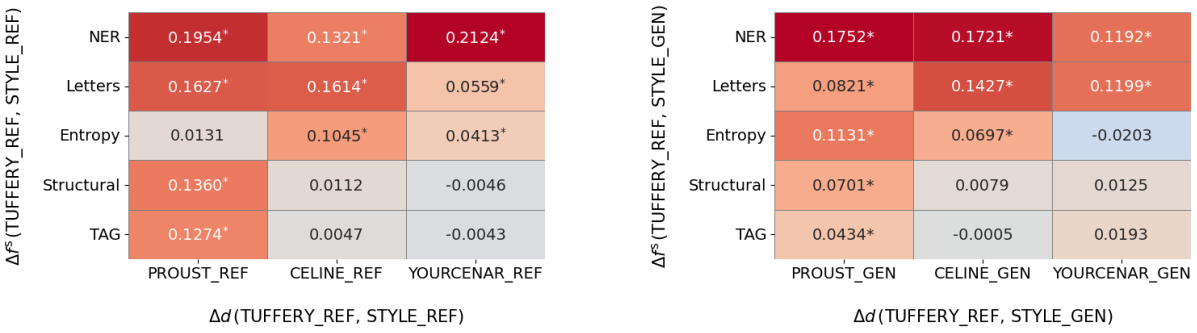
$$\Delta f^s(\text{TUFFERY\_REF}, Y) = f^s_{\text{TUFFERY\_REF}}(i) - f^s_Y(j) \quad (4)$$

where the comparison class Y is either a reference class $Y \in$ STYLE_REF or a generated class $Y \in$ STYLE_GEN, i is the i-th vector of TUFFERY_REF, and j is the j-th vector of class Y.

To evaluate embedding sensitivity to writing style, whether human (STYLE_REF) or model-generated (STYLE_GEN), we compute Pearson correlations, written r, between Δd and Δf^s for each stylistic feature s, relative to TUFFERY_REF.

Interpretation. We interpret the Δd–Δf^s correlations as distribution-level associations between dispersion shifts and stylometric shifts across corpora, rather than as causal estimates of rewriting effects (i.e., the indices i and j need not refer to paired source–rewrite items). In particular, for STYLE_GEN, the topic is controlled by construction via rewriting of TUFFERY_REF, so remaining variation is expected to reflect stylistic transfer and generator-specific rewriting behaviour specifically.

[5] The complete list and description of the eight feature families are available at: https://github.com/EnzoFleur/style_embedding_evaluation/.



(a) (b)

Figure 3: Pearson correlations (r) between embedding dispersion shifts in 2D UMAP reduction (Δd) and stylistic feature shifts (Δf^s) for each author-labeled corpus, comparing TUFFERY_REF with (a) the three human-authored corpora in STYLE_REF, and (b) the three style-imitated corpora in STYLE_GEN. Asterisks indicate $p < 0.01$ after Bonferroni correction.

4.3 Embedding Sensitivity to Authorial Style

To assess embedding sensitivity to human authorial style, we computed the 2D UMAP Δd–Δf^s correlations for comparisons between TUFFERY_REF and each author corpus PROUST_REF, CELINE_REF, and YOURCENAR_REF. Figure 3a reports the sensitivity correlations obtained for each author corpus, broken down by stylistic feature family. For transparency, we report the FullD correlations in Appendix A.3 (Figure 6a).[6]

For each human author, we observe moderate-to-weak yet significant correlations between embedding-dispersion shifts and stylistic features. PROUST_REF shows broad sensitivity across NER ($r = 0.195^*$), Letters ($r = 0.163^*$), Structural features ($r = 0.136^*$), and TAG ($r = 0.127^*$). CELINE_REF concentrates on Letters ($r = 0.161^*$), NER ($r = 0.132^*$), and Entropy ($r = 0.105^*$). Finally, YOURCENAR_REF is dominated by sensitivity to NER ($r = 0.212^*$) with minimal Letters and Entropy contributions. Overall, these correlations align with the expected authorial stylistic profiles of Proust (syntactic and structural complexity, morphosyntactic richness, proper-names allusions), Céline (spoken register with lexical unpredictability and nonstandard forms), and Yourcenar (strong referential density) described in subsection 3.1.1.

However, topic variation between STYLE_REF and TUFFERY_REF may confound these results. To isolate style from topic, we now examine STYLE_GEN, where topic is aligned with TUFFERY_REF while style varies through generative rewriting.

[6] The full set of raw Pearson correlations for 3D and 10D UMAP, in addition to 2D UMAP and FullD, is available in our GitHub repository.

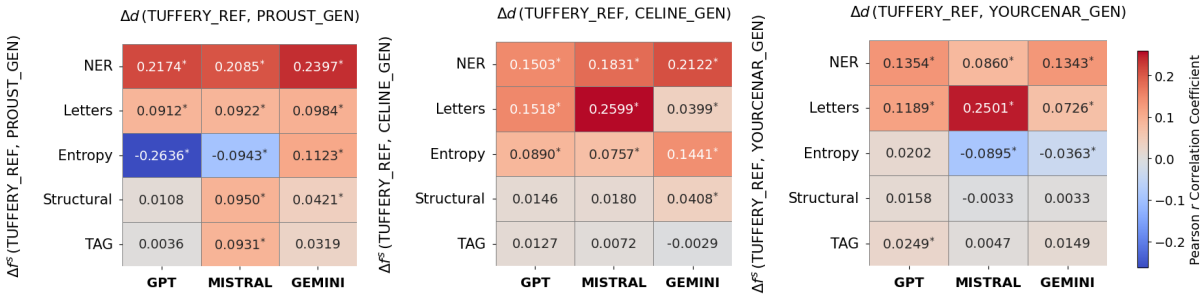
4.4 Sensitivity After LLM Rewriting

To assess embedding sensitivity in the context of LLM imitations, we computed the 2D UMAP Δd–Δf^s correlations for comparisons between TUFFERY_REF and each imitated-author corpus PROUST_GEN, CELINE_GEN, and YOURCENAR_GEN. Figure 3b reports the sensitivity correlations for each imitated-author corpus, broken down by stylistic feature family. For transparency again, we report the FullD sensitivity correlations in Appendix A.3 (Figure 6b).[7]

We observe that PROUST_GEN shows highest embedding sensitivity to NER ($r = 0.175^*$) and Entropy ($r = 0.113^*$), with weaker effects for Letters and Structural. CELINE_GEN concentrates on NER ($r = 0.172^*$) and Letters ($r = 0.143^*$). YOURCENAR_GEN emphasizes Letters ($r = 0.120^*$) and NER ($r = 0.119^*$).

Overall, STYLE_GEN exhibits moderate stylistic fidelity to the main authorial stylistic dimensions observed in STYLE_REF, but these dimensions are attenuated and reweighted. PROUST_GEN maintains the broad PROUST_REF stylistic profile in compressed form, with NER only slightly reduced and stronger reductions in Letters, Structural, and TAG. CELINE_GEN shows the highest stylistic fidelity among the generated classes, preserving spoken-style register of CELINE_REF (persistent Letters and weaker Entropy) while overweighting referentiality (NER). YOURCENAR_GEN retains the NER-based referential anchoring of YOURCENAR_REF, but more weakly, with partial dilution and a shift toward Letters.

[7] As before, the full set of raw Pearson correlations for all target dimensions is available in our GitHub repository.



(a) (b) (c)

Figure 4: Pearson correlations (r) between embedding dispersion shifts in 2D UMAP reduction (Δd) and stylistic feature shifts (Δf^s) per LLM imitator, comparing TUFFERY_REF with (a) PROUST_GEN, (b) CELINE_GEN, and (c) YOURCENAR_GEN. Asterisks indicate $p < 0.01$, with Bonferroni correction.

5 Discussion

LLM Stylistic Fidelity. Figure 4 reports LLM-level sensitivity for the three STYLE_GEN corpora and reveals author-specific preservation patterns hidden in the pooled results (Figure 3).

Figure 4a shows that MISTRAL achieves the highest stylistic fidelity for PROUST_GEN, preserving Proust's broad profile as measured in PROUST_REF. (NER, Letters, Structural, TAG). By contrast, Figure 4b reveals that GPT achieves higher stylistic fidelity for CELINE_GEN, best preserving Céline's core spoken-style pattern as observed in CELINE_REF (NER, Letters, Entropy). Finally, Figure 4c shows that GPT yields greater style fidelity to Yourcenar's defining features in case of YOURCENAR_GEN (NER, Letters).

Validation vs Stylistic Fidelity. These results on LLMs' fidelity to human style only partially align with the style transfer validation results reported in subsection 3.3. When evaluating style transfer from STYLE_REF to STYLE_GEN (Table 1), we obtained the highest transfer accuracy for Proust using MISTRAL, for Céline using MISTRAL and GEMINI, and for Yourcenar using GEMINI. However, when assessing stylistic fidelity to the target author profile by LLM (Figure 4), we find that while MISTRAL still best preserves Proust's broad stylistic profile, GPT shows higher stylistic fidelity for Céline than both MISTRAL and GEMINI, and higher stylistic fidelity for Yourcenar than GEMINI.

Explaining Metric Mismatch. The divergence stems from the two evaluations emphasizing different, though overlapping, stylistic features. Stylistic fidelity measures embeddings preservation of the broad author stylistic profile, while the 3-5-gram-based validator measures surface features, most directly Letters and indirectly NER and TAG.

For Proust, rankings align because MISTRAL best preserves the Letters-level features that dominate the validator, as well as the target author profile (NER, Letters, Structural, TAG). For Céline, MISTRAL and GEMINI obtain higher style transfer accuracy, consistent with their emphasizing Letters-level surface markers that are highly validator-salient, while GPT better preserves Céline's author-defining pattern (NER, Letters, Entropy). For Yourcenar, GEMINI similarly achieves higher validator accuracy by matching Letters-level features, whereas GPT better preserves Yourcenar's defining combination of NER and Letters.

6 Conclusion

Our study reveals moderate-to-weak embedding sensitivity to authorial style in French. Comparing literary texts from three authors with their LLM imitations, we find that embeddings broadly reflect author-characteristic stylistic patterns, although the preservation of specific features varies across authors and LLMs. By analyzing stylistic shifts in embedding space, our approach complements style transfer evaluation by tracing how specific features are preserved or altered by generative rewriting.

These findings remain preliminary and require replication across additional languages, broader author sets, and alternative stylistic features. Further analyses should also move beyond aggregated representations to examine more directly how individual embedding architectures encode stylistic features. More broadly, our results suggest that deviations from the human baseline after generative rewriting may help identify LLM-based imitation of authorial style through stylistic transformation patterns. We leave these questions for future work.

Limitations

This study has four main limitations. First, the analyses rely primarily on 2D UMAP because embedding dimensionality varies across models in FullD and because 2D yields the best alignment with both authorial style preservation and FullD purity scores. However, the distortions of authorial features observed in these dimensions require further investigation.

Second, the inventory of stylistic features we consider (NER, Letters, Entropy, Structural, POS Tags) is tailored to the current author sample rather than exhaustive, which may limit coverage of authorial stylistic signals beyond surface features.

Third, the stylistic correlations observed are consistently significant but moderate in magnitude and sometimes limited, calling for broader validation to ensure robustness.

Finally, our study focuses on French literary texts, so generalization to other languages and to non-literary materials in which style is also salient (e.g., correspondence, personal journals, speeches, periodicals), is needed.

Ethical Considerations

This work follows the principles of open science, AI transparency, and sustainability, with a strong emphasis on reproducibility and public access to results (lawful use under the EU text-and-data mining exception, Directive 2019/790/EC, Art. 3;[8] in the U.S. context, comparable research use would fall under transformative fair use[9]).

To support open science while respecting copyright regarding the text materials, we release only the 864 LLM-generated rewritings of Tufféry's texts *in-the-style-of* the three other authors (Proust, Céline, Yourcenar). These constitute transformative stylistic imitations consistent with fair use principles. We do not distribute the 384 original literary texts due to copyright restrictions. We do, however, release the full set of 16,224 vector embeddings (= 1,248 texts $\times$ 13 embedding models) (research-only license; reconstruction or re-identification attempts prohibited). These embeddings are provided for research use only, and we do not evaluate embedding inversion risk in this work.

To advance AI transparency, our GitHub repository releases all code, the 864 LLM-generated rewritings of Tufféry's texts, full set of 16,224 embeddings, and analytical results, in particular raw unadjusted Pearson correlations, with documentation for reproducibility.

To promote sustainability, we used the MTEB leaderboard (Hugging Face) to combine smaller and larger open-source pretrained models in a way that helps limit carbon emissions.

Acknowledgements

We thank Stéphane Tufféry for authorizing us to use the material from his book *Le style mode d'emploi* in our study, as well as two anonymous reviewers for helpful comments and feedback. This work was supported by the programs THEMIS (grant agreements n°DOS022279400 and n°DOS022279500) and TRUSTEDNEWS (ANR-25-ASM2-0003). EZ acknowledges Infopro Digital for supporting her PhD research, alongside her work.

Declaration of Contribution

BI conceptualized the research problem and designed the experiment. LS and AB managed the data collection and generation processes. LS was responsible for coding and testing the selected generation models, while BI managed the style evaluation models. BI analyzed and discussed the results, with LS and JGG. BI wrote the paper, which was read and revised collaboratively by all authors. Correspondence: benjamin.icard@lip6.fr.

References

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 759–766, Online. Association for Computational Linguistics.

Laurent Broche. 2022. Yourcenar et *Mémoires d'Hadrien* parmi les historiens. *Anabases*, 36:131–154.

Étienne Brunet. 1982. Le style de proust dans la recherche du temps perdu. etude quantitative. In *VII International Symposium of the Association for Literary and Linguistic Computing*, volume 3, pages 51–76. Giardini Editori.

[8] https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019L0790

[9] https://www.law.cornell.edu/uscode/text/17/107

Étienne Brunet. 2014. La lexicométrie française : naissance, évolution et perspectives. *Revue de l'Université de Moncton*, 45(1–2):13–33.

Quang Anh Bui, Muriel Visani, Sophea Prum, and Jean-Marc Ogier. 2011. Writer identification using tf-idf for cursive handwritten word recognition. In *2011 International Conference on Document Analysis and Recognition*, pages 844–848. IEEE.

John Burrows. 2002. 'delta': a measure of stylistic difference and a guide to likely authorship. *Literary and linguistic computing*, 17(3):267–287.

Florian Cafiero and Jean-Baptiste Camps. 2019. Why molière most likely did write his plays. *Science Advances*, 5(11):eaax5489.

William B. Cavnar and John M. Trenkle. 1994. N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, pages 161–175, Las Vegas, NV, USA.

Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, and 1 others. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

Haoyang Chen, Zhongyuan Han, Zengyao Li, and Yong Han. 2023. A writing style embedding based on contrastive learning for multi-author writing style analysis. In *CLEF 2023 Working Notes*. CEUR Workshop Proceedings.

Michael E. Colvin. 2005. *Baroque Fictions: Revisioning the Classical in Marguerite Yourcenar*, volume 271 of *Faux Titre*. Brill / Rodopi, Leiden.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20(3):273–297.

Louis-Ferdinand Céline. 1932. *Voyage au bout de la nuit*. Denoël et Steele.

Antoine Silvestre de Sacy. 2025. Hypersegmentation du discours chez louis-ferdinand céline. *Corpus*, (27).

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Géraud Faye, Benjamin Icard, Morgane Casanova, Julien Chanson, François Maine, François Bancilhon, Guillaume Gadek, Guillaume Gravier, and Paul Égré. 2024. Exposing propaganda: an analysis of stylistic cues comparing human annotations and machine classification. In *Proceedings of the Third Workshop on Understanding Implicit and Underspecified Language*, pages 62–72, Malta.

J Ganesh, Soumyajit Ganguly, Manish Gupta, Vasudeva Varma, and Vikram Pudi. 2016. Author2vec: Learning author representations by combining content and link information. In *WWW (Companion volume)*, pages 49–50.

John A Hartigan and Manchek A Wong. 1979. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108.

J Berenike Herrmann, Arthur M Jacobs, and Andrew Piper. 2021. Computational stylistics. *Handbook of Empirical Literary Studies*, pages 451–486.

Rebecca M. M. Hicke and David Mimno. 2025. Looking for the inner music: Probing llms' understanding of literary style. *Computational Humanities Research*, 1:e3.

Zachary Horvitz, Ajay Patel, Kanishk Singh, Chris Callison-Burch, Kathleen McKeown, and Zhou Yu. 2024. TinyStyler: Efficient few-shot text style transfer with authorship embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13376–13390.

Carole Hough, editor. 2016. *The Oxford Handbook of Names and Naming*. Oxford Handbooks in Linguistics. Oxford University Press, Oxford.

Baixiang Huang, Canyu Chen, and Kai Shu. 2024. Can large language models identify authorship? In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 445–460, Miami, Florida, USA. Association for Computational Linguistics.

Baixiang Huang, Canyu Chen, and Kai Shu. 2025. Authorship attribution in the era of llms: Problems, methodologies, and challenges. *ACM SIGKDD Explorations Newsletter*, 26(2):21–43.

Benjamin Icard, Evangelia Zve, Lila Sainero, Alice Breton, and Jean-Gabriel Ganascia. 2025. Embedding style beyond topics: Analyzing dispersion effects across different language models. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3511–3522, Abu Dhabi, UAE. Association for Computational Linguistics.

Vlado Kešelj, Fuchun Peng, Nick Cercone, and Calvin Thomas. 2003. N-gram-based author profiles for authorship attribution. In *Proceedings of the conference pacific association for computational linguistics, PACLING*, volume 3, pages 255–264.

Junghwan Kim, Haotian Zhang, and David Jurgens. 2025. Leveraging multilingual training for authorship representation: Enhancing generalization across languages and domains. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 34867–34892, Suzhou, China. Association for Computational Linguistics.

Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2011. Authorship attribution in the wild. *Language Resources and Evaluation*, 45(1):83–94.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.

Chang Liu, Zhongyuan Han, Haoyang Chen, and Qingbiao Hu. 2024. Team liuc0757 at pan: A writing style embedding method based on contrastive learning for multi-author writing style analysis. *Working Notes of CLEF*.

Véronique Magri-Mourgues. 2006. Corpus et stylistique. *Corpus*, (5). Special issue.

Suraj Maharjan, Deepthi Mave, Prasha Shrestha, Manuel Montes, Fabio A. González, and Thamar Solorio. 2019. Jointly learning author and annotated character n-gram embeddings: A case study in literary text. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*.

Inderjeet Mani. 2022. *Computational modeling of narrative*. Springer Nature.

Christopher D Manning. 2008. *Introduction to information retrieval*. Cambridge University Press.

Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861.

George Mikros. 2025. Beyond the surface: stylometric analysis of gpt-4o's capacity for literary style imitation. *Digital Scholarship in the Humanities*, 40(2):587–600.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. 2022. Mteb: Massive text embedding benchmark. *arXiv preprint arXiv:2210.07316*.

Ajay Patel, Delip Rao, Ansh Kothary, Kathleen McKeown, and Chris Callison-Burch. 2023. Learning interpretable style embeddings via prompting LLMs. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 15270–15290, Singapore. Association for Computational Linguistics.

Marcel Proust. 1913. *Du côté de chez Swann*, volume 1, chapter 1. Gallimard.

Raymond Queneau. 1947. Exercices de style: Edition gallimard. *Collection Folio*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Germán Ríos-Toledo, Juan Pablo Francisco Posadas-Durán, Grigori Sidorov, and Noé Alejandro Castro-Sánchez. 2022. Detection of changes in literary writing style using n-grams as style markers and supervised machine learning. *Plos one*, 17(7):e0267590.

Daniel Rockmore, Jiayi Chen, Mohammad Javad Latifi Jebelli, Allen Riddell, and Harrison Stropkay. 2025. On the literary landscapes of vector embeddings. *Computational Humanities Research*, 1:e18.

Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.

Raphaël Sarfati, Haley Moller, Toni J. B. Liu, Nicolas Boulle, and Christopher Earls. 2025. What's in a prompt? language models encode literary style in prompt embeddings. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 24059–24068, Suzhou, China. Association for Computational Linguistics.

Jacques Savoy. 2012a. Authorship attribution: A comparative study of three text corpora and three languages. *Journal of Quantitative Linguistics*, 19(2):132–161.

Jacques Savoy. 2012b. Authorship attribution based on specific vocabulary. *ACM Transactions on Information Systems*, 30(2):12:1–12:30.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Urvashi Soni and Sunita Dwivedi. 2024. Clutching of clustering validation criteria. *International Journal of Future Computer and Communication*, 13(1).

Efstathios Stamatatos. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3):538–556.

Alexander Styhre. 2011. Céline and the aesthetics of hyperbole: Style, points, parataxis and other literary devices. *ephemera: theory & politics in organization*, 11(3):258–270.

Enzo Terreau, Antoine Gourru, and Julien Velcin. 2021. Writing style author embedding evaluation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 84–93. Association for Computational Linguistics.

Stéphane Tufféry. 2000. Le style mode d'emploi. *Paris: Cylibris*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Gaurav Verma and Balaji Vasan Srinivasan. 2019. A lexical, syntactic, and semantic perspective for understanding style in text. *arXiv preprint arXiv:1909.08349*.

Zhengxiang Wang, Nafis Irtiza Tripto, Solha Park, Zhenzhen Li, and Jiawei Zhou. 2025. Catch me if you can? not yet: LLMs still struggle to imitate the implicit writing styles of everyday authors. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 10040–10055, Suzhou, China. Association for Computational Linguistics.

Anna Wegmann, Marijn Schraagen, and Dong Nguyen. 2022. Same author or just same topic? towards content-independent style representations. In *Proceedings of the 7th Workshop on Representation Learning for NLP*, pages 249–268. Association for Computational Linguistics.

Marguerite Yourcenar. 1951. *Mémoires d'Hadrien*. Plon.

Hviezdoslava Zábojniková. 2006. Louis-ferdinand céline et l'oral populaire. *Verbum: Analecta Neolatina*, 8(1):117–125.

Validation Set		PROUST_REF	CELINE_REF	YOURCENAR_REF	STYLE_REF (held-out 20%)
Corpus-Level	Macro-F1	—	—	—	0.825
	Accuracy	0.737	0.789	0.950	**0.828**

Test Set		PROUST_GEN	CELINE_GEN	YOURCENAR_GEN	STYLE_GEN (100%)
Corpus-Level	Macro-F1	—	—	—	0.528
	Transfer Accuracy	0.403	0.438	0.760	**0.534**
Per-Class	*when* GPT *is used*	0.260	0.385	0.854	0.500
	when MISTRAL *is used*	0.604	0.646	0.625	0.625
	when GEMINI *is used*	0.344	0.281	0.802	0.476

Table 3: Style transfer results with Function words Frequencies + `LinearSVC`. We report corpus-level performances on STYLE_REF (held-out 20%) and STYLE_GEN (100%), and per-class transfer accuracy for each target author label. We indicate corpus-level accuracy in bold, and the highest per-class transfer accuracy across LLMs in blue.

A Appendix

A.1 Style Transfer Evaluation Using Function Words

Table 3 reports style transfer performances obtained on the corpus with the validator using function word frequencies in combination to `LinearSVC`.

At the corpus level, this validator yields weaker but overall consistent corpus-level results compared to 3-5-gram-based validator, with lower performance on both STYLE_REF (0.828 vs. 0.966 accuracy; 0.825 vs. 0.965 Macro-F1) and STYLE_GEN (0.534 vs. 0.664 transfer accuracy; 0.528 vs. 0.669 Macro-F1), while preserving the same contrast between validation on reference texts and weaker transfer on generated texts.

Across reference authors (PROUST_REF, CELINE_REF, YOURCENAR_REF), both validators agree in assigning the highest accuracy to Yourcenar (0.950 vs. 1.000), although the function-word-based model is notably less accurate for Proust and Céline.

Across imitated author (PROUST_GEN, CELINE_GEN, YOURCENAR_GEN) and LLM combinations, the picture is partly stable and partly shifted: MISTRAL remains the best model for PROUST, and also remains best for CELINE, though without the earlier tie with GEMINI. For YOURCENAR, however, the best system shifts from GEMINI under the character *n*-gram validator to GPT under the function-word-based validator, while GEMINI remains strong.

A.2 Confusion Matrix for the 3-5-Gram-Based Validator

Figure 5 complements the aggregate performance results given in Table 1 for the 3-5-gram-based validator, showing the class-specific error structure on STYLE_GEN. The dominant errors are mutual confusions between PROUST_GEN and YOURCENAR_GEN (107 PROUST_GEN instances predicted as YOURCENAR_REF, and 73 YOURCENAR_GEN instances predicted as PROUST_REF), whereas misclassifications into CELINE_REF are comparatively rare for the other two styles (8 and 22).

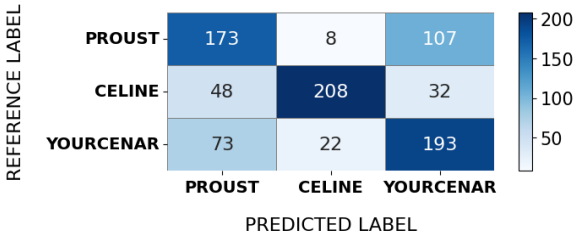
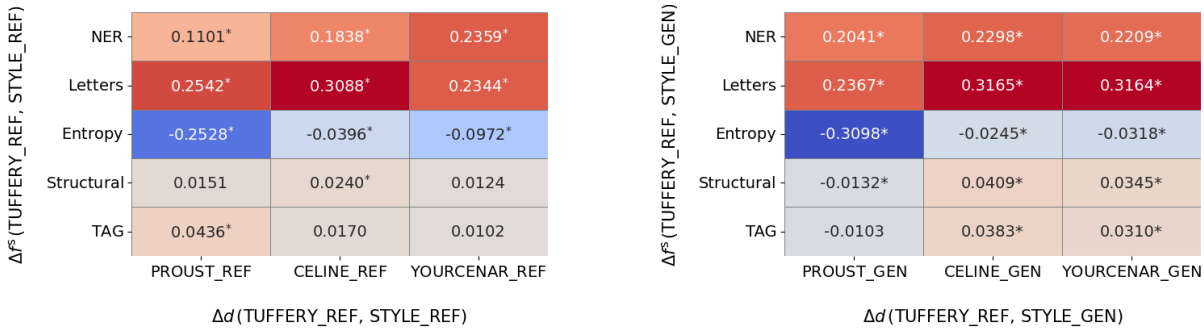


Figure 5: Confusion matrix for the STYLE_GEN evaluation of the TF-IDF character 3-5-gram + `LinearSVC` classifier trained on STYLE_REF, shown by imitated authorial style. Rows correspond to true author labels and columns to predicted labels.

A.3 FullD Sensitivity Correlations to Style

Figure 6 reports sensitivity correlations between changes in stylistic features and changes in FullD embedding dispersion across authors, comparing TUFFERY_REF with both human-written corpora (STYLE_REF) and style-imitated (STYLE_GEN) corpora.

Compared to 2D UMAP (Figure 3a), Figure 6a shows that many feature-wise correlations remain significant in FullD and some are actually stronger than in 2D UMAP, but the resulting profiles are less faithful to the authorial characteristics of the human authors. For PROUST_REF, the expected broad profile becomes distorted: Letters dominates, while Structural is nearly absent and NER is reduced. For CELINE_REF, the expected concentration on Letters, NER, and Entropy is weakened by a reversal



(a) (b)

Figure 6: Pearson correlations (r) between embedding dispersion shifts in FullD (Δd) and stylistic feature shifts (Δf^s) for each author-labeled corpus, comparing TUFFERY_REF with (a) the three human-authored corpora in STYLE_REF, and (b) the three style-imitated corpora in STYLE_GEN. Asterisks indicate $p < 0.01$ after Bonferroni correction.

of Entropy and the appearance of Structural. For YOURCENAR_REF, the expected referential dominance of NER becomes less distinctive because Letters rises to a comparable level and Entropy turns negative. Overall, compared with 2D UMAP, the FullD correlations differentiate the three human authors less clearly because they overweight shared features, especially Letters and, to a lesser extent, NER, while weakening the author-specific balance of stylistic features.

Compared to 2D UMAP (Figure 3b), Figure 6b shows that these distortions become even stronger after LLM rewriting. In PROUST_GEN, the profile narrows to NER and Letters, while Entropy becomes strongly negative and Structural and TAG largely disappear. In CELINE_GEN, Letters and NER remain prominent, but Entropy again turns negative and Structural and TAG become more visible than expected for Céline. In YOURCENAR_GEN, the profile becomes flatter and less specific than in 2D, with strong Letters and NER accompanied by positive Structural and TAG and negative Entropy.

FullD therefore preserves authorial characteristic features less well than 2D UMAP, both for human-authored texts and after LLM rewriting, because the more selective feature-level preservation observed in 2D is replaced by a more generic pattern shared across corpora. Our GitHub repository shows similar distortions with the 3D and 10D UMAP reductions.