The Influence of AI Tools on University Students' Writing Style: A Stylometric Analysis of Narrative Texts

Dimitris Bilianos1, Katerina Florou2,
1 National and Kapodistrian University of Athens, GREECE – DBILIANOS@ILL.UOA.GR
2 National and Kapodistrian University of Athens, GREECE – kathyflorou@ill.uoa.gr

ABSTRACT (ENGLISH)

This study investigates the impact of ChatGPT on the writing styles of university students learning Italian as a foreign language. Using a subset of the UniC corpus, which includes narrative texts from 2023 (produced with ChatGPT assistance) and 2024 (produced without AI tools under supervised conditions), we employed stylometric techniques such as Principal Component Analysis (PCA) and K-means clustering to analyze stylistic patterns. Results revealed that while stylistic groupings were evident, they did not align distinctly with ChatGPT usage or year of authorship. Texts from both years appeared in overlapping clusters, suggesting that individual writing habits and pedagogical influences may obscure the stylistic impact of AI assistance. These findings underscore the complexity of distinguishing AI-generated content from human-authored texts and highlight the nuanced interplay between AI tools and student writing practices. Limitations, including a small dataset and variability in ChatGPT usage, are discussed, along with suggestions for future research to further explore AI's role in educational contexts.
Keywords: Stylometric Analysis, AI-Generated Texts, Writing Style in Higher Education

ABSTRACT (ITALIANO)

L'impatto degli strumenti di intelligenza artificiale sullo stile di scrittura degli studenti universitari: Un'analisi stilometrica di testi narrativi

Questo studio analizza l'impatto di ChatGPT sullo stile di scrittura di studenti universitari che apprendono l'italiano come lingua straniera. Utilizzando un sottoinsieme del corpus UniC, che comprende testi narrativi del 2023 (redatti con l'assistenza di ChatGPT) e del 2024 (prodotti senza strumenti di intelligenza artificiale in condizioni supervisionate), sono state applicate tecniche stilometriche come l'Analisi delle Componenti Principali (PCA) e il clustering K-means per esaminare i modelli stilistici. I risultati mostrano che, sebbene siano emerse alcune aggregazioni stilistiche, queste non corrispondono in modo netto all'uso di ChatGPT o all'anno di composizione. Testi di entrambi gli anni si trovano in cluster sovrapposti, suggerendo che abitudini di scrittura individuali e influenze pedagogiche possano offuscare l'impatto stilistico degli strumenti di intelligenza artificiale. Questi risultati evidenziano la complessità di distinguere i contenuti generati dall'AI da quelli scritti interamente dagli studenti, sottolineando l'interazione sfumata tra strumenti di AI e pratiche di scrittura accademica. Sono discusse le limitazioni dello studio, inclusa la dimensione ridotta del campione e la variabilità nell'uso di ChatGPT, insieme a suggerimenti per ricerche future volte a indagare più a fondo il ruolo dell'intelligenza artificiale nei contesti educativi.
Parole chiave: Analisi Stilometrica, Testi Generati dall'Intelligenza Artificiale, Stile di Scrittura nell'Istruzione Universitaria

1. INTRODUCTION

Recent advancements in artificial intelligence (AI) have sparked increasing interest in their potential to transform educational environments. Among these, OpenAI platforms, particularly ChatGPT, have garnered significant attention for their role in enhancing distance and online learning (Bozic & Poola, 2023; Firat, 2023). The growing incorporation of AI tools in education underscores their potential to reshape teaching methodologies and learning experiences, offering new opportunities for both educators and students. A specialized body of research is emerging that focuses on ChatGPT's application as a natural language processing (NLP) tool in higher education, shedding light on its capacity to support academic research and its broader implications for pedagogical practices (Fuchs, 2023).

As students increasingly integrate ChatGPT into their language learning routines, they employ it for a wide range of tasks. These include generating original text, clarifying definitions, verifying spelling, and exploring complex syntactic structures. This shift suggests a significant departure from traditional language resources, with ChatGPT serving as a modern alternative to conventional tools such as dictionaries, grammar guides, and automated translation services like Google Translate (Aydin, 2024; Baskara & Mukarto, 2023). Moreover, ChatGPT's interactive nature offers learners an engaging platform to

practice and refine their speaking and comprehension skills, fostering a low-stress environment where students can freely experiment with vocabulary and grammar without fear of judgment from peers (Klimova et al., 2024; Vo & Nguyen, 2024).
However, despite its numerous advantages, ChatGPT is not without limitations. Ethical concerns have emerged, particularly regarding its potential for misuse in academic settings. One major issue is the risk that students may rely on ChatGPT to complete assignments by copying and pasting generated responses, circumventing academic integrity and teacher oversight (Nugroho et al., 2023). This limitation is central to our investigation, as it raises questions about how the tool may influence students' writing practices, particularly when used for both primary production and as a supplementary aid. As AI tools like ChatGPT continue to be integrated into academic practices, it also becomes increasingly important to develop reliable methods for distinguishing between AI-generated and human-written content (see [Author 1] among others).
In this study, we aim to explore how ChatGPT affects the style of students' writing, focusing on the potential shifts in language use, structure, and originality when it is employed as either a primary writing tool or an assistive resource.

2. CORPUS DEVELOPMENT AND DATA PRESENTATION

Each sub-corpus represents a small collection of texts developed over the course of an academic year as part of a larger corpus, UniC (University Corpus). UniC was created primarily for studies on Greek learners of Italian as a foreign language and their interlanguage development (Author 2). The project aimed to document the written language of university students and future teachers in Italian as a foreign language. The topics and sizes of the sub-corpora vary, but the texts consistently adopt a narrative genre each year. For this study, we selected data from the last two academic years.

Sub Corpus	N. Words	N. Learners	Use of online tools (for corrective purposes)
UniC 2021	42.751	43	Google translate
Unic 2022	27.069	23	Google translate
Unic 2023	31.215	31	Chat GPT
Unic 2024	4.214	7	NONE

Table 1: Uni Corpus

For the purposes of our study, we analyzed a set of short narratives written by students in two distinct years: 2023 and 2024. The purpose was to explore stylistic patterns and determine whether texts written with the aid of large language models, like ChatGPT, differ stylistically from those written without such assistance.
The dataset consists of narratives written by students across two different years. In 2023, a total of 31 narratives were produced, while in 2024, only 7 narratives were collected. A key distinction between these two years lies in the conditions under which the narratives were composed. In 2023, students were encouraged to use ChatGPT to assist in their writing, resulting in texts that potentially blend human and machine-generated stylistic features. By contrast, in 2024, students were explicitly instructed not to use any assistance from large language models (LLMs). These narratives were written at the university under supervised conditions, ensuring they reflect purely human-written styles.
All texts were saved in plain .txt format to maintain consistency across both years. Text files from 2023 were labeled sequentially (e.g., 001.txt, 002.txt), while those from 2024 included the year in their filenames (e.g., 2024-001.txt). This labeling system facilitated easy differentiation between the two datasets.

While students in 2023 were encouraged to use ChatGPT as an assistive tool, the extent and nature of their interaction with the AI (e.g., for generating entire drafts, for specific suggestions, for editing) were

not systematically recorded. Similarly, the specific prompts used by students were not collected, which introduces a degree of variability in the AI-assisted texts.

3. PURPOSE OF THE ANALYSIS

For this analysis, a set of word n-grams (specifically, unigrams and bigrams) were extracted from the texts and weighed using Term Frequency-Inverse Document Frequency (TF-IDF). This method assigns a higher weight to words that appear frequently in a specific text but infrequently across the entire corpus, thus highlighting terms that are characteristic of individual writing styles. We considered the top 1000 most salient of these TF-IDF weighed n-grams as features for our analysis. These features were chosen as word usage patterns, including combinations of adjacent words, which are commonly used in stylometric studies to capture variations in writing style.

To investigate through clustering whether stylistic differences in narrative writing can be attributed to the use of LLMs such as ChatGPT, we made use of the following techniques:

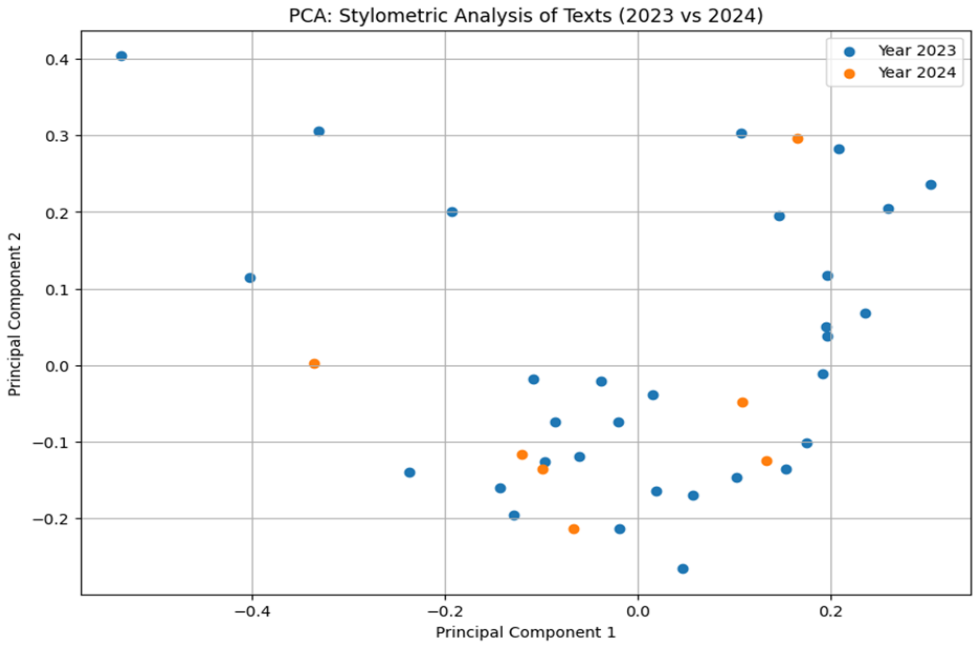
- Principal Component Analysis (PCA) for dimensionality reduction and visualization of stylistic variance.
- K-means clustering to identify groupings of texts based on stylistic similarities or differences.

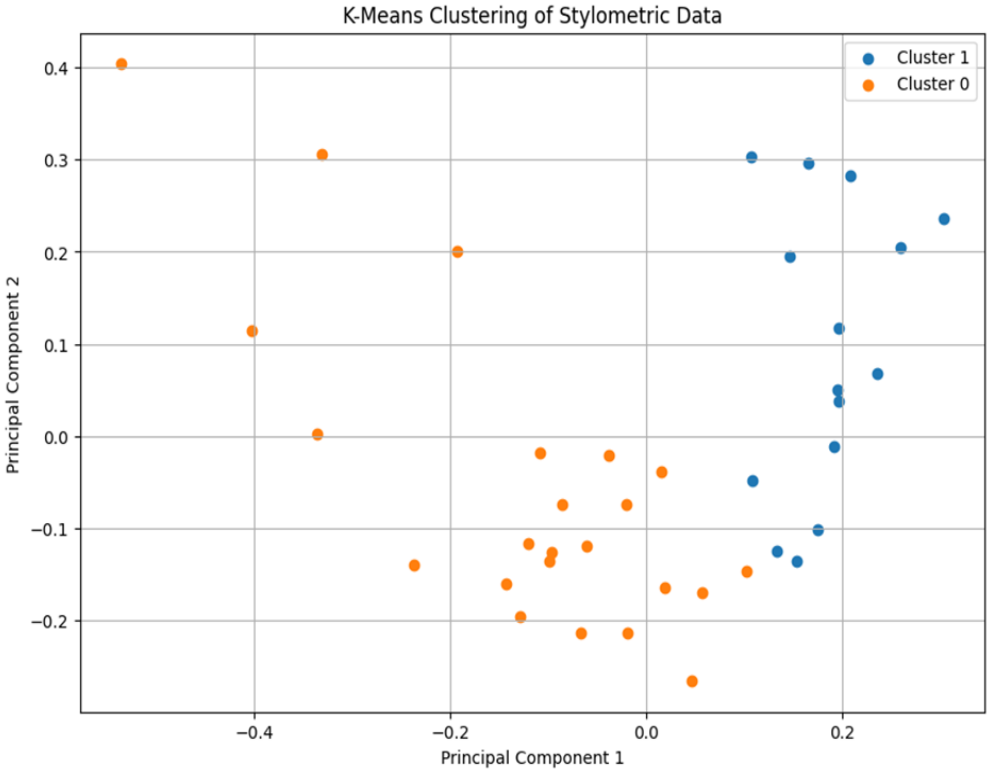
While Principal Component Analysis (PCA) and K-means clustering are machine learning techniques, they were applied here to analyze patterns within a set of extracted linguistic features relevant to writing style. Our approach aimed to identify potential groupings based on the overall distribution of these features across the texts.

PCA is a statistical technique used to reduce the dimensionality of datasets while preserving as much variance as possible. It transforms the original data into a set of new, uncorrelated variables called principal components. The first principal component captures the greatest variance in the data, while the second principal component captures the next highest variance, and so on. By projecting high-dimensional data onto these components, PCA simplifies the visualization and interpretation of complex datasets (Jolliffe, 2002). In this analysis, PCA enables us to map the stylistic features of the texts onto a two-dimensional space, making it easier to visually inspect potential groupings or patterns based on their stylistic variance.

K-means clustering is an unsupervised machine learning algorithm used to partition data into distinct clusters. The algorithm works by initializing a set number of cluster centroids and then iteratively assigning data points to the nearest centroid based on their distance, recalculating the centroids until the optimal configuration is reached. The goal is to minimize the variance within each cluster while maximizing the variance between clusters. In this study, K-means clustering allows us to group the narratives based on their stylistic features, helping to identify potential differences in style between the 2023 and 2024 texts.

4. RESULTS





Based on the clustering results, it appears that the stylistic differences between the 2023 and 2024 texts do not strongly correlate with the use of ChatGPT. While the K-means clustering algorithm divides the texts into two clusters, these clusters do not strictly correspond to the year of authorship or the use of AI assistance. Both 2023 and 2024 texts are present in both Cluster 0 and Cluster 1, indicating that stylistic similarities and differences are distributed across years rather than being exclusively tied to whether ChatGPT was used.

Cluster 1 contains a higher proportion of 2023 texts, suggesting that some texts written with ChatGPT share certain stylistic features that were captured by the clustering. However, Cluster 0 also includes a substantial number of 2023 texts, which highlights variability in how students used or edited AI-assisted content. Interestingly, some 2024 texts also appear in Cluster 1, implying that purely human-written texts can sometimes resemble AI-generated styles. This could be due to individual writing habits, common educational influences, or even shared narrative structures that students employed across both years.

The overlap between the clusters suggests that the stylistic impact of ChatGPT is not uniform or easily distinguishable from human writing through clustering alone. Instead, the clustering may reflect broader stylistic differences, such as variations in sentence complexity, use of specific vocabulary, or narrative structure, which are present in both AI-assisted and non-AI-assisted texts.

In summary, the clustering results demonstrate that while there are detectable stylistic groupings, these do not clearly separate the texts based on ChatGPT usage. This indicates that other factors, such as individual writing style or pedagogical influences, may play a significant role in shaping the stylistic characteristics of the narratives.

5. CONCLUSION

The analysis of the narratives from 2023 and 2024 aimed to explore whether the use of ChatGPT resulted in identifiable stylistic differences in the students' writing. Through stylometric techniques, including Principal Component Analysis (PCA) for dimensionality reduction and K-means clustering for grouping, we examined patterns across the texts. While there were observable stylistic groupings, these clusters did not distinctly separate texts based on the use of ChatGPT or the year of authorship. Both AI-assisted (2023) and non-AI-assisted (2024) texts were present in both clusters, suggesting that the impact of ChatGPT on writing style is neither uniform nor easily distinguishable from human-authored content. The findings highlight the complexity of attributing stylistic features to the use of AI, as individual writing habits, narrative structures, and pedagogical influences may overshadow the direct effects of AI assistance.

This study underscores the importance of analyzing the nuanced relationship between AI-assisted and human-written texts. While some stylistic trends were observed in 2023 texts, likely influenced by ChatGPT, the overlap between years and clusters suggests that human intervention, such as editing AI outputs, can blur distinctions.

Even though these results contribute to ongoing discussions about the role of AI in education and writing, particularly regarding its impact on students' stylistic development, it is important to note that our use of PCA and K-means clustering served to explore patterns in the distribution of extracted linguistic features, offering a broad overview of potential stylistic differences rather than a deep dive into specific stylistic markers.

6. LIMITATIONS

Despite the insights gained, this study has several limitations. First, the dataset is relatively small, with only 31 texts from 2023 and 7 from 2024. This limited sample size may reduce the robustness of the clustering analysis and the generalizability of the findings. A larger and more diverse dataset might reveal clearer patterns or more nuanced stylistic differences. Second, the analysis did not consider deeper semantic or syntactic structures. Incorporating advanced linguistic features or additional computational techniques could provide more comprehensive insights.

Another limitation lies in the variability of how students may have used ChatGPT in 2023. Some may have relied heavily on AI-generated outputs, while others might have used it sparingly or extensively edited its suggestions. This heterogeneity complicates efforts to draw clear stylistic distinctions. Furthermore, while 2024 texts were written without AI assistance under supervised conditions, individual writing habits and external influences may still have contributed to stylistic overlap with 2023 texts.

A significant limitation of this study is the lack of detailed content analysis of the narratives. It is possible that the clustering patterns observed are, to some extent, driven by differences in the topics or themes explored by students in 2023 and 2024, rather than solely by variations in their writing style due to AI assistance. Future research should consider incorporating content analysis methodologies to control for this potential confounding factor.

These limitations suggest avenues for future research, including larger datasets, expanded stylistic features, and more nuanced clustering methods, to deepen our understanding of the intersection between AI and human authorship. More specifically, future research could benefit from incorporating dedicated stylometric tools to extract a wider range of linguistic features before applying machine learning techniques. This could provide a more granular and potentially more accurate stylistic analysis. However, preliminary data suggests that while stylometric methods can identify patterns, the boundaries between AI and human stylistics are not always clear-cut.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the students of the Department of Italian Language and Literature for granting us permission to use their written works. We are also grateful to the personnel and the director of the Laboratory of Linguistic Analysis and Computational Processing of Romance Languages, Prof. Gianna Giannoulopoulou. We would like to thank the anonymous reviewers for their comments.

REFERENCES

Aydin, S. (2024). Using ChatGPT in Foreign Language Research: An Overview. DOI: https://doi.org/10.29329/irelt.2024.1045.2

Baskara, R., Mukarto, M. (2023). Exploring the Implications of ChatGPT for Language Learning in Higher Education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343- 358

Božić V, Poola I. (2023). ChatGPT and education. *Preprint*.. https://www.researchgate.net/profile/Velibor-Bozic-2/publication/369926506_Chat_GPT_and_education/links/64350844ad9b6d17dc4d3a79/Chat-GPT-and-education.pdf

Firat M. (2023). How ChatGPT can transform autodidactic experiences and open education?. https://osf.io/preprints/osf/9ge8m

Florou, K. (2024). The Effect of English as a Second Foreign Language On Learning Italian as a Third Foreign Language: A learner corpus-based research in written speech. *International Journal of Language and Literary Studies*, 6(3), 407-419.

Fuchs K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is ChatGPT a blessing or a curse?. In *Frontiers in Education* 2023 May 17 (Vol. 8, p. 1166682). Frontiers Media SA.

Jolliffe, I. T. (2002). Principal Component Analysis. Springer.

Klimova B, Pikhart M and Al-Obaydi LH (2024). Exploring the potential of ChatGPT for foreign language education at the university level. Front. Psychol. 15:1269319. doi:10.3389/fpsyg.2024.1269319

Mikros, G. K., Koursaris, A., Bilianos, D., & Markopoulos, G. (2023). AI-Writing Detection Using an Ensemble of Transformers and Stylometric Features. In *IberLEF*@ SEPLN.

Nugroho, A., Putro, N. H. P. S., & Syamsi, K. (2023). The potential of ChatGPT for language learning: Unpacking its benefits and limitations. *Register Journal*, 16(2), 224-247.