

AbjadStyleTransfer: Authorship Style Transfer for Arabic-Script Languages at AbjadNLP 2026

Shadi Abudalifa¹, Saad Ezzini¹, Ahmed Abdelali², Mustafa Jarrar^{3,4},
Mo El-Haj⁵, Nadir Durrani³, Hassan Sajjad⁶, Farah Adeeba⁷

¹King Fahd University of Petroleum & Minerals, ²Humain,

³Hamad Bin Khalifa University, ⁴Birzeit University, ⁵VinUniversity,

⁶Dalhousie University, ⁷University of Engineering & Technology

Abstract

Authorship style transfer aims to rewrite a given text so that it reflects the distinctive style of a target author while preserving the original meaning. Despite growing interest in text style transfer, most existing work has focused on English and other high-resource languages, with limited attention to languages written in the Arabic script. In this paper, we present an overview of **AbjadStyleTransfer**, a shared task organised as part of the AbjadNLP workshop at EACL 2026, which targets authorship style transfer for Arabic-script languages with a strong focus on literary text.

The shared task covers Modern Standard Arabic and Urdu, and is designed to encourage research on controllable text generation in morphologically rich and stylistically diverse languages. Participants are required to generate text that conforms to the writing style of a specified author, given a semantically equivalent formal input. We describe the task motivation, dataset construction, evaluation protocol, and participation statistics, and provide an initial discussion of the challenges associated with authorship style transfer in Arabic-script languages. **AbjadStyleTransfer** establishes a new benchmark for literary style transfer beyond Latin-script settings and supports future research on culturally grounded and linguistically informed text generation.

1 Introduction

Text style transfer has become an important research direction within Natural Language Processing, aiming to modify stylistic properties of text while preserving its underlying semantic content (Hu et al., 2022). Among its various subproblems, *authorship style transfer* (AST) constitutes a particularly demanding setting, as it requires modelling fine-grained and often implicit authorial traits, including lexical choice, syntactic preferences, and narrative voice, rather than broad stylistic dimen-

sions such as sentiment or formality (Hu et al., 2017).

Most prior work on AST has focused on English and other high-resource languages, where large parallel datasets and established benchmarks are available. In contrast, languages written in the Arabic script remain comparatively underexplored, despite their long literary traditions and stylistic diversity. Arabic and Urdu introduce additional challenges arising from rich morphology, flexible word order, and strong genre- and period-specific conventions, all of which complicate controllable generation and reliable evaluation (Habash, 2010; Abudalifa et al., 2024; El-Haj et al., 2018). Similar challenges have been documented across Arabic corpus construction and stylistic analysis efforts, particularly for literary and cultural text collections (El-Haj, 2020). These properties make direct transfer of existing AST techniques non-trivial and highlight the need for dedicated evaluation frameworks.

The **AbjadStyleTransfer** shared task builds on earlier efforts introduced in the AraGenEval shared task at the Third Arabic Natural Language Processing Conference in 2025 (Abudalifa et al., 2025). AraGenEval provided the first large-scale benchmark for Arabic authorship analysis, encompassing authorship style transfer, authorship identification, and AI-generated text detection. Results from that task demonstrated the potential of transformer-based and prompting-driven approaches for modelling Arabic authorial style, while also revealing persistent limitations in stylistic consistency and semantic preservation for generative systems.

AbjadStyleTransfer extends this line of work by isolating authorship style transfer as a standalone generative task and by broadening the scope to multiple Arabic-script languages. In contrast to earlier benchmarks, the task places a strong emphasis on literary text and content fidelity, encouraging systems to model authorial style beyond surface-level lexical variation. By providing parallel data aligned

to a standardised formal style and evaluating generation quality using established automatic metrics, AbjadStyleTransfer aims to advance research on controllable, author-aware text generation in morphologically rich and underrepresented language settings (Abdu et al., 2025; Mughaus et al., 2026; Almujaivel et al., 2025).

Hosted as part of the AbjadNLP workshop at EACL 2026 (El-Haj, 2025, 2026), this shared task offers a unified evaluation framework, curated datasets, and an open competition platform. It is intended to support reproducible research and to provide deeper empirical insight into the challenges of authorship style transfer for languages that share the Arabic script but differ substantially in linguistic structure and stylistic conventions.

2 Related Work

Authorship Style Transfer (AST) is a specialised text generation task that aims to rewrite a given input so that it reflects the distinctive writing style of a target author while preserving the underlying semantic content. Style in this context encompasses a range of linguistic properties, including lexical preferences, syntactic constructions, discourse organisation, and rhetorical devices that collectively characterise an author’s voice. AST differs from broader *Text Style Transfer* (TST), which typically focuses on modifying high-level attributes such as sentiment, formality, or politeness, without necessarily modelling writer-specific characteristics.

Recent progress in deep learning and large language models (LLMs) has substantially advanced research in TST and AST, enabling more fine-grained and controllable stylistic transformations. A common line of work adopts supervised learning with parallel data, framing style transfer as a sequence-to-sequence problem in which encoder–decoder architectures are trained to map input text to a target style (Hu et al., 2022; Gong et al., 2019). Within this paradigm, several studies have introduced copy or pointer mechanisms to explicitly preserve content that should remain unchanged, such as named entities or rare words, thereby reducing semantic drift during generation (Pan et al., 2024; Chawla and Yang, 2020; Merity et al., 2016).

Beyond purely supervised settings, latent-variable models have been explored to disentangle content and style representations. In particu-

lar, approaches based on Variational Autoencoders (VAEs) and Denoising Autoencoders (DAEs) learn continuous latent spaces in which stylistic attributes can be manipulated while attempting to maintain semantic consistency (Hu et al., 2017). These methods typically rely on auxiliary style classifiers to guide generation towards the desired stylistic target. Although effective in some settings, they often struggle to capture the subtle and highly individualised features required for authentic authorship style transfer, especially in low-resource or literary domains, such as those represented in Arabic literary corpora.

3 Task Description

Hosted as part of the AbjadNLP workshop at EACL 2026 (El-Haj, 2025, 2026), AbjadStyleTransfer is formulated as a controlled text generation task. Given a formal input text and the name of a target author, systems are required to generate a rewritten version of the text that reflects the stylistic characteristics of the specified author while preserving the original meaning.

The shared task focuses on literary text and is organised into two language-specific tracks, each evaluated independently using the Codabench platform.

3.1 Arabic Authorship Style Transfer

This track targets Modern Standard Arabic. The dataset is constructed from publicly accessible literary works by 21 authors, with ten books per author. Each book is segmented into semantically coherent paragraphs. Selected paragraphs are rephrased into a standardised formal style using an automated paraphrasing process, resulting in parallel pairs consisting of a formal input and an author-styled reference text. All text pairs are limited to a maximum length of 1900 tokens. The data is split into training, validation, and test sets.

3.2 Urdu Authorship Style Transfer

The Urdu track mirrors the Arabic setting in task formulation and data structure. It uses the same overall corpus design, author coverage, and paragraph-level segmentation strategy, enabling comparative analysis across languages that share a script but differ in linguistic structure and stylistic conventions. As with Arabic, the dataset is divided into training, validation, and test splits.

3.3 Input and Output

For both tracks, the model input consists of:

- a source text written in a formal, neutral style, and
- the name of the target author.

The system output is a generated text that expresses the content of the source text in the style of the specified author. Ground-truth author-styled references are provided for training and validation.

3.4 Evaluation Metrics

System outputs are evaluated using automatic text generation metrics:

- **BLEU**, measuring n-gram overlap with the reference text, and
- **chrF**, capturing character-level similarity and better accommodating morphological variation.

In addition to these primary metrics, qualitative analysis is encouraged to assess robustness across different author styles and literary registers.

4 Data Collection and Paraphrase Generation

4.1 Corpus Collection for Arabic

A textual dataset was constructed using works by 21 distinct writers whose publications are in the public domain. From each writer, ten volumes were chosen for inclusion. The selected texts were divided into coherent paragraph-level units, yielding matched paragraph pairs for analysis. The resulting corpus encompasses material from the following authors: A. Amin, A. T. Pasha, A. Shawqi, A. Rihani, T. Abaza, G. K. Gibran, J. Zaydan, H. Hanafi, R. Barr, S. Moussa, T. Hussein, A. M. Al-Aqqad, A. G. Makawi, G. Le Bon, F. Zakaria, K. Kilani, M. H. Heikal, N. Mahfouz, N. El Saadawi, W. Shakespeare, and Y. Idris.

Next, the GPT-4o mini LLM was used to rewrite the chosen passages in a more standardized and formal register. Because the resulting outputs were comparatively lengthy, only source–target text pairs that fit within the context window limits of selected language models were retained. The finalized corpus was subsequently partitioned into three subsets for training, validation, and evaluation purposes.

4.2 Corpus Collection for Urdu

We curated an Urdu literary corpus by crawling publicly available textual content from the Rekhta digital library. The corpus consists of prose articles and short literary texts authored by a diverse group of prominent Urdu writers, covering multiple literary movements and stylistic traditions.

Texts were collected for the following authors: Qurat-ul-Ain Haider, Saadat Hassan Manto, Rajinder Singh Bedi, Ghulam Abbas, Ismat Chughtai, Prem Chand, Krishan Chander, Mumtaz Mufti, Muhammad Hameed Shahid, and Ahmad Nadeem Qasmi.

For each author, multiple articles were extracted to ensure adequate thematic coverage and linguistic diversity. The collected texts were preprocessed to remove metadata, formatting artifacts, and non-content elements, resulting in a clean corpus containing only Urdu text.

4.3 Two-Step Paraphrase Generation via Back-Translation

To generate paraphrased variants of the original Urdu texts, we employed a two-step back-translation pipeline. In the first step, each Urdu text segment was translated into English. In the second step, the English translation was translated back into Urdu, yielding a paraphrased version of the original text.

Both translation steps were performed using GPT-4o. The back-translation process was guided by explicit instructions to neutralize authorial style. During the Urdu regeneration phase, the model was instructed to remove stylistic markers associated with the original author, such as characteristic sentence structures, lexical choices, and narrative voice, while preserving the underlying semantic content.

This translation–back-translation strategy serves a dual purpose: (i) introducing controlled lexical and syntactic variation for paraphrase generation, and (ii) reducing stylistic bias by systematically minimizing identifiable author-specific features in the generated text. After sentence segmentation, the corpus comprised approximately 10,000 sentences per author, resulting in a balanced dataset across all writers.

5 Results

5.1 Arabic Track Results

Table 1 is reserved for the official results of the Arabic authorship style transfer track. The task attracted 10 registered participants and 28 submissions during the development and final evaluation phases.

Participant ID	BLEU	chrF
HCMUS_The Fangs	19.77	55.74
hurryte	15.07	47.45

Table 1: Results for the Arabic authorship style transfer track.

5.2 Urdu Track Results

The Urdu track recorded no registered participants. Consequently, there were no eligible submissions within the assessment window, and therefore no official results appear on the leaderboard for this track.

6 System Overview

This work introduces a zero-shot framework for Arabic authorship style transfer proposed by the HCMUS_The Fangs team, which secured the top rank in this shared task. The authors effectively redefining the challenge as a strictly zero-shot setting rather than a data-scarce scenario. They design a dual-encoder model trained with contrastive learning, aiming to capture generalizable stylistic patterns instead of encoding author-dependent signals. The method builds on pre-trained Arabic language models and exploits contrastive loss functions to activate implicit stylistic information already present in these representations. Experimental results demonstrate notable performance gains and clearly surpassing baselines based on retrieval-augmented generation and multi-task learning.

7 Discussion

This study demonstrates that Arabic models, even before task-specific training, internally capture rich stylistic cues associated with canonical writers such as Naguib Mahfouz and Taha Hussein. Instead of adding explicit stylistic features or redesigning the model architecture, the approach leverages these latent representations and refines them through contrastive objectives. This strategy yields stronger

results than more elaborate interventions. Notably, methods that attempt direct style encoding or architectural enhancement tend to harm performance rather than improve it. These outcomes run counter to prevailing trends in style transfer research and indicate that, particularly in zero-shot scenarios, success depends less on acquiring new stylistic knowledge and more on activating what the model already knows, even when available data is sparse or misaligned.

8 Limitations

This study is subject to a number of constraints. First, experiments were conducted solely on Arabic data, so the conclusions may not transfer to languages with distinct morphological or writing-system characteristics. In addition, the evaluation dataset is limited to well-known historical figures whose writings are heavily represented in large pre-training corpora. Performance could vary considerably for modern or less-documented authors, where prior stylistic signals are weaker.

Owing to limited computational resources available, the participants were unable to investigate LLMs that may capture more nuanced stylistic patterns. Furthermore, assessment was based entirely on automatic similarity measures (BLEU and chrF), without incorporating style-aware evaluation methods—such as stylometric representations, authorship attribution surrogates, or human judgment—which restricts our ability to assess deeper stylistic consistency beyond surface form.

Lastly, the proposed contrastive framework depends on the availability of author-specific examples during training, which constrains its use in scenarios where no reference texts exist for a target author.

9 Conclusion

This shared task shows the top-performing submission to the AbjadStyleTransfer shared task, which ranked first by obtaining 19.77 BLEU and 55.74 chrF. The system addresses zero-shot Arabic authorship style transfer using a contrastive style learning framework. Its central premise is that pre-trained models already internalize stylistic patterns of well-known authors. Therefore, performance gains come from explicitly eliciting this latent knowledge via contrastive learning, rather than introducing architectural changes that may interfere with it.

Acknowledgements

We thank all the participating teams for their hard work and contributions. We also acknowledge the support of the AbjadNLP workshop organizers and the EACL 2026 conference for hosting this shared task.

References

- Fahad J Abdu, Raed Mughaus, Shadi Abudalfa, Moataz Ahmed, and Ahmed Abdelali. 2025. An empirical evaluation of arabic text formality transfer: a comparative study. *Language Resources and Evaluation*, 59(4):4093–4153.
- Shadi Abudalfa, Saad Ezzini, Ahmed Abdelali, Hamza Alami, Abdessamad Benlahbib, Salmane Chafik, Mo El-Haj, Abdelkader El Mahdaouy, Mustafa Jarar, Salima Lamsiyah, and Hamzah Luqman. 2025. [The AraGenEval shared task on Arabic authorship style transfer and AI generated text detection](#). In *Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks*, pages 1–13, Suzhou, China. Association for Computational Linguistics.
- Shadi I Abudalfa, Fahad J Abdu, and Maad M Alowaiifeer. 2024. Arabic text formality modification: A review and future research directions. *IEEE Access*.
- Sultan Almujaawel, Damith Premasiri, Tharindu Ranasinghe, Mo El-Haj, and Ruslan Mitkov. 2025. Complex concept-based readability estimation from arabic curriculum. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 24(11):1–21.
- Kunal Chawla and Diyi Yang. 2020. [Semi-supervised formality style transfer using language model discriminator and mutual information maximization](#). *Preprint*, arXiv:2010.05090.
- Mahmoud El-Haj. 2020. Habibi-a multi dialect multi national arabic song lyrics corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1318–1326.
- Mahmoud El-Haj, Paul Rayson, and Mariam Aboelezz. 2018. Arabic dialect identification in the context of bivalency and code-switching. In *Proceedings of the 11th International Conference on Language Resources and Evaluation, Miyazaki, Japan.*, pages 3622–3627. European Language Resources Association.
- Mo El-Haj. 2025. Proceedings of the 1st workshop on nlp for languages using arabic script (abjadnlp 2025). In *Proceedings of the 1st Workshop on NLP for Languages Using Arabic Script, held at COLING 2025*, Abu Dhabi, United Arab Emirates.
- Mo El-Haj. 2026. Proceedings of the 2nd workshop on nlp for languages using arabic script (abjadnlp 2026). In *Proceedings of the 2nd Workshop on NLP for Languages Using Arabic Script, held at EACL 2026*, Rabat, Morocco.
- Hongyu Gong, Suma Bhat, Lingfei Wu, JinJun Xiong, and Wen-mei Hwu. 2019. [Reinforcement learning based text style transfer without parallel training corpus](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3168–3180, Minneapolis, Minnesota. Association for Computational Linguistics.
- Nizar Y Habash. 2010. *Introduction to Arabic natural language processing*. Morgan & Claypool Publishers.
- Zhiqiang Hu, Roy Ka-Wei Lee, Charu C. Aggarwal, and Aston Zhang. 2022. [Text style transfer: A review and experimental evaluation](#). *Association for Computing Machinery*, 24(1):14–45.
- Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. Toward controlled generation of text. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17*, page 1587–1596. JMLR.org.
- Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. 2016. [Pointer sentinel mixture models](#). *Preprint*, arXiv:1609.07843.
- Raed Mughaus, Shadi Abudalfa, Hamzah Luqman, Fahad Abdu, Mohammed AlAli, Nawaf Al-Dowayan, and Ahmed Abdelali. 2026. [Ma'aks: manually-curated parallel dataset for arabic text sentiment swap](#). *Language Resources and Evaluation*, 60(1):1.
- Lei Pan, Yunshi Lan, Yang Li, and Weining Qian. 2024. [Unsupervised text style transfer via llms and attention masking with multi-way interactions](#). *Preprint*, arXiv:2402.13647.