Advances in Engineering Innovation Vol.16 Issue 9 EWA Publishing
Available Online: 24 September 2025 DOI: 10.54254/2977-3903/2025.27219

Detecting the strange: automatic recognition and evaluation of defamiliarization in modernist poetry

Ziyi Yang

Department of Computer Science, Central South University, Changsha, China

8208240708@csu.edu.cn

Abstract. This study addresses the long-standing gap between literary theory and computational modeling by focusing on defamiliarization, a central technique in modernist poetics. While defamiliarization has been extensively theorized in literary studies, its computational treatment remains limited due to its semantic complexity and subjective nature. To operationalize this phenomenon, the research introduces MelPoet, a poetics-informed neural architecture adapted from MelBERT. Two tasks are formulated: defamiliarization identification as binary classification, and defamiliarization scoring as regression of estrangement intensity. A curated dataset of modernist poetry and control texts was annotated for both presence and degree of defamiliarization. Experimental results demonstrate that MelPoet substantially outperforms strong baselines in both classification accuracy and scoring correlation, confirming the efficacy of its theory-driven design. This work not only advances computational methods for modelling figurative language but also provides a systematic framework for integrating literary concepts into natural language processing, thereby opening new avenues for large-scale, data-driven analysis of poetic style.

Keywords: Natural Language Processing(NLP), defamiliarization, MelBERT, modernist poetry

1. Introduction

Defamiliarization, a literary technique that makes the familiar appear strange, is central to modernist poetics. As defined by Russian formalist Viktor Shklovsky, this technique exemplifies a disruption of linguistic norms to renew perception [1]. T hrough ambiguity, syntactic dislocation, and unexpected imagery, poets such as T. S. Eliot and Ezra Pound deliberately challenge readers' habitual perceptions [2,3]. Consider Eliot's famous line from The Waste Land, "April is the cruellest month", which inverts seasonal expectations to evoke emotional unease [2]. Pound's compressed metaphor in In a Station of the Metro— "The apparition of these faces in the crowd; / Petals on a wet, black bough" —which juxtaposes urban modernity with organic beauty [3].

However, defamiliarization has received relatively little attention in the field of computational linguistics. This gap stems in part from key methodological challenges: defamiliarization is difficult to simplify into surface text features, as it relies on multiple subversions of syntax, semantics, and reader expectations, and often runs through different levels of interpretation. In addition, the inherent subjectivity of this phenomenon also makes annotation and evaluation difficult. Therefore, there is still a lack of benchmark datasets or clear computational task definitions for this phenomenon.

To meet this challenge, the study focus on imagistic defamiliarization, characterized by the displacement and juxtaposition of semantically distant images. This technique is pervasive in modernist poetry and displays recognizable regularities, making it suitable for computational modeling. To operationalize this focus, the study proposes an empirical framework comprising two tasks: defamiliarization identification, formulated as a binary classification task to detect the presence of imagistic estrangement, and defamiliarization scoring, formulated as a regression task to estimate its perceived intensity. Beyond that, the proposed approach draws on MelBERT, a contextualized model originally developed for metaphor detection, whose semantic sensitivity makes it well-suited for capturing the image disjunctions central to this study [4]. MelBERT is adapted and extended into MelPoet,a poetics-informed model incorporating three targeted innovations. .

By adapting this architecture to modernist poetics, the research not only advances computational methods for analyzing defamiliarization but also contributes to the broader integration of data-driven approaches into literary studies, offering new possibilities for systematic and large-scale analysis of poetic style.

2. Related work

2.1. Defamiliarization

The concept of defamiliarization (also known as ostranenie) was first introduced by Russian formalist Viktor Shklovsky in his seminal 1917 essay Art as Technique [1]. Roman Jakobson furthered this linguistic-literary perspective, emphasizing the foregrounding of language in defamiliarization [5]. In literary practice,modern poets like T. S. Eliot and Ezra Pound operationalized these ideas in English-language poetry, emphasizing perceptual renewal and creative disruption.

In literary studies, defamiliarization in modernist poetry is commonly theorized along four dimensions. First, imagistic defamiliarization disrupts perceptual habits through unconventional image pairings. Second, linguistic defamiliarization involves syntactic and lexical deviation. Third, narrative and logical defamiliarization break temporal or causal coherence. Finally, thematic and philosophical defamiliarization challenges established metaphysical or moral frameworks. Building on these theoretical foundations, the present study focuses on imagistic defamiliarization, particularly the technique of image displacement, which represents a common and systematically recognizable manifestation of estrangement. It integrates this focus into the proposed computational framework.

2.2. Natural language processing for literary studies

Parallel efforts in rhetorical device recognition have made massive progress, particularly in metaphor detection, which has evolved from early rule-based systems to transformer-based models such as MelBERT and MiceCL [6]. These models capture figurative expressions at both token and phrase levels, often using curriculum learning or semantic clustering to improve generalization [7,8]. Beyond metaphor, recent research has extended to other non-literal forms such as sarcasm, irony, and hyperbole, frequently within social media and dialogue contexts [9,10]. Notably, Hossain et.'s humor benchmark introduces a layered framework of recognition, ranking, and explanation, offering a valuable methodological parallel for the work on modeling defamiliarization in poetry [11].

In subjective language evaluation, tasks such as naturalness scoring, fluency assessment, and style classification aim to capture linguistic quality through human or model-based judgments [8]. Techniques like content-style disentanglement and BERT-based metrics, widely used in style transfer evaluation, offer useful parallels for assessing defamiliarization strength in poetic language.

3. Methodology

3.1. Dataset

To explore sentence-level defamiliarization recognition, the study constructs a dataset comprising both defamiliarized and non-defamiliarized text samples.

Defamiliarized text is drawn from around 20 modernist poems, mainly by T. S. Eliot and Ezra Pound. Individuals with a background in literature annotate each sentence. Annotators labeled whether the sentence contains defamiliarization and rated its intensity on a scale from 1 to 5, focusing on image displacement and conceptual leap.

Non-defamiliarized text is sampled from factual and expository genres, including news reports and textbook excerpts. These texts typically employ direct, unambiguous language and serve as contrastive controls.

The annotated dataset, consisting of both defamiliarized and non-defamiliarized sentences, was randomly divided into training (80%), validation (10%), and test (10%) subsets. The training set was used to fit MelPoet and baseline models, the validation set guided hyperparameter tuning and early stopping, and the test set was reserved for final evaluation only. This standard partitioning ensures fair comparison across models and provides a reliable assessment of generalization to unseen data.

All sentences are split at punctuation boundaries and filtered for completeness. Annotation agreement is evaluated using Cohen's Kappa for binary labels and Krippendorff's alpha for intensity scores. In cases of disagreement, annotators discussed the sentence collaboratively to reach a consensus. The final dataset contains binary defamiliarization labels and corresponding intensity scores for each sentence.

3.2. MelBERT

MelBERT is a contextualized neural architecture originally developed for metaphor detection. Its core principle is the late interaction of contextualized embeddings, enabling the model to capture semantic dissonance between a target word and its surrounding context. This property makes MelBERT a suitable foundation for the present study, since both metaphor and poetic

defamiliarization involve divergence between a word's contextual and literal meanings. By leveraging pre-trained BERT embeddings, MelBERT thus offers a strong foundation for modelling figurative language phenomena.

3.3. Task setups

The detection of poetic defamiliarization in this study is formulated as two complementary tasks: a classification task that identifies whether a poetic line contains imagistic estrangement, and a regression task that estimates the intensity of this estrangement along a continuous scale.

The MelPoet architecture is developed to computationally model poetic defamiliarization, focusing on both its detection and its graded evaluation. As shown in Figure 1, in this framework, each poetic line is first processed at the input layer, where raw text from modernist poetry serves as the foundation for feature extraction. The architecture integrates a semantic encoder, adapted from MelBERT, with a poetic feature extractor designed to capture stylistic and imagistic disruptions. A feature fusion module then combines these signals, enabling the model to account for both linguistic regularities and the unexpected imagery that characterizes defamiliarization.

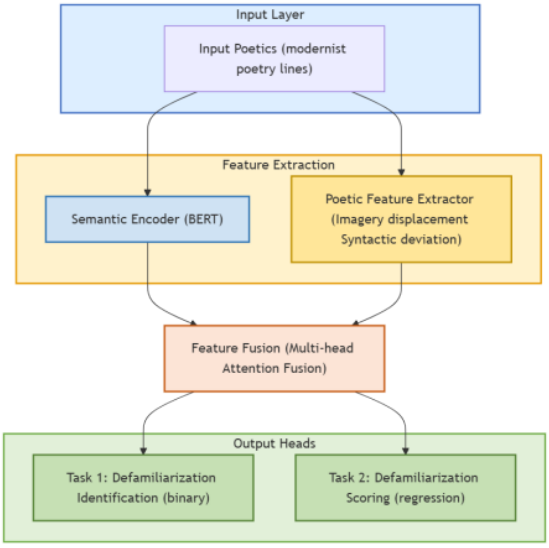

Figure 1. MelPoet architecture diagram

This architectural adaptation from MelBERT to MelPoet is not merely a change in domain but a fundamental reengineering to address the unique computational challenges posed by poetic language. This transition is marked by three core innovations as follows, each designed to capture a distinct facet of poetic estrangement.

Dual-View Optimization. MelPoet adopts a dual-view fine-tuning strategy, in which training is performed under two parallel configurations: a conservative view with three epochs at a learning rate of 2e-5, and an exploratory view with five epochs at 3e-5. The outputs from both views are ensembled during evaluation, balancing stability against adaptability. This duality enables the model to generalize across the irregular structures of modernist poetry while mitigating the risk of overfitting to highly idiosyncratic imagery.

Type-Constrained Attention. The model incorporates Jakobson's theory of poetic function by imposing lexical type constrains on the self-attention module [12]. This model emphasizes the role of content words (e.g., nouns, verbs, and adjectives) serve as the primary carriers of imagery and semantic disruption, while function words provide grammatical scaffolding.This computationally models poetic principles and enhances the linguistic representation of defamiliarization.

Defamiliarization-Aware Pretraining. Beyond conventional masked language modelling, MelPoet is further pretrained on a curated corpus of modernist poetry emphasizing imagistic estrangement. This additional pretraining is conducted for 10 epochs with a masked token probability of 15%. By repeatedly exposing the model to disjunctive image pairings, this stage enhances sensitivity to semantic dissonance, improving both detection and scoring of poetic estrangement in downstream tasks.

4. Experiments

4.1. Baselines

To benchmark the performance of MelPoet and validate the efficacy of its specialized architecture, it was compared against several established models representing different methodological approaches to natural language processing and figurative language detection. All baseline models were trained and evaluated on the same dataset splits and using the same metrics as MelPoet to ensure a fair comparison:

Vanilla BERT: BERT-base fine-tuned on the dataset for both tasks.

CNN-LSTM: A hybrid model for metaphor detection, capturing local n-gram and long-range dependencies [13].

Cross-Lingual Transfer: A multilingual transfer model for figurative language detection [14].

4.2. Experimental parameters and evaluation

MelPoet extends the MelBERT framework with a shared mBERT encoder and two task-specific heads:

Binary is classified for defamiliarization identification.

Regression is for defamiliarization intensity scoring.

The model is trained on sentence-level annotations from modernist English poetry and non-poetic control texts. Sentences are tokenized with WordPiece, truncated/padded to 64 tokens, and optimized with AdamW (lr = 2e-5, batch size = 16, epochs = 5, dropout = 0.1). Loss functions are Binary Cross-Entropy (identification) and Mean Squared Error (scoring).

Evaluation uses Accuracy, Precision, Recall, F1 for classification, and Pearson, Spearman, MAE for regression. All experiments run on a single NVIDIA RTX 3090 using HuggingFace Transformers.

4.3. Results

Table 1 and Table 2 present the performance of all models.

The experimental results show that MelPoet consistently outperforms all baselines across both tasks.

Table 1. Performance on defamiliarization identification

Model	Accuracy(%)	Macro F1-Score(%)
Vanilla	82.5	81.2
CNN-LSTM	78.1	77.5
Cross-Lingual Transfer	80.3	79.9
MelPoet(Ours)	89.4	88.7

According to the results, MelPoet achieves the best performance in defamiliarization identification, with 89.4% accuracy and an F1-score of 88.7%, clearly surpassing all baseline models.

Table 2. Performance on defamiliarization scoring

Model	Pearson Correlation	RMSE
Vanilla BERT	0.65	1.15
CNN-LSTM	0.58	1.28
Cross-Lingual Transfer	0.62	1.19
MelPoet(Ours)	0.84	0.87

According to the results, MelPoet also leads in defamiliarization scoring, attaining the highest correlation (0.84) and the lowest error (0.87), demonstrating robust capacity in modeling the intensity of poetic estrangement.

For Task 1, MelPoet's performance in defamiliarization identification significantly surpasses other models, with its strong Accuracy and Macro F1-score demonstrating the effectiveness of combining MelBERT-based contextual embeddings with poetic-structure-aware features. The clear gap between MelPoet and earlier deep learning methods like CNN-LSTM highlights the benefit of contextualized representations for this task.

In Task 2, MelPoet achieves a superior Pearson correlation and lower RMSE, indicating its ability to not only detect defamiliarization but also precisely model its degree. The substantial improvement over both Vanilla BERT and the Cross-Lingual Transfer baseline confirms that MelPoet's specialized architecture, which incorporates poetic and linguistic features, is more effective than general-purpose or purely transfer-based methods.

5. Discussion

Overall, the results confirm the effectiveness of MelPoet in modeling defamiliarization. Compared with strong baselines, MelPoet demonstrates clear improvements in both identification accuracy and intensity scoring, highlighting the value of integrating literary theory into computational design. In particular, the dual-view optimization enhanced robustness across noisy poetic inputs, the defamiliarization-aware pretraining strengthened the model's sensitivity to imagistic disjunction, and the type-constrained attention allowed more precise capture of estrangement signals by prioritizing content words. These targeted innovations jointly account for the observed performance gains and underscore the necessity of tailoring neural architectures to the unique demands of poetic language.

6. Conclusion

This study introduces MelPoet, a poetics-informed neural architecture for detecting and scoring defamiliarization in poetry. By integrating contextualized language modeling with poetic-structural features,MelPoet operationalizes the formalist concept of ostranenie and demonstrates clear performance gains over strong baselines. The findings underscore the value of embedding literary theory into computational models,offering a systematic framework for poetic language analysis. However, limitations include the relatively small scale of the labeled corpus,the high cost of manual annotation. Future work will expand the corpus and explore multilingual, cross-genre applications to enhance the model's robustness and applicability. Thus, the objective is not only to refine computational methods, but also to bring readers closer to the very moment when language, like in modernist verse, shifts and shimmers into the unfamiliar, renewing perception itself.

References

[1] Shklovsky, V. (1965). Art as technique. In *L. T. Lemon & M. J. Reis (Trans.), Russian formalist criticism: Four essays* (pp. 3-24). University of Nebraska Press.

[2] Eliot, T. S. (2003). The waste land. In *F. Kermode (Ed.), The waste land and other poems* (pp. 51-85). Penguin Classics.

[3] Pound, E. (1990). Personae: The Collected Poems of Ezra Pound. *New Directions*.

[4] Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D., & Lee, J. (2021). MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. arXiv. https: //arxiv.org/abs/2104.13615.

[5] Jakobson, R. (1960). Linguistics and poetics. In *T. A. Sebeok (Ed.), Style in language* (pp. 350-377). MIT Press.

[6] Jia, K., & Li, R. (2024). Metaphor detection with context enhancement and curriculum learning. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1: Long Papers) (pp. 2726-2737). Association for Computational Linguistics.

[7] Stowe, K., Luan, Y., Bhatia, A., Gung, J., & Peng, S. (2024). Curriculum Learning for Metaphor Detection. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2024)* (pp. 7392-7406). Association for Computational Linguistics.

[8] Jia, K., Wu, Y., Liu, M., & Li, R. (2024). Curriculum-style data augmentation for llm-based metaphor detection. arXiv. https: //arxiv.org/abs/2412.02956.

[9] Memon, M. Q., Banbhrani, S. K., Akhter, M. N., Noureen, F., & Mehreen, F. (2024). Detecting sarcasm in social media posts using transformer-based language models with contextual and sentiment-aware features. *Spectrum of Engineering Sciences*, 2(5), 534–549.

[10] Badathala, N., Kalarani, A. R., Siledar, T., & Bhattacharyya, P. (2023). A match made in heaven: A multi-task framework for hyperbole and metaphor detection. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics.*

[11] Hessel, J., Marasović, A., Hwang, J. D., Lee, L., Da, J., Zellers, R., Mankoff, R., & Choi, Y. (2022). Do androids laugh at electric sheep? Humor "understanding" benchmarks from the new yorker caption contest. arXiv. https: //arxiv.org/abs/2209.06293

[12] Jakobson, R. (1987). Language in literature (K. Pomorska & S. Rudy, Eds.). Harvard University Press.

[13] Gao, G., Choi, E., Choi, Y., & Zettlemoyer, L. (2018). Neural metaphor detection in context. arXiv. https: //arxiv.org/abs/1808.09653.

[14] Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., & Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1: Long Papers) (pp. 248-258). Association for Computational Linguistics.