IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

# A Survey on SLM for stylized Marathi Prose and Poetry Generation

**Manasi Bhole[1], Riddhi Lahare[2], Veda Kimbahune[3], Vedika Kute[4], P. P. Joshi[5]**

Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India[1-5]

**Abstract:** Text can be generated using Small Language Models with less computational power but it is doubtful that how well literature text can be created using these on low resource Indic Languages. Marathi Language is a good case study for this. While generation good Marathi text using SLMs , not only correct Marathi is required but also the quality of text , cultural fit and uniqueness of author should be depicted in the generated text.

This survey is about understanding how well SLMs can create text , poems , stories given an author style. This survey focuses on recent studies on this. Three main aspects of this survey are as mentioned below 1. SLMs perform good on specific tasks but it doesn't apply when it comes to low resource language. 2. The quality of literature depends on multiple aspects and automatic evaluation does not assure accuracy. 3. Author style does not indicate copying only keywords they commonly use and SLMs try to capture only surface level feature of author instead of the depth.
We conducted research on small language models and found out that model generates repeated words and Gujarati/ English text upon providing prompt to generate Marathi text.
Small Language Models , Marathi Natural Language Processing , Low resource Languages , Text generation ,Prose and Poetry Generation , Author style , QLoRa , Indic NLP , Fine tuning.

## 1 INTRODUCTION

There is a great improvement in automatic text generation by Large Language Models , but they are quite expensive to run , train and fine tune. This causes rise in interest to study Small Language Models as they inexpensive compared to Large Language Models. Such models tend to use less computational cost , energy and resources. Also they can operate on local hardware. BART-Large was fine tuned by Marco *et al.* for generating movie synopses and were rated by humans. The results generated were more creative than GPT-4o indicating large language models do not always necessarily produce better result in the field of creative writing.

Literature writing is comparatively harder that normal text generation as it needs accurate language and correct semantics. Along with that it needs emotions , creativity and imagination. Poem generation is even tougher as it needs rhymes , rhythm , metre and a proper stucture. (**koziev_sinev_2026?**) Koziev and Sinev studied that there is no single automatic way to evaluate the poem quality as it also needs human touch.

It is even more challenging for Marathi which is low resource language. Most LLMs are trained on the languages that have extensive resources available. Consequently they produce better quality text for languages like English or Hindi but fail to perform similarly on Marathi . LLMs struggle on low resource languages with respect to grammar , vocabulary and cultural references. Just producing Devnagri text is not sufficient but it should also consider the correct grammar , and not mix it with Hindi, Gujarati or simialr languages.

Thus, this survey emphasizes on connecting bridge between SLMs , fine tuning using efficient parameter , Marathi NLP, literature and poem generation , style of author. The main goal is to understand what has previously been researched , what issues are still remaining and the need of a small yet targeted model especially designed for stylised Marathi Prose and Poetry Generation.

## 2 BACKGROUND AND PROBLEM DEFINITION

### 2.1 Small Language Models

As mentioned earlier, SLMs are concise models which use less computation power and resources as compared to LLMs. SLMs perform well on local hardware and are useful for specific jobs. Earlier research has shown that SLMs are good in creative writing (**marco_slm_2025?**) but the study is mostly done for English language and does not assure its accuracy for low resourced language like Marathi .

 This work is licensed under a Creative Commons Attribution 4.0 International License 

2.2 Literary Text Generation

Writing literature does not just include correct grammatical formatting but it also needs creativity and coherence for stories while it also needs rhyme and rhythm for poems. (**chhun_2024?**) According to earlier studies it has been observed that BLEU which is a traditional metrics is not enough for measuring the quality of literature as linguistic accuracy and human supervision are also necessary.

2.3 Low Resource Marathi NLP

Marathi is rich in literature but the digital literature resources available for Marathi are far less as compared to high resource languages like English. The ultimate goal of this project is to curate a corpus nearly 200 authors and poets of Marathi using Marathi WikiSource as the main data source. The text collected will be preprocessed and organized according author's information.

2.4 Authorial Style and Stylometry

Every author and poet posses different writing style considering word selection , structure of the sentences and linguistic features. According to previous studies (**osullivan_2025?**) it has been found out that the above mentioned feature can be used to distinguish human writers and the text that AI has generated. Our goal is to understands these features for SLMs to create text that depicts style of selected Marathi author.

2.5 Problem Definition

So the important research aspect is : How efficiently can a pretrained SLM adapt to producing correct Marathi Pros and Poetry while depicting the selected Marathi author's and poet's style.

The two main objectives of the study are :

1. Language fidelity : Generating accurate and substantial Marathi text without any unnecessary other language interfernce or mixing.
2. Literary and stylistic fidelity : Generating appropriate prose and poetry that demonstrate the distinctive writing style of the considered author.

## 3 LITERATURE REVIEW

The curated studies are assessed based on the proffered taxonomy :

3.1 SLMs and creative Generation

Macro , Rello and Gonzalo illustrated that refined BART Large model can execute well on brief innovative writing jobs and can fabricate even more astonishing concepts than LLMs(**marco_slm_2025?**). However, the research narrows down to synopses of English movie without considering specific styles of authors.

Thennal and Hatzel discovered stories generated by LLMs conceal a cramped span of narrative trends than the stories written by humans(**thennal_2026?**). This showcases the neccesity to maintain variety when creating text for distinct authors of Marathi language

Gao*et al.* put forward an LLM SLM framework where an SLM verifies and filters the produced expressions(**gao_2026?**). However, it mostly emphasised on Chinese but it can be used to relate with Marathi language in our case

3.2 Poetry and Literary Generation

Koziev and Sinev studied automatic poetry generation and featured the hardness of asssesing the poetry across multiple languages and their styles(**koziev_sinev_2026?**). There is also a discussion about maintaing a balance between the form and meaning of a poem.

Ferreria *et al.* analyzed serveral semantics and rhetorical categories of a poem and tried to interpret the flow of the generated poetry(**ferreria_2026?**). This study indicated that syllable count and rhyme are not the parameters to evaluate the poem.

Jobanputra *et al.* used developmental refinement and search techniques to ameliorate haiku according to syllables, rhmye and emotions(**jobanputra_2025?**). However this technique is constrained to a particular form of poem.

Ma, Yao and Haensch assessed Classical Chinese poetry using LLM evaluation, automatic metric and human verification.(**ma_2026?**).

Yazan *et al.* studied the innovativeness in literature using concepts such as novelty, emotions and semantics(**yazan_2026?**) .

3.3 Evaluation of creative writing

Different assessment dimensions such as coherence, relevance and engagement are required for creative writing and not only simple vocabulary and lexical metrics. This was shown by Chhun (**chhun_2024?**).

The study of Marco, Gonzalo, and Frenzo depicted that different readers prioritize variety of aspects of the creative writing. Themes, expressive style and rhetorical variety are considered significant by the experts (**marco_reader_2025?**).

Haverals and Martin illustrated that the author's identity plays an important role in the evaluation of the literature (**haverals_2025?**). Hence if the identity of the author is known to the evaluator then the results of the evaluation might get influenced. Therefore to avoid any bias, evaluation should be blind meaning only the literary content should be presented and be visible to the evaluator while keeping the identity of the author secret.

Chakrabarty, Laban, and Wu explored common issues in the creative writings generated by the AI (**chakrabarty_2025?**). They tried to demonstrate the prominence of involving humans in the editing and evaluation of the creative literary writing. But, this research is mainly concentrated on English.

## 4 COMPARATIVE ANALYSIS

From the reviewed studies we can derive that there is strong existing work for individual components of the problem, but no one study tackles all of them together. Table [tab:comparison] compiles most relevant literature, and it unveils three observations.

| study | Year | Language / Domain | Approach | Key Contribution | Limitation for the Present Problem |
|---|---|---|---|---|---|
| Marco et al. (**marco_slm_2025?**) | 2025 | English creative writing | Fine-tuned SLM | Indicates that a small model can perform effectively on short creative generation. | Does not include Marathi, poetry, or multi-author style modelling. |
| Koziev & sinev (**koziev_sinev_2026?**) | 2026 | Poetry | Survey | Identifies the important generation and multidimensional poetry evaluation challenges. | Has limited focus on low-resource author-specific generation. |
| chhun (**chhun_2024?**) | 2024 | story generation | Multi-dimensional evaluation | Highlights the limits of BLEU-style evaluation for stories. | Lacks the ablility model authorial style. |
| Thennal & Hatzel (**thennal_2026?**) | 2026 | English stories | Narrative embeddings | Shows that LLM outputs often exhibit similarities in their storytelling styles and structures . | Does not address low-resource settings or author-conditioned generation. |
| O'Sullivan (**osullivan_2025?**) | 2025 | Creative writing | Burrows' Delta / stylometry | Highlights the differences in stylistic clustering between human-written and AI-generated text. | Focuses on distinguishing human-written text from AI-generated content rather than Marathi author imitation. |
| Mikros (**mikros_2025?**) | 2025 | Literary prose | stylometric style imitation | Demonstrates that surface imitation does not guarantee deep stylistic fidelity. | Limited to a single model and two authors, focusing exclusively on English-language text. |
| wang et al. (**wang_2025?**) | 2025 | Everyday writing | Attribution / verification | Highlights the limitations of implicit style imitation using few-shot examples. | Does not focus on Marathi literary texts or fine-tuning approaches. |
| Icard et al. (**icard_2026?**) | 2026 | French literature | Embedding + explicit features | Highlights that embeddings capture only a part of authorial style. | Does not include Marathi SLM |
| shi et al. (**shi_2025?**) | 2025 | Multilingual literature | Syntax, metaphor, phonetics | Highlights the value of deeper structural features, particularly | Focuses on genre classification instead of generation. |

| study | Year | Language / Domain | Approach | Key Contribution | Limitation for the Present Problem |
|---|---|---|---|---|---|
| | | | | rhythm. | |
| Ma et al. (**ma_2026?**) | 2026 | classical chinese poetry | Computational + LLM + human evaluation | Shows the need for hybrid evaluation and identifies LLM-judge bias. | Focuses on Chinese poetry rather than Marathi literary generation. |
| Vinaixa Roselló (**vinaixa_2026?**) | 2026 | Spanish poetry | Llama 3 8B fine-tuning | Highlights that small-data literary-style adaptation is feasible. | Uses Informal evaluation and focuses on one author. |

First, though SLM research satisfactorily shows that it can be effective for specialized tasks, most of these research in focused on English or niche task like movie synopsis(**marco_slm_2025?**). Second is regarding the quality. Quality evaluation is a multidimensional process, which highlights the importance of human evaluation(**koziev_sinev_2026?**; **ma_2026?**). TThird, stylometric research demonstrates that the style of an author can be measured and identified computationally. Yet, the current models struggle to consistently reproduce deeper stylistic signatures and properties or an author(**mikros_2025?**; **wang_2025?**; **icard_2026?**).

These findings strongly define the makings of a successful Marathi literary model. The goal is not to only create an ordinary language model, but to satisfy the language correctness, subject relevance, literary quality, and author- specific style. It should also strive for diversity across generated work, thus resisting convergence to a single generic literary voice.

Our proposed solution starts from a sequential evaluation of models before moving on to the fine-tuning process.

4.1 Stage I: Five-Dimensional Baseline Evaluation

Initially, we focused our evaluation on Phi-4-mini, Llama 3.3 3B, Gemma 3 4B, and Qwen 3 4B. The outputs generated on these models were evaluation by humans on five dimensions: Language Quality, Style Adherence, Subject Adherence, Coherence, and Cohesion. The results demonstrated that though the existing SLMs could usually, albeit partially, follow subject and stylistic instructions, the quality of Marathi language itself remained inconsistent. Llama 3.3 70B achieved the highest overall score of 3.08/5 and subject adherence was consistently the strongest dimension, reaching a 3.8/5. The other metrics trailed behind with the maximum observed scores being approximately 2.69/5, 3.0/5 and 2.69/5 for style adherence, coherence and cohesion respectively. Language quality remained consistently low, with its maximum reaching approximately 2.6/5.

4.2 Stage II: Strict Marathi Language Gate

Moving forward, a stricter evaluation process was introduced to distinguish literature quality from the basic language validity or correctness. Four compact models (Gemma 3 1B, Llama 3.2 1B, Phi-4-mini, and Qwen3 1.7) were evaluated on 26 Marathi prose and poetry prompts. Placed under this strict criterion, 0 of the total 104 responses managed to pass the Marathi Language Gate. Llama 3.2 1B suffered from Gujarati substitution and repletion loops of certain words in the text. Phi-4-mini and Qen3 1.7B also demonstrated the same issue of repetition loops. Qwen3 1.7B also fall backed to English in some cases. These observation show the importance of checking and verifying actual language correctness and quality before moving on to assigning literary- quality scores.

This failure pattern also revealed particular struggles the models were facing. For instance, potry, satire, abstract, reflection, Marathi dialogue and rural/folk stye prompts were challenging. This motivates explicit language aware evaluation and stronger target language adaption.

4.3 Stage III: Expanded Model Evaluation

The pool of models considered was expanded to include and compare models of different parameter scales. The average scores were as follows:

*EXPANDED PRELIMINARY MODEL EVALUATION*

| Model | Average Score / 5 |
|---|---|
| Gemma 3 4B | 2.99 |
| Llama 3.2 1B | 0.04 |
| Llama 3.2 3B | 1.42 |
| Phi-4-mini 3.8B | 1.33 |
| Qwen3 1.7B | 0.08 |
| Qwen3 4B | 1.57 |
| Llama 3.3 70B | 3.08 |

 This work is licensed under a Creative Commons Attribution 4.0 International License 

The expanded evaluation leads us to the conclusion that though model size and family influence the quality of Marathi generated, the parameter count alone isn't enough to predict and explain performance. This justifies and motivates selection of the base model that will be used for the final fine tuning process through controlled experimentation rather than jumping to the conclusion that the smallest or the largest model will automatically perform most optimally.

## 5 RESEARCH GAP AND FUTURE SCOPE

The appraised studies and provisional experiments emphasize the following research gaps:

### 5.1 Low resource Marathi Literary Generation

Highly exhaustive work is performed on high resource languages like English. However , the analysis of prose and poetry for Marathi is restricted to small data creating a crystal clear research need for the same.

### 5.2 Author specific styles in compact models

The pre-existing work depicts that LLMs imitate the style of authors but it is confined in such a way that that it becomes difficult to distinguish between different Marathi authors for the LLM. Hence the intended design tries to use a corpus of multiple authors with meta data of the Marathi authors.

### 5.3 Unified Prose and Poetry Generation

Many platforms and application targets on either stories or poetry. Whereas a literature model of Marathi language should substantiate both.

### 5.4 Multi dimensional style representation

Vocabulary by itself is not enough to preserve the style of author. It should embrace sematic , syntactic , lexical , emotional , rhetorical and genre-specific features(**jones_style_space?**; **icard_2026?**)

### 5.5 Proposed Research Direction

Based on the study conducted following solution is proposed in Fig. 1:
*Proposed research pipeline*

### 5.6 Parameter efficient adaptation

LoRA and QLoRA provide streamlined ways to adjust pretrained models without modernizing all the parameters (**lora_2021?**; **qlora_2023?**). But it is still doubtful how competent they are in the case of Marathi literature.

## 6 CONCLUSION

The above survey scrutinized the SLMs for literature and poetry generation , stylogenetics , authorial style and fine tuning methods using efficient parameters. Similar words are not enough for evaluating the literature quality. This becomes critical in the case of poetry which has important aspects of rhyme , rhythm, structure and creativity.
The study conducted shows that the existing models only express the surface level regarding the writing style of author. The introductory experiment conducted by us portrayed the issue in generation of Marathi text as the output has repetitive words and language mixing
These results of the study call attention to the need for a concise model which is particularly customized for the generation of Marathi literature. The suggested approach aims to use corpus of multiple authors of Marathi language and QLoRA to formulate Marathi prose and poetry with better language coherence and correctness.

## REFERENCES

[1]. G. Marco, L. Rello, and J. Gonzalo, "Small language models can outperform humans in short creative writing: A study comparing SLMs with humans and LLMs," in Proc. 31st Int. Conf. Computational Linguistics (COLING), 2025, pp. 6552-6570.
[2]. D. K. Thennal and H. O. Hatzel, "Do Large Language Models Always Tell The Same Stories?," arXiv preprint arXiv:2606.17350, 2026.
[3]. P. Sui, Y. Zhu, T. Cheng, P. West, R. J. So, H. Long, and A. Holtzman, "Spoiler Alert: Narrative Forecasting as a Metric for Tension in LLM Storytelling," arXiv preprint arXiv:2604.09854, 2026.
[4]. W. Haverals and M. Martin, "Everyone prefers human writers, including AI," arXiv preprint arXiv:2510.08831, 2025.
[5]. G. Marco, J. Gonzalo, and V. Fresno, "The Reader is the Metric: How Textual Features and Reader Profiles Explain Conflicting Evaluations of AI Creative Writing," in *Proc. 63rd Annu. Meeting of the Association for Computational Linguistics (ACL)*, 2025, pp. 25432-25449.

 This work is licensed under a Creative Commons Attribution 4.0 International License 

I. Koziev and L. Sinev, “Computational Approaches to Automatic Poetry Generation and Evaluation: A Survey," *Journal of Artificial Intelligence Research*, vol. 86, Art. 4, 2026, doi: 10.1613/jair.1.20584.

[6]. H. Gao, C. Song, P. Zhang, J. Zhang, C. Yang, and L. Ge, “LLM- SLM Collaborative Framework of Idiomatic Expression Generation," in *Findings of ACL*, 2026.

[7]. C. Chhun, “Meta-Evaluation Methodology and Benchmark for Auto- matic Story Generation,” Ph.D. dissertation, Institut Polytechnique de Paris, Paris, France, 2024.

[8]. T. Chakrabarty, P. Laban, and C.-S. Wu, ‘Can AI writing be salvaged? Mitigating Idiosyncrasies and Improving Human-AI Alignment in the writing Process through Edits,’’ 2025.

[9]. J. O’Sullivan, stylometric comparisons of human versus AI-generated creative writing,” *Humanities and Social sciences Communications*, vol. 12, Art. 1708, 2025, doi: 10.1057/s41599-025-05986- 3.

[10]. Z. Wang, N. I. Tripto, S. Park, Z. Li, and J. Zhou, Catch Me If You Can? Not Yet: LLMS still struggle to Imitate the Implicit writing styles of Everyday Authors,” in *Findings of EMNLP*, 2025.

[11]. G. Mikros, Beyond the surface: stylometric analysis of GPT-40’s capacity for literary style imitation, *Digital Scholarship in the Humanities*, vol. 40, no. 2, pp. 587- -600, 2025, doi: 10.1093/11c/fqaf035.

[12]. S. Jones, J. F. Naveda, P. Roetling, J. Harrington, and S. Thakkar, On the structure of style space for Documents, in *AAAI Fall Symposium*, 2004.

[13]. B. Icard, L. Sainero, A. Breton, E. Zve, and J.-G. Ganascia, ‘Measuring Embedding Sensitivity to Authorial style in French: Comparing Literary Texts with Language Model Rewritings,’ arxiv preprint arxiv:2605.10606, 2026.

[14]. W. Shi, Z. Zhang, S. Yan, and Y. Yang, ‘LLMS Know More Than words: A Genre study with Syntax, Metaphor & Phonetics,’’ arxiv preprint arxiv: 2512.04957, 2025.

[15]. B. Yazan, A. Akdag, and D. Dell’Anna, ‘what Makes Literary Quotes Stand Out? A Multidimensional Analysis of Literary Creativity,’’ in *Proc. Int. Conf. Computational Creativity (ICCC)*, 2026.

[16]. P. Ferreira, C. silva, A. Alves, and H. Gonçalo oliveira, The Flow of LLM-Generated Poetry, in *Proc. Int. Conf. Computational Creativity (ICCC)*, 2026.

[17]. V. D. Jobanputra, B. T. Reddy, S. G. Bhojanapalli, K. A. S. V. Krishna Aditya, B. Chandrasekara, and R. Murali, LLM-aided Evolutionary Algorithms for Haiku Generation, in *Proc. Genetic and Evolutionary Computation Conference (GECCO ’25 Companion)*, 2025, doi: 10.1145/3712255.3734358.

[18]. B. Ma, Y. Yao, and A.-C. Haensch, ‘Capabilities and Evaluation Biases of Large Language Models in classical Chinese Poetry Generation: A Case Study on Tang Poetry, in *Findings of the Association for Computational Linguistics: ACL 2026*, 2026, pp. 16933–16957.

[19]. X. Vinaixa Roselló, ‘Evaluating LLM Generalization of Literary style: Fine-tuning on Federico García Lorca’s Poetry, 2026.

[20]. E. J. Hu, Y. Shen, P. wallis, z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and w. Chen, ‘LORA: Low-Rank Adaptation of Large Language Models, arxiv preprint arxiv:2106.09685, 2021.

[21]. T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “QLORA: Efficient Finetuning of Quantized LLMS,” arxiv preprint arxiv:2305.14314, 2023.

 This work is licensed under a Creative Commons Attribution 4.0 International License 