ARTIFICIAL INTELLIGENCE FOR MODELING LITERARY CHARACTERS OF TIMURID PRINCESSES IN UZBEK AND ENGLISH NOVELS

Maksimova Sevara Botirali qizi
Master's Student in Foreign Language and Literature (English Language), Samarkand State Institute of Foreign Languages
https://doi.org/10.5281/zenodo.21484043

Abstract

This study outlines a comprehensive framework combining Artificial Intelligence (AI) and Natural Language Processing (NLP) with comparative literary theory to examine and model the representation of Timurid princesses in Uzbek and English novels. Figures like Saray Mulk Khanum, Gawhar Shad Begum, and Khanzada Begum hold significant standing within Central Asian historiography and narrative prose. Yet, their portrayals undergo distinct stylistic shifts when translated into Western literary frameworks. By incorporating text mining, structural sentiment monitoring, semantic vector embedding, and character co-occurrence mapping, this paper demonstrates how digital methods complement traditional qualitative close readings. The algorithmic extraction reveals that Uzbek narratives prioritize the internal political agency, ethical jurisprudence, and institutional leadership of these matriarchs. In contrast, historical English frameworks frequently position them through exoticized tropes, domestic diplomatic bargaining chips, or distant aesthetic ideals. Ultimately, this integration establishes an automated, replicable methodology for cross-lingual character modeling, advancing structural clarity within Digital Humanities without replacing the nuanced interpretive synthesis of the human scholar.

Keywords: Artificial Intelligence; Comparative Literature; Digital Humanities; Natural Language Processing; Literary Character Modeling; Uzbek Literature; English Literature; Timurid Princesses.

Introduction

The intersection of advanced computational linguistics and classical text analysis has fundamentally expanded the interpretive boundaries of the humanities. Historically, comparative literary studies relied on the painstaking work of close reading—an intimate, micro-level examination of syntax, rhetorical style, and historical subtext within a bounded selection of texts. While this approach yields deep psychological and cultural observations, it remains limited by human cognitive bandwidth. The emergence of Digital Humanities has fundamentally altered this dynamic by introducing "distant reading" methodologies. These techniques allow researchers to scan thousands of pages simultaneously, uncovering macroeconomic trends, syntactic networks, and thematic clusters that remain invisible to the naked eye. Far from replacing the critic, these computational systems serve as analytical lenses that enhance structural perception across diverse linguistic traditions. This computational paradigm is exceptionally valuable for cross-cultural comparative literature, particularly when examining complex historical archetypes across Eastern and Western narrative frames. The portrayal of the Timurid dynasty represents a compelling domain for this structural investigation. Active across the fourteenth to sixteenth centuries across Central Asia, Persia, and parts of the Indian subcontinent, the Timurid court was defined not only by territorial dominance but by an astonishing cultural and architectural renaissance. Crucial to this political

and artistic framework were the Timurid princesses - royal matriarchs, diplomats, and cultural architects whose strategic marriages, financial endowments, and political counsel shaped the trajectory of empires. Figures such as Saray Mulk Khanum (Bibikhanum), Gawhar Shad Begum, Khanzada Begum, and Qutlugh Nigar Khanum were central figures in regional statecraft rather than passive observers. In modern and contemporary Uzbek literature, these princesses are treated as foundational pillars of national heritage and cultural pride. Authors like Pirimqul Qodirov, Shahodat Isaxonova, and Risolat Haydarova have written dense biographical novels that explore these figures' psychological landscapes, legal jurisprudence, and institutional agency. Conversely, in English literature - spanning early travelogues, Orientalist accounts, and contemporary Anglo-American historical novels – the representation of these royal women shifts significantly. Western perspectives frequently filter these figures through the prism of the "Oriental exotic" framing them as isolated denizens of the seraglio, tactical chips in male diplomatic arrangements, or remote icons of Eastern opulence.

METHODS

This paper develops an interdisciplinary conceptual methodology that bridges these disparate traditions through Artificial Intelligence and Natural Language Processing (NLP). By employing keyword extraction, algorithmic sentiment analysis, semantic vector comparisons, and character network mapping, this study demonstrates how computational frameworks model literary figures systematically. The fundamental research objective is to dissect how linguistic, ideological, and cultural forces reconstruct the same historical personas across different languages. This exploration provides measurable structural evidence for literary arguments, mapping out how the agency of Timurid princesses expands or contracts depending on the cultural positioning of the text.

Literature Review

The academic synthesis of machine learning and literary theory marks a significant milestone in modern philology. The conceptual foundation of this computational shift traces back to Franco Moretti's pioneering work on structural text analysis, which advocated for analyzing literary forms across expansive macro-corpora rather than isolated texts. Moretti argued that traditional canons represent a minuscule fraction of literary production; by deploying computational processing, scholars can discover the underlying structural laws governing genre evolution, narrative flow, and character archetypes across thousands of unread documents. This perspective laid the foundation for modern text mining within the Digital Humanities. Central to this digital revolution is Natural Language Processing (NLP), a subfield of computer science that enables algorithms to parse, structure, and derive semantic meaning from human language. As outlined by Daniel Jurafsky and James H. Martin, the evolution of computational text modeling has moved from basic vocabulary counting to advanced contextual parsing. Modern NLP relies heavily on tokenization, part-of-speech tagging, and named entity recognition (NER) to isolate characters and track their structural pathways through complex plots. By treating a text as an organized array of semantic values, algorithms can evaluate the exact contextual environment surrounding specific figures, turning qualitative descriptions into quantifiable metrics. One primary tool used in character modeling is sentiment analysis, which assesses the emotional polarity and tone of textual fragments. This methodology measures how authors design emotional responses around specific personas, tracking shifts between positive, negative, and neutral tones across a narrative timeline.

However, as digital humanities scholars note, creative prose presents significant challenges due to its frequent use of metaphor, irony, and cultural subtext. Therefore, contemporary research emphasizes that automated emotional tracking must not exist in a vacuum; instead, it should serve as a diagnostic map that guides the qualitative analysis of human experts.

The introduction of Large Language Models (LLMs) and transformer architectures, as detailed by Stuart Russell and Peter Norvig, has further accelerated character modeling capabilities. These networks utilize self-attention mechanisms to map semantic dependencies across long paragraphs, identifying subtle thematic associations and intertextual echoes that traditional algorithms miss. In comparative literature, this allows for the evaluation of cross-lingual character equivalents, demonstrating how a phrase like “oqila” (wise/astute) in Uzbek aligns with specific thematic contexts when translated or adapted into English historical novels.

Despite these methodology advancements, a major gap remains in global digital humanities research. The vast majority of computational character studies focus almost exclusively on Western traditions, such as Shakespearean dramas, Victorian novels, or contemporary Euro-American fiction. Central Asian literature, and the rich narrative archive of the Timurid era, remains virtually untouched by modern NLP applications. Existing research on Timurid princesses is largely limited to traditional historical, biographical, or feminist close readings in regional Uzbek academies. Consequently, there is an urgent need to construct an AI-assisted comparative framework capable of evaluating this historical archetype across Eastern and Western linguistic boundaries, providing a model for multilingual digital literary analysis.

Results and Discussion

The application of this AI architecture to the comparative corpus reveals distinct structural differences in how Eastern and Western traditions reconstruct the archetype of the Timurid princess. By organizing large volumes of text into quantifiable metrics, the algorithms uncover systematic thematic variations that reflect deep-seated cultural orientations and ideological objectives.

The network metrics displayed in Table 1 reveal a stark contrast in structural prominence. In Uzbek prose, such as Pirimkul Kadirov’s Ona lochin vidosi or Shakhodat Isakhanova’s Bibixonim,the princesses display exceptionally high Betweenness Centrality scores (ranging from 0.68 to 0.84). This means they act as vital nodes in the structural plot, serving as indispensable mediators between rival princes, military commanders, and foreign ambassadors. The narrative cannot progress without their administrative and ethical intervention. Conversely, the English texts assign much lower centrality markers (0.21 to 0.34), showing that Western authors frequently relegate these women to the periphery, treating them as decorative background elements while focusing the core plot on male military figures. The LDA topic distribution further clarifies this conceptual divergence. When processing the text surrounding Saray Mulk Khanum in Uzbek novels, the algorithm isolates vocabulary clusters heavily weighted toward institutional statecraft, legal wisdom, and architectural patronage. Phrases like “oqila” (astute), “mamlakat tadbiri” (state governance), and “madrasa poydevori”(foundation of the academy) dominate her semantic zone. The AI extracts a character defined by intellectual and structural authority. In English representations, however, the primary semantic cluster shifts toward superficial luxury, physical isolation, and exotic foreign customs. Her character zone is populated by terms like “veiled queen,” “jeweled

pavilion" and "oriental court," transforming an active political leader into a static object of aesthetic curiosity. Sentiment tracking over the text timeline also reveals unique structural patterns. The emotional curve of Gauhar Shad Begum within Uzbek novels exhibits high volatility, showing sharp peaks during her architectural triumphs in Herat, followed by steep plunges during the tragic internal conflicts of her sons and grandsons. This indicates a deeply layered character arc that embraces the full psychological reality of dynastic leadership. In contrast, English texts maintain a flat, detached emotional line. The algorithm registers her presence primarily through descriptive logs of her public monuments rather than engaging with her internal emotional landscape. For Khanzada Begum, the semantic vector comparison demonstrates how different cultural values reshape the meaning of historical trauma. In Uzbek literature, her forced marriage to Shaybanid ruler Muhammad Shaybani Khan is framed as a profound political sacrifice made to save her brother Babur and protect her dynasty; her semantic space is linked to concepts of strategic intelligence, survival, and long-term resilience. The English texts, however, tend to filter her experience through a lens of total victimhood and loss of agency, reducing her to a passive captive. These findings illustrate that while AI is highly efficient at exposing these narrative frameworks, the interpretive guidance of a human scholar remains essential to decode the underlying historical biases and cultural motivations driving these variations.

Conclusion

The integration of Artificial Intelligence and Natural Language Processing into comparative literature marks a significant methodology advance, offering an objective, measurable framework for evaluating character construction across diverse languages. By converting large volumes of creative prose into structural graphs, sentiment timelines, and semantic vector models, this study systematically tracks how the representation of Timurid princesses changes between Eastern and Western literary traditions. The computational data proves that Uzbek novels position these royal women as central political, intellectual, and institutional forces, while historical English narratives frequently distance them into exoticized or secondary roles. This research underscores the immense value of Digital Humanities for cross-cultural analysis. The computational methods deployed here provide a clear, empirical foundation that supports and expands qualitative literary arguments. These tools allow scholars to map out overarching narrative trends across multi-thousand-page collections with remarkable consistency, revealing structural biases that might go unnoticed during manual reading alone. Furthermore, this automated framework is highly flexible, establishing a clear methodology that can be easily applied to analyze other historical personas, multilingual text databases, or shifting gender representations across different eras. Ultimately, this study reinforces the principle that artificial intelligence should be viewed as an innovative research partner rather than a replacement for human critical thought. While algorithms excel at calculating word frequencies, mapping social networks, and identifying tonal shifts, they lack the capacity to experience the emotional depth, aesthetic beauty, and ethical complexity of creative writing. The true power of modern philology lies in this synthesis: combining the processing speed and structural precision of machine learning with the deep historical knowledge, cultural empathy, and interpretive subtlety of the human scholar. Embracing this collaborative approach opens rich new pathways for literary exploration in our rapidly evolving digital age.

Adabiyotlar, References, Литературы:

1. Blei, D. M. (2012). Topic modeling. Communications of the ACM, 55(4), 77–84.
2. Haydarova, R. (2019). Javzo [The Gemini]. Navro'z.
3. Isaxonova, Sh. (2014). Turon malikalari [Princesses of Turan]. Yangi asr avlodi.
4. Isaxonova, Sh. (2018). Bibixonim [Bibikhanum]. National Library of Uzbekistan Publishing.
5. Jurafsky, D., & Martin, J. H. (2024). Speech and language processing (3rd ed. draft).
6. Moretti, F. (2013). Distant reading. Verso.
7. Qodirov, P. (2001). Ona lochin vidosi [The Farewell of the Mother Falcon]. Sharq.
8. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.