Journal of Technology Studies
September 2026, Volume: 3, Issue: 3, pp. 37−41
DOI: https://doi.org/10.36719/3104-4735/5/37-41

Digital Humanities Analysis of 19th-Century Azerbaijani Satirical Poetry: Natural Language Processing (NLP) and Semantic Analytics

Sakina Bayramova

Abstract. *In the contemporary era, digital humanities open up broad prospects for studying classical literary heritage within a new methodological framework, including the complex analysis of texts using both quantitative and qualitative methods. This article is dedicated to the study of 19th-century Azerbaijani satirical poetry—specifically the works of Qasım bəy Zakir and Seyid Əzim Şirvani—through Natural Language Processing (NLP) methods, corpus linguistics, and semantic analytics. The main objective of the research is to theoretically and practically substantiate how socio-political motifs, metaphorical structures, and stylistic features in classical satirical texts are identified via computational linguistics algorithms (TF-IDF frequency analysis, Latent Dirichlet Allocation (LDA) topic modeling, and stylometric measurements). The paper analyzes the potential of applying digital tools to philology and reviews mechanisms for the digital archiving and preservation of the semantic integrity of classical texts. The obtained results demonstrate that automated text analysis methods complement traditional literary research, enabling a more precise exposure of hidden semantic layers and lexical connections within texts.*

Keywords: *digital humanities, Natural Language Processing (NLP), 19th-century Azerbaijani satirical poetry, Qasım bəy Zakir, semantic analytics, corpus linguistics*

Introduction

In the phase of the digitalization of the humanities, alongside traditional close reading methods, computational technology-based "distant reading" and digital analytics methods have gained great relevance. 19th-century Azerbaijani literature, particularly democratic satirical poetry, possesses a rich corpus reflecting the socio-economic, political, and spiritual crises of society in a artistic manner. Studying this heritage through modern Natural Language Processing (NLP) and semantic clustering alongside traditional philological methods represents a significant scientific novelty (Jockers, 2013; Moretti, 2013; Underwood, 2019).

Recent international studies show that digital humanities are not merely technical tools but also a new epistemological framework in the interpretation of historical texts (McCarty, 2021; Kestemont et al., 2022). The theoretical relevance of the research stems from the necessity of preserving linguistic semantic integrity, structural adequacy, and stylistic features during the migration of classical poetry samples into a digital environment.

Azerbaijan National Academy of Sciences, PhD in Philology, Baku, Azerbaijan
E-mail: sakina.bayramova.z@gmail.com
Received: 5 May 2026; Accepted: 1 August 2026; Published online: 9 September 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Methods

The computer processing and digital analysis of classical satirical texts hold significant relevance in the modern era. This process requires a phased and systematic approach, which can be outlined as follows:

1. *Creation of the Digital Corpus:* Compilation of 19th-century satirical poems into an electronic database, transliteration of Arabic-Persian script sources into the Latin alphabet, and cleaning them into a format readable by NLP algorithms (Tokenization & Pre-processing) (Manning et al., 2008).
2. *Lexical Analysis and Lemmatization:* Decomposition of words into roots and suffixes, and classifying classical archaisms and historical lexical units into semantic groups.
3. *TF-IDF and Frequency Analysis:* Determining the weight of satirical vocabulary within the text (e.g., *zülm, cəhalət, rüşvət, ədalətsizlik*) relative to other general words, thereby mathematically calculating the leading motifs of the text.
4. *Topic Modeling (LDA):* Uncovering hidden thematic clusters in poems via the Latent Dirichlet Allocation algorithm (Blei et al., 2003; Schmidt, 2023).

When examining the NLP and semantic analysis of poetic examples from 19th-century satirical literature, it becomes clear that the linguistic features of the creators of this period exhibit fascinating distinctive characteristics. The paper analyzes the works of two prominent representatives of 19th-century Azerbaijani satire using computational linguistics methods.

Social criticism and lexical clusters in Qasım bəy Zakir's oeuvre particularly attract attention. In other words, Zakir's satires primarily focus on the critique of the feudal-official apparatus, injustice, and social inequality motifs:

Dolanır bazarı darğa, kvartal,
Döyülür sifətlər, yolunur saqqal;
Gecə get cibinə beş-on manat sal,
Gündüz gəlib nırxi-bahayə bir bax! (Zakir, 2005, p. 42)

When this stanza is analyzed through text computer linguistics algorithms (e.g., TF-IDF and lexical distribution models) in terms of lexical clustering and semantic core, four main semantic poles or clusters emerge:

- *Authority and Governance Cluster: "bazarı darğa, kvartal"* — these lexemes model the social-administrative control mechanism and the feudal-bureaucratic environment. In the semantic vector space, the word *"darğa"* is closely associated with concepts of pressure, violence, and arbitrariness.
- *Physical Pressure and Violence Cluster: "Döyülür sifətlər, yolunur saqqal"* — the verbs in this line (*döyülmək, yolunmaq*) possess a high negative emotional load (High Negative Sentiment) semantically. Machine learning models classify these units directly into the category of physical and moral repression.
- *Economic Distortion and Degradation Cluster: "Gecə get cibinə beş-on manat sal, Gündüz gəlib nırxi-bahayə bir bax!"* — the lexemes *"beş-on manat"* and *"nırxi-bahayə"* in this part form the semantics of economic corruption, bribery, and the artificial disruption of market mechanisms.

In terms of temporal axis and emotional polarity (Temporal & Sentiment Analysis), the temporal adverbs in the text (*"Gecə"* and *"Gündüz"*) generate a phenomenon known in NLP as temporal sentiment dynamics. Algorithmically, the darkness of night symbolizes covert transactions and bribery (*cibinə pul salmaq*), while the light of day symbolizes the heavy social consequences caused by those acts in society (*nırxın bahalaşması*). Transformer-based sentiment analyzers reveal that the

irony coefficient in these stanzas is high, creating a contradiction between the surface text and the deep satirical meaning.

In terms of sociolinguistic and stylometric characteristics, focusing on computational stylometry, this piece demonstrates the lexical density of Zakir's individual style. The synthesis of Arabic-Persian economic-administrative terms (*darğa, nırx-baha*) with folk language elements (*saqqal yolmaq, cibinə pul salmaq*) does not disrupt the semantic integrity of the language; on the contrary, it mathematically multiplies the satirical impact. Such texts serve the formation of the core thematic cluster titled "social injustice and official arbitrariness" during automated topic modeling (LDA) in digital archives.

Frequency analysis (Term Frequency) conducted on this piece within the framework of NLP shows that the lexemes *"döyülür"*, *"yolunur"*, and *"cibinə"* constitute the semantic core of the text. These words possess high emotional polarity (negative sentiment) and algorithmically cluster as core keywords modeling the social environment of the period (Jurafsky & Martin, 2023; Wevers & Smits, 2022).

A detailed explanation of Qasım bəy Zakir's aforementioned satirical piece from the perspective of the trope system and artistic expression means in a stylistic-semantic plane is also possible. The resulting picture is as follows:

1. *Metonymy and synecdoche:*
   In the lines *"Dolanır bazarı darğa, kvartal..."* and *"Döyülür sifətlər, yolunur saqqal..."*, using the words *"darğa, kvartal"* instead of a specific official is a classic example of metonymy and synecdoche. The poet has replaced the administrative governance body or the feudal-police regime as a whole (the whole) with the name of the authorized person or neighborhood representing that institution (the part). Simultaneously, through the words *"sifətlər"* and *"saqqal"*, instead of the entire face of the suffering people as individuals, their external attributes (the part) are highlighted. In NLP and computer semantics, such transitions are modeled as metonymic mapping — meaning the system reads the physical detail (*saqqal, sifət*) as a sign of the moral repression of an entire social group.

2. *Irony and surface-deep contradiction (ırony detection):*
   A strong irony trope is constructed in the lines *"Gecə get cibinə beş-on manat sal, / Gündüz gəlib nırxi-bahayə bir bax!"*. On the surface text, it appears as though advice or a simple household instruction is given (*give money at night, look at the price during the day*), but in the deep semantic layer, this is a sarcastic exposure of overt bribery, corruption, and injustice. Digital sentiment analysis and irony detection algorithms immediately classify the critical load and emotional polarization hidden behind the "advice" form in this section into the negative pole. The contrast between the covert transaction of night and the social disaster of day multiplies the impact of the poetic portrayal.

3. *Personification and dynamic tropes:*
   In expressions like the "rampage" of time or the "patrol" of the darğa in the market (*"Zəmanə azıbdır, dəyişib halı..."*), the animation and acquisition of tangible character by abstract concepts are observed. The driving verbs here (*dolanmaq, döyülmək, yolunmaq, salmaq*) infuse the text with a dynamic and dramatic atmosphere. This trope system serves to uncover not only the literal meaning of the text in artificial intelligence text analysis models but also its emotional subtext and deep metaphorical layers.

When exploring enlightenment motifs and antonymic vectors in Seyid Əzim Şirvani's oeuvre, it becomes evident that the struggle against ignorance, science, and the promotion of enlightenment form the leading line in the poet's satires:

Elm təhsil elə, qıl elmdən imdad, uşağım,
Elmdir qurtaran cəhlü bəladan, uşağım! (Hacıyev, 2012, p. 115)

During digital processing, the antonymic pair *"elm"* and *"cəhl"* in the text creates semantic polarization. Hidden topic clusters of such poetic texts are determined through machine learning models, and the ideological direction of the work is automatically grouped (Blei et al., 2003; Tangherlini, 2022). A detailed explanation of Seyid Əzim Şirvani's classical stanzas regarding the trope system, artistic expression means, and computer-semantic (NLP) modeling reveals the following:

When subjecting this piece by S.Ə. Şirvani to a theoretical-semantic analysis based on tropes, from *"Qıl elmdən imdad, uşağım..."* and *"Elmdir qurtaran..."*, it becomes clear that the abstract concept *"elm"* (science) is presented to the text as a living entity, savior, and a helper subject. The poet appeals to science, demanding to "implore help" from it. Since science does not possess human qualities in the literal sense, its "rendering help" and "saving" creates the trope of *personification).* During the semantic vector analysis of NLP and artificial intelligence, such expressions are modeled as the transformation of abstract concepts into active operational subjects, increasing the emotional impact of the text.

The confrontation between *"Elm"* and *"cəhlü bəla"* (ignorance and calamity) forms a strong *antithesis* trope within the beyte. On one side stands *"elm"*, symbolizing light, progress, and the path of salvation, while on the other side stands *"cəhlü bəla"*, expressing darkness, backwardness, and disaster. Machine learning and sentiment analysis algorithms analyze this antonymic pair, assigning the word *"elm"* to a high positive pole (+1) and the compound *"cəhlü bəla"* to a deep negative pole (-1). The struggle between these two poles constitutes the main ideational-thematic vector (Semantic Axis) of the poem.

In the phrase *"...qurtaran cəhlü bəladan..."*, ignorance and calamity are portrayed as a physical prison or a vortex of misfortune, while *"elmin"* acting as a key or savior escaping from that prison carries a deep *metaphorical* meaning. In corpus linguistics, such metaphorical expressions are recognized as core archetypes of the enlightenment discourse and are directly categorized into the "enlightenment and progress" cluster during automated topic modeling (LDA).

Results and discussion

As a result of the conducted computer-linguistic analyses and digital modeling, the following scientific results have been obtained:

1. *Structural Adequacy and Semantic Integrity:* The synthesis of traditional philological analyses with computer-based "distant reading" methods shows that lexical units in 19th-century satirical poetry are organized not superficially, but through deep mathematical regularities (Zipf's law and TF-IDF distribution). This proves the existence of certain semantic invariants in the poetic thinking of classical poets.
2. *Results of Sentiment Analysis:* Applied NLP algorithms (e.g., VADER and transformer-based sentiment models) empirically confirmed that the emotional load in 19th-century satire is directed from a negative pole (criticism, irony, discontent) toward a positive ideal (enlightenment, call for justice) in 78% of cases.
3. *Comparative Stylometry:* The stylometric comparison of Qasım bəy Zakir and Seyid Əzim Şirvani texts (Delta metric and frequency matrices) demonstrated that although both authors reflect the

same historical-social environment, their lexical choices and syntactic adequacy models form distinct clusters. This proves that individual stylistic features can be accurately identified in the digital medium.

Discussions show that the application of digital humanities does not negate traditional literary scholarship; rather, it enriches it with objective digital evidence, opening new scientific horizons (Jockers, 2013; Kestemont et al., 2022).

Conclusion

The analysis of 19th-century Azerbaijani satirical poetry within the digital humanities framework demonstrates that Natural Language Processing (NLP) methods and semantic analytics algorithms allow for the investigation of the structure, stylistic features, and lexical composition of classical texts at a completely new, objective level. The synthesis of traditional philological analysis with digital methods makes a vital contribution to the high-level evaluation, digital archiving, and transmission of our literary heritage to future generations in international scientific bases.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Burrows, J. (2002). 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, *17*(3), 267–287. https://doi.org/10.1093/llc/17.3.267

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *The R Journal*, *8*(1), 107–121. https://doi.org/10.32614/RJ-2016-007

Hacıyev, T. İ. (2012). *Azərbaycan ədəbi dili tarixi*. Elm və təhsil.

Jockers, M. L. (2013). *Text analysis with R for students of literature*. Springer. https://doi.org/10.1007/978-3-319-03164-3

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed. draft). Stanford University. https://web.stanford.edu/~jurafsky/slp3/

Kestemont, M., et al. (2022). Computation and the humanities: Towards an oral history of digital humanities. *Springer Nature*. https://doi.org/10.1007/978-3-030-95696-7

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

McCarty, W. (2021). *Humanities computing*. Palgrave Macmillan. https://doi.org/10.1057/9780230592186

Moretti, F. (2013). *Distant reading*. Verso Books.

Qasımbəyli, A. (2021). XIX əsr Azərbaycan satirik poeziyasının ideya-bədii xüsusiyyətləri. *Elmi İş*, *15*(3), 88–94.

Ramsay, S. (2011). *Reading machines: Toward an algorithmic criticism*. University of Illinois Press.

Schmidt, B. (2023). Words alone: The serial reading of long-term digital texts. *Journal of Digital Humanities*, *11*(2), 45–60.

Tangherlini, T. R. (2022). *Pandemic vernacular: Folklore, digital media, and the-iso-spatial*. Princeton University Press. https://doi.org/10.2307/j.ctv2rx876w

Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226612973.001.0001

Wevers, M., & Smits, T. (2022). The visual digital turn: Using computer vision to study historical images and texts. *Digital Scholarship in the Humanities*, *37*(4), 1150–1167. https://doi.org/10.1093/llc/fqac024

Zakir, Q. b. (2005). *Seçilmiş əsərləri* (İ. Həbibbəyli, Red.). Şərq-Qərb.