



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

A SURVEY OF COMPUTATIONAL MODELS IN CONTEMPORARY LITERARY STUDIES AND THEIR LIMITS

Isaeva Gulnora Abdukadirovna

Senior Lecturer, Department of Russian Language and Literature Bukhara State University

Bakhtiyarova Mokhinur Feruz qizi

Second-year student, Philology and Language Teaching (Russian) Bukhara State University

Abstract

The article surveys the principal computational models applied to literary texts and weighs their analytical value. Stylometry, distant reading, macroanalysis, topic modelling, sentiment based plot analysis, character networks and machine learning are examined through their founding studies and measurable results. Each method is described by its unit of analysis, its representative tool and the empirical scale at which it has been tested, from the 150 word feature set of Burrows's Delta to corpora of several thousand novels. The critique advanced by N.Z.Da is set against the constructive replies of T.Underwood and A.Piper. The survey shows where quantitative evidence genuinely extends literary history and where its claims remain fragile.

Keywords: computational literary studies, stylometry, Burrows's Delta, distant reading, macroanalysis, topic modelling, Latent Dirichlet Allocation, sentiment analysis, emotional arcs, character networks, machine learning, digital humanities.



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

Introduction

Counting words to read a novel once sounded like a contradiction. The practice now has its own journals, packages and controversies, and it carries several competing names. F. Moretti gave the project its founding provocation when he argued that scale itself forces a change of instrument. The history of literature, he held, cannot be enlarged simply by reading further into the archive. His formulation remains the most cited statement of the field's motive. "world literature is not an object, it's a problem, and a problem that asks for a new critical method: and no one has ever found a method by just reading more texts". The sentence reframes the archive as a problem of method rather than of coverage, and it set the terms for everything that followed. Two decades later the field still lacks a single label. N.Z. Da, its sharpest internal critic, catalogues the proliferation directly, listing the subfield as it travels under many headings.

That instability of naming is not cosmetic. It marks genuine disagreement about whether the enterprise is a branch of statistics, a wing of cultural history or a new descriptive poetics. This article maps the main models against one another, gives each its founding study and its measurable scale, and asks a single question. Where does the arithmetic earn its keep, and where does it merely restate what close reading already knew? Stylometry comes first because it is the oldest and the most testable. The survey then moves through distant reading, macroanalysis, topic modelling, plot arcs, character networks and predictive learning, before turning to the debate that N.Z. Da reopened in 2019.

Methods and literature review

The study used comparative analysis of founding publications, descriptive synthesis of method parameters such as feature sets and corpus size, source



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

criticism applied to primary monographs and journal articles and citation tracing across the controversy that surrounds the field. Quotations were checked against publisher texts rather than secondary summaries, and two widely circulated paraphrases were excluded once they proved untraceable in the printed originals. The literature falls into founders, consolidators and critics. J.F.Burrows established the quantitative baseline for authorship work with his Delta measure. D.L.Hoover later tested that measure on larger samples and confirmed its dependence on the most frequent words of the comparison set. M.Eder, J.Rybicki and M.Kestemont turned the procedure into shared software through the stylo package. F.Moretti supplied the programmatic vocabulary of abstract models for literary history. M.L.Jockers extended the argument from manifesto to corpus practice under the name macroanalysis. D.M.Blei introduced the probabilistic machinery of topic models to a wide computing readership. T.Underwood reframed machine learning as a way of comparing historical perspectives rather than erasing them. A.Piper positioned computation as a partner in the making of literary meaning rather than its replacement.

Results

Stylometry offers the clearest case of a method that does what it claims. Burrows defined the reach of his Delta with unusual candour about its limits. “The procedure offers a simple but comparatively accurate addition to our current methods of distinguishing the most likely author of texts exceeding about 1,500 words in length. It is of even greater value as a method of reducing the field of likely candidates for texts of as little as 100 words in length”. The honesty about thresholds is itself a sign of a mature instrument. Implementation matters as much as theory, and the field acquired a common toolkit in 2016. “This software paper



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

describes ‘Stylometry with R’ (stylo), a flexible R package for the high-level analysis of writing style in stylometry. Stylometry (computational stylistics) is concerned with the quantitative study of writing style, e.g. authorship verification, an application which has considerable potential in forensic contexts, as well as historical research” The same family of techniques famously unmasked J. K. Rowling behind the pseudonym Robert Galbraith in 2013, a result the Sunday Times confirmed and the author conceded.

Distant reading rests on a deliberate exchange of detail for pattern. Moretti stated the trade openly in the opening page of *Graphs, Maps, Trees*. “‘Distant reading’, I have once called this type of approach; where distance is however not an obstacle, but a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models” Distance becomes a kind of vision rather than a loss of it. Where Moretti theorised, M. L. Jockers built. His *Macroanalysis* treats the digital library as a new evidentiary condition rather than a convenience. “Today, in the age of digital libraries and large-scale book-digitization projects, the nature of the ‘evidence’ available to us has changed, radically. Which is not to say that we should no longer read books looking for, or noting, random ‘things,’ but rather to emphasize that massive digital corpora offer us unprecedented access to the literary record and invite, even demand, a new type of evidence gathering and meaning making” The book grounds that claim in a corpus of more than three thousand nineteenth century novels. Topic modelling answers a different need, the discovery of themes nobody labelled in advance. D. M. Blei described the appeal in plain terms. “Topic modeling algorithms are statistical methods that analyze the words of the original texts to discover the themes that run through them, how those themes are connected to each other, and how they change over time” A.



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

Goldstone and T. Underwood applied the method reflexively, modelling 5,940 articles from PMLA at one hundred topics to read the history of literary scholarship itself.

Sentiment analysis gave plot a measurable shape. A. J. Reagan and colleagues reduced thousands of stories to a small set of recurring emotional trajectories. “by classifying the emotional arcs for a filtered subset of 1,327 stories from Project Gutenberg’s fiction collection, we find a set of six core emotional arcs which form the essential building blocks of complex emotional trajectories” Figure 1 reproduces the six shapes their study isolated.

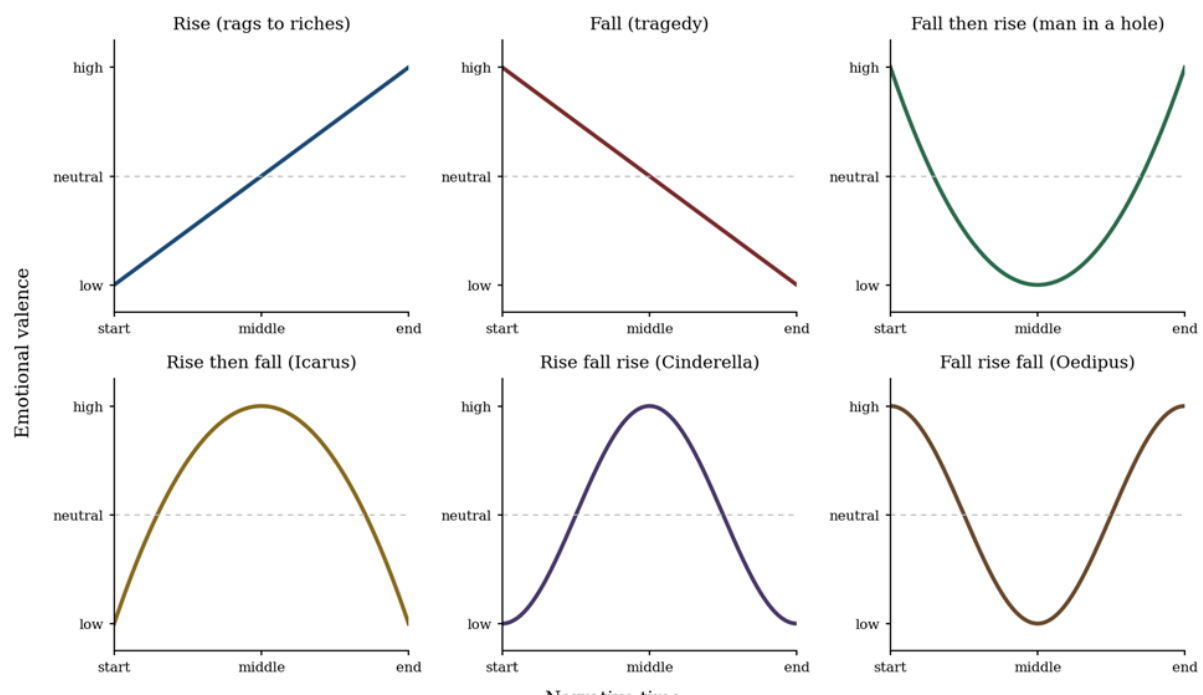


Figure 1. The six core emotional arcs of narrative after Reagan

Plot can also be read as structure rather than feeling. F. Moretti recast drama as a graph in which figures are nodes and their exchanges are edges. “the analysis of



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

how vertices are linked by edges has revealed many unexpected features of large systems, the most famous one being the so-called ‘small-world’ property, or ‘six degrees of separation’: the uncanny rapidity with which one can reach any vertex in the network from any other vertex” Removing the Hamlet node nearly splits the play’s network into two halves. Table 1 sets these methods side by side, with the unit each one measures and the scale at which it has been demonstrated.

Table 1. Comparative overview of computational models in literary studies

Method	Principal proponent	Representative technique or tool	Unit of analysis	Empirical anchor
Stylometry and authorship attribution	J. F. Burrows; M. Eder	Burrows’s Delta; the stylo package in R	150 most frequent words	Reliable above c. 1,500 words
Distant reading	F. Moretti	Graphs maps and trees as abstract models	Genres and national corpora	Stanford Literary Lab pamphlets
Macroanalysis	M. L. Jockers	Theme and influence modelling at scale	Whole literary record	3,346 nineteenth century novels
Topic modelling	D. M. Blei; A. Goldstone and T. Underwood	Latent Dirichlet Allocation	Latent themes across documents	5,940 PMLA articles at 100 topics
Sentiment and plot arcs	A. J. Reagan and colleagues	Matrix factorisation and clustering of valence	Emotional trajectory	1,327 Project Gutenberg works
Character networks	F. Moretti	Graph of speech act edges between figures	Dramatis personae	Network reading of Hamlet
Predictive modelling and embeddings	T. Underwood; A. Piper	Regularised classification; word vectors	Genre prestige and gender	Two centuries of prose



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

Predictive modelling closes the sequence. T. Underwood argues that machine learning is valuable precisely because it absorbs the assumptions buried in its training evidence, and he turns that property from a defect into a research instrument. “By training models on evidence selected by different people, we can crystallize different social perspectives and compare them rigorously to each other” On that view a classifier trained on Victorian reviewers and one trained on modern anthologists become two readable positions rather than one neutral verdict. The same instrument can be turned on the discipline that wields it. Why topic modelling drew literary scholars at all is best put by D. M. Blei, who stressed that the themes are not supplied beforehand. “Topic modeling algorithms do not require any prior annotations or labeling of the documents — the topics emerge from the analysis of the original texts. Topic modeling enables us to organize and summarize electronic archives at a scale that would be impossible by human annotation” Read against the PMLA experiment of A. Goldstone and T. Underwood, the claim acquires a reflexive edge, since the archive being summarised is the record of literary criticism itself.

Discussion

The strongest objection comes from inside the field. N. Z. Da reads a large sample of published work and reaches a blunt diagnosis. “In a nutshell the problem with computational literary analysis as it stands is that what is robust is obvious (in the empirical sense) and what is not obvious is not robust, a situation not easily overcome given the nature of literary data and the nature of statistical inquiry” Her charge is not that numbers lie but that the surprising findings tend to be the unstable ones, while the stable findings tend to restate the familiar. A. Piper answers from a different premise, treating computation as a maker of meaning



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

rather than a verifier of facts. “It is less a handbook and more an extended demonstration of the ways in which computation, when applied critically and creatively, can confirm, revise, but also invent new narratives about literary history” Between these positions the survey suggests a working rule. Stylometry and topic modelling repay their cost when the question is descriptive and the corpus is large, while aesthetic or evaluative claims still need the reader that A. Piper keeps in view.

Conclusion

Across seven families of method the pattern is consistent. Computation extends literary history most reliably where the unit of analysis is countable and the corpus is large, as in authorship attribution, thematic discovery and the mapping of emotional trajectory, and it grows fragile where the object is judgement, value or singular interpretation. The models reviewed here do not replace reading. They redraw the scale at which a question can be asked, and they make the analyst’s assumptions visible enough to be argued about, which is the condition N. Z. Da demands and the opportunity T. Underwood and A. Piper describe.

References

1. Blei D. M. Probabilistic Topic Models // Communications of the ACM. New York: Association for Computing Machinery, 2012. Vol. 55, No. 4. P. 77-84.
2. Burrows J. F. ‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship // Literary and Linguistic Computing. Oxford: Oxford University Press, 2002. Vol. 17, No. 3. P. 267-287.



International Conference on Computing, Artificial Intelligence and Information Systems

Hosted Online from Warsaw, Poland

Date: 11th June, 2026

Website: <https://econferencia.com>

-
3. Da N. Z. The Computational Case against Computational Literary Studies // *Critical Inquiry*. Chicago: The University of Chicago Press, 2019. Vol. 45, No. 3. P. 601-639.
 4. Eder M., Rybicki J., Kestemont M. Stylometry with R: A Package for Computational Text Analysis // *The R Journal*. Vienna: R Foundation for Statistical Computing, 2016. Vol. 8, No. 1. P. 107-121.
 5. Goldstone A., Underwood T. What Can Topic Models of PMLA Teach Us about the History of Literary Scholarship? // *Journal of Digital Humanities*. 2012. Vol. 2, No. 1. P. 39-48.
 6. Hoover D. L. Testing Burrows's Delta // *Literary and Linguistic Computing*. Oxford: Oxford University Press, 2004. Vol. 19, No. 4. P. 453-475.
 7. Jockers M. L. *Macroanalysis: Digital Methods and Literary History*. Urbana; Chicago; Springfield: University of Illinois Press, 2013. 224 p.