Vol. 4 No. 1 (2026)

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895
https://llrjournal.com/index.php/11

DIGITAL HUMANITIES AND COMPUTATIONAL STYLISTICS IN CONTEMPORARY ENGLISH LITERARY STUDIES

Namra Bashir[*1], Khadija Zahid[2], Rafea Bukhari[3], Hafiz Haqnawaz[4], Talha[5]

*[*1]Department of English Literature, Riphah international University, Lahore, Pakistan*
[2]Department of English, Abdul Wali Khan University Mardan
[3]Department of English Literature, University of Balochistan, Quetta.
[4]Department of English Literature, University of Makran, Panjgur
[5]University of Makran

[*1]namrabashir75@gmail.com, [2]mmusknkkhan@gmail.com, [3]syyeda.rafea.bukhari@gmail.com, [4]haqnawazkhaliq514@gmail.com, [5]talha@uomp.edu.pk

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

DOI: http://doi.org/10.5281/zenodo.21947855

Abstract

This study provides a comprehensive examination of digital humanities and computational stylistics within contemporary English literary studies, tracing the evolution of stylistic analysis from classical rhetoric to twenty-first-century computational methodologies. The paper systematically surveys the theoretical foundations, mathematical frameworks, and methodological innovations that define this rapidly expanding interdisciplinary field, with particular emphasis on the dialectical relationship between traditional hermeneutic practices and quantitative text analysis. Through critical analysis of Burrows' Delta metric and its variants including Quadratic Delta and Cosine Delta the review elucidates how mathematical modeling of high-frequency function words enables robust authorship attribution and stylistic distance measurement across large digital corpora. The paper explores the paradigm of scalable reading as a mixed-methods approach that integrates macro-level distant reading with micro-level close textual analysis, demonstrating how computational tools function as heuristic mechanisms for literary discovery rather than replacements for human interpretation. Additionally, the review addresses corpus-driven narratology, structural parsing, and computational character modeling, highlighting both the achievements and persistent challenges of applying natural language processing to literary texts. Critical controversies, particularly Nan Z. Da's computational case against computational literary studies, are examined alongside defenses of exploratory data analysis as a valid critical practice. The review concludes by investigating frontier developments in generative AI and cognitive stylometry, including LLM-induced stylistic normalization and perplexity-based modeling of reader surprise. By synthesizing theoretical debates, methodological innovations, and emerging research directions, this paper demonstrates how computational approaches are reshaping literary scholarship while underscoring the necessity of maintaining critical rigor in quantitative text analysis.

Keywords: *computational stylistics, digital humanities, distant reading, authorship attribution, Burrows' Delta, scalable reading, natural language processing, cognitive stylometry, corpus linguistics, literary analysis*

1. Introduction

The intersection of linguistics, literary studies, computer science, and data visualization has birthed computational stylistics, a rapidly expanding discipline within computational literary studies and the broader digital humanities. Computational stylistics is systematically concerned with modeling the multi-dimensional, complex phenomenon of literary style using statistical, quantitative, and computational methods (Burrows, 2002). By operating on large digital corpora and applying transparent, repeatable workflows, this empirical field provides contemporary literary scholarship with new scales of observation, allowing critics to test long-standing aesthetic theories and formulate novel interpretative frameworks (Hoover et al., 2014).

Historically, stylistics emerged as a formal discipline at the beginning of the twentieth century, acting as an intellectual bridge between linguistics and literary criticism. The systematic study of style, however, extends back to the literary scholarship of the Greeks and Romans in the fifth century B.C., where rhetoric reigned as the dominant art. In the third book of his *Rhetoric*, Aristotle famously expressed suspicion of style, conceptualizing it as an accretion or an ornate enlargement added to elemental, basic textual material (Simpson, 2004).

Over the centuries, the definition of style has continuously evolved, shifting between views of collective linguistic norms and individual expression. Ben Jonson and his contemporaries echoed the notion that style is a highly individual quality, deeply rooted in the psychology of a specific person or sub-group rather than

the wider speech community (Bloch, 1953). This was famously encapsulated by Bernard le Bovier de Buffon's aphorism, *"Le style c'est l'homme même"* (Style is the man himself), suggesting that a writer's output is an intimate reflection of cognitive precision and hard labor. Jonathan Swift offered a more minimalist, functionalist perspective, defining style simply as *"proper words in proper places,"* while Matthew Arnold argued that the secret of style lay in having something to say and saying it with maximum clarity. In modern linguistics, Bernard Bloch defined style as the specific quality that distinguishes an individual's use of language from its general usage within a speech community (Aristotle, trans. 2007).

As the discipline matured, stylistics fractured into multiple specialized branches, including cognitive, corpus, critical, feminist, formalist, functionalist, historical, multimodal, pedagogical, and pragmatic stylistics (Jeffries & McIntyre, 2010). Roger Fowler particularly advanced the field by emphasizing the ideological function of style, demonstrating that linguistic choices are never neutral but are inherently bound to socio-cultural power dynamics. Similarly, Alan Batson argued that literary works present both a grammatical word order and a stylistic word order, with the latter serving as the proper object of literary criticism by utilizing grammatical rules to achieve heightened poetic expression (Stockwell, 2019).

2. Theoretical Frameworks and the Scale of Textual Exploration

The introduction of computational methodologies has generated significant debate regarding the nature of reading and text interpretation. Traditional literary scholarship relies heavily on hermeneutics the interpretive reconstruction of textual meaning through the close reading of individual canonical works. Close reading prioritizes the microscopic analysis of semantic ambiguity, historical context, and formal aesthetic structures (Fowler, 1996).

In contrast, Franco Moretti introduced the paradigm of "distant reading," arguing that the sheer volume of published literature the "great unread" requires scholars to step back from individual texts to analyze macroscopic patterns, structural trends, and genre developments over centuries. Under this framework, individual texts are treated as comparable data units, enabling the mapping of literary history as an interconnected system. However, this systemic perspective inevitably sacrifices the granular nuance of close textual analysis (Leech & Short, 2007).

To reconcile these opposing scales, Martin Mueller conceptualized "scalable reading," a mixed-methods approach that integrates qualitative hermeneutics with quantitative analysis. Scalable reading utilizes the metaphor of "zooming" to navigate fluidly between macro-level computational discoveries and micro-level close readings. In a typical scalable workflow, distant reading of a large corpus identifies anomalous patterns, statistical shifts, or structural outliers, which then serve as targeted entry points for human interpretation and close reading (Wales, 2014).

This dialectical movement has been refined by scholars like Katherine Bode, who cautions against treating computational data as a direct, unmediated representation of the historical record. Bode argues that both close and distant reading can exhibit an ahistorical bias if they ignore textual scholarship and bibliography. To counter this, contemporary scalable reading models emphasize that data curation, document metadata, and publication histories must be critically integrated into the computational workflow (Leech & Short, 2007).

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

Figure 1: The Epistemological Lifecycle of Traditional Scholarly Criticism and the Analog Data Matrix Hierarchy.

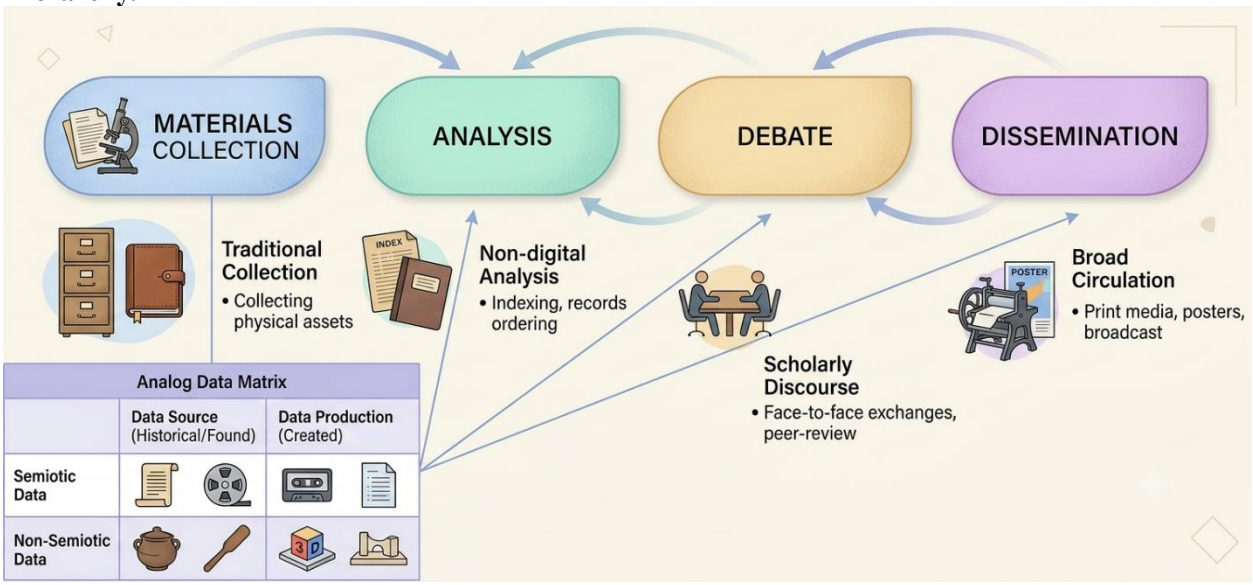


Table 1: Methodological Taxonomy of Reading Scales in Digital Humanities

Reading Paradigm	Scale of Observation	Primary Objective	Theoretical Grounding	Typical Toolkits and Techniques
Close Reading	Micro-scale (isolated passages, single texts, lines of verse)	Hermeneutic interpretation, exposing semantic ambiguity, historical nuances, and aesthetic forms	Traditional literary criticism, New Criticism, classical philology	Manual qualitative annotation, concordance tracking, direct textual engagement
Distant Reading	Macro-scale (large-scale corpora, global literary systems, multi-century trends)	Detecting historical trends, structural patterns, and systemic shifts across hundreds of texts	Literary sociology, systems theory, evolutionary model of genres, macroanalysis	Token frequency modeling, topic modeling, network analysis, metadata mining
Scalable Reading	Multi-scale (dynamic zooming between macro-phenomena and micro-contexts)	Synthesizing quantitative patterns with contextual interpretation to build verifiable models	Mixed-methods, iterative data-driven hermeneutics, digitally assisted text analysis	Integrated visualization platforms (e.g., Voyant, Palladio, Gephi), document clustering, text viewers

This scalable workflow is illustrated in Andrew Piper's monograph *Enumerations* (2018), which outlines a three-step cycle. First, based on established reading experience, a scholar formulates a qualitative presumption or belief. Second, the scholar tests this presumption across a broad corpus using distant reading measurements. Third, the resulting data points especially statistical anomalies are subjected to close reading to interpret, verify, or falsify the original belief (Band, 2020).

For example, a scholar might investigate the relationship between Augustine's late-antique correspondence and his quoting behavior, hypothesizing that Augustine quoted the Bible more frequently or explicitly to female correspondents under the assumption that they required simpler, highly structured arguments. However, macro-scale measurement of every citation across his letters reveals that while Augustine did indeed quote the Bible more frequently when writing to women, close reading of the surrounding epistolary context refutes the hypothesis of intellectual condescension (Jockers, 2013).

Instead, late-antique cultural codes of modesty dictated that pious women speak and write primarily through biblical allusion. Augustine was not simplifying his writing, but rather adopting the stylistic and linguistic register of his addressees. This case study demonstrates how empirical measurement and close textual analysis work in tandem to refine historical arguments (Eder et al., 2016).

3. Methodological Foundations and the Mathematical Modeling of Style

The quantitative modeling of style rests on the premise that an author's unique voice or "stylome" is encoded in the distribution of high-frequency words, specifically function words such as articles, prepositions, and pronouns. Because these words are topic-robust and largely escape conscious authorial control, their relative frequencies remain stable across an author's oeuvre while varying significantly between different authors (Holmes, 1998).

In his foundational work, John Burrows identified three distinct regions in the word frequency lists of literary corpora. The first region consists of the most frequent words, which are analyzed using distance metrics like Delta. At the opposite end of the frequency spectrum, the lowest-frequency words (hapax legomena and dis legomena) are analyzed using Iota. The intermediate region which Burrows described as the "large area between the extremes of ubiquity and rarity" is targeted by metrics like Zeta, which track word vocabulary choices that distinguish specific genres, genders, or authors (Craig & Kinney, 2009).

The most influential metric in contemporary stylometry remains Burrows' Delta (Δ_B), which measures the stylistic distance between texts by comparing standard-deviation-normalized mean word frequencies (Burrows, 2006). To calculate Delta, each text in a comparison corpus is represented as a feature vector of the relative frequencies of the n_w most frequent words (MFWs). For each word w_i, a standard z-score is computed to scale the feature's variance:

$$z_i(D) = \frac{f_i(D) - \mu_i}{\sigma_i}$$

where f_i(D) is the relative frequency of word w_i in document D, and μ_i and σ_i are the mean and standard deviation of w_i across the comparison corpus. This transformation centers each feature to a mean of zero and a standard deviation of one (Kestemont, 2014). The classical Burrows' Delta (Δ_B) is formulated as the Manhattan (L_1) distance between the resulting z-score vectors:

$$\Delta_B(D, D') = \frac{1}{n_w} \sum_{i=1}^{n_w} | z_i(D) - z_i(D') |$$

As a ranking and classification metric, the constant fraction $\frac{1}{n_w}$ can be omitted. This allows the formula to be simplified algebraically as:

$$\Delta_B^{(n)}(D, D') = \sum_{i=1}^{n} \frac{1}{\sigma_i} | f_i(D) - f_i(D') |$$

To address the limitations of Manhattan distance, which treats all features as orthogonal and equally weighted, several modifications have been proposed. Shlomo Argamon introduced Quadratic Delta (Δ_Q), which corresponds to the squared Euclidean (L_2) distance (Argamon, 2008):

$$\Delta_Q(D, D') = \sum_{i=1}^{n_w} (z_i(D) - z_i(D'))^2$$

A more significant improvement is Cosine Delta (Δ_Z), proposed by Peter Smith and W. Aldridge, which calculates the angular distance between the standardized vectors. Cosine Delta measures the direction of stylistic deviation rather than magnitude, which reduces sensitivity to document length and prevents a few highly variable words from dominating the distance calculation (Smith & Aldridge, 2011). Mathematically, it is derived from the cosine similarity of the z-score vectors $\mathbf{x} = \mathbf{z}(D)$ and $\mathbf{y} = \mathbf{z}(D')$:

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

$$\cos\alpha = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} = \frac{\sum_{i=1}^{n_w} z_i(D) z_i(D')}{\sqrt{\sum_{i=1}^{n_w} z_i(D)^2} \cdot \sqrt{\sum_{i=1}^{n_w} z_i(D')^2}}$$

Table 2: Comparative Mathematical Formulas and Geometric Properties of Stylometric Delta Metrics

Metric Variant	Formula Type and Distance Metric	Mathematical Formulation	Geometric Shape of Iso-Delta Curves	Probabilistic Assumptions
Burrows' Delta (Δ_B)	Manhattan (L_1) distance over standardized z-scores	$\frac{1}{n_w}\sum_{i=1}^{n_w} \mid z_i(D) - z_i(D') \mid$ [cite: 22]	Multi-dimensional diamonds (highly elongated along axes if standard deviations vary)	Standard Laplace distribution over independent indicator word frequencies
Quadratic Delta (Δ_Q)	Squared Euclidean (L_2) distance over standardized z-scores	$\sum_{i=1}^{n_w}(z_i(D) - z_i(D'))^2$ [cite: 22]	Multi-dimensional spheres or axis-parallel ellipses	Multi-dimensional Gaussian (normal) distribution over variables
Cosine Delta (Δ_Z)	Angular Cosine distance of normalized vectors	$1 - \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}$ [cite: 22]	Unit hypersphere projection (measures angular trajectory independent of vector length)	Directional variance-stabilized distribution across dimensions

These geometric and probabilistic interpretations clarify why different distance measures provide distinct perspectives on literary texts. For example, Manhattan distance (\Delta_B) assumes a Laplace probability distribution, making it robust against minor stylistic variations. Conversely, Quadratic Delta (\Delta_Q) assumes a Gaussian distribution, meaning it heavily penalizes large coordinate differences and is highly sensitive to outliers (Evert et al., 2017).

The choice of metric is also closely tied to the morpho-syntactic properties of the target language. In their "Deeper Delta" investigations, Jan Rybicki and Maciej Eder demonstrated that while standard Delta procedures yield near-perfect results for English and German prose, their accuracy degrades when applied to highly inflected, agglutinative, or synthetic languages such as Polish, Latin, and Hungarian.

In highly inflected languages, stylistic information is distributed across a wider grammatical base rather than being concentrated in a few high-frequency function words. Consequently, achieving stable authorship attribution in these languages requires adjusting the MFW window and systematically omitting the absolute top of the frequency rank list (Jannidis et al., 2015).

Recent work has also extended Delta to translation studies. While early applications demonstrated that Delta usually identifies the original author of a text rather than its translator, recent studies have repurposed the metric to evaluate translation fidelity. By calculating an inverted, min-max normalized Burrows' Delta (invΔ_B) over 500 MFWs within segmented five-sentence windows (k = 5), researchers can measure the stylistic distance between machine-translated outputs and human reference translations. This approach provides a reference-free diagnostic tool that correlates with semantic-embedding baselines while capturing stylistic features that standard lexical-overlap metrics (such as BLEU or chrF2) often miss (Milanesi, 2024).

Finally, the discipline continues to develop new metrics that generalize classical distance measures by modeling word frequencies as probabilistic distributions. Notable among these are Rank-Turbulence Delta (which evaluates the turbulent displacement of word ranks across texts) and Jensen–Shannon Delta (which measures the divergence between two probability distributions). Empirical evaluations show that Rank-Turbulence Delta achieves classification accuracy comparable to Cosine Delta, while Jensen–Shannon Delta consistently matches or exceeds the performance of classical Burrows' Delta on historical corpora (Anwar et al., 2014).

4. Corpus-Driven Narratology and Structural Parsing

The integration of corpus linguistics into literary studies has enabled quantitative narrative theory, allowing researchers to test traditional structuralist concepts at scale across large collections of prose fiction. This corpus-driven narratology uses statistical modeling to track structural and stylistic changes over time (Jacobs, 2018).

A prominent example of this scale is the Corpus of British Narrative Fiction (CBNF), which comprises 150 novels written by authors born between 1800 and 1970. The corpus is balanced to represent subgenres such as domestic realism, Gothic fiction, detective novels, and social problem novels, with equal representation of male and female authors (75 each) (Fischer-Starcke, 2010). Statistical analysis of the CBNF has revealed clear diachronic shifts:

- **Narrative Perspective:** First-person narration increased from 14% of sampled texts in 1850 to 41% in 1999, indicating a long-term stylistic shift toward more personal, subjective narrative voices.
- **Lexical Diversity:** Lexical diversity decreased significantly in twentieth-century prose ($\beta = -0.18, p < 0.01$), reflecting a shift toward simpler, more colloquial vocabulary.
- **Pronoun Ratios:** Changes in pronoun ratios (e.g., first-person plural we versus singular I) correlate with shifts in narrative empathy and focalization, providing an empirical basis for tracking the development of character-focused narration.

In addition to diachronic trends, computational narratology has focused on structural parsing, particularly the automated detection and classification of literary events (Zeb et al., 2026).

Table 3: Classification Categories of Literary Events in Structural Narratology

Event Class Code	Event Class Name	Semantic Definition and Scope	Illustrative Literary Examples
CMS	Cognitive Mental State	Captures internal cognitive transitions, shifts in knowledge, or emotional states of characters	*He realized his mistake*; *She decided to leave the manor*
COM	Communication	Encompasses verbal, written, or symbolic dialogue and acts of transmission	*The detective whispered the clue*; *He wrote an urgent letter*
CON	Conflict	Represents physical, emotional, social, or structural clashes between agents	*They fought over the inheritance*; *The systemic strikes paralyzed the town*
GA	General Activity	Routine, non-conflictual, and non-communicative actions within the narrative world	*The clock struck midnight*; *She walked slowly down the corridor*
LE	Life Event	Major biological, relational, or socio-legal transitions of characters	*The protagonist was born*; *They were married in the village church*
MOV	Movement	Physical dislocation, spatial traversal, or geographical movement	*The carriage fled into the dark forest*; *He ran toward the harbor*
OTH	Others	Miscellaneous event mentions that do not fit into the primary six categories	*The candle flickered out*; *The rain stopped suddenly*

While large language models (LLMs) are highly effective at general-domain text processing, they often struggle with specialized tasks like narrative event extraction and coreference resolution in creative writing. In benchmark evaluations on the Vrittanta-EN corpus, zero-shot and prompt-based LLMs performed significantly worse than task-specific supervised baselines (Liu et al., 2019).

To bridge this gap, contemporary workflows combine dense contextual embeddings (such as those from RoBERTa) with a linguistically informed "Expert Index". This Expert Index integrates seven stylistic features tailored to narrative structure, allowing hybrid classification models (such as the Situation-Task-Action-

Consequence, or STAC, framework) to outperform pure LLM approaches, particularly in detecting complex narrative dynamics like CONFLICT (Kumar et al., 2024).

Managing the integrity of literary corpora also requires robust preprocessing tools, particularly when assembling large poetry datasets. To remove duplicate poems, researchers apply substring Levenshtein distance metrics to evaluate textual similarity. The similarity between two poems $A$and B(where $| A | \geq | B |$in character length) is calculated as:

$$\text{sim}(A, B) = 1 - \frac{\min(\text{lev}(a_1, B), \text{lev}(a_2, B), \ldots, \text{lev}(a_n, B))}{| B |}$$

where $\{a_1, a_2, \ldots, a_n\}$represents the set of all possible substrings of poem A, and $\text{lev}(a_x, B)$is the standard Levenshtein distance between substring $a_x$and poem B(Levenshtein, 1966).

In large-scale poetry databases, similarity scores typically exhibit a strongly bimodal distribution: a major peak between 0.4 and 0.5 corresponds to completely unrelated texts, while a minor peak near 1.0 identifies duplicates or minor text variants (Törnberg, 2024). Using a similarity threshold of 0.75, researchers can build undirected graphs where nodes represent poems and edges represent duplicate relationships, enabling the automated selection of primary variants and the removal of duplicate records. This structural parsing is supported by open-source digital infrastructure such as Poetree, a dedicated multilingual poetry database deposited at Zenodo, which provides APIs and specialized packages for computational poetics (Plecháč et al., 2024).

5. Critical Controversies and Epistemological Friction

The growth of computational stylistics has generated significant debate within literary studies, challenging scholars to critically examine the limits of quantitative modeling (Brown et al., 2020).

5.1. The Computational Case Against CLS

In 2019, Nan Z. Da published "The Computational Case against Computational Literary Studies" in *Critical Inquiry*, sparking widespread debate across the humanities. Da argued that computational literary analysis often suffers from a fundamental mismatch between its statistical tools and the complex nature of literary data. Her central critique was that "what is robust is obvious (in the empirical sense) and what is not obvious is not robust" (Haider & Kuhn, 2018).

According to Da, quantitative studies in CLS frequently use complex algorithms to confirm trivial literary facts, or rely on fragile statistical setups that fail under replication. She specifically targeted the use of term frequency-inverse document frequency (TF-IDF) scaling, Latent Dirichlet Allocation (topic modeling), and network analysis, claiming that researchers often "metaphorize" coding and statistics without adhering to mathematical rigor. Da also criticized the subfield's "modesty" or "incrementality," where methodological failures are reframed as opportunities to adjust models rather than acknowledging the limitations of the quantitative approach (Da et al., 2019).

The digital humanities community responded with detailed counter-arguments. Prominent scholars like Ted Underwood and Andrew Piper argued that Da's critique was built on a disciplinary category error, as she evaluated CLS strictly through the lens of Confirmatory Data Analysis (CDA) using statistics solely to confirm or reject pre-formulated hypotheses (Herrmann et al., 2023.

In contrast, contemporary computational stylistics primarily operates under the paradigm of Exploratory Data Analysis (EDA). Within EDA, visualizations, clustering algorithms, and dimensionality reductions are not designed to prove a single "right" or "wrong" interpretive thesis. Instead, they function as heuristic mechanisms that reshape textual structures to reveal new relationships and anomalies, directing human readers back to the text. Statistical modeling is thus defended not as an objective replacement for interpretation, but as a subjective, iterative, and transparent critical practice (Zhou, 2025).

6. NLP Pipelines and Computational Character Modeling

To analyze the structural complexity of narratives, contemporary computational stylistics relies on specialized natural language processing (NLP) pipelines. Standard NLP tools, which are typically trained on modern journalistic or technical prose, often fail when applied to literary texts due to historical vocabulary shifts, complex syntactic structures, and creative linguistic deviations (Babenko & Athavale, 2024).

To address these domain-specific challenges, David Bamman and his team developed *BookNLP*, a specialized pipeline trained on *LitBank* a dataset of 19th- and early 20th-century English fiction. *BookNLP* integrates several sequential modules to perform structural analysis at scale:

- **Named Entity Recognition (NER):** Detects character names, locations, and institutions, adapting to non-standard proper nouns and historical expressions.

- **Coreference Resolution:** Groups different textual mentions of a character (e.g., "Mr. Darcy," "Fitzwilliam," "he," "himself") into a single entity chain. This task is challenging in long-form narratives, as standard sliding-window neural networks tend to break coreference chains when a character does not appear for several chapters (Paulson & Leonard, 2025).

- **Quotation Attribution:** Automatically identifies the speaker of each direct quotation, linking dialogues directly to the correct character entity.

- **Character Attribute Extraction:** Associates characters with their physical actions (verbs), spoken words, and descriptive adjectives (modifiers) (Cortal, 2026).

Once these features are extracted, they can be utilized to construct multi-dimensional computational representations of literary characters. The structural and semantic validity of these representations has been evaluated using benchmarks such as *AustenAlike*. *AustenAlike* tests how well NLP pipelines capture character similarities across three distinct dimensions:

1. **Narrative Role Similarity:** Groups characters who fill similar structural functions in Jane Austen's plots (e.g., siblings of the protagonist).

2. **Social Similarity:** Groups characters based on demographic and sociological variables (e.g., gender, age, economic status, or profession).

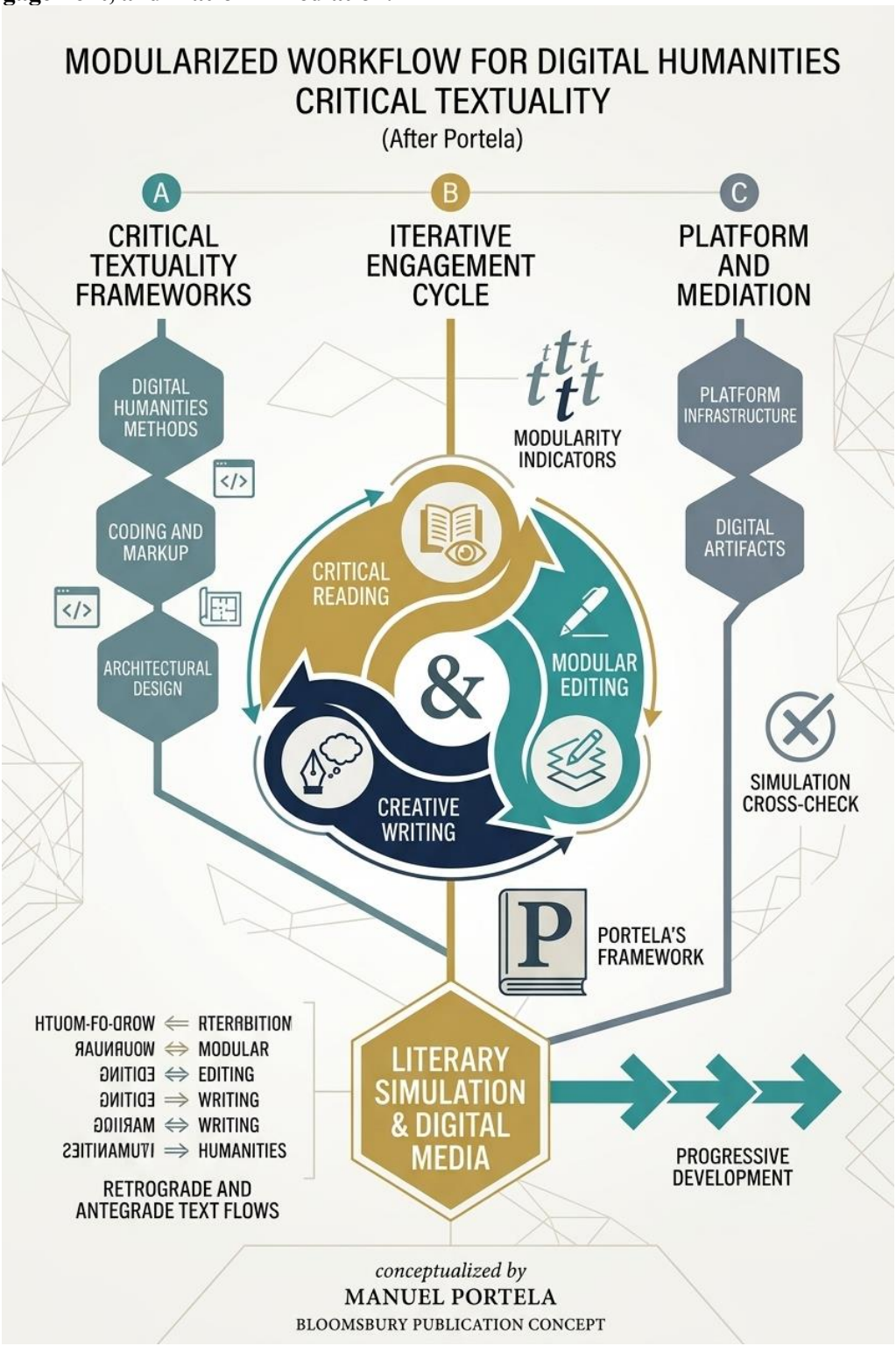
3. **Expert Scholarly Similarity:** Evaluates characters using comparisons extracted from decades of peer-reviewed scholarship in *Persuasions*, a journal dedicated to Austen studies (Graham & Graham, 2025).

The results of these evaluations reveal a persistent "expert gap" in computational character modeling. While features such as event-based actions and descriptive modifiers can capture broad demographic and narrative roles, they exhibit only moderate correlation with the more nuanced psychological and thematic alignments identified by expert human readers. Even advanced generative models like GPT-4 struggle to replicate these expert comparisons, placing the expert-identified most similar character in a top-ten list only about 50% of the time (Lane & Dyshel, 2025).

Additionally, the computational cost of tracking characters over long contexts remains high. Automatic coreference systems frequently break long-distance chains due to the use of fixed-size sliding windows, which fail to link character mentions across hundreds of thousands of tokens. To address this scale, researchers have introduced *BOOKCOREF*, a book-scale coreference benchmark containing fully annotated narrative texts with an average length exceeding 200,000 tokens (Boukhaled, 2026).

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

Figure Title: Figure 2: Modularized Workflow Architecture for Digital Humanities Critical Textuality, Iterative Engagement, and Platform Mediation.



7. The Frontier of Generative AI and Cognitive Stylometry

The rapid development of Large Language Models (LLMs) has introduced a paradigm shift in computational stylistics, moving the field from descriptive token-counting to cognitive and collaborative modeling (He et al., 2026).

7.1. LLM-Induced Stylistic Normalization

As LLMs are increasingly used for writing, editing, and stylistic translation, stylometrists have begun analyzing the specific linguistic changes these models introduce during text revision. Recent empirical studies have evaluated how frontier generative models (including GPT, Claude, and Gemini) alter the stylistic properties of personal narratives. Across different models and prompt conditions ranging from generic "improvement" prompts to explicit "voice-preserving" instructions LLM rewriting produces a consistent pattern of stylistic normalization (Klüwer, 2025).

When asked to revise a text, these models systematically suppress highly individual, informal, and vernacular markers. Specifically, first-person pronouns, contractions, and direct emotional words decrease, while vocabulary richness (type-token ratio), mean word length, and punctuation complexity increase.

This "directional convergence" flattens the linguistic variance of texts in feature space, rendering the rewritten texts stylistically uniform and making them significantly harder to match back to their original human sources. This normalization has profound implications for digital humanities, as it suggests that the widespread integration of generative AI text-processing tools could erode the linguistic diversity of contemporary digital archives and complicate authorship attribution (Stockwell, 2019).

Table 4: Statistical Evidence of Stylistic Shift Under Frontier LLM Generic Improvement Prompts

Evaluation Model	Number of Significant Markers	Mean Absolute Effect Size (|d|)	Predominant Directional Trends Across 13 Markers
GPT-5.4	13 / 13 (100% after FDR correction)	0.60 (Medium effect range)	Decreases function words and contractions; increases vocabulary diversity and word length
Claude Sonnet 4.6	13 / 13 (100% after FDR correction)	1.29 (Large effect range)	Decreases first-person pronouns; increases punctuation elaboration and average word length
Gemini 3.1 Pro	13 / 13 (100% after FDR correction)	1.45 (Large effect range)	Shifts narrative from immediate, embedded perspective to formal, distanced abstraction

7.2. Cognitive Stylometry and the Modeling of Surprise

Rather than treating LLMs merely as text generators, the paradigm of "cognitive stylometry" utilizes transformer-based architectures to simulate historically situated readers. Under this framework, a language model is first pre-trained on a large baseline corpus to establish a normative linguistic standard. It is then fine-tuned on a highly specialized historical or ideological subcorpus (e.g., state propaganda or a specific literary school) to simulate how a reader becomes habituated to a particular linguistic style (Leech & Short, 2007).

By tracking the model's perplexity which measures its predictive uncertainty on a token-by-token basis scholars can model the psychological dynamics of reading. Perplexity is defined mathematically as the exponentiated cross-entropy of the model's predictions:

$$P(X) = \exp\left(-\frac{1}{N}\sum_{i=1}^{N} \log P(x_i \mid x_{<i})\right)$$

where x_i represents the i-th token and $x_{<i}$ represents the preceding context.

When the model encounters highly formulaic and repetitive language, its perplexity decreases, reflecting "cognitive overfitting" and the automated processing of familiar structures. Conversely, when the model is applied to experimental literary fiction, it registers sharp spikes in perplexity (Milanesi, 2024).

These high-perplexity moments identify occurrences of "non-anomalous surprise" such as creative metaphors, unconventional syntax, and sudden linguistic register shifts that characterize literary defamiliarization. Cognitive stylometry thus provides a quantitative method to trace how literary styles challenge reader expectations, formalizing the Russian Formalist concept of *ostranenie* (making strange) as a measurable cognitive signature (Eder et al., 2016).

Conclusion

The integration of computational methodologies into English literary studies represents a transformative development that fundamentally reconfigures how scholars conceptualize, analyze, and interpret literary style. This review has demonstrated that computational stylistics, far from being a reductive replacement for traditional hermeneutic practices, offers powerful complementary tools that expand the scales of observation available to literary critics while maintaining the centrality of human interpretation. The mathematical modeling of stylistic features particularly through Burrows' Delta and its variants has proven remarkably effective for authorship attribution and stylistic distance measurement, though the choice of metric requires careful consideration of linguistic properties and corpus characteristics. The paradigm of scalable reading emerges as a particularly productive framework, enabling scholars to navigate fluidly between macro-level detection of structural patterns and micro-level close reading of textual anomalies. This dialectical movement between computational discovery and human interpretation exemplifies the iterative, transparent, and critically reflexive practice that defines contemporary digital literary scholarship. Furthermore, corpus-driven narratology and computational character modeling have extended the reach of quantitative methods to structural analysis and character representation, though persistent challenges particularly in long-distance coreference resolution and capturing expert-level literary nuance indicate areas requiring continued methodological refinement. Critical controversies surrounding computational literary studies, particularly those articulated by Nan Z. Da, have productively challenged practitioners to examine the epistemological foundations of their work. The defense of computational methods as exploratory rather than confirmatory tools clarify their proper role within literary scholarship: not as objective arbiters of truth but as heuristic mechanisms that reshape textual structures to reveal new relationships and generate interpretative insights. This epistemological stance acknowledges that computational models are inherently subjective, shaped by the decisions of data curation, feature selection, and algorithmic design. The frontier of generative AI and cognitive stylometry introduces both opportunities and challenges for the field. The documented tendency of large language models toward stylistic normalization raises concerns about the erosion of linguistic diversity in digital archives, while cognitive stylometry's modeling of reader surprise through perplexity measurements offers novel methods for quantifying literary defamiliarization. These developments suggest that the future of computational stylistics lies not in replacing human readers but in developing collaborative frameworks where computational models serve as historically situated reader-simulators that enhance our understanding of literary experience. Ultimately, this review affirms that computational stylistics, when practiced with critical awareness of its methodological limitations and epistemological assumptions, enriches rather than diminishes literary scholarship. By providing new scales of observation, testing long-standing aesthetic theories, and formulating novel interpretative frameworks, computational approaches contribute to a more comprehensive understanding of literary style as a multi-dimensional phenomenon. The continued dialogue between computational and traditional methods promises to generate insights that neither approach could achieve independently, ensuring that literary studies remain a vibrant, self-reflexive, and methodologically diverse discipline equipped to engage with the complexities of textual meaning in the digital age.

References

Anwar, H., Brown, B. J., Campbell, E. T., & Browne, D. E. (2014). Fast decoders for qudit topological codes. *New Journal of Physics*, *16*(6), 063038.

Argamon, S. (2008). Interpreting Burrows's Delta: Geometric and probabilistic foundations. *Literary and Linguistic Computing, 23*(2), 131–147.

Aristotle. (2007). *On rhetoric: A theory of civic discourse* (2nd ed., G. A. Kennedy, Trans.). Oxford University Press. (Original work published ca. 4th century BCE)

Babenko, I., & Athavale, V. A. (2024, September). Methods of literary text analysis with the help of artificial intelligence. In *International Conference on Distance Education Technologies* (pp. 81-95). Cham: Springer Nature Switzerland.

Band, J. (2020). Long comment regarding a proposed exemption under 17 USC 1201.Bloch, B. (1953). Linguistic structure and linguistic analysis. *Word, 9*(1), 40–44.

Boukhaled, M. A. (2016). *On computational stylistics: mining literary texts for the extraction of characterizing stylistic patterns* (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language models are few-shot learners*. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Burrows, J. F. (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

Burrows, J. F. (2006). All the way through: Testing for authorship in different frequency strata. *Literary and Linguistic Computing, 22*(1), 27–47.

Cortal, G. (2026). *Natural language processing for subjectivity analysis in personal narratives* (Doctoral dissertation, Université Paris-Saclay).

Craig, H., & Kinney, A. F. (Eds.). (2009). *Shakespeare, computers, and the mystery of authorship*. Cambridge University Press.

Da, N. Z. (2019). The computational case against computational literary studies. *Critical inquiry*, *45*(3), 601-639.

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *The R Journal, 8*(1), 107–121.

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *The R Journal, 8*(1), 107–121.

Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., & Vitt, T. (2017). Understanding and explaining Delta measures for authorship attribution. *Digital Scholarship in the Humanities*, *32*(suppl_2), ii4-ii16.

Fischer-Starcke, B. (2010). Corpus linguistics in literary analysis.

Fowler, R. (1996). *Linguistic criticism* (2nd ed.). Oxford University Press.

Graham, O., & Graham, O. (2025). Leveraging Natural Language Processing for the Computational Generation of Creative Writing.

Haider, T., & Kuhn, J. (2018). Supervised rhyme detection with Siamese recurrent networks. *Proceedings of the Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 81–86.

Plecháč, P., Cinková, S., Kolár, R., Šeļa, A., De Sisto, M., Nugues, L., ... & Kočnik, N. (2024). Poetree: Poetry treebanks in czech, english, french, german, hungarian, italian, portuguese, russian, slovenian and spanish. *Research Data Journal for the Humanities and Social Sciences*, *9*(1), 1-17.

He, G., Wang, C., & Yang, Y. (2026). Beyond Surface Complexity: Measuring Contextual Coherence in Academic Writing via Conditional Perplexity.

Herrmann, J. B., Bories, A. S., Frontini, F., Jacquot, C., Pielström, S., Rebora, S., ... & Sinclair, S. (2023). Tool criticism in practice. On methods, tools and aims of computational literary studies. *DHQ: Digital Humanities Quarterly*, *17*(3).

Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing, 13*(3), 111–117.

Hoover, D. L., Culpeper, J., & O'Halloran, K. (Eds.). (2014). *Digital literary studies: Corpus approaches to poetry, prose, and drama*. Routledge.

Jacobs, A. M. (2018). The gutenberg english poetry corpus: exemplary quantitative narrative analyses. *Frontiers in Digital Humanities*, *5*, 5.

Jannidis, F., Pielström, S., Schöch, C., & Vitt, T. (2015, June). Improving Burrows' Delta. An empirical evaluation of text distance measures. In *Digital Humanities Conference* (Vol. 11, p. 10).

Jeffries, L., & McIntyre, D. (2010). *Stylistics*. Cambridge University Press.

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Kestemont, M. (2014). Function words in authorship attribution: From black magic to theory? *Proceedings of the 3rd Workshop on Computational Linguistics for Literature*, 59–66.

Klüwer, N. (2025). *Context over Categories-Implementing the Theory of Constructed Emotion* (Doctoral dissertation, Technische Universität Wien).Kumar, A., Singh, A., Sharma, R., & Sharma, D. M. (2024). *Narrative event extraction with hybrid expert-guided language models: The Vrittanta-EN corpus*. Proceedings of the Language Resources and Evaluation Conference.

Lane, H., & Dyshel, M. (2025). *Natural language processing in action*. Simon and Schuster.

Leech, G. N., & Short, M. (2007). *Style in fiction: A linguistic introduction to English fictional prose* (2nd ed.). Pearson Longman.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*(8), 707–710.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://arxiv.org/abs/1907.11692

Milanesi, S. (2024). Mathematical Modelling in Biosciences and Discrete Optimization.

Paulson, D., & Leonard, S. (2025). Applications of NLP in Computational Poetics and Literary Analysis.

Simpson, P. (2004). *Stylistics: A resource book for students*. Routledge.

Smith, P. W. H., & Aldridge, W. (2011). Improving authorship attribution: Optimizing Burrows' Delta method. *Journal of Quantitative Linguistics, 18*(1), 63–88.

Stockwell, P. (2019). *Cognitive poetics: An introduction*. routledge.

Törnberg, P. (2024). Large language models for computational social science: Opportunities and challenges. *Computational Communication Research, 6*(1), 1–30.

Wales, K. (2014). *A dictionary of stylistics* (3rd ed.). Routledge.

Zeb, K., Alam, P., & Khattak, A. (2026). Algorithmic Narrative Structures: How Large Language Models Mimic and Disrupt Classical Structuralist Literary Syntax. *Liberal Journal of Language & Literature Review*, *4*(2), 2088-2114.

Zhou, S. J. (2025). *Decoding the Multiverse: New Methods for Strengthening the Robustness of Scientific Inference* (Master's thesis, The University of North Carolina at Charlotte).