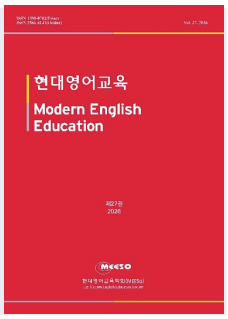
Modern English Education, Vol. 27, (2026) pp. 213-232
doi: https://doi.org/10.18095/meeso.2026.27.1.213

Modern English Education

ISSN: 1598-0782 / e-ISSN: 2586-6141
http://journal.meeso.or.kr



Computational Literary Analysis of Charlotte Brontë's *Jane Eyre*: A Natural Language Processing Approach

Young Kyo Oh (Chonnam National University)

Received: 3 March 2026
Revised: 29 March 2026
Accepted: 13 April 2026

Oh, Young Kyo. (2026). Computational literary analysis of Charlotte Brontë's *Jane Eyre*: A Natural Language Processing approach. *Modern English Education*, *27*, 213-232.

Keywords
Natural Language Processing,
Jane Eyre,
network text analysis,
topic modeling,
sentiment analysis
자연어처리,
제인에어,
네트워크 텍스트 분석,
토픽모델링,
감성분석

Young Kyo Oh
Lecturer
Department of Education
Chonnam National University
5young-kyo@hanmail.net

Abstract

This study utilizes Natural Language Processing (NLP) to analyze Charlotte Brontë's *Jane Eyre*, exploring how computational text analysis can make Victorian fiction more accessible to non-specialist readers. By employing an integrated framework that combines network text analysis, Latent Dirichlet Allocation (LDA) topic modeling, and sentiment analysis, the research offers a multi-dimensional view of the novel's narrative structure. A detailed examination using TF-IDF and network centrality reveals that the discourse in *Jane Eyre* centers around Jane's narrated consciousness, particularly through verbs of perception, cognition, and speech that shape how events are perceived, evaluated, and socially negotiated. The LDA topic modeling identifies three key thematic clusters: cognitive-relational dynamics, domestic enclosure as a prerequisite for ethical action, and class-conscious interactions. Furthermore, sentiment analysis uncovers a predominantly neutral-to-negative emotional baseline, interrupted by positive peaks in romantic passages and specific spikes of fear during Gothic sequences. Collectively, these computational insights shed light on Jane's developmental journey through constraint, crisis, and self-determined resolution. The study concludes that computational evidence effectively enhances close reading by providing clear, replicable indicators of narrative structure, thus serving as valuable instructional tools for English language learners and enriching Digital Humanities pedagogy.

INTRODUCTION

The digitization of literary texts has transformed scholarly approaches to literary study. Computational methods now allow researchers to analyze large textual datasets using statistical models to identify patterns that traditional reading methods may overlook (Mahadevkar et al., 2024). In English literary studies, these developments have raised questions about how best to analyze and teach classic texts, leading to new interdisciplinary work that combines literary criticism with computational linguistics and artificial intelligence (Khan et al., 2025). Early digital humanities research concentrated on corpus linguistics and stylometric authorship attribution. Since the early 2000s, however, the disciplinary scope has expanded to encompass Digital Literary Studies, focusing on narrative structure, genre taxonomy, and literary historiography (Jockers, 2013;

Ramsay, 2011). Contemporary scholarship increasingly advocates synthesizing "close reading"—the hermeneutic foundation of traditional literary criticism—with computational "distant reading," proposing complementary rather than oppositional frameworks (Salmi, 2024; Sekar, 2024).

Recently, the integration of computational methods into literature teaching has emerged as a significant strand within English language education research. Empirical studies show that corpus-based activities—including concordance analysis, keyword extraction, and collocation study—support vocabulary development, enhance stylistic awareness, and promote learner autonomy by enabling students to discover patterns in authentic texts (Lusta et al., 2023; Rodriguez & Csomay, 2024). Sharma (2025) surveys machine learning tools—including sentiment analysis, topic modeling, and character network visualization—and argues that these methods can uncover patterns that enhance student engagement while acknowledging methodological limitations in capturing literary nuance. Frei (2025) provides practical LLM-based prompts for didacticizing literary texts through summarization, simplification, and scenario generation, but warns of risks including factual inaccuracies and cultural insensitivity in automated outputs. In the field of English education, AI-based digital research such as text mining, text-network analysis, and topic modeling is also being undertaken (T. Y. Kim et al., 2023; Maeng et al., 2023). This emerging literature suggests that computational tools offer valuable pedagogical affordances when mediated by informed teaching practice but cannot replace interpretive depth and critical judgment (Frei, 2025; Sharma, 2025).

Text analysis, grounded in NLP technologies, facilitates systematic examination of large-scale textual datasets. These methodologies focus on keyword extraction and relational mapping among lexical units, enabling researchers to uncover novel relationships and knowledge structures embedded within textual data (Mahadevkar et al., 2024). By correlating extracted keywords with diverse informational sources, computational approaches reveal patterns that remain invisible to traditional close reading methods. This study applies three principal NLP frameworks—network text analysis, topic modeling, and sentiment analysis—to Charlotte Brontë's *Jane Eyre* (1847/1998). Text network analysis visualizes interpersonal relationships as nodes and links, quantifying connectivity through centrality measures (Amalvy et al., 2025). LDA topic modeling identifies thematic structures by extracting representative keywords (Jiang et al., 2025). Sentiment Analysis evaluates emotional polarity and affective intensity, quantifying narrator and character emotional attitudes (Yuri & Pascale, 2024).

Most computational literary studies analyze multiple texts to identify patterns across genres or historical periods (Jockers, 2013; Underwood, 2019). The present study takes a different approach, focusing on a single text—*Jane Eyre*—to demonstrate how computational methods can reveal patterns within individual novels. This "microscopic" approach (Y. K. Oh, 2024) complements large-scale distant reading by providing detailed analysis of specific texts. This approach may be particularly useful for teaching, as it can help students who find Victorian prose difficult to understand (Schabbach, 2024). By showing quantitative patterns in the text, computational analysis offers concrete evidence to support literary interpretation. The present study extends this pedagogical scholarship by applying an integrated NLP framework to *Jane Eyre*, demonstrating how computational patterns can be translated into instructional resources. By focusing on a single canonical text rather than large-scale corpus comparison, this microscopic approach models a pedagogically tractable workflow that teachers and students can replicate without specialized technical expertise.

Although previous studies have demonstrated the value of computational approaches in literary and educational research, relatively few have integrated keyword analysis, network text analysis, topic modeling, and sentiment analysis within a single pedagogically oriented reading of a Victorian novel. Addressing this gap, the present study examines how these complementary NLP techniques can illuminate the narrative structure, thematic organization, and emotional progression of *Jane Eyre* while also offering a replicable framework for literary pedagogy. To guide the analysis, the study addresses the following research questions: (1) What are the most important keywords in *Jane Eyre*, as measured by TF-IDF analysis? (2) How are these keywords related to one another, as measured by network centrality? (3) What are the main thematic patterns in the novel, as identified through LDA topic modeling? and (4) How do emotional patterns change across the novel's chapters, as tracked by sentiment analysis?

LITERATURE REVIEW

Computational Approaches to English Literature

The digitization of literary corpora has expanded computational literary studies from macro-scale pattern detection to theoretically informed interpretive experimentation. Foundational work established key rationales and paradigms:

macroanalysis demonstrated how topic modeling and network analysis can trace thematic evolution across large corpora (Jockers, 2013), algorithmic criticism argued that computational constraints may productively intensify interpretive play (Ramsay, 2011), and large-scale evidence-based literary history showed how digital methods reveal patterns of change that remain largely invisible to traditional close reading (Underwood, 2019).

More recent scholarship extends these commitments through microscopic, single-text analyses and methodologically plural integrations: NLP pipelines have been used to quantify character relations, themes, and affect in canonical nineteenth-century fiction (Y. K. Oh, 2024), computer-assisted approaches have operationalized focalization and reader address in *Jane Eyre* and *Villette* (Schabbach, 2024), distant reading has been applied to gender politics and body representation in Dickens's works (Chen & Xu, 2024), and action-based paradigms have been proposed to balance technical precision with humanistic interpretation in British literary history (Thomas, 2024).

Across these strands, contemporary debates increasingly emphasize that computational "distant reading" and traditional "close reading" are complementary rather than oppositional, with distant methods enabling cultural-historical amplification while close reading remains indispensable for interpretive depth (Salmi, 2024; Sekar, 2024).

Methodologically, network text analysis, topic modeling, and sentiment analysis form three core toolsets for modeling narrative structure, thematic organization, and affective dynamics. Network text analysis represents textual entities (e.g., characters or keywords) as nodes and their relations (e.g., co-occurrence or dialogue links) as edges, enabling quantitative tests of narrative structure through automated extraction and graph modeling (Amalvy et al., 2025; Elson et al., 2010; Stiller et al., 2003). Centrality metrics—degree, closeness, betweenness, and eigenvector—provide interpretable indicators of narrative prominence and structural mediation within such networks (Borgatti & Everett, 2006; Scott & Carrington, 2011).

Topic modeling, especially LDA, infers latent thematic structure by modeling documents as mixtures of topics and topics as probability distributions over words (Blei et al., 2003), with early literary applications highlighting both interpretive promise and challenges posed by figurative language and disciplinary discourse (Mimno, 2012; Rhody, 2012). Recent work further demonstrates LDA's cross-domain versatility and its applicability to nineteenth-century women's writing, including comparative topic patterns in novels by female authors (Jiang et al., 2025; Rha & Silver, 2021), while emphasizing that topic-number selection should balance statistical fit and human interpretability, typically via perplexity and coherence (Chang et al., 2009).

Sentiment analysis complements structural and thematic modeling by quantifying affective polarity and intensity in text. Traditional lexicon-based approaches rely on curated resources that assign positivity, negativity, and neutrality to lexical items (Baccianella et al., 2010), whereas applied studies show both the utility and limitations of sentiment modeling for literary language: computational affect classification can detect stylistic signatures, yet models may miss sentiment where human readers perceive strong affect, especially in low-arousal, highly concrete literary expression (Yuri & Pascale, 2024). Historical texts pose an additional challenge because emotional lexicons shift over time, making historically informed interpretive frameworks essential when tracking sentiment trajectories across nineteenth-century narrative arcs, including *Jane Eyre* (Gilbert & Gubar, 2020; H. Kim, 2022).

Linking Literary Text Analysis to English Language Education

Adolphs (2006) establishes a comprehensive pedagogical framework for integrating electronic text analysis into English language and literary education, demonstrating how computational techniques such as frequency analysis, concordancing, and collocation studies illuminate vocabulary patterns, discourse structures, and register variation within literary texts. By rendering implicit linguistic patterns explicit and observable, these corpus-based methods support language-focused learning and facilitate essential form-meaning connections for language acquisition, while grounding instruction in authentic language use embedded in meaningful literary contexts that complement traditional teaching approaches.

Extending this computational framework to advanced applications, Navarro-Colorado (2017) demonstrates the effectiveness of introducing NLP techniques to humanities students through a focused intensive course covering word frequencies, part-of-speech tagging, and topic modeling. The curriculum design not only achieved high student satisfaction and perceived utility for corpus analysis but also equipped learners with transferable analytical competencies, illustrating how NLP methods can be made accessible to non-technical students while fostering data literacy, enhancing understanding of linguistic pattern contributions to meaning construction, and building confidence in systematic textual analysis applicable to both literary interpretation and broader language learning objectives.

Jane Eyre: Feminist, Gothic, and Bildungsroman Scholarship

Jane Eyre has generated extensive scholarship across feminist, Gothic, and Bildungsroman frameworks. Gilbert and Gubar's (2020) *The Madwoman in the Attic* established the foundational feminist reading, interpreting the novel as a feminist Bildungsroman in which Jane's assertive gaze and moral autonomy challenge Victorian patriarchal norms. Gao (2013) likewise frames Jane's pursuit of equality and independence as a form of proto-feminist consciousness, tracing her gradual transformation across the novel's major settings. This Bildungsroman trajectory can be further theorized through genre debates in Korean scholarship: J. H. Oh (2012) argues that women's Bildungsroman in nineteenth-century Britain is driven not by smooth reconciliation but by an intensified conflict between self-realization and the socially restricted endpoint of integration (often marriage) under the ideology of separate spheres, and that the genre's dynamism emerges when novels expose contradictions and fissures within that closure. Consistent with this view, recent digital-humanities work has also operationalized *Jane Eyre*'s growth arc at the chapter level: using text mining on the novel's 38 chapters, Ban et al. (2022) reconstruct three dominant topics—love, independence, and change—and quantify plot development through chapter-wise positive/negative sentiment trajectories, suggesting that computational evidence can supplement interpretive accounts of Jane's movement from oppression toward conditional autonomy.

Recent scholarship continues to explore feminist dimensions in ways that resonate with both genre-theoretical and empirical perspectives. Thierauf (2023) examines *Jane Eyre* through the lens of "white feminism," showing how Victorian romance and feminist inheritance can clash with contemporary classroom politics, while Abdulla and Ahmed (2023) compare madness and feminism in *Jane Eyre* and *Girl, Interrupted*, foregrounding women's struggles across temporal contexts. Al-Dabool (2025) similarly emphasizes how Brontë's narrative challenges Victorian gender inequality and male domination.

Alongside these readings, psychoanalytic and identity-focused studies examine character formation; for instance, Anees et al. (2021) interpret patriarchal ideology as gender discrimination by mapping how major male characters function as embodiments of Victorian patriarchal systems that Jane's perspective contests. These approaches collectively underscore that feminist resistance in *Jane Eyre* is articulated not only through explicit claims to equality but also through the protagonist's negotiation of socially sanctioned forms of virtue, discipline, and self-authoring—precisely the kind of constrained "growth" that women's Bildungsroman scholarship highlights (J. H. Oh, 2012). Gothic scholarship further clarifies how narrative space and affect intensify this developmental tension. Tine (2024) analyzes Gothic experiences across *Jane Eyre*, *Wuthering Heights,* and *The Tenant of Wildfell Hall*, emphasizing how architectural enclosure and psychological terror register gendered constraints in nineteenth-century contexts. From a Bildungsroman perspective, such Gothic episodes can be read not as digressions but as pressure points that force ethical decisions and reorient self-understanding; correspondingly, Ban et al.'s (2022) chapter-level sentiment mapping provides a quantitative avenue for identifying narrative inflection points where crisis and transition co-occur with shifts in affective polarity. In a related vein, Bildungsroman scholarship traces Jane's maturation across five key settings, and recent work has refined spatial accounts of subject formation: Nie et al. (2024) apply Foucauldian spatial analysis to theorize "space escape," while Schiavone (2025) reads the novel as hagiographic romance that opens space for transgressive gender identity. Schabbach (2024) complements these accounts through computer-assisted focalization analysis, demonstrating how narrative strategies shape protagonist subjectivity across Brontë's fiction. Together, spatial-Gothic and computational approaches reinforce the view that Jane's "growth" is structurally uneven, marked by recurrent constraints and episodic breakthroughs rather than linear progress (Ban et al., 2022; J. H. Oh, 2012).

Postcolonial and class-based readings, finally, examine Jane's ambiguous social position. Van den Bossche (2005) argues that *Jane Eyre* produces rather than merely reproduces ideologies and identities, with Jane's agency negotiating class through personal virtue that can supersede class origin. Read through the lens of women's Bildungsroman, this negotiation highlights how "integration" is never purely personal but mediated by political-economic and ideological conditions that delimit what counts as legitimate female autonomy (J. H. Oh, 2012). In this respect, computational reconstructions of thematic emphasis—such as the prominence of "independence" and "change" as recurrent topic dimensions in chapter-level modeling—can be understood as an additional evidentiary layer for tracking how the novel repeatedly stages the contradiction between self-making and social containment across its narrative phases (Ban et al., 2022).

METHOD

Data Collection and Preprocessing

The primary analysis and modeling were conducted using the Python-based TEXTOM solution and Python coding. The specific research procedure is illustrated in Figure 1.

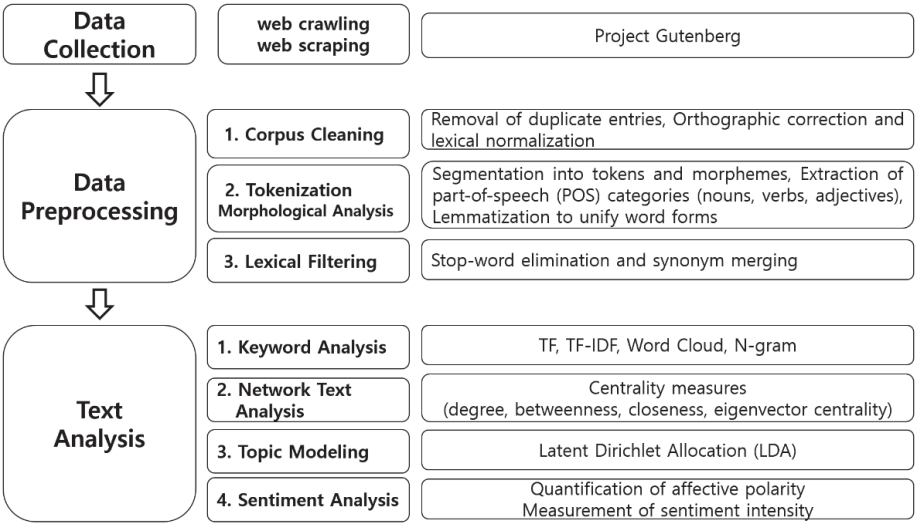


FIGURE 1
Research Procedure

Text analysis in the humanities is particularly active in fields where textual data has already been accumulated, such as linguistics and literature. Large-scale humanities texts have undergone digitization to enable the application of data analysis technologies (Chen & Chang, 2019). The Gutenberg Project (n.d.), a representative initiative for digitizing humanities texts, provides approximately 59,000 classic texts as e-books, including literary works such as Shakespeare's plays and Charlotte Brontë's *Jane Eyre*. Research attempts have been made to analyze these humanities text corpora through text analysis methods (Chen & Chang, 2019). Such text analysis is applied not only to large-scale humanities projects but also to individual literary studies. *Jane Eyre* serves as an ideal case study for demonstrating how integrated computational analysis can enhance the accessibility and pedagogical utility for English language learners. The novel's canonical status, narrative complexity, thematic richness, and curricular prevalence provide robust material for illustrating how NLP methods can complement traditional literary pedagogy while developing transferable analytical skills essential for academic English proficiency.

For the text analysis of Charlotte Brontë's *Jane Eyre*, the research was conducted in the following sequence: data collection → data preprocessing → data analysis. Python programming language and the text analysis solution TEXTOM were used as analytical tools. In the data analysis process, NLP techniques were applied in the following order to extract the core content of the novel *Jane Eyre*: Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) analysis, network centrality analysis, topic modeling, and sentiment analysis.

Indexing is used to structure natural language text data so that computers can analyze it. In this processing stage, keywords that reflect the core content of the text are extracted, and unnecessary words that perform simple grammatical functions are removed as stopwords. Subsequently, a morphological analyzer is applied to divide sentences by part of speech, and important elements such as nouns, verbs, and adjectives are extracted using libraries. Data preprocessing—the process of decomposing, refining, and filtering text—is the most fundamental and critical step in NLP.

The data preprocessing process consists of three stages: Stage 1, data cleaning; Stage 2, tokenization; Stage 3, word refinement. In this study, the data preprocessing process was conducted using a cleaning/morphological analyzer. First, normalization was performed by converting all uppercase letters to lowercase, and a common standard for characters was established to define generality in order to obtain meaningful analysis results. Morphological analysis was performed for tokenization and part-of-speech extraction. For this purpose, the CoreNLP analysis module—software developed to analyze morphemes—was utilized. Next, in the word refinement stage, the words extracted through data cleaning and morphological analysis often contained typographical errors, residual particles, stopwords, and duplicates. Next, a secondary refinement process was conducted to eliminate single-letter words, honorifics such as "miss" and "mr," and meaningless words.

Examples of stopword removal and duplicate integration are as follows: In the autobiographical novel *Jane Eyre*, the protagonist's first-person perspective "I" appears 7,226 times. To balance the frequency analysis, only objective third-person references were used in the analysis, and "I" was treated as a stopword.

TABLE 1
Examples of Stopword Removal and Duplicate Integration

Process	Examples
Stopwords	be, have, do, one, mr, miss, sir, mrs, t, st, c, e, d, j, m, n, s, ere, ma, i, il, je, l, y, ad, ail, p, ce, am
Duplicate Integration	jane→janeeyre, eyre→janeeyre, elliott→janeeyre

Data Analysis

The data analysis consisted of four complementary techniques applied sequentially to extract core content and semantic structures from the novel. First, keyword analysis using Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) was conducted to identify salient terms within the text. A Document-Term Matrix (DTM) was constructed to calculate term frequencies, and TF-IDF weighting was applied to prioritize terms that are highly frequent in the target document but rare across a broader corpus, thereby highlighting contextually significant keywords rather than generic high-frequency terms (Chen & Chang, 2019).

Second, network text analysis was performed to examine structural relationships among key terms. A co-occurrence matrix was generated for the top 50 terms, and centrality measures—including degree centrality (the number of direct connections a node has), closeness centrality (the proximity of a node to all others via shortest paths), betweenness centrality (the frequency with which a node lies on shortest paths between other nodes), and eigenvector centrality (the influence of a node based on the centrality of its neighbors)—were calculated to quantify the positional importance of terms within the semantic network (Scott & Carrington, 2011).

Third, LDA topic modeling was employed to uncover latent thematic structures within the text corpus. The LDA model assumes that each document is a mixture of latent topics, and each topic is a probability distribution over words (Blei et al., 2003). To determine the optimal number of topics (k), perplexity and coherence metrics were evaluated across a range of k values; lower perplexity indicates better predictive performance, while higher coherence reflects greater semantic similarity among words within topics (Chang et al., 2009). Hyperparameters α (document-topic density) and β (topic-word density) were set to 0.5 and 0.01, respectively, and the model was trained over 500 iterations. The top 30 words per topic were extracted, and results were visualized using the pyLDAvis library to assess topic distribution and word-topic associations.

Finally, sentiment analysis was implemented in Python using the Natural Language Toolkit (NLTK) for sentence segmentation and the Hugging Face Transformers library for emotion classification. Specifically, we applied a pretrained DistilRoBERTa-based emotion classifier (j-hartmann/emotion-english-distilroberta-base) via the pipeline API to obtain probabilistic scores for seven emotion categories (joy, sadness, fear, anger, surprise, disgust, and neutral) at the sentence level. Together, these four methods—TF-IDF keyword analysis, network centrality analysis, LDA topic modeling, and sentiment analysis—provided a multi-layered computational framework for literary text analysis, enabling triangulation of lexical, thematic, structural, and affective features.

FINDINGS

RQ1: Text Frequency Analysis of *Jane Eyre*

This study employed Term Frequency–Inverse Document Frequency (TF-IDF) to identify lexically salient terms in *Jane Eyre*. The *Jane Eyre* novel corpus was systematically segmented into 16,599 units through the automated processing features of TEXTOM. TF-IDF weights combine raw term frequency with inverse document frequency, thereby highlighting words that are both frequent and contextually distinctive to the target text. Previous studies identified words with high TF-IDF values, such as 'rochest', 'john', 'jane', 'miss', 'day', 'sir', and 'eye' (Ban et al., 2022).

As shown in Table 2, the term 'janeeyre' (TF-IDF = 4,463.37) exhibited the highest weight, followed by action and perception verbs: 'say' (2,480.31), 'see' (1,906.66), 'go' (1,841.18), 'look' (1,741.57), 'come' (1,640.89), 'think' (1,629.08), and 'know' (1,464.64). Character names including 'rochester' (1,389.79), 'reed', 'helen', and 'john' also ranked highly, alongside the spatial term 'room' (1,147.46) and evaluative words such as 'good' (1,489.09), 'heart', and 'love'.

These results indicate that *Jane Eyre* is linguistically centered on the protagonist's subjective experience, with dialogue ('say'), perception ('see', 'look'), cognition ('think', 'know'), and interpersonal relationships (character names) constituting the narrative's lexical core. The prominence of spatial and affective vocabulary further reflects the novel's thematic engagement with autonomy, enclosure, and moral introspection, corroborating feminist readings that emphasize Jane's subjective gaze as an instrument of agency (Gao, 2013; Gilbert & Gubar, 2020).

TABLE 2
Word Count and Term Frequency-Inverse Document Frequency (TF-IDF)

Rank	Term	TF-IDF	Rank	Term	TF-IDF	Rank	Term	TF-IDF
1	janeeyre	4,463.37	26	ask	972.92	51	hour	760.47
2	say	2,480.31	27	get	952.96	52	pass	757.45
3	see	1,906.66	28	find	914.38	53	more	755.33
4	go	1,841.18	29	face	913.79	54	life	753.85
5	look	1,741.57	30	sit	898.58	55	put	733.16
6	come	1,640.89	31	own	897.23	56	long	732.16
7	think	1,629.08	32	wish	896.22	57	want	731.55
8	good	1,489.09	33	door	893.85	58	way	725.88
9	know	1,464.64	34	love	892.92	59	don	724.26
10	take	1,405.51	35	john	888.44	60	place	688.09
11	rochester	1,389.79	36	two	876.88	61	head	681.65
12	make	1,326.42	37	house	853.30	62	return	676.92
13	little	1,282.36	38	speak	846.41	63	thing	670.44
14	day	1,211.91	39	word	836.04	64	open	665.68
15	eye	1,211.83	40	heart	815.15	65	like	661.57
16	feel	1,164.69	41	turn	814.14	66	keep	659.16
17	give	1,155.57	42	stand	803.62	67	live	656.40
18	room	1,147.46	43	answer	803.62	68	let	651.91
19	hear	1,136.23	44	rise	798.59	69	something	650.61
20	time	1,122.88	45	lady	796.08	70	light	647.82
21	seem	1,096.07	46	other	794.06	71	call	647.82
22	hand	1,088.83	47	man	790.03	72	fairfax	646.82
23	leave	1,084.60	48	such	783.52	73	child	646.06
24	tell	1,032.04	49	last	768.19	74	reed	643.28
25	night	981.24	50	nothing	768.19	75	voice	632.60

RQ2: Network Text Analysis of *Jane Eyre*

To examine the structural relationships among keywords in *Jane Eyre*, a co-occurrence network was constructed based on word proximity within a sliding window. Table 3 presents the overall network statistics, and Figure 2 visualizes the resulting network graph.

TABLE 3
Overview of the Jane Eyre Network

Network Measures	Value
Nodes	100
Total Edges	3,781
Diameter	2
Average Degree	75.62
Average Path Length	1.23616
Degree Centralization	9.27891
Closeness Centralization	0.37373
Betweenness Centralization	0
Connected Components	1
Graph Density	0.76383

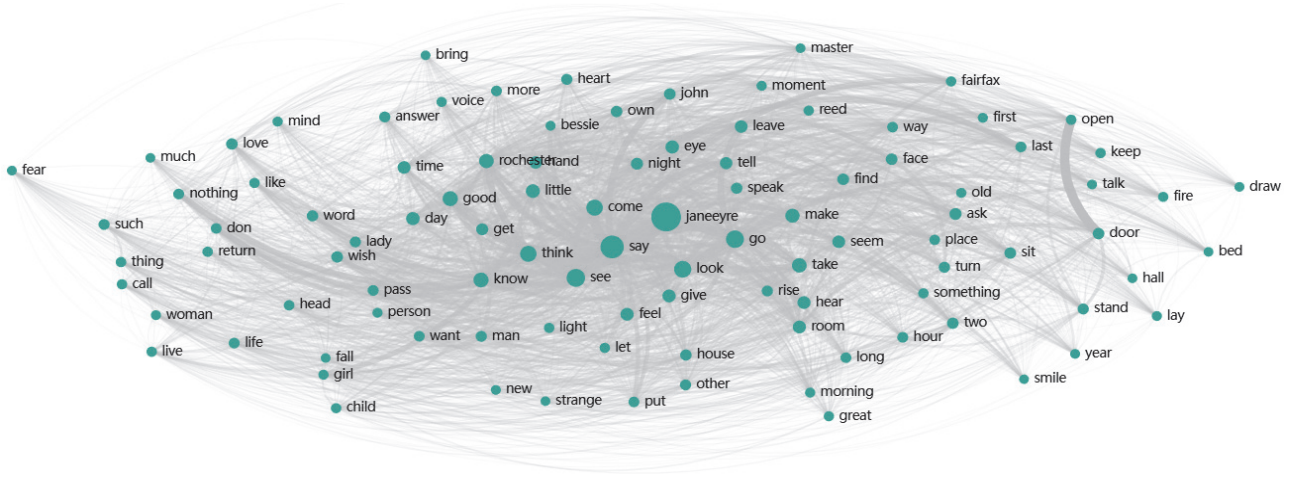


FIGURE 2
Jane Eyre Text Network Visualization

Figure 2 presents a visual representation of the co-occurrence network. Central nodes—including 'janeeyre', 'say', 'good', 'look', 'think', and character names such as 'rochester', 'john', and 'reed'—are positioned at the network's core, reflecting their high degree centrality and frequent co-occurrence with other terms. The dense clustering of edges around these central nodes illustrates the protagonist-centric narrative structure, in which Jane's actions, perceptions, and relationships dominate the lexical landscape.

Peripheral nodes include less frequent terms such as 'child', 'girl', 'woman', 'life', 'mind', and 'heart', which are nonetheless connected to the core, indicating their thematic relevance. The high overall density (0.764) and short average path length (1.236) suggest that the novel's vocabulary is semantically cohesive, with minimal isolated clusters or disconnected subgraphs.

This network structure aligns with the narrative's focus on *Jane Eyre* as the central focalizer, whose voice, actions, and relationships are the primary organizing principles of the text (Schabbach, 2024). The absence of betweenness centralization further indicates that no single term functions as an exclusive semantic bridge, reinforcing the interpretation that the novel's discourse is uniformly centered around the protagonist's subjective experience (Nie et al., 2024).

To examine the relational structure and the relative importance of terms within the text, six centrality measures were calculated for the *Jane Eyre* co-occurrence network. Table 4 presents the quantitative centrality analysis of the top 50 words, with each metric reflecting the influence and structural significance of terms from different perspectives within the novel's lexical network (Scott & Carrington, 2011).

The centrality pattern in Table 4 is significant not merely because it identifies frequent or well-connected words, but because it clarifies how the novel organizes experience. The prominence of 'say', 'see', 'look', 'think', and 'know' suggests that the narrative is structured through acts of perception, reflection, and verbal exchange rather than through externally detached description. In other words, events become meaningful in the novel as Jane observes them, interprets them, and recounts them. This lexical pattern is consistent with the first-person design of the text, in which narration is inseparable from consciousness. The network therefore supports the claim that *Jane Eyre* is not simply a sequence of plotted incidents, but a narrative system shaped by the heroine's ongoing cognitive and moral processing of social experience.

The same pattern also has implications for the novel's character system. The strong visibility of 'janeeyre' and 'rochester', together with high-centrality verbs of speech and perception, indicates that relationships in the novel are built less through stable social labels than through interaction, address, and interpretive exchange. Characters matter not simply as fixed roles in a Victorian social hierarchy, but as figures who are continuously evaluated through conversation, observation, and emotional negotiation. This is why the network repeatedly foregrounds lexical items associated with speaking, seeing, and thinking: they mark the processes through which intimacy, authority, resistance, and dependence are narratively produced. In this sense, the network structure quantitatively supports a reading of *Jane Eyre* as a Bildungsroman in which selfhood emerges through relational testing rather than through isolated inwardness alone.

Specifically, Table 4 presents six centrality measures for the top 50 terms in the *Jane Eyre* co-occurrence network: degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, PageRank, and clustering coefficient. Each metric captures a distinct dimension of structural importance within the network.

TABLE 4

Top 50 Network Centrality Measures for Jane Eyre

Term	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient
janeeyre	4.586	0.934	0.004	0.119	0.017	0.776
say	11.616	1.000	0.005	0.127	0.041	0.759
see	7.091	0.971	0.004	0.124	0.025	0.767
go	6.626	0.952	0.004	0.121	0.024	0.765
look	5.980	0.961	0.004	0.123	0.022	0.771
come	5.899	0.990	0.005	0.126	0.021	0.764
think	5.081	0.961	0.004	0.122	0.019	0.768
good	5.152	0.952	0.004	0.121	0.019	0.767
know	4.566	0.917	0.004	0.117	0.017	0.777
take	4.606	0.961	0.004	0.123	0.017	0.771
rochester	4.556	0.943	0.004	0.120	0.017	0.770
make	4.040	0.971	0.004	0.124	0.015	0.766
little	4.343	0.934	0.004	0.119	0.016	0.775
eye	3.192	0.876	0.003	0.110	0.012	0.776
day	3.202	0.900	0.003	0.114	0.012	0.781
feel	3.051	0.876	0.003	0.111	0.012	0.790
give	3.010	0.884	0.003	0.112	0.012	0.787
room	4.040	0.900	0.003	0.114	0.015	0.782
hear	3.727	0.925	0.004	0.118	0.014	0.773
time	3.061	0.868	0.003	0.109	0.012	0.787
seem	2.828	0.925	0.004	0.118	0.011	0.771
hand	3.323	0.917	0.004	0.117	0.013	0.775
leave	3.253	0.892	0.003	0.113	0.012	0.778
tell	3.253	0.900	0.003	0.115	0.012	0.785
night	3.172	0.853	0.003	0.107	0.012	0.781
ask	2.919	0.884	0.003	0.112	0.011	0.785
get	2.788	0.846	0.003	0.106	0.011	0.795
find	2.030	0.839	0.003	0.105	0.008	0.790
face	2.505	0.811	0.002	0.100	0.010	0.798
sit	2.808	0.853	0.003	0.107	0.011	0.783
own	2.485	0.846	0.003	0.106	0.010	0.799
wish	2.586	0.832	0.002	0.104	0.010	0.802
door	3.051	0.786	0.002	0.094	0.012	0.784
john	2.889	0.846	0.003	0.106	0.011	0.789
love	2.202	0.780	0.002	0.094	0.009	0.809
two	2.404	0.818	0.003	0.100	0.010	0.783
house	2.364	0.811	0.002	0.100	0.009	0.802
speak	2.495	0.853	0.003	0.107	0.010	0.782
word	2.364	0.825	0.002	0.102	0.009	0.798
heart	1.828	0.798	0.002	0.098	0.008	0.809
turn	2.343	0.818	0.002	0.101	0.009	0.797
stand	2.222	0.786	0.002	0.094	0.009	0.795
answer	2.030	0.825	0.002	0.102	0.008	0.796
rise	2.051	0.818	0.003	0.100	0.009	0.787
other	2.000	0.825	0.003	0.102	0.008	0.786
lady	2.222	0.839	0.003	0.105	0.009	0.793
man	2.040	0.805	0.002	0.098	0.008	0.795
such	1.646	0.773	0.002	0.092	0.007	0.799
last	2.556	0.805	0.002	0.098	0.010	0.799
nothing	2.040	0.825	0.002	0.103	0.008	0.802

Degree Centrality: The term 'say' exhibited the highest degree centrality (11.616), indicating that it co-occurs with the greatest number of other terms. This reflects the dialogue-driven nature of the novel, where speech acts are central to character interaction and narrative progression. Other high-degree terms include 'see' (7.091), 'go' (6.626), 'look' (5.980), and 'come' (5.899), all action and perception verbs that structure the protagonist's subjective experience. The protagonist's name, 'janeeyre' (4.586), and the male lead 'rochester' (4.556) also rank highly, confirming their centrality to the narrative.

Closeness Centrality: The term 'say' achieved a closeness centrality of 1.000, indicating that it has the shortest average path to all other nodes in the network. High closeness values for 'come' (0.990), 'see' (0.971), 'make' (0.971), and 'look' (0.961) suggest that these terms are semantically proximate to the broader lexical field, functioning as key organizing concepts within the text's discourse.

Betweenness Centrality: All terms exhibited very low betweenness centrality (ranging from 0.002 to 0.005), with 'say' (0.005) and 'come' (0.005) showing marginally higher values. The near-zero betweenness across the network indicates an absence of critical bridging nodes, consistent with the high graph density observed in Table 3. This suggests that the novel's vocabulary forms a uniformly interconnected semantic space rather than a modular or hierarchical structure.

Eigenvector Centrality: The term 'say' (0.127) achieved the highest eigenvector centrality, followed by 'come' (0.126), 'make' (0.124), and 'see' (0.124). Eigenvector centrality reflects connections to other highly connected nodes; thus, these terms are not only central themselves but are also linked to other central terms, amplifying their structural influence within the network.

PageRank: PageRank values ranged from 0.007 to 0.041, with 'say' (0.041), 'see' (0.025), and 'go' (0.024) exhibiting the highest scores. PageRank weights nodes by both their direct connections and the importance of their neighbors, reinforcing the centrality of action and cognition verbs in the narrative's lexical architecture.

Clustering Coefficient: Clustering coefficients were uniformly high across all terms (ranging from 0.759 to 0.809), indicating that terms tend to form tightly interconnected triads. This high clustering reflects the semantic cohesion of the novel's vocabulary, where co-occurring terms share overlapping contexts and thematic associations.

Overall, the centrality analysis reveals that *Jane Eyre* is structured around a small set of highly connected action, perception, and dialogue verbs, alongside the protagonist and key character names. The absence of betweenness centrality and the high clustering coefficient suggest a densely interconnected, protagonist-centered narrative network with minimal structural fragmentation, corroborating interpretations of the novel as a subjective, first-person narrative focused on Jane's voice and perceptual experience (Gilbert & Gubar, 2020; Schabbach, 2024).

RQ3: Topic Modeling Analysis Results

LDA topic modeling was used to identify recurring lexical constellations within the novel. The present study treated segmented intra-text units as analytic documents and used the topic model to detect recurrent patterns of word co-occurrence across the narrative. In this study, topics are understood not as self-sufficient literary themes but as statistically recurring word clusters that require interpretive validation through close reading. Topic number was examined through coherence and perplexity across candidate models, and the three-topic solution was retained because it provided the best balance between interpretability and model fit. Since topic assignments in a single-text setting remain sensitive to segmentation and initialization, the results are interpreted as exploratory but meaningful thematic patterns rather than definitive partitions of the novel.

To determine the optimal number of topics for LDA modeling of *Jane Eyre*, perplexity and coherence scores were computed for topic numbers ranging from 2 to 10. The analysis was conducted using the default parameter settings: alpha = 0.5, beta = 0.01, 500 iterations, 10 topics, and 30 words per topic. These parameter settings were used to identify the most appropriate topic structure for the dataset. Table 5 presents the optimal topic number selection metrics.

TABLE 5

Perplexity and Coherence Scores

# of Topic	Perplexity Score	Coherence Score
2	-30.263	-5.472
3	-30.484	-6.524
4	-30.635	-8.709
5	-30.713	-9.092
6	-30.762	-9.878
7	-30.797	-10.334
8	-30.844	-12.131
9	-30.903	-11.941
10	-30.912	-12.713

Figure 3 illustrates the trade-off between perplexity and coherence across topic numbers from 2 to 10. While perplexity scores decreased gradually with increasing topic numbers, coherence scores declined sharply after $k = 3$, indicating a loss of semantic interpretability. At $k = 3$, the coherence score was −6.524, significantly better (less negative) than at $k = 4$ (−8.709) and beyond, while perplexity (−30.484) remained comparable to lower topic counts. This inflection point, where coherence degradation accelerates while perplexity improvements plateau, suggests that three topics provide the optimal balance between model fit and thematic interpretability. Consequently, $k = 3$ was selected for the LDA analysis of *Jane Eyre*.

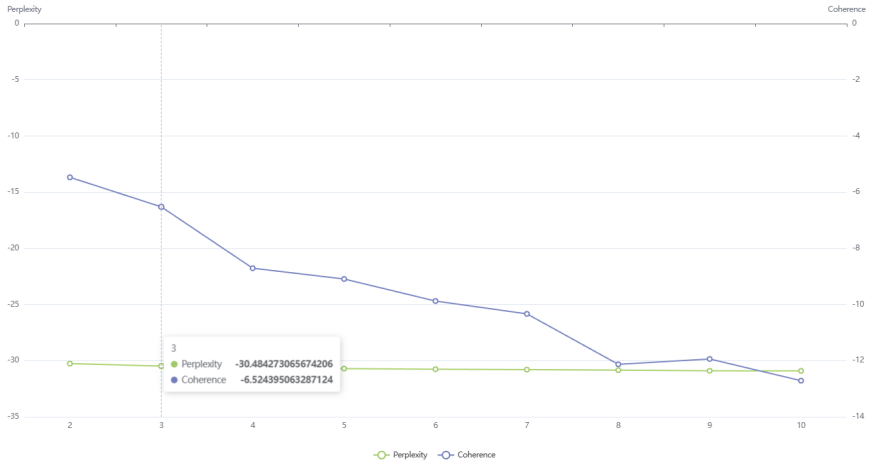


FIGURE 3
Visualization of Perplexity and Coherence Scores across Topic Numbers

Topic Identification and Interpretation

TABLE 6
Top Terms by Topic in LDA Analysis

Topic	Words with Weights
Topic 1	say(0.017), janeeyre(0.011), look(0.011), come(0.009), see(0.009), think(0.009), rochester(0.008), go(0.008), hear(0.008), know(0.007), find(0.005), john(0.005), give(0.005), time(0.005), more(0.004), eye(0.004), feel(0.004), good(0.004), don(0.004), wish(0.004), word(0.004), long(0.003), head(0.003), heart(0.003), let(0.003), rise(0.003), such(0.003), get(0.003), god(0.003), two(0.003)
Topic 2	janeeyre(0.033), good(0.01), take(0.009), day(0.008), think(0.007), go(0.007), say(0.007), own(0.005), see(0.005), night(0.005), come(0.005), room(0.005), leave(0.005), get(0.004), rochester(0.004), love(0.004), stand(0.004), tell(0.004), like(0.004), hand(0.004), ask(0.004), other(0.004), open(0.004), door(0.004), seem(0.003), life(0.003), know(0.003), house(0.003), eye(0.003), wish(0.003)
Topic 3	make(0.01), say(0.01), little(0.009), see(0.008), go(0.007), look(0.006), hand(0.006), seem(0.006), know(0.006), face(0.005), give(0.005), eye(0.005), feel(0.005), room(0.004), sit(0.004), come(0.004), tell(0.004), time(0.004), leave(0.004), speak(0.004), put(0.004), last(0.004), lady(0.003), man(0.003), fire(0.003), pass(0.003), heart(0.003), bessie(0.003), hall(0.003), great(0.003)

Based on the weighted word distributions in Table 6, three thematic topics were identified. While common verbs of motion appear across all clusters, distinct noun-verb associations reveal the novel's thematic structure.

Topic 1: Cognitive Processing and Relational Dynamics

Topic 1 is characterized by cognitive verbs like 'think' (0.009) and 'know' (0.007) co-occurring with key figures such as 'Rochester' (0.008) and 'John' (0.005). This cluster captures Jane's internal psychological processing, reflecting her negotiation of authority and affection through introspection. The data supports interpretations of the protagonist constructing an autonomous self through internal deliberation (Abdulla & Ahmed, 2023; Nie et al., 2024).

Topic 2: Domestic Enclosure and Temporal Existence

Topic 2 focuses on the material reality of the setting, featuring spatial markers like 'room' (0.005) and 'door' (0.004) alongside temporal terms like 'day' (0.008). The dominance of 'janeeyre' (0.033) in this cluster strongly links the protagonist to the domestic sphere. This aligns with Gilbert and Gubar's (2020) analysis of Gothic enclosure, where moral agency is

tested within the confines of physical space and daily routines (Schiavone, 2025).

Topic 3: Embodied Interaction and Class Consciousness

Topic 3 highlights physical presence and social hierarchy, dominated by body-part terms (hand, face) and class indicators (lady, man, Bessie). The specific mention of Bessie alongside general social terms suggests that abstract class distinctions are experienced through concrete, embodied interactions (Van den Bossche, 2005). This corroborates the view that Jane negotiates social mobility through direct interpersonal exchange (Gao, 2013).

Collectively, the LDA analysis demonstrates that *Jane Eyre* is structured around subjective consciousness (Topic 1), domestic spatiality (Topic 2), and embodied social relations (Topic 3). These findings confirm the novel's status as a Bildungsroman that integrates internal psychological growth with the material realities of Victorian society (Gilbert & Gubar, 2020; Nie et al., 2024).

Specifically, the topic clusters can be further grounded in brief textual evidence. For cognitive processing and relational dynamics, Jane explicitly resists inherited authority: "I don't think, sir, you have a right to command me" (Brontë, 1847/1998, Ch. 14), while also regulating her own desire through self-reflective restraint: "keep to your caste" (Ch. 17). The cluster of domestic enclosure and temporal existence is equally visible in passages of spatial confinement and arrested duration, such as Jane's repetitive movement "along the corridor of the third storey" (Ch. 12) and her description of a return to "stagnation" and "viewless fetters" (Ch. 12). Embodied interaction and class consciousness emerge in scenes where hierarchy is felt as bodily and verbal pressure, from Jane's denunciation of John Reed as a "slave-driver" (Ch. 1) to her ironic observation about "masters" and "paid subordinates" (Ch. 14).

Taken together, the three topics describe the novel not as a set of disconnected motifs but as a coherent structure of development. Topic 1 foregrounds Jane's reflective negotiation of attachment, authority, and self-respect; Topic 2 maps the material and temporal constraints of domestic enclosure; and Topic 3 shows that class relations are experienced through embodied social interaction rather than through abstraction alone. Read together, these topics suggest that the novel's central movement is the formation of moral and psychological agency under conditions of spatial restriction, emotional conflict, and unequal social power. The thematic structure revealed by LDA therefore aligns with critical accounts of *Jane Eyre* as a novel of self-formation in which inward consciousness, domestic discipline, and classed embodiment are tightly interdependent.

Topic-Specific Characteristics

Figure 4 presents an Intertopic Distance Map that visualizes the semantic relationships among the three topics in the PC1-PC2 coordinate space. Topic 2, located in the upper-left area, constitutes the largest cluster at 34.2% of the total tokens, indicating its thematic prominence within the corpus. Topic 3, positioned in the upper-right area, accounts for 32.6% of the tokens, whereas Topic 1, situated in the lower-central area, comprises 33.2%. The clear spatial separation among the three clusters indicates minimal thematic overlap and supports the interpretive distinctiveness of the identified topics.

FIGURE 4
Intertopic Distance Map for the Three Topics

The bar chart in Figure 5 displays the top 30 most relevant terms for Topic 1 ($\lambda = 1$). While 'janeeyre' exhibits the highest overall frequency in the corpus (represented by the blue bar), the estimated term frequencies within Topic 1 (red bars) reveal a strong prevalence of sensory and cognitive verbs such as 'say', 'look', 'see', 'think', and 'hear'. This distribution, alongside character references like 'rochester', corroborates the label "Cognitive Processing and Relational Dynamics" assigned to this topic. Instead of physical navigation, these terms highlight the protagonist's internal narrative voice and psychological evaluation of her surroundings and relationships, aligning with psychoanalytic interpretations of her development (Abdulla & Ahmed, 2023; Nie et al., 2024).

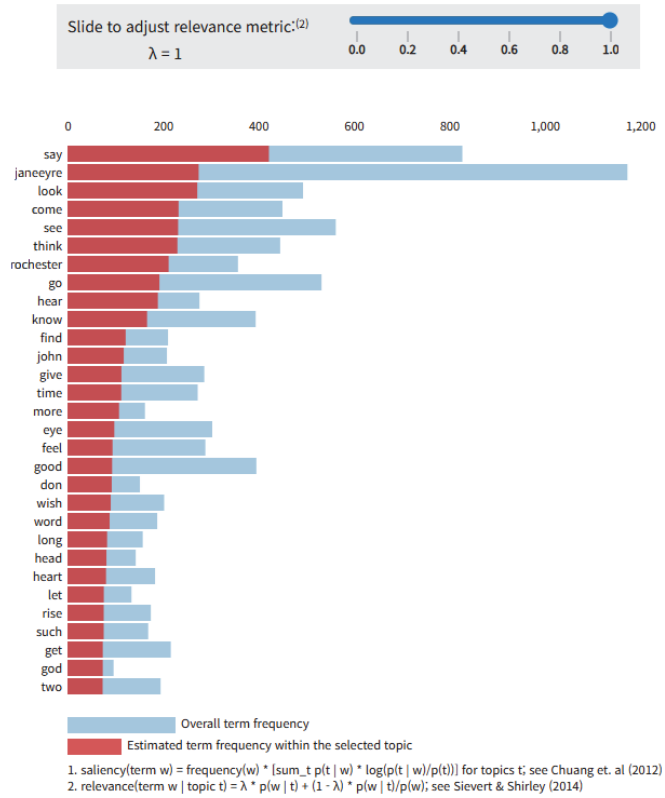


FIGURE 5
Top-30 Most Relevant Terms for Topic 1

The bar chart in Figure 6 displays the top 30 most relevant terms for Topic 2 ($\lambda = 1$). While 'janeeyre' shows the highest frequency overall, the red bar indicates its specific prevalence within this topic is significant (approx. 900 occurrences), strongly linking the protagonist to the accompanying vocabulary. Unlike the cognitive focus of Topic 1, this list features spatial and temporal markers such as 'room', 'door', 'house', 'day', and 'night', alongside action verbs like 'leave', 'take', and 'stand'. These terms corroborate the label "Domestic Enclosure and Temporal Existence" assigned to Topic 2. Rather than abstract moral concepts, this cluster captures the protagonist's embodied experience within the physical boundaries of the Victorian household, supporting Gilbert and Gubar's (2020) analysis of the novel as a drama of enclosure and the struggle for movement within domestic spaces.

The bar chart in Figure 7 shows the top 30 most relevant terms for Topic 3 ($\lambda = 1$). The verbs 'make' and 'say' exhibit the highest within-topic frequency (red bars), followed by 'little', 'see', 'look', and 'go'. Crucially, this topic is distinguished by the clustering of somatic terms like 'hand', 'face', and 'eye' alongside interpersonal markers including 'lady', 'man', and the specific character 'Bessie'. These terms corroborate the label "Embodied Interaction and Class Consciousness." Unlike the internal focus of Topic 1, the presence of speech acts (speak, tell, say) combined with setting markers (room, hall, fire) suggests concrete scenes of social transaction and dialogue within the domestic sphere. This validates the three-topic model's capacity to differentiate between the protagonist's internal cognition (Topic 1), her domestic spatial existence (Topic 2), and her embodied negotiation of social class (Topic 3).

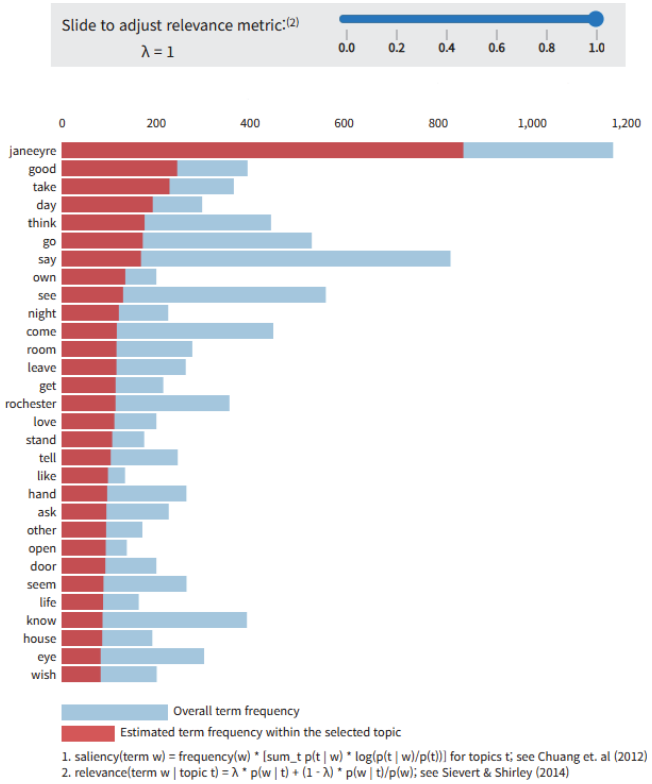


FIGURE 6
Top-30 Most Relevant Terms for Topic 2

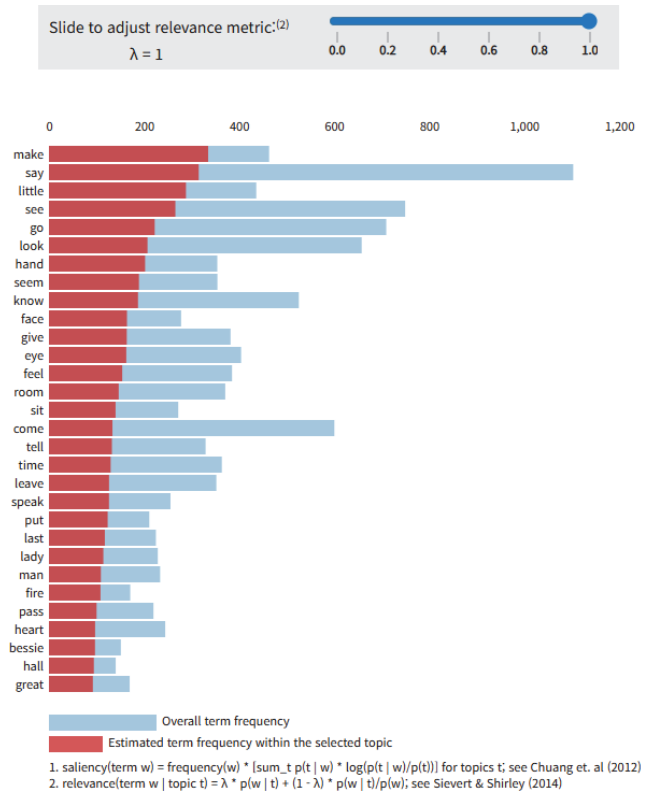


FIGURE 7
Top-30 Most Relevant Terms for Topic 3

RQ4: Sentiment Analysis Results

Sentiment trends in *Jane Eyre* are significant because the novel is structured through Jane's focalized consciousness and retrospective self-narration. Emotional shifts across chapters can be read as indicators of narrative tension and affective emphasis. However, the results should be interpreted cautiously, since retrospection, irony, and Victorian diction may complicate automatic emotion classification. In addition, narrator text and dialogue were not separately modeled in the present study, so the scores reflect the combined emotional profile of narrated discourse and character speech.

Figure 8 presents the chapter-by-chapter emotional trajectory of *Jane Eyre* across seven affect categories, derived from sentiment classification. The x-axis represents the 38 chapters of the novel, while the y-axis indicates emotion intensity.

Neutral (gray line) maintains the highest baseline intensity throughout the narrative, peaking at approximately 0.33 around Chapter 30. This dominance reflects the novel's detailed descriptive passages and Jane's retrospective narration, which often requires emotional equilibrium to recount past trauma (Gao, 2013).

Surprise (sky blue line) exhibits the most distinct anomaly in the dataset, showing a dramatic spike between Chapters 10 and 12, reaching intensities near 0.26. This surge empirically captures a pivotal narrative transition: Jane's decision to advertise for a governess position, her departure from the stagnation of Lowood, and her first unexpected encounter with Rochester at Hay Lane, marking the shift from passive endurance to dynamic interaction.

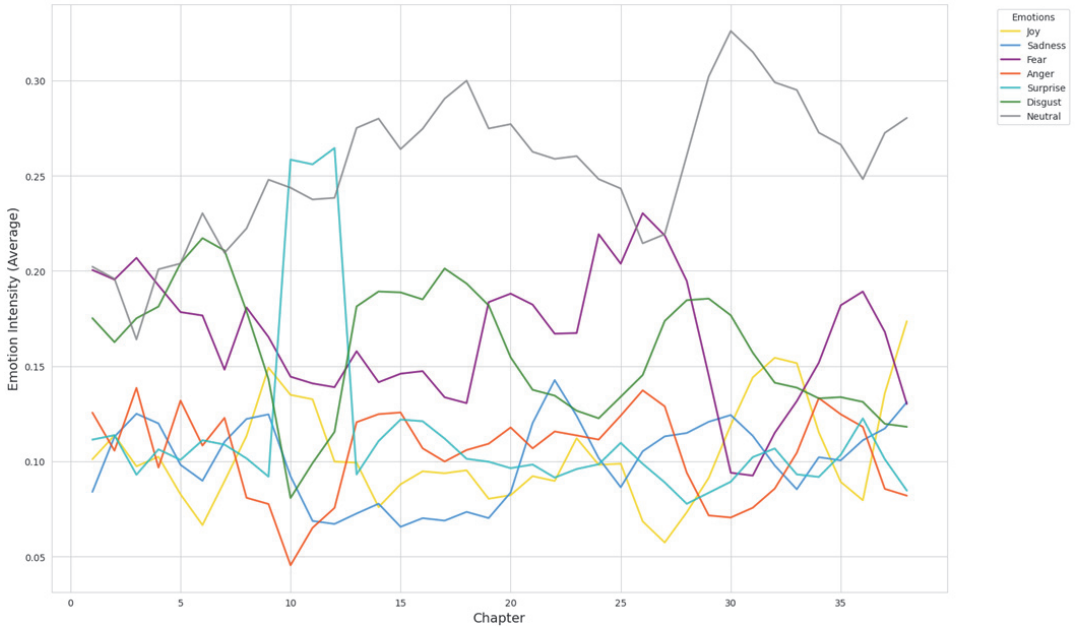


FIGURE 8
Emotional Trends in Jane Eyre (7 Emotions)

Fear (purple line) demonstrates significant volatility, with a notable peak around Chapter 25 (approx. 0.23). This elevation coincides with the Gothic climax of the novel—the bed-curtain tearing incident and the terrifying revelation of Bertha Mason on the wedding day—validating Gilbert and Gubar's (2020) interpretation of the text's underlying psychological terror.

Disgust (green line) emerges as a prominent negative emotion, particularly in the early chapters (Chapters 1–8), where it consistently hovers around 0.18–0.20. This high trajectory aligns with Jane's visceral reaction to the moral hypocrisies of the Reed family and Mr. Brocklehurst, corroborating analyses of the novel's social critique (Abdulla & Ahmed, 2023; Nie et al., 2024).

Joy (yellow line) remains suppressed for the majority of the text but exhibits a sharp upward trajectory in the final chapter (Chapter 38). This terminal rise captures the resolution in "Reader, I married him," marking the final transition from struggle to domestic fulfillment.

Meanwhile, Anger (red line) and Sadness (blue line) maintain relatively stable intensities, generally fluctuating between 0.05 and 0.15. This suggests that while Jane experiences these emotions, the narrative's primary psychological tension is driven more by the interplay of Disgust (social aversion), Fear (existential threat), and Surprise (novelty/shock), rather than

sustained melancholy or rage (Schiavone, 2025).

The sentiment trajectories are meaningful not only as emotional fluctuations but also as indicators of narrative pressure and transition. Peaks in surprise correspond to moments when Jane's world is structurally reoriented, especially in the movement from Lowood to Thornfield and in her first meaningful encounters with Rochester. Fear intensifies when the text shifts into Gothic uncertainty, showing that emotional escalation is tied to epistemic instability as much as to sensational plot. By contrast, the delayed rise of joy near the ending suggests that emotional resolution in the novel is neither immediate nor continuous, but achieved only after prolonged moral testing, separation, and reconfiguration of social relations. For example, emotional spikes are also legible in the novel's local language. Romantic attachment intensifies in Jane's declaration that Rochester is her "home—my only home" (Brontë, 1847/1998, Ch. 22), whereas Gothic fear is marked by "some coming vision from another world" (Brontë, 1847/1998, Ch. 2) and by the nocturnal scream that rends the silence at Thornfield (Brontë, 1847/1998, Ch. 20). The sentiment results therefore reinforce the broader argument that *Jane Eyre* develops through alternating phases of confinement, disruption, and self-authored resolution.

DISCUSSION

This study demonstrates that an integrated NLP pipeline—combining network text analysis, LDA topic modeling, and sentiment analysis—can operationalize the lexical, thematic, and affective organization of *Jane Eyre* in a manner that is both replicable and theoretically interpretable. By treating the novel as a structured linguistic system rather than solely an interpretive artifact, the analysis contributes to the methodological pluralism advocated in digital literary studies, where computational evidence is positioned as an amplification of humanistic reading rather than its replacement.

This study advances methodological development in digital humanities and English language pedagogy by demonstrating how network analysis, topic modeling, and sentiment analysis function as complementary analytical frameworks that reveal narrative structures difficult to isolate through traditional close reading alone. Network centrality measures quantify discourse organization by identifying lexical hubs such as 'janeeyre,' 'rochester,' and cognitive verbs ('think,' 'know,' 'feel') that structure the novel's semantic architecture, thereby operationalizing intuitions about protagonist-centered focalization and social power dynamics (Schabbach, 2024). Concurrently, LDA topic modeling decomposes the text into three empirically grounded thematic clusters—cognitive-relational dynamics, domestic enclosure, and embodied class interaction—making latent semantic structures explicit and providing scaffolding for interpreting the novel's ideological concerns regarding autonomy, spatial constraint, and social mobility (Gilbert & Gubar, 2020; Van den Bossche, 2005). Sentiment analysis further contributes by mapping affective trajectories across the 38-chapter narrative arc, revealing how emotional pacing aligns with Bildungsroman developmental stages and psychological realism, while TF-IDF profiles of cognitively loaded verbs support testable hypotheses about Brontë's representation of consciousness and moral reasoning (Ban et al., 2022; J. H. Oh, 2012).

Together, these computational procedures generate what Ramsay (2011) terms 'interpretive prompts'—quantitative patterns that productively guide subsequent close reading by surfacing statistically salient features that may remain implicit in unaided textual engagement. This multi-method triangulation aligns with Bednarek's (2007) argument that corpus-driven techniques foreground interpretable textual patterns, including thematic emphasis, character construal, and rhetorical strategy, thereby extending traditional literary analysis with empirical evidence. For English language education, this framework offers pedagogical affordances by transforming linguistically challenging Victorian prose into visible, teachable structures: network visualizations clarify lexical relationships and narrative centrality, topic distributions provide thematic organizers for pre-reading scaffolding, and sentiment curves identify emotionally salient chapters for targeted discussion of narrative turning points (Adolphs, 2006; Navarro-Colorado, 2017). The transparency and replicability of these methods support collaborative learning activities where students can examine shared quantitative evidence, formulate data-driven interpretive hypotheses, and negotiate meanings through structured inquiry—competencies transferable to academic writing, critical reading, and research literacy across disciplines (Lusta et al., 2023; Rodriguez & Csomay, 2024).

Overall, the computational results are most valuable when read as evidence of form rather than as isolated measurements. TF-IDF and network centrality show that the novel is lexically organized around Jane's acts of seeing, thinking, and speaking; topic modeling reveals that this consciousness is developed within recurring structures of relational negotiation, domestic enclosure, and classed embodiment; and sentiment analysis traces how these structures are emotionally lived across the narrative. What emerges, therefore, is a quantitatively supported account of *Jane Eyre* as a first-person Bildungsroman in which consciousness, social relation, and moral agency are inseparable. This interpretive synthesis is important because it demonstrates how computational analysis can extend, rather than replace, literary reading by making large-scale narrative

patterns more explicit.

Integrating Topic Modeling with Literary Reading

The LDA results provide empirical grounding for long-established literary claims about the novel's psychological interiority, ethical self-fashioning, and socially mediated embodiment. The selection of a three-topic solution, supported by the balance of perplexity and coherence, aligns with prior topic-modeling research emphasizing interpretability as a core validity criterion in literary applications (Chang et al., 2009; Rhody, 2012). In this respect, the present study follows the principle that topic models are not self-explanatory outputs but analytic instruments requiring interpretive calibration to remain meaningful for literary scholarship.

Substantively, Topic 1 "Cognitive Processing and Relational Dynamics" foregrounds cognition and perception verbs alongside relationship-signaling entities (e.g., rochester, john), reinforcing narratological accounts of *Jane Eyre* as a text where focalized consciousness structures story-world access. This is consistent with prior arguments that the novel's subjectivity is constructed through spatially and psychologically mediated experience (Nie et al., 2024) and through narrative strategies that manage openness, focus, and reader address (Schabbach, 2024). Topic 2 "Domestic Enclosure and Temporal Existence" prioritizes domestic spatial terms (room, door, house) and temporal markers (day, night), offering quantitative support for readings that treat domestic architecture as a mechanism through which agency is negotiated and moral decisions are staged. Such results are compatible with scholarship framing the novel as a developmental trajectory in which autonomy is repeatedly tested within institutional and domestic constraints (J. H. Oh, 2012). Topic 3 "Embodied Interaction and Class Consciousness" highlights bodily terms (hand, face, eye) alongside social-position markers (lady, man, bessie), echoing arguments that Victorian textual bodies function as ideological sites where gender politics and social hierarchy become legible (Chen & Xu, 2024; Van den Bossche, 2005).

Importantly, these topic structures also speak to the growing body of work demonstrating how computational modeling can support "microscopic" single-text analysis—distinct from macro-scale genre mapping—by revealing internal thematic segmentation and recurrence patterns that may be difficult to stabilize through close reading alone (Ramsay, 2011; Salmi, 2024; Underwood, 2019). The present findings therefore extend prior pedagogically oriented NLP case studies in English literary education by showing that topic models can be used not only for summary but for theory-consistent thematic argumentation (Y. K. Oh, 2024; Sekar, 2024).

Sentiment Trajectories as Evidence of Narrative Affect and Genre Dynamics

The sentiment analysis, conducted across 38 chapters with seven affect categories, provides a chapter-level affective profile that is interpretable in relation to narrative dynamics. The predominance of Neutral affect is consistent with the descriptive and retrospective narration typical of nineteenth-century realist prose, while the patterned fluctuations of Surprise, Fear, Disgust, and Joy indicate that affect is not uniformly distributed but strategically concentrated at narratively consequential junctures. This supports recent arguments that sentiment analysis can be productively adapted to literary texts when treated as an interpretive cue rather than a definitive psychological measurement (Yuri & Pascale, 2024).

The observed spike of Disgust in early chapters plausibly corresponds to scenes of institutional and domestic oppression, which feminist interpretations identify as foundational to Jane's resistance and self-definition (Al-Dabool, 2025; Anees et al., 2021; Gilbert & Gubar, 2020). Similarly, the prominent rise of Fear around the mid-to-late narrative aligns with Gothic intensification and crisis—an arc consistent with comparative Gothic readings of *Jane Eyre* within the Brontë literary field (Tine, 2024). The delayed upward trajectory of Joy toward the ending suggests that affective resolution is narratively postponed, echoing Bildungsroman accounts in which emotional stabilization is a late-stage outcome of moral and social negotiation rather than an early romantic payoff (J. H. Oh, 2012). Together, these affective signatures extend prior *Jane Eyre*–specific text mining studies by articulating how emotion categories can be linked to interpretive claims about oppression, crisis, and resolution without reducing the novel to a single dominant mood (Ban et al., 2022).

Methodologically, the findings also connect to sentiment-research infrastructure in computational linguistics (e.g., lexicon resources for polarity and affect), while highlighting why literary-domain considerations remain crucial for model validity in historical prose (Baccianella et al., 2010; Yuri & Pascale, 2024). In this sense, sentiment outputs function best as structured evidence for triangulation with other NLP layers (topics and networks), rather than as stand-alone proof of literary meaning.

Triangulation Across Networks, Topics, and Sentiment: Interpretive Value

A central contribution of this study lies in triangulating three analytical layers: (1) network-based representations of lexical association and discourse prominence, (2) probabilistic topic structure, and (3) affective progression. Such triangulation corresponds to recommendations in digital humanities to avoid single-method determinism and to build convergent validity through complementary computational lenses (Jockers, 2013; Thomas, 2024; Underwood, 2019).

From a computational narratology standpoint, the network perspective is particularly valuable because literary character and concept relations can be formalized as graphs, enabling reproducible structural claims about narrative organization (Borgatti & Everett, 2006; Scott & Carrington, 2011). This approach is consistent with earlier work on extracting and analyzing literary social networks (Elson et al., 2010) and with more recent NLP discussions of automatic character network construction as a scalable literary-analytic task (Amalvy et al., 2025). When integrated with LDA, network outputs help specify whether high-frequency elements are merely common tokens or structurally central discourse anchors, while sentiment trajectories indicate whether structurally central segments also correspond to affectively salient narrative moments.

In pedagogical terms, this multi-layer convergence has direct utility for explainable literary instruction: educators can locate teachable thematic clusters (topics), identify lexically central nodes (networks), and select emotionally distinctive chapters (sentiment) to reduce cognitive load for non-specialist readers. This aligns with the pedagogical shift in distant-reading-informed teaching that emphasizes visualization, pattern recognition, and interpretive scaffolding as supports for engagement with linguistically challenging canonical texts (Sekar, 2024).

Although much literary scholarship privileges plot- and story-centered explication, a segmented computational approach (e.g., keywords/collocations, network structure, topic distributions, and sentiment trajectories) can deepen textual understanding by externalizing how narrative meaning is organized across lexical recurrence, relational architecture, thematic clustering, and affective pacing, thereby making interpretive claims more transparent and empirically accountable. Pedagogically, this evidential, multi-layer workflow is effective for English language education because it converts stylistically dense literary prose into explicit, teachable patterns that support scaffolded reading, data-informed discussion, and evidence-based academic writing, while also cultivating learners' transferable analytical and digital literacy skills.

CONCLUSION

Theoretically, the study supports an interdisciplinary claim: computational evidence can strengthen theory-guided literary interpretation by supplying measurable patterns that constrain or refine qualitative assertions (Ramsay, 2011; Salmi, 2024). In particular, the topic structure offers a data-driven account of how cognition, enclosure, and embodiment co-exist as separable yet interlocking dimensions of meaning-making, which can be connected to feminist and class-oriented arguments about agency and ideology in the novel (Gilbert & Gubar, 2020; Van den Bossche, 2005).

Methodologically, the work contributes to the "microscopic" end of computational literary studies by showing that rigorous interpretive payoffs are possible even without multi-novel corpora, provided that analytic decisions (e.g., topic number selection) are transparent and theory-aligned (Chang et al., 2009; Underwood, 2019). For future research, applying the same pipeline to comparative sets of nineteenth-century women's novels—an approach already advocated in topic-based comparisons of female-authored fiction—would allow researchers to distinguish text-specific signatures from period-level regularities (Rha & Silver, 2021).

Pedagogically, the study recommends a structured instructional design model: (a) introduce topic summaries as pre-reading organizers, (b) use network visualizations to map character/idea prominence and relational clusters, and (c) use sentiment curves to guide discussion of turning points and genre shifts. Such sequencing resonates with recent discussions of digital humanities pedagogy that frame computational outputs as bridges between distant reading and close reading practices (Jockers, 2013; Salmi, 2024; Sekar, 2024).

However, two limitations should be noted. First, topic models and co-occurrence networks rely on distributional patterns and can underrepresent figurative or context-sensitive meanings, an issue repeatedly discussed in the topic-modeling literature for humanities texts (Rhody, 2012). Second, sentiment models trained on contemporary language may imperfectly capture Victorian affective semantics, supporting calls for domain-sensitive adaptation when analyzing literary prose (Yuri & Pascale, 2024).

This study shows that an integrated NLP workflow can produce a replicable, theory-relevant account of *Jane Eyre*'s thematic structure and affective progression, thereby strengthening the methodological bridge between computational analysis and literary interpretation. By combining topic modeling with sentiment trajectories and network-informed structural evidence, the analysis illustrates how computational methods can make canonical Victorian fiction more

accessible for educational contexts while remaining compatible with interpretive traditions in literary studies.

At the level of disciplinary contribution, the findings support the view that computational reading should be understood as an evidential extension of interpretive practice: algorithms stabilize patterns; scholars explain significance (Ramsay, 2011; Salmi, 2024; Underwood, 2019). At the level of pedagogy, the results align with recent arguments that distant-reading tools can function as interpretive scaffolds—especially for non-specialist readers—by transforming complex texts into visible thematic and emotional structures (Y. K. Oh, 2024; Sekar, 2024).

Future research should empirically validate this integrated framework through controlled pedagogical interventions while simultaneously extending the pipeline to cross-linguistic comparative analysis of translated Victorian novels, thereby establishing both educational efficacy and methodological generalizability across diverse linguistic and cultural contexts.

In sum, this study makes three contributions. First, it demonstrates how network analysis, topic modeling, and sentiment analysis can be integrated to reveal narrative patterns in Victorian fiction. Second, it provides a methodological framework that can be applied to other literary texts. Third, it offers English teachers practical tools for helping students engage with literature that might otherwise appear distant or difficult. By making computational methods accessible to non-specialists, the study illustrates how such approaches can enrich literary inquiry without requiring advanced technical expertise.

References

Abdulla, P. S., & Ahmed, A. A. (2023). Madness and feminism: An analysis of women's struggles in Bronte's *Jane Eyre* and Kaysen's *Girl, Interrupted*. *Journal of Philology and Educational Linguistics*, *2*(2), 1–18. https://doi.org/10.53898/jpes2023221

Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies*. Routledge. https://doi.org/10.4324/9780203087701

Al–Dabool, A. R. M. (2025). A feminist interpretation of the novel *Jane Eyre* by author Charlotte Brontë. *International Journal of Social Science and Humanities*, *9*(2), 55–71.

Amalvy, A., Labatut, V., & Dufour, R. (2025). The role of natural language processing tasks in automatic literary character network construction. *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 8462–8473). Abu Dhabi, UAE: Association for Computational Linguistics.

Anees, M., Ali, A., Iqbal, D. L., Ahmad, S., Khan, I. U., & Khan, A. (2021). Patriarchal ideology as gender discrimination: A feminist study of *Jane Eyre*. *International Journal of Innovation*, *15*(5), 1077–1103.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: *An enhanced lexical resource for sentiment analysis and opinion mining*. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (LREC'10) (pp. 2200–2204). European Language Resources Association (ELRA).

Ban, Geonwoo, Ahn, Somi, & Park, Soyoung. (2022). Topic analysis and sentiment analysis of *Jane Eyre* through text mining. *The Journal of Humanities and Social Science*, *13*(2), 2415–2428.

Bednarek, M. (2007). Teaching English literature and linguistics using corpus stylistic methods. *Bridging Discourses: ASFLA 2007 Online Proceedings*, 129–138.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466–484.

Brontë, C. (1998). *Jane Eyre*: An autobiography. Project Gutenberg. (Original work published 1847)

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, *22*, 1–9.

Chen, C. M., & Chang, C. (2019). A Chinese ancient book digital humanities research platform to support digital humanities research. *The Electronic Library*, *37*(2), 314–336. https://doi.org/10.1108/EL-10-2018-0213

Chen, H., & Xu, Q. (2024). Gender politics and Victorian literary representation of the body: A distant reading of the body in Charles Dickens's works. *Humanities and Social Sciences Communications*, *11*(1), 1–11.

Elson, D. K., Dames, N., & McKeown, K. R. (2010). *Extracting social networks from literary fiction*. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 138–147). Uppsala, Sweden: Association for Computational Linguistics.

Frei, G. V. (2025). AI-powered teaching of literature in foreign language classrooms: Practical prompts for the didacticization of literary works. *Philologia Mediana*, *17*(17), 649–663. https://doi.org/10.46630/phm.17.2025.45

Gao, H. (2013). Reflection on feminism in *Jane Eyre*. *Theory and Practice in Language Studies*, *3*(6), 926–931. https://doi.org/10.4304/tpls.3.6.926-931

Gilbert, S. M., & Gubar, S. (2020). *The madwoman in the attic: The woman writer and the nineteenth-century literary imagination* (2nd ed.). Yale University Press.

Jiang, C., Zhang, L., & Wang, Y. (2025). Integrating natural language processing techniques with LDA topic modeling: Deconstructing the evolution of media discourse on educational modernization ideology. *Education and Information Technologies*, *30*, 23419–23466. https://doi.org/10.1007/s10639-024-13289-4

Jockers, M. L. (2013). *Macroanalysis*: *Digital methods and literary history*. University of Illinois Press. https://doi.org/10.5406/

illinois/9780252037528.001.0001
Khan, A. A., Minhas, N., & Kaloi, M. A. (2025). From text to tech: Exploring the impact of digital humanities on literary interpretation. *Review Journal of Social Psychology & Social Works*, *3*(2), 620–636.
Kim, Hoyeol. (2022). *Computational approaches in the humanities: From sentiment analysis to deep learning colorization* [Unpublished doctoral dissertation]. Texas A&M University.
Kim, Tae-Young, Oh, Shinyu, & Lee, Eun Jin. (2023). Analysis of research trends in English learning motivation in Korea using text mining. *Modern English Education*, *24*, 317–340. https://doi.org/10.18095/meeso.2023.24.1.317
Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, *9*(12). https://doi.org/10.1016/j.heliyon.2023.e22731
Maeng, Unkyoung, Ko, Ho Kyoung, & Son, Bok Eun. (2023). Utilizing digital · AI in English education: Research trends through text network analysis and topic modeling. *Modern English Education*, *24*, 156–170.
Mahadevkar, S. V., Patil, S., Kotecha, K., & Soong, L. W. (2024). Exploring AI-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, *11*(1), 92. https://doi.org/10.1186/s40537-024-00948-z
Mimno, D. (2012). Computational historiography: Data mining in a century of classics journals. *Journal on Computing and Cultural Heritage (JOCCH)*, *5*(1), 1–19. https://doi.org/10.1145/2160165.2160168
Navarro-Colorado, B. (2017). A quick intensive course on Natural Language Processing applied to literary studies. *Teach4DH@ GSCL*, 37–42.
Nie, Z., Mani, M., & Kaur, H. (2024). Unveiling *Jane Eyre*: Space escape and the construction of subjectivity through a Foucauldian lens. *Journal of Literary Studies*, *40*(1), 1–19. https://doi.org/10.25159/1753-5387/15716
Oh, Jung-hwa. (2012). Female development in nineteenth-century England and dynamics of the Bildungsroman. *Women's Studies Review*, *29*(2), 3–35. https://doi.org/10.18341/wsr.2012.29.2.3
Oh, Young Kyo. (2024). AI-powered text analysis of English literature: A Natural Language Processing (NLP) study centered on Jane Austen's novel *Emma*. *Korean Journal of English Language and Linguistics*, *24*, 979–1010. https://doi.org/10.15738/kjell.24..202409.979
Project Gutenberg. (n.d.). About Project Gutenberg. Retrieved January 2, 2026, from https://www.gutenberg.org/about/
Ramsay, S. (2011). *Reading machines: Toward an algorithmic criticism*. University of Illinois Press. https://doi.org/10.5406/illinois/9780252036415.001.0001
Rha, L., & Silver, S. (2021). Topic modeling and analysis: Comparing the most common topics in 19th-Century novels written by female writers. *Aresty Rutgers Undergraduate Research Journal*, *1*(3).
Rhody, L. M. (2012). Topic modeling and figurative language. *Journal of Digital Humanities*, *2*(1), 19–35. http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/
Rodriguez, E., & Csomay, E. (2024). Using Corpus linguistics and language analysis tools for ESL vocabulary instruction: Two case studies. *The CATESOL Journal*, *35*(1). https://doi.org/10.5070/B5.35860
Salmi, H. (2024). Textuality as amplification: Reconsidering close reading and distant reading in cultural history. *Rethinking History*, *28*(2), 261–277. https://doi.org/10.1080/13642529.2024.2360318
Schabbach, G. R. (2024). *Focus, openness and reader addresses in Jane Eyre and Villette: A study of narrative strategies*. [Unpublished master's thesis]. Universidade Federal do Rio Grande do Sul.
Schiavone, M. (2025). Reading *Jane Eyre* as a Hagiographic Romance. *Brontë Studies*, *51*(1), 28–43. https://doi.org/10.1080/14748932.2025.2564940
Scott, J., & Carrington, P. J. (2011). *The SAGE handbook of social network analysis*. SAGE publications.
Sekar, J. J. (2024). How distant is 'Distant Reading'? A paradigm shift in pedagogy. *Asian Journal of Languages, Literature and Cultural Studies*, *7*(1), 84–99.
Sharma, R. C. (2025). Reading between the lines: Exploring English literature through AI and machine learning. *Journal of Informatics Education and Research*, *5*(2), 5958–5978. https://doi.org/10.52783/jier.v5i2.3123
Stiller, J., Nettle, D., & Dunbar, R. I. M. (2003). The small world of Shakespeare's plays. *Human Nature*, *14*(4), 397–408. https://doi.org/10.1007/s12110-003-1013-1
Thierauf, D. (2023). Cancel Jane? *Jane Eyre*, romance, and the lure of white feminism. *Nineteenth-Century Gender Studies*, *19*(3), 1–15. https://m.ncgsjournal.com/issue193/PDFs/thierauf.pdf
Thomas, J. B. (2024). *Digital reading*: *A computational approach to literary studies with applications in British literary history*. [Unpublished doctoral dissertation]. The University of North Carolina at Chapel Hill. https://cdr.lib.unc.edu/downloads/2j62sg86h
Tine, E. P. (2024). Gothic experiences in *Jane Eyre*, *Wuthering Heights*, and *The Tenant of Wildfell Hall*. *ESI Preprints (European Scientific Journal, ESJ)*, *20*(11), 69–85. https://doi.org/10.19044/esj.2024.v20n11p69
Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226612973.001.0001
Van den Bossche, C. R. V. (2005). What did *Jane Eyre* do? Ideology, agency, class and the novel. *Narrative*, *13*(1), 46–66.
Yuri, B., & Pascale, F. (2024). Sentiment analysis for literary texts: Hemingway as a case-study. *Journal of Data Mining & Digital Humanities*, 1–20. https://doi.org/10.46298/jdmdh.13155