



Lexical Patterns in Hawthorne's *The House of the Seven Gables*: A Corpus Stylistic Analysis

Mamona Yasmin Khan¹ & Sajda Abuzar²

¹Professor of English, Department of English, The Women University Multan, Email: mamona.6231@wum.edu.pk

²Scholar, Department of English, The Women University Multan, Email: waasfaqureshi786@gmail.com

ARTICLE INFO

Article History:

Received: January 20, 2025
Revised: February 26, 2025
Accepted: March 02, 2025
Available Online: March 04, 2025

Keywords:

Corpus stylistics, lexical patterns, keywords, KWIC, semantic domains

Corresponding Author:

Mamona Yasmin Khan

Email:

mamona.6231@wum.edu.pk



ABSTRACT

In recent years, there has been remarkable progress in computer science and technology, leading to the emergence of Computational Linguistics as a new approach for analyzing linguistic elements. This study is a mixed method approach and aims to analyze keywords and key semantic domains in Hawthorne's selected novel shedding light on the author's stylistic fingerprints and narrative techniques. The whole text of the novel acts as a sample and is compared to reference corpora of 19th Century Literature. Data collection is facilitated by user-friendly corpus software, Antconc is used for generating keywords and USAS and Voyant tools are used collectively for the extraction of semantic domains. This approach introduces corpus stylistic analysis, which examines language usage in specific texts, providing fresh insights into previously interpreted data. Findings reveal that the corpus approach to literary texts primarily uses quantitative data, but how those results are interpreted and assessed depends greatly on the analyst's familiarity with and understanding of the texts being studied. Ultimately, the study advocates for the utilization of corpus stylistic analysis to explore diverse aspects of literary texts but also reveals that that computerized analysis of texts is not contrary to conventional analytical methods.

Introduction

The field of corpus stylistics is becoming more and more important in language analysis. It combines elements from two distinct fields. A means of identifying different stylistic elements is made possible by corpus tools and software. Literary texts are analyzed using electronic tools by researchers. The discipline came about as a result of innovation. Corpus Stylistics is a tool that blends machine learning and human intelligence to identify linguistic style patterns. Researchers and pioneers in the field regard this discipline as an additional complement to fill in any gaps. It draws attention to the language paradigms present in the literary discourse. As a result, it strengthens the two sectors' reliant interdependence. It is possible to state that Corpus Stylistics

serves as a link between the three different domains of language, literature, and computers. According to Mahlberg (2012), the Corpus Stylistics examination of literary texts requires both human perception and a computerized data analysis method. Various analytical points for examining the literary texts are made possible by corpus stylistics analysis. Language analysis can be done on the text by using corpus software. Similar considerations can be made for both brief and lengthy suspensions, quotations and non-quotes, tagging, part-of-speech analysis, etc. in corpus stylistic analysis. It is crucial to note that the analytical theories and procedures that are already in use are neither rejected nor replaced by this developing field. With a focus on both language and literature, this study uses corpus-based stylistic analysis to showcase new findings on electronically accessible literary texts. The researcher emphasizes the method's objectivity and correctness in comparison to traditional literary analysis by using corpus tools for stylistic analysis. This method complements existing approaches by offering additional insights into lexical patterns in the novel and verifies manual interpretations, demonstrating the value of corpus tools in providing extensive and objective analysis. The research centers on the selection of lexicons and keywords, contrasting them with other 19th-century literary works. The main objective of the study is:

1. To investigate how the corpus techniques can derive dominating semantic areas from the literary texts. The study answers the following research question,

Q. How can corpus-stylistic analysis provide insights about the characters and themes in Hawthorne's *The House of the Seven Gables*, through keywords, KWIC, collocation and semantic domains?

This study is unique because it applies literary analysis to the comparatively underutilized corpus stylistic method. Through the use of corpus software to derive and examine keywords, the study emphasizes their significance for comprehending the themes and characters of the novel. By applying concordance lines to analyze these keywords, one can gain a greater knowledge of the text's patterns and further decipher the connotative meanings within it.

Synopsis of *The House of Seven Gables*

Hawthorne's work is set in an old mansion that represents the misdeeds and sorrows of the Pyncheon family's ancestors. The Pyncheon family is the focus of the narrative; they are cursed by Matthew Maule, a man who was falsely hanged for witchcraft by the patriarch of the family. The story follows the lives of several people who live in the mansion: their young, gregarious cousin Phoebe, who infuses the gloomy home with life and hope; their elderly, reclusive brother Clifford, who was recently freed from prison after being falsely convicted of murder; and Hepzibah, an elderly, lonely woman who struggles to maintain her dignity. The mysterious daguerreotypist Holgrave, who discloses the sinister past of the Pyncheon family and eventually falls in love with Phoebe, is a pivotal character in the novel. As the narrative progresses, mysteries are revealed, the real antagonist is revealed, and the curse is ultimately lifted, opening the door for the remaining Pyncheons to start afresh in hope. The novel offers a vivid, evocative portrayal of 19th-century New England life while delving into issues of guilt, revenge, and the unavoidable impact of the past on the present.

Literature Review

An effective method for analyzing literary texts is corpus stylistics, an interdisciplinary technique that combines corpus linguistics and literary analysis. Large electronic corpora and computer methods are used in this approach to find patterns and features in texts that might be missed by

more conventional literary analysis. Corpus stylistics, which emphasizes word and phrase frequency, distribution, and context, offers unbiased insights into the stylistic decisions, thematic concerns, of writers. Recent research has shown how this method might improve our comprehension of a range of literary works by providing new insights on both well-known and lesser-known pieces. As the field develops further, it promises to close the gap between qualitative literary interpretation and quantitative linguistic analysis, enhancing literary analysis in new manners. The Corpus Stylistics analysis method is used by Jaafar (2017) in "Corpus Stylistics Analysis of Thomas Harris" to assess keywords from the novel "The Silence of the Lamb." He also makes an effort to use "key semantic domain analysis" to ascertain the connection between the extracted terms. The keyword analysis identified the work's 'murder' theme. The most significant character in the book is likewise reflected in the quantity of Jaafar's occurrences. In a related investigation, Starcke (2009) examined the terms and phrases that are frequently used by various characters in Jane Austen's "Pride and Prejudice." To achieve her research objective, Starcke also employed the corpus stylistic analysis method. Analyzing terms and phrases facilitates the investigation of components that are frequently disregarded by critics who apply conventional methods to evaluate different literary genres.

Shakespeare's style in *Romeo and Juliet* is examined by Culpeper (2002) using corpus stylistic approaches with an emphasis on lexis. His main focus was on character depiction, so he used a method known as "style-markers," or "words whose occurrences vary substantially from their patterns in a normative corpus," to find them. An essay by Archer, Culpeper, and Rayson (2009) discusses Shakespeare's comedies and tragedies and offers semantic depth to the more popular keyword strategy. They say that by looking for keywords within of carefully chosen, pertinent semantic domains, they can identify concepts that a typical keyword search would not have brought up. Joseph Conrad's "Heart of Darkness" has several primary themes that are revealed by Stubbs's (2005) analysis of word and lexeme distribution. Stubbs contends that there are other factors to take into account besides word frequency. The underlying relationships between themes and the words that relate to them can be found using collocation analysis.

Theoretical Underpinnings

Stylistics is the systematic study of the literary value that is present in a particular text. The idea that is being discussed has its roots in the late 19th and early 20th century Russian formalist understanding of the meaning of literature. This approach sought, primarily through formal analysis, to identify the particular textual features that result in distinctive literary effects. Because of this, stylistics has devoted most of its history to studying style and, by extension, the meaning of literary works. The genesis of stylistics can be traced back to the initial enthusiasm that surrounded the possible contributions of linguistics to the field of literary study, as well as the introduction of core ideas from Russian formalism, such as the notion of "de-familiarization." Drawing on the work of literary stylisticians such as Roman Jakobson, the foregrounding hypothesis provides a theoretical framework for identifying distinctive language components that depart from the norm. In the context of corpus stylistics, foregrounding theory offers a conceptual framework for recognizing and evaluating language norm deviations, shedding light on authors' intentional stylistic choices. The lexical feature-focused analytical tool developed by Leech and Short presented in their book *Style in Fiction*, offers a thorough and systematic set of criteria for the analysis of lexical choices in a text. Among the many criteria covered by this comprehensive checklist are affective meaning, particularity, range, and frequency. We have an organized method for analysis that allows us to look more closely at the lexical elements of stylistic expression by incorporating this checklist into our process of analysis.

A word cloud generated with the Voyant tools from the novel THOSG is seen in the image. Each word's size and prominence in the word cloud indicate how frequently and critically it occurs in the text. Prominent terms like "house," "Phoebe," "little," and "said" are displayed prominently in the image along with a colorful selection of words that are presented at various sizes and angles. The primary concepts or points made in the literary work are summed up in these sentences. The linguistic style, vocabulary choice, and central themes of the works are contrasted with those of other 19th-century novels by keyword extraction.

The Keyword List of THOSG

The Antconc software's list of keywords is included in the table below, but it's important to remember that it only includes content words, function words are excluded. According to Culpeper (2009), keywords are "style markers" that show text-based stylistic patterns derived from particular data.

Table

keywords list generated for THOSG

Rank	Type	Freq_Tar	Freq_Ref	Keyness (Likelihood)	Keyness (Effect)
1.	Hepzibah	439	0	2081.344	0.008
2.	Phoebe	378	0	1791.937	0.007
3.	Pyncheon	364	0	1725.525	0.007
4.	Clifford	328	4	1512.248	0.006
5.	Judge	222	103	666.371	0.004
6.	Maule	106	0	502.252	0.002
7.	Old	331	630	454.858	0.006
8.	Cousin	107	23	390.167	0.002
9.	Gables	82	2	370.008	0.002
10.	Holgrave	77	0	364.824	0.001
11.	Shop	115	58	335.601	0.002
12.	House	245	533	296.375	0.005
13.	Little	261	607	294.617	0.005
14.	Alice	82	21	288.460	0.002
15.	Man	257	673	253.671	0.005
16.	Colonel	62	14	223.884	0.001
17.	Venner	46	0	217.935	0.001
18.	Window	99	104	208.209	0.002
19.	Carpenter	48	3	205.181	0.001
20.	Garden	75	58	184.542	0.001
21.	Matthew	47	9	175.066	0.001
22.	Gentleman	52	19	167.614	0.001
23.	Sunshine	48	14	163.926	0.001
24.	Jaffrey	33	0	156.341	0.001

Moreover, the log-likelihood (keyness) value is a factor that shapes this list. The significance of character names to the novel's storyline can be seen via an examination of the keyword list, which takes into account both their frequency in the target corpus and their keyness (log-likelihood) value. Hepzibah's character is the most prevalent with 439 appearances. The novel's keywords list highlights the names of the characters, emphasizing their significance to the plot. This realization emphasizes how crucial characterization is to the novel's plot. Moving ahead, we shall separate the keyword list into two categories.

Table 2

keyword categories of THOSG

Category	Keywords
Proper Noun:	(fictional characters and titles related keywords)
Characters' Names	Hephzibah, Phoebe, Pyncheon, Clifford, Judge, Maule, Holgrave, Alice, Colonel, Venner, Matthew, Jaffrey, Alice
Title	Judge, Colonel, Daguerreotypist, Wizard, Gentlewoman, Maiden, Miss, Lady, Kinsman, Ancestor, Maid, Guest, Esquire, Baron, Secretary, President, Governor
Contextual words:	(related to the aboutness of the text)
	Hereditary, Misfortune, Sorrow, Sympathy, Deed, Death, Ancestor, Substance, Charm, Eminent, Conscience, Wrought, Stern, Happiness, Purpose, Utmost, Venerable, Wretched

The table that is presented shows two distinct keyword groupings. Nouns in the first category are further divided into titles and names of fictitious characters inside the keywords list. Meanwhile, terms pertaining to the novel's setting are included in the second group. These keyword classifications within the novel are based on pertinent linkages that suggest a broad classification into "fictional world signals" and "thematic signals." phrases such as "fictional world" here refer to Hepzibah, Phoebe, Pyncheon, Clifford, Judge, Maule, Holgrave, and so on, whereas phrases such as "thematic signal" in the work refer to Hereditary, Misfortune, Sorrow, Sympathy, Deed, Death, Ancestor, and so on.

Hepzibah as node

Hepzibah Pyncheon is shown as leading a solitary and frugal life, heavily impacted by her love for her brother Clifford. She is also noted for her social distancing and emotional sensitivity. The descriptions of Hepzibah's garden, which yields exceptional qualities, highlight her isolation and

the particular attention she gives it. Her need to rent out a piece of the garden highlights her financial struggles, while the garden itself represents her refuge from society and the outside world. Her interactions seem awkward and unfavorable because of this necessity, as demonstrated by a cousin's shock to see that someone else is taking care of the garden. One of Hepzibah's main traits is her intense love for Clifford.

Hepzibah, a character in the story, is portrayed as a sensitive and anxious woman who is deeply attached to her cousin Clifford. Her lack of affection for children and unfulfilled dreams contribute to her complex personality. Her "unmalleable" mentality and aversion to change further exacerbate her isolation. The text uses a balance of function and content words to enhance readability and coherence. The title "Miss" highlights her solitary status, while the terms "cousin" and "poor" reflect her social status and financial struggles. The mention of Clifford often emphasizes his importance and closeness to Hepzibah, shaping her character as an unmarried, working-class woman. The text's use of collocates helps to enhance the narrative's coherence and readability.

Phoebe as node

Conversely, Phoebe Pyncheon is portrayed as innocent and pure, possessing a "transparent shadow" and a grin that embodies her unadulterated essence. Her "heart's language" reveals her warmth and emotional sensitivity in conversations, particularly with Clifford. Phoebe's charisma is enhanced by her beautiful voice and musical skills, which improve the mood. Her "spiritual quality" and lively demeanor stand in stark contrast to the surrounding darkness, demonstrating her beneficial influence. Phoebe's love of the outdoors and her involvement in the garden represent her connection to nature and her capacity to infuse her surroundings with vitality and joy. In her dealings with Chanticleer's second wife, Phoebe, a merchant with a keen sense of business, shows empathy and understanding. Her use of language to improve her living environment and employment demonstrates her optimism and inventiveness. Her active participation in talks, her connections to Clifford and Hepzibah, and her allusion to "Holgrave," which may or may not imply romantic relationship, all influence Phoebe's role in the narrative. The term "little" might also refer to her age or height, which could affect how other people view her and her character. In general, Phoebe's personality and social interactions influence her narrative.

Clifford as node

Clifford is said to possess a quiet and "unsatisfactory elegance," which suggests a sophisticated mind that could be challenging for others to understand. His brief "tremulous exhilaration" after being granted freedom is indicative of his nervous disposition. He is a sensitive and kind person who finds contentment in the little things in life, like spending time with hens. His philosophical convictions and intellectual interests give his character depth, and his yearning for the past is clear when he hears the harpsichord.

Clifford's story is characterized by his longing for independence and comfort, as well as his complex relationship with society. His interactions and decisions are shaped by his strong relationship with his cousin, Hepzibah, and his struggle with resources and social exclusion. The term "cousin" highlights his relationship with Hepzibah, which affects his standing within the Pyncheon family. Overall, Clifford's experiences highlight his unique experiences and the importance of his relationship with Hepzibah in shaping his character.

Holgrave as node

Another important character, Holgrave, is presented as an idealistic and hopeful person who envisions a "golden era." He is proud of his individuality and independence and does not take pride

in his ancestry. Holgrave's appreciation in gardening reflects a grounded side of himself and his connection to nature. His emotional dependence on Phoebe for happiness and his love interest in her give his character a more intimate side. In his chats, Holgrave demonstrates his analytical reasoning and perceptiveness, yet there's a hint of mystery and restraint that make you wonder if he's hiding anything.

Holgrave's persona is complex due to his skeptical attitude towards youth and philosophical outlook on life. Words like "Mr. Holgrave," "Holgrave's demeanor," and "Miss Hepzibah tells me" are used to describe his interactions and function in the story. Collocate analysis highlights Holgrave's engaged dialogue, formality, and respect in dealings with Phoebe. His relationship with Phoebe, referred to as "Phoebe," suggests a romantic relationship. These collocates show Holgrave as a person with a specific social status and engaged in dialogue.

Semantic domain analysis

The field range of the semantic field is the set of meanings that fall under this combination. Lexical fields, often referred to as semantic fields, are groups of related words that are bound together by a common notion (Zhou, 2001). The majority of the phrases in the work are classified under the "X" category, which deals with "psychological actions, states, and processes," therefore it's clear that the text explores the innermost feelings and ideas of the characters. Hawthorne skillfully explores repentance, atonement, and the fallout from previous actions. The main characters struggle with their psychological states, which are influenced by memories, wants, and guilt.

Conscience and guilt are two themes that are closely related to the "X" terms. A lot of characters suffer with guilt because of past wrongdoings or the misdeeds of their ancestors. Hepzibah Pyncheon, for instance, struggles with guilt due to her family's past and her own perceived inadequacies. The characters' consciences serve as a strong motivator for much of the narrative.

The novel examines **intellect and spirituality** on multiple occasions. While figures such as Judge Pyncheon strive for wealth and power, spiritual awareness is the path to ultimate atonement. A recurrent theme as the individuals negotiate their complicated lives is the pursuit of wisdom, which is shown as elusive but necessary.

The subject of **memories from the past** also comes up often. Characters are formed by familial influences and their upbringing; the home itself is a representation of the past, full of startling truths and sinister secrets. The Pyncheon Mansion's architecture mirrors the moral deterioration that exists within its walls, symbolizing the weight of shame and history. Numerous characters go out on journeys of **self-realization and atonement**. After being unfairly imprisoned for years, Clifford Pyncheon longs to leave the oppressive atmosphere of his house. Hepzibah too wants to be freed from her family's past and finds atonement. Their decisions and behaviors are influenced by sensibility and perception, which shapes interactions and relationships. Through the diverse perspectives of the individuals, the novel delves into topics of **appearance versus reality** and truth versus illusion. The heritage of the Pyncheon family is closely linked to prominence and reflects questions of social class obligations, lineage, and inheritance. The choices and consequences made by the heroes are greatly influenced by their legacy.

Table 3

Words and semantic domains

Semantic domain	Example
Names and grammar	Hepzibah, Pyncheon, phoebe, Maule, Holgrave, Jefferey...
Psychological actions, states and processes	Perception, spirituality, wisdom, reminiscences, conscience, quest, grasp, sensibility, eminent,
Architecture and kinds of houses and buildings	Sidewalk, chimney, shop, door, doorstep, neighborhood, roof, threshold, garden, build, interior, window, portal chamber...
Social actions, states and processes	Governor, keeper, artist, housewife, puritan, wizard, uncle, community, public, founder, generation, deportment, praiseworthy, witchcraft, mystic, ancestors, heredity, ancestral, colonel, powerful, tradition
Numbers and measurements	Seven, figure, depth, breath, deeper, vast, upward, miniature...
Life and living things	Death, dead, life, deceased, crow, hens, chicken, monkey, elephant, elm, flowers, garden, poises, tree...
Substances, material, objects and equipment	Substance, iron, gold, soil, commodity, sphere, hue, gubbins, ...
Language and communication	Responded, claim, croak,
Movement, location, travel and transport	Race, gush, impelled, drawn, upward, forth, aside...
The body and the individual	Visage, figure, heart, joints, breath, physical...
Arts and crafts	Artist, figure, picture, portrait, emblem...
Government and public	governor, representative, public, judge, colonel, puritan, political...
Emotions	Melancholy, pathos, airy, sunny, smile...
Food and farming	Chicken, gingerbread, pear, beans

The novel explores the Pyncheon family's ancestral home, focusing on its physical environment and architectural features. The house symbolizes the weight of the past, shame, and the history of the Pyncheon family. The house's secrets and moral deterioration are reflected in its ominous environment, with thresholds, gateways, and portals symbolizing limits and changes in status. Architectural aspects like the chimney and roof hint at the coziness of the home and concealed forces. The "S" category, which stands for "Social actions, states, and processes," highlights the social dynamics and interpersonal relationships in the characters. The dark past of the Pyncheon family affects later generations, and the novel's anguish is largely caused by conflicts between societal rules and human wants arising from social positions and expectations.

There are allusions to societal injustice and hierarchical organizations, along with important issues of **power and influence**. Tradition and change draw attention to the ways in which the past and present interact, and they also mirror broader changes in New England society during the 1800s.

Overall, these themes produce an intricately detailed story that explores the relationship between the past and the present, the individual and the group, and the visible and invisible forces that shape human fate. Reading Hawthorne encourages readers to consider ageless questions of identity, morality, and the lasting effects of the past because of his rich descriptions and nuanced characters. This study offers quantitative evidence to complement qualitative assessments by demonstrating how well corpus stylistics can be used to detect important themes in literary works.

Conclusion

The study discusses how Hawthorne's characters fit the lexicons precisely. Based on frequency, the software Antconc categorized the characters' meanings. In order to fulfill the suggested study goals, they were then chosen for further examination. Specifically, the novel's texts were examined from a literary standpoint using the keyword analysis, which also revealed connections between the keywords and the topics chosen for the novel. To do a comprehensive study of the chosen keywords, the KWIC analysis of every chosen node is also carried out. The KWIC analysis was performed on the nodes "Hepzibah," "Pheobe," "Clifford," and "Holgrave". This work is a parody the strict social norms, theological hypocrisy, and the Puritan way of life in 17th-century Puritan New England. Additionally, THOSG investigates how the Pyncheon family is impacted by Puritan values within a Puritanical context. In THOSG, Clifford Pyncheon's quest for spiritual and emotional liberation is examined, but individual salvation journeys are also examined. We also observed that the narrative employed a symbol of nature. The garden and other natural elements in THOSG convey concepts of optimism and rejuvenation, while the woodland serves as a symbol of freedom and a contrast to the oppressive civilization. When it comes to the characters, we have strong female characters like Hepzibah, a determined woman who must overcome social expectations and Clifford, a male character suffers from feelings of guilt and struggle with suppressed remorse, the consequences of their past actions, and their desire for atonement. Because of his upbringing, Clifford suffers from social disapproval and inherited guilt. faces the challenges of reintegrating into society and seeks out emotional and spiritual liberation Phoebe, a young woman from THOSG, is a positive influence she brings life and hope into the otherwise gloomy surroundings. Holgrave is a strange, reclusive man who is closely associated with the Pyncheon family. reveals hidden aspects of the Pyncheon family's history. In addition, he speaks with strength in every word that he says. The semantic domain analysis show That the themes of the novel are basically related to the psychological and emotional aspect like the theme of conscience and guilt, spirituality, past and memories, self-realization and atonement, appearance versus reality and power and influence are prevalent in the work of Hawthorne.

In summary, our study supports the notion that the past several years have truly marked the dawn of computer technology. In light of this, it is possible to consider corpus stylistic analysis as a useful methodology for analyzing literary text features that are resistant to manual or traditional text analysis techniques. However, the idea that computerized text analysis is incompatible with traditional analytical techniques is incorrect. It's just a modern way of finding the novelties in both non-literary and literary texts. Similarly in future, any text by Hawthorne can be analysed in the context of more recent writings in order to highlight the differences between language used at the time and language used now. It is possible to look at characters from different sections of Hawthorne's works to find the relationships between different characters.

References

1. Archer, D., Culpeper, J. & Rayson, P., (2009). Love - "a familiar or a devil"? An Exploration of Key Domains in Shakespeare's Comedies and Tragedies. In: D. Archer, ed. What's in a Word list? (pp. 137-157). Farnham: Ashgate.

2. Culpeper, J. (2002). Computers, language, and characterization: an analysis of six characters in *Romeo and Juliet*. <https://doi.org/10.1177/0963947005048873>
3. Culpeper, J., 2009. Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics*, 14(1), pp. 29-59.
4. Jaafar, E. A. (2017). Corpus Stylistic Analysis of Thomas Harris' *The Silence of the Lambs*.
5. Leech, G. N., & Mick S. (2007). *Style in Fiction: a Linguistic Introduction to English Fictional Prose* (2nd ed.). Pearson Longman.
6. Mahlberg, M. (2012). The corpus stylistic analysis of fiction—or the fiction of Corpus Stylistics? In *Corpus Linguistics and Variation in English* (pp. 77-95). Brill Rodopi.
7. Mahlberg, M., & McIntyre, D. (2011). A case for corpus stylistics. *English Text Construction*, 4(2), 204–227. <https://doi.org/10.1075/etc.4.2.03mah>
8. Stubbs, M. (2005). Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature*, 14(1), 5–24.
9. Zhou, Weijie. (1997). A Research on English Semantic Field. *Shandong Foreign Languages Journal*, 68, 21-23.