

EXISTENCE, EXPECTATION, AND INTERFERENCE: A CORPUS STYLISTICS STUDY OF WORD CLUSTERS IN JANE AUSTEN'S WORKS

Suhandoko

UIN Sunan Ampel, Surabaya, Indonesia
Corresponding email: suhandoko@uinsa.ac.id

ABSTRACT - The study examines the thematic and stylistic elements in Jane Austen's novels by examining the word clusters. The computational analysis using AntConc concordance software focuses on finding the recurring bi-grams and tri-grams that potentially explain Austen's narrative style. The finding of the study suggests that word clusters like "to be", "it would be," and "could not be" are key features in Austen's narrative technique to portray the existence, expectations, and interference women encountered within patriarchal society during the era. This corpus stylistics approach helps confirm the significance of Austen's writings in understanding the societal construct in the era and provides quantitative evidence to support traditional literary analysis.

Keywords: Corpus Stylistics, Jane Austen, Thematic Analysis, Word Clusters

ABSTRAK – Penelitian ini meneliti elemen tematik dan stilistik dalam karya-karya Jane Austen dengan mengkaji kelompok kata. Analisis komputasi menggunakan piranti lunak konkordansi AntConc berfokus pada pencarian bi-gram dan tri-gram yang berulang yang berpotensi menjelaskan gaya naratif Austen. Penelitian menemukan bahwa kelompok kata seperti "to be", "it would be", dan "could not be" adalah fitur utama dalam teknik Austen untuk menggambarkan eksistensi, ekspektasi, dan interferensi yang dihadapi wanita dalam masyarakat patriarki pada era tersebut. Pendekatan stilistika korpus ini tidak hanya membantu mengonfirmasi signifikansi tulisan Austen dalam memahami konstruksi sosial pada zamannya, tetapi juga menyediakan bukti kuantitatif untuk mendukung analisis sastra tradisional.

Kata Kunci: Stilistika Korpus, Jane Austen, Analisis Tematik, Kelompok Kata

INTRODUCTION

The recent adoption of computational models in literary analysis marks a significant advancement in the field. Such an approach may enhance understanding of text complexities and authors' styles. These tools also allow researchers to analyze large quantities of text to reveal patterns that are hard to notice in traditional readings. Moretti (2013) introduced the idea of "distant reading" to refer to the use of computational analysis to help researchers explore large volumes of literature to identify literary trends and developments. Jockers (2013) expanded on this by using computational techniques to quantify themes and emotional development in literature. Thus, this technique provides a more objective basis for what were once largely subjective evaluations. These methods are also crucial in revisiting canonical works by offering new insights into narrative techniques and character development.

The integration of computational methods into literary analysis has expanded the understanding of literary texts, particularly in exploring canonical authors whose works have been extensively researched through traditional readings. Corpus stylistics, a method that utilizes large digital collections of texts (corpora) to analyze language patterns, can detect subtle stylistic features and themes across large bodies of work. Hoover (2007) applied corpus stylistic techniques to distinguish between the writing styles of different authors. He has effectively identified a unique authorial narrative style based on word frequency and variance. Such studies challenge the traditional literary interpretation by providing empirical evidence to support or refute long-held assumptions about authorship. Similarly, researchers like Underwood (2019) have employed these methods to identify the changes in literary genres. He found that societal shifts may shape literary themes and word choice.

In the context of canonical authors, Burrows (2002) applied principal component analysis to Jane Austen's works. He demonstrates how Austen meticulously picked and chose words and sentence structure to reflect style maturation previously acknowledged by literary critics but never quantified to this extent. The use of corpus stylistics in studies on Shakespeare has provided insights into how Shakespeare uses functional words and phrase structures to distinguish his works and emphasize his authorship (Demmen, 2012; Hung et al., 2020; Murphy, 2006). Similarly, the work of Oakes and Ji (2012) applied these techniques to modern literature and found the evolution of linguistic complexity that characterizes the movement of the era. They also offered quantitative support to traditional qualitative assessments.

Jane Austen's novels are well-known for their complex narrative styles and diverse themes. Scholars have garnered significant attention from scholars to approach them with creative and innovative analysis methods. Barchas (2012) explores how Austen enriches the realism and credibility of her narratives by incorporating actual geographical and historical references. This strategy not only grounds her stories in reality but also actively engages readers, making them feel part of Austen's world. Similarly, Flavin (2004) examines Austen's use of free indirect discourse, revealing how it enhances the emotional and psychological complexity of her

characters. Moreover, thematic studies, such as Johnson's (2014) analysis of *Jane Austen's Cults and Cultures*, investigate the integration of religion, ethics, and morality into the fabric of Austen's social critiques. Wiltshire (1992) highlights how themes of health, illness, and bodily conditions are woven into character development and plot progression. From a feminist and gender perspective, Johnson (1995) critiques the constraints of Austen's patriarchal society on women, while Kaplan (2005) discusses how Austen's interactions with other women not only influenced her narratives but also positioned her as a progressive advocate for female empowerment and solidarity.

Despite the rich body of scholarship on Austen's narrative techniques and thematic explorations, there remains a significant gap in the quantitative analysis of her textual structures. Jane Austen's careful use of third-person perspective and well-crafted dialogues between characters shape unique patterns and make them amenable for computational analysis (Rybicki et al., 2016). This study aims to fill this gap by providing empirical data on Austen's use of language, which can ground subjective interpretations in measurable evidence and potentially challenge or expand existing literary interpretations. Such analysis not only highlights Austen's skill in depicting complex emotional states and social dynamics through minimalistic language but also shows how these thematic elements are woven consistently across her novels (Rybicki et al., 2016).

LITERATURE REVIEW

Corpus stylistics, a methodological approach within the broader field of corpus linguistics, focuses on the systematic study of literary texts through large digital text collections or corpora. This approach applies quantitative analysis techniques to identify patterns of language use, stylistic features, and thematic elements across extensive data sets, offering insights often imperceptible through traditional literary analysis. The utility of corpus stylistics lies in its ability to provide empirical evidence to support interpretations of texts, facilitating a deeper understanding of an author's style and thematic consistency (Stubbs, 2005). This approach allows scholars to quantitatively analyze linguistic features that characterize an author's style, contributing significantly to the field of digital humanities. One work in this area by McEnery and Hardie (2012) details how corpus stylistics bridges the gap between linguistic analysis and literary criticism, enabling the study of how language in use shapes meaning and affects the reader's perception.

Specific studies have applied corpus stylistics to various literary texts to uncover deeper narrative and thematic patterns. For instance, Stubbs (2005) used corpus techniques to examine key themes in a large corpus of British novels, revealing underlying patterns of word usage that reflect broader cultural and societal norms. In another study, Fischer-Starcke (2009) applied detailed stylistic analysis to Jane Austen's texts, showing how certain lexical bundles could reveal nuances in character development and social interaction. Corpus stylistics has also proven instrumental in comparative literature studies. O'Halloran (2007) successfully applied

these methods to compare the narrative techniques used by different authors within the same genre, providing a statistical basis for stylistic similarities and differences. Mahlberg (2013) further extended this by exploring themes and absences in Dickens's novels, demonstrating how repeated word clusters across a literary corpus could signal thematic significance. Further, the application of corpus stylistics in studying character speech and development was exemplified by McIntyre and Walker (2019), who analyzed speech and thought presentation in a corpus of modern English novels to show how different narrative strategies affect reader engagement and character perception.

Word clusters, also known as n-grams or lexical bundles, are sequences of contiguous words used consistently across texts and can reveal significant insights when applied to literary analysis. The theoretical basis for using word clusters lies in corpus linguistics, where they are analyzed to understand syntax, semantics, and stylistic functions within a corpus. Biber et al. (2021) discussed how word clusters could reveal patterns of habitual language use in spoken and written communication, providing a foundation for their application in literary contexts.

In literary analysis, word clusters help illuminate an author's stylistic signatures and thematic preoccupations. Sinclair (2004) suggested that such clusters are not just random aggregations but are selections that reflect the writer's intentions and thematic focuses. By examining these clusters, researchers can discern the underlying structures that shape narrative techniques and influence thematic development. For instance, Scott (2008) demonstrated how word clusters could be used to analyze thematic coherence across a literary text, suggesting how repeated word combinations align with recurring themes.

The potential for discovering new insights into Jane Austen's work through analyzing word clusters is significant. Austen's precision in language choice suggests that her texts are likely to exhibit distinctive clustering patterns that serve specific narrative functions. Mahlberg (2013), in her study on Dickens, showed how an analysis of word clusters could identify characteristic narrative voices and thematic elements unique to an author. Moreover, analyzing word clusters can lead to deeper insights into Austen's thematic structures. Hoover (2007) used cluster analysis to distinguish personal and emotional subtexts in literary texts, offering a method that could similarly illuminate the subtleties of Austen's depictions of social interactions and moral critiques. By identifying and analyzing the frequency and distribution of specific word clusters in Austen's novels, researchers can better understand how she constructs her narratives and themes, potentially challenging traditional interpretations and offering new perspectives on her works.

METHODOLOGY

The corpus for this study comprises all six major novels by Jane Austen: "Sense and Sensibility," "Pride and Prejudice," "Mansfield Park," "Emma," "Northanger Abbey," and "Persuasion." These texts are sourced from Project Gutenberg to ensure they are free from

copyright restrictions. Prior to analysis, the texts underwent preprocessing to remove any extraneous content such as headers, footers, and editorial notes to ensure that the analysis focuses solely on the original words of Jane Austen.

The analysis begins by loading the cleaned texts into AntConc to create a comprehensive corpus of Austen’s novels. The configuration of AntConc is set to identify n-grams ranging from two to three words, with a frequency threshold adjusted to the corpus size to ensure the statistical significance of the results. This tool scans the corpus, identifying recurring word clusters and ranking them by frequency to determine their prominence in the texts.

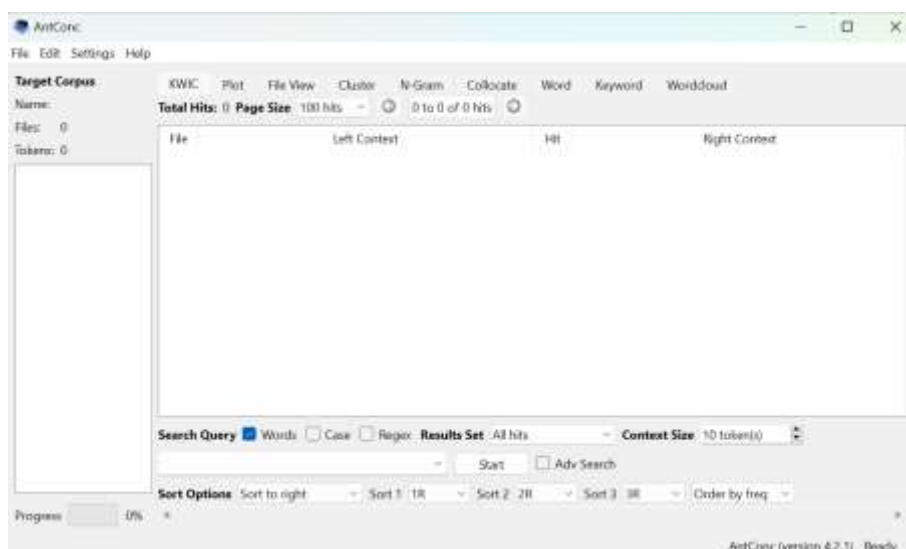


Figure 1. A Screenshot of the AntConc Interface

The methodology section begins with the cluster identification process, utilizing AntConc to conduct a detailed cluster analysis across the complete corpus of Jane Austen’s novels. This step is essential as it systematically identifies the most frequently occurring bi-grams and tri-grams within the texts. By pinpointing these recurring phrases, the analysis aims to uncover patterns in Austen’s language use that are significant for interpreting her stylistic and thematic choices. The selection of bi-grams and tri-grams, which consist of two and three contiguous words, respectively, helps reveal underlying language structures that Austen consistently employs across her literary works.

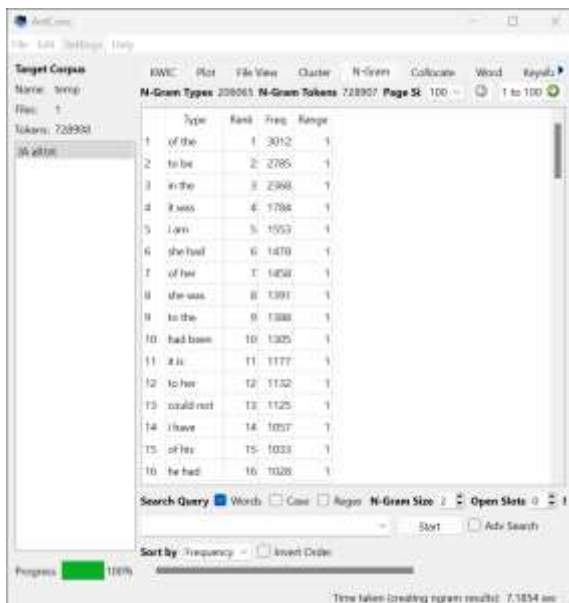


Figure 2. A Screenshot of the Result of Bi-Grams (Two-Word Cluster) in AntConc

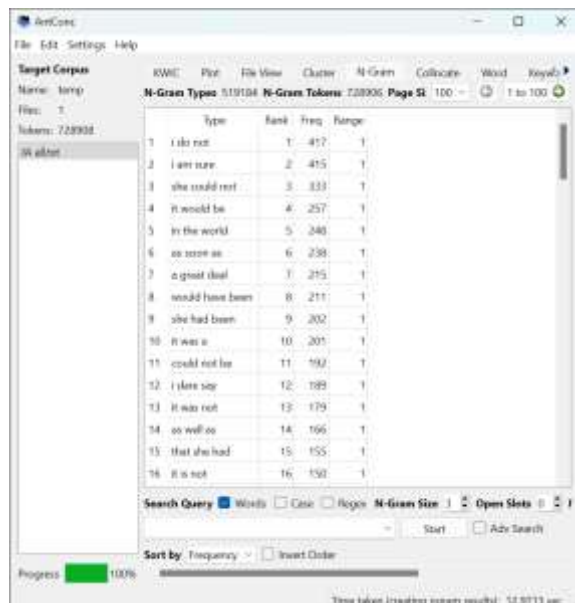


Figure 3. A Screenshot of the Result of Tri-Grams (Three-Word Cluster) in AntConc

Following the identification of these word clusters, the next step is context analysis. Each cluster identified in the previous step is subjected to a concordance analysis using AntConc. This involves examining the clusters within their broader textual environments, providing a richer context for each occurrence. The concordance tool in AntConc allows us to see where and how these word clusters appear within the novels, making it possible to assess their contribution to key thematic elements and the narrative style. This detailed examination helps understand these clusters' role in crafting narratives, developing characters, or enhancing thematic depth, thus offering insights into Austen's literary techniques and her approach to building complex literary contexts.

Finally, the frequency and distribution analysis focuses on quantifying how often and in what distribution these word clusters appear across Austen's novels. This statistical analysis is crucial for determining the prevalence of specific clusters and identifying if certain phrases are particularly favored in specific thematic or narrative contexts. For instance, some clusters might be more predominant in romantic scenes or descriptions of social settings, indicating a stylistic preference or thematic focus. This step not only aids in mapping out the stylistic signatures of Austen but also in highlighting potential variations in language use among different novels, which could indicate evolving themes or shifts in narrative focus throughout her writing career.

Table 1 Top 20 Two-Word Clusters in Jane Austen's Work

Emma		Mansfield Park		Northanger Abbey		Persuasion		Pride and Prejudice		Sense and sensibility	
Type	Freq	Type	Freq	Type	Freq	Type	Freq	Type	Freq	Type	Freq
to be	608	of the	748	of the	379	of the	429	of the	460	to be	436
of the	566	to be	645	to be	280	to be	378	to be	438	of the	430
it was	449	in the	579	in the	277	in the	323	in the	384	in the	359
in the	446	it was	412	i am	190	had been	255	i am	305	it was	280
i am	395	she had	374	of her	179	she had	227	mr darcy	273	of her	274
she had	334	of her	346	it was	169	it was	220	of her	266	to the	242
she was	331	she was	334	she had	154	captain wentworth	196	it was	254	mrs jennings	234
had been	308	to the	330	it is	153	he had	190	to the	252	to her	229
it is	301	sir thomas	328	to the	150	to the	176	of his	234	i am	224
mr knightley	299	i am	316	she was	141	mr elliot	174	she was	212	she was	209
i have	283	to her	312	of a	127	she was	164	she had	205	of his	204
could not	278	had been	311	to her	121	could not	148	had been	204	i have	192
of her	265	could not	266	and the	112	lady russell	147	it is	193	it is	184
mrs weston	256	miss crawford	266	the general	109	he was	134	i have	186	she had	184
have been	243	it is	258	do not	105	sir walter	131	to her	181	could not	163
he had	242	he had	254	had been	103	of her	128	that he	178	at the	160
to the	238	and the	249	in a	102	all the	127	could not	169	have been	158
do not	236	have been	232	could not	101	i have	125	he had	165	on the	158
mr elton	229	of a	229	for the	101	i am	123	for the	160	in a	155
and the	224	he was	216	such a	97	have been	117	he was	159	and the	153

Table 2 Top 20 Three-Word Clusters in Jane Austen's Work

Emma		Mansfield Park		Northanger Abbey		Persuasion		Pride and Prejudice		Sense and sensibility	
Type	Freq	Type	Freq	Type	Freq	Type	Freq	Type	Freq	Type	Freq
i do not	136	i do not	99	i am sure	56	she could not	54	i am sure	62	i am sure	72
i am sure	109	she could not	88	i do not	46	she had been	36	i do not	62	as soon as	60
she could not	73	i am sure	81	in the world	41	i am sure	35	as soon as	56	in the world	58
a great deal	64	it would be	66	she could not	28	a great deal	34	she could not	51	i do not	46
it would be	63	she had been	62	it would be	27	he had been	34	that he had	37	could not be	42
would have been	60	it was a	59	i dare say	26	it would be	32	in the world	34	she could not	39
do not know	55	would have been	54	i would not	26	could not be	29	it would be	33	her sister s	37
it was not	55	miss crawford s	51	the general s	24	lady russell s	29	i am not	32	it would be	36
it was a	53	as well as	50	as soon as	23	i do not	28	it was not	30	would have been	36
she had been	53	a great deal	49	the pump room	23	it was a	28	mr darcy s	30	in spite of	35
i dare say	50	in the world	45	it was not	21	mr elliot s	27	that he was	30	mrs jennings s	35
mr frank churchill	50	as soon as	43	a great deal	20	captain wentworth s	26	could not be	29	i dare say	34
in the world	49	that she had	42	one of the	18	as soon as	23	i dare say	29	to be sure	31
i assure you	47	sir thomas s	41	out of the	17	had not been	23	of mr darcy	28	it was not	29
could not be	45	i dare say	40	part of the	17	and captain wentworth	22	that it was	28	that she had	27
it will be	42	it is not	39	the rest of	17	he did not	21	as well as	27	the rest of	27
and mrs weston	40	there was no	39	which she had	17	in the world	21	on the subject	27	at the park	26
it is not	39	and she was	37	and mrs allen	16	that he had	21	would have been	27	it was a	26
mr elton s	39	mr crawford s	37	as well as	16	the miss musgroves	21	by no means	26	it is not	25
she did not	39	not to be	37	could not be	16	there had been	21	that she had	26	mrs john dashwood	25

RESULTS AND DISCUSSION

“To be” existing and recognized

Our analysis of the two-word cluster in Jane Austen’s works found that “to be” is the most frequently occurring word in all the works. At its core, “to be” is inherently existential. It is a fundamental verb to discuss identity, status, and existence. In the context of Austen’s novels, it could define or question a character’s place in society, relationships, or sense of self. Syntactically, “to be” is used in several ways: as an auxiliary verb in passive constructions, linking verb that connects subjects to their descriptions, and as a base for expressing hypothetical conditions and future plans.

As an auxiliary verb in passive constructions, “to be” emphasizes the action being done to the subject rather than by the subject. This syntactic use subtly shifts the focus from the doer of the action to the receiver, as seen in the following. This shift is especially unique as it often highlights the roles and identities of characters within their social settings.

Mansfield Park	as everybody was ready, there was nothing act of kindness which Fanny felt at her heart.	to be to be	done but for Mrs. Grant to alight and the spared from her aunt Norris’s
Sense and Sensibility	now urged of the allowances which ought I supply it.” Elinor had not needed this	to be to be	made for him, and it is my wish to be assured of the injustice to which her sister
Northanger Abbey	sudden removal of her own. But as it was not a variety of amusements, a variety of things	to be to be	supposed that Captain Tilney, whenever seen and done all day long, which I can
Persuasion	as Charles, for they want me excessively its environs—for young men are, sometimes	to be to be	acquainted with Captain Wentworth, and I met with, strolling about near home—was their
Emma	into love; but, alas! she was not so easily be lost to them for Harriet’s sake; if he were	to be to be	talked out of it. The charm of an object to thought of hereafter, as finding in Harriet’s
Pride and Prejudice	Mr. Darcy, with grave propriety, requested to the pleasures of youth as the first. And	to be to be	allowed the honour of her hand, but in vain. kept back on _such_ a motive! I think it would

As a linking verb that connects subjects to their descriptions, “to be” connects the subjects or characters with descriptions that define who they are or how others see them. This usage of “to be” shapes each character’s identity and place in society by attaching qualities, roles, or societal expectations to them.

Mansfield Park	And Sir Thomas’s wishing just at first here he had been acting as he ought to do.	to be To be	only with his family, is so very natural, that she the friend of the poor and the oppressed!
Sense and sensibility	eyes full of meaning, “there seemed to me and always openly acknowledged,	to be to be	a coldness and displeasure in your manner that a favourite child. They settled in town, received
Northanger Abbey	for _you_ are just the kind of girl no children. But when a young lady is	to be to be	a great favourite with the men.” a heroine, the perverseness of forty
Persuasion	you are married, that’s all. Let him know me again, meaning, for the rest of my life,	to be to be	a friend of yours, and then he will think little of only yours truly, “WM. ELLIOT.” Such a letter
Emma	Mr. Martin again without perceiving him return the compliment, and discover her	to be to be	a very inferior creature—and rather wondering a lady. I could not have believed it! And to
Pride and Prejudice	that her charming daughter seemed born her and liked her, and pronounced her	to be to be	a duchess; and that the most elevated rank, a sweet girl, and one whom they should not

As a base, “to be” serves to express hypothetical conditions and future intents. When “to be” is used to express conditions, it often frames a character’s existence in terms of possibilities and potential. This hypothetical use reflects on what could be, influencing how characters see themselves and their potential roles in the world. When “to be” is used to describe future plans in Jane Austen’s novels, it clearly outlines what a character is expected to do or become. This usage does not just predict what will happen; it also prescribes what should happen (according to societal norms or personal ambitions).

Mansfield Park.	for writing.” “No, I dare say, nor if he were arrive here till Tuesday evening.” “And it is	to be to be	gone a twelvemonth, would you ever write to a two months’ visit, is not?” “Yes. My uncle talked
Sense and sensibility	well he may. I should not wonder, if he was “Oh, ho!—I understand you. Mr. Ferrars is	to be to be	in the greatest passion!—and Mr. Donovan thinks the man. Well, so much the better for him. Ay, to
Northanger Abbey	attention, returned it saying, “Well, if it is was convinced of his father’s believing it	to be to be	so, I can only say that I am sorry for it. Frederick an advantageous connection, it was not till the
Persuasion	good woman, and if her second object was four horses, and whether they were likely	to be to be	sensible and well-judging, her first was to see situated in such a part of Bath as it might suit Miss
Emma	society all that he wanted; if Harriet were I see Jane every day:—but as you like. It is	to be to be	the chosen, the first, the dearest, the friend, the a morning scheme, you know, Knightley; quite a
Pride and Prejudice	her displeasure; and if you find it likely of speaking, “whether he is likely	to be to be	raised by your coming to us again, which I should in this country much longer.” “I do not at all know;

The analysis highlights that the frequent use of “to be” in Jane Austen’s works serves a purpose beyond sentence structure. Jane Austen’s use of the verb “to be” in her novels serves as a subtle yet powerful tool to critique and illuminate the societal constraints faced by women in her time. Particularly through its application as an auxiliary verb in passive constructions, Austen showcases how societal actions and decisions often act upon women rather than originate from them. This grammatical choice is not merely a stylistic one but a deliberate narrative technique that highlights the passive societal roles women are relegated to. For instance, phrases like “to be invited” or “to be seen” underline that women’s social participation and visibility are not autonomous but rather dictated by others’ permissions and judgments. This portrayal effectively critiques the limited agency granted to women, suggesting that their roles and identities are often defined by external forces rather than by their own choices or desires.

Jane Austen uses the verb “to be” as a linking verb in her novels to deepen her exploration of gender roles, focusing on the attributes that define her female characters’ identities within society. This approach is crucial as it brings these characters into the spotlight, urging readers to see them as individuals with unique qualities and roles. When Austen describes a character as “is to be a great favourite,” she emphasizes the need to appreciate women for more than just their beauty or suitability for marriage. Using “to be” in this way confirms the presence of these qualities in her characters and challenges readers to recognize and value the deeper contributions women make to the social fabric. This linguistic technique affirms women’s identities and prompts a reevaluation of their roles in society, encouraging a more profound appreciation of their personal attributes and societal contributions.

Jane Austen also uses “to be” to explore hypothetical situations and future possibilities, which helps open discussions about potential changes in society and personal dreams. She uses phrases like “are to be” and “could be” to think about a future where women have more control over their lives and more influence in their communities. This way of using “to be” does not just accept what the future holds as a given; instead, it creates room for questioning and hoping for different outcomes. When Austen’s characters think about their futures, she challenges the restrictive roles society expects women to fill and hints at the possibility of a wider range of life experiences and choices. This invites readers to consider how society limits women and to imagine what it would be like if those limits were removed.

The frequent use of “to be” in Austen’s works reflects her narrative style in constructing identities and exploring existential issues. The corpus analysis demonstrates that “to be” is mainly used in three ways: as an auxiliary verb in passive constructions, as a linking verb, and as a base. Austen often uses “to be” in passive constructions. This style seems to mirror the passive roles typically assigned to women during the Regency period. This usage emphasizes actions performed by the characters rather than actions they initiate themselves. This particularly affects female characters depicted as passive and dependent on men (Tannen, 1996). As a linking verb, “to be” is frequently followed by the description. This word serves to connect characters with societal expectations of how women should be described and defined. This suggests that Austen uses “to be” particularly to criticize the societal norms in her era. Furthermore, “to be” creates hypothetical scenarios where characters are described to think about “what could be.” Furthermore, “to be” creates hypothetical scenarios where characters are described to think about “what could be.” This feature in Austen’s works frames and explores the possibilities for women within the limitations of the existing social order. This approach enables Austen to challenge the societal expectations in her era subtly and opens a discussion of potential changes where these limitations are loosened or removed.

“It would be” possible to happen?

In Jane Austen’s works, the phrase “it would be” is frequently found across all her novels, appearing with varying frequency. In the novel “Emma,” the phrase appears 63 times. In “Mansfield Park,” it appears 66 times. In “Northanger Abbey,” it appears 27 times. In “Persuasion,” it appears 32 times. In “Pride and Prejudice” it appears 33 times, and in “Sense and Sensibility” it appears 36 times. The phrase “it would be” helps characters consider “what if” scenarios. This means the characters can discuss potential outcomes based on their decisions without the narrative committing to a specific direction. It allows characters to explore different possibilities and think deeply about their choices.

Syntactically, “it would be” utilizes the modal verb “would,” which is closely related to the future-tense verb “will” but is used for a different purpose. While “will” often suggests future actions that are likely to happen, “would” is used to explore possibilities or hypothetical scenarios contingent on certain conditions and not guaranteed. In this construction, “would”

partners with the verb “be” to establish a speculative or conjectural mood rather than stating something definitive. For example, when someone says, “it would be very wrong,” the adjective “difficult” describes a possible scenario. The use of “would be” creates an atmosphere of speculation and conjecture rather than stating something definitive. This verb structure links the subject, often “it,” to a description or potential state. For example, when someone says “it would be difficult,” the adjective “difficult” describes a possible scenario. This type of phrasing adds depth to the narrative by allowing characters and readers to consider various outcomes and dilemmas.

Mansfield Park.	he should ever get well of his gouty complaints, you. What is there to surprise you in it?” “I think	it would be it would be	a different matter: she should then very wrong. In a general light, private
Sense and sensibility	ength Lucy exclaimed with a deep sigh, “I believe he should have a numerous family, for instance,	it would be it would be	the wisest way to put an end to the a very convenient addition.” “To be
Northanger Abbey	ay now; for, as I am known to be in the country, a large acquaintance here.” “I wish we had any;	it would be it would be	taken exceedingly amiss; and it is a somebody to go to.” “Very true, my
Persuasion	not the trustiest, steadiest creature in the world, Even if he did not come to Camden Place himself,	it would be it would be	enough to spoil her; for she tells me, in her power to send an intelligible
Emma	women to marry. Were I to fall in love, indeed, and connexion were in her favour.—For Harriet,	it would be it would be	a different thing! but I never have advantageous and delightful indeed.
Pride and Prejudice	ecause if it were to get round to the lady’s family have a dangerous fit of illness if she should die	it would be it would be	an unpleasant thing.” “You may a comfort to know that it was all in

In Jane Austen’s novels, this phrase establishes a connection between the subject, typically an impersonal “it,” and the complement, which describes a potential state or outcome. This could be an adjective, a noun, or a more complex phrase. “Mansfield Park” talks about “it would be a different matter.” Those words indicate that things could change depending on better health. The phrasing hints something may happen to shift the current circumstances. Similarly, “it would be very wrong” introduces moral judgment, indicating the character’s belief in the potential wrongness of an action, thereby setting an expectation of ethical standards within the story. The phrase “it would be” in Jane Austen’s books is a clever way to talk about things that could happen or what people expect to happen. It lets her show what is going on inside her characters’ minds, such as hopes, worries, and the rules of society they have to deal with.

Mansfield Park.	likely; and I thought so too. We both thought If I could suppose my aunt really to care for me,	it would be it would be	a comfort to you. But if you do not consequence to anybody. _Here_, I
Sense and sensibility	her answer were sent. Elinor had always thought a mother, the question could not give offence.	it would be It would be	more prudent for them to settle at the natural result of your affection
Northanger Abbey	such a companion and friend as Isabella Thorpe, ria is gone. She was quite wild to go. She thought	it would be it would be	impossible for you to be otherwise; something very fine. I cannot say I
Persuasion	not the trustiest, steadiest creature in the world, better go back and change the box for Tuesday.	it would be It would be	enough to spoil her; for she tells me, a pity to be divided, and we should
Emma	that many of them wanted a better. Emma said “If he would be so good as to read to them,	it would be it would be	awkward; Mrs. Weston was in a kindness indeed! It would amuse
Pride and Prejudice	be much in company. But really, ma’am, I think and uncommon I hope it will continue, for	it would be it would be	very hard upon younger sisters that a great loss to me to have many

The prevalence of the phrase “it would be” in Jane Austen’s works subtly implies uncertainty and emotional vulnerability in a refined and controlled manner. This phrase begins with the impersonal pronoun “it,” which does not refer to any specific object, followed by “would,” a

form of the past tense of “will” used here to express speculation or possibility, not certainty. The word “be” is a linking verb that connects the subject with a hypothesized trait or state. For example, when a character says, “it would be a comfort if” in the context of Austen’s work, they are imagining a pleasant possibility, but without any certainty that it will occur. This indicates a desire or hope rather than a definite prediction. On the other hand, the phrase “it would be unpleasant if...” expresses concern about the possibility of an adverse outcome. This does not indicate outright pessimism but rather a consideration of potential negative consequences that one wishes to avoid.

By using “it would be,” Austen’s characters reflect on potential future scenarios or different outcomes indefinitely. This mode of expression suits the restrained and polite society of her novels, where direct statements might be socially unacceptable or too forward. It allows characters to explore their thoughts and feelings without committing to a particular stance or outcome, which adds a layer of realism to their interactions. Furthermore, Austen’s use of the phrase “it would be” also reflects her clever comment on the social and cultural rules of her time. This phrase often highlights the restrictions faced by people, especially women, who were not expected to express their complaints openly. By using this hypothetical language, Austen’s characters can safely share their true feelings or criticize society in a way that was acceptable during that period. This method not only fits within the social norms but also encourages readers to think about these restrictions and what they mean for the characters.

The phrase “it would be” is also a key tool throughout Jane Austen’s novels. It enables characters to explore potential outcomes and think about right and wrong. The use of this feature reflects the personal challenges and societal pressure exposed to the characters and demonstrates tentativeness. Fischer-Starcke (2009) argues how such hedging presents “statements [that] do not seem to be desired in the characters’ communicative behaviours” (p. 513). This reflects Austen’s voice in examining the conditional moods and subjunctive wishes of her characters and emphasizing their constrained agency within highly structured social environments. This modal verb may also contain evaluative meaning where speakers present their point of view (Makhloof, 2020). The clever employment of this phrase makes Austen’s works more interesting because it allows characters to engage in conjectural conversations and express their view in a subtle way. This subtlety is crucial, considering that Austen’s works are made during a time when society is governed mostly by a patriarchal system that may inhibit women from expressing their thoughts and desires freely.

“Could not be” achieving

Jane Austen’s prevalent use of the phrase “could not be” in her works helps highlight the things that are impossible for the characters, whether due to personal limits or the strict norms of the society they live in. The phrase “could not be” involves a modal verb paired with negation. “Could” is the past tense of “can” indicating ability or possibility. When combined with “not,” it changes into a denial of this ability or possibility. In other words, the phrase suggests that

something is an impossibility or prohibition of something. The main verb “be” that follows the modal serves as a linking element to complement that defines the state of condition that is considered impossible. This construction effectively portrays a sense of limitation within the narrative.

Mansfield Park	Thomas, getting up, said that he found that he not have had to follow her if she had staid.” This	could not be could not be	any longer in the house without just denied, and Fanny was silenced.
Sense and sensibility	the probability of an entire recovery. Elinor valley, and chiefly in silence, for Marianne’s mind	could not be could not be	cheerful. Her joy was of a different controlled, and Elinor, satisfied with
Northanger Abbey	sense added some bitter emotions of shame. She be at, was beyond her comprehension. Isabella	could not be could not be	mistaken as to the room; but how aware of the pain she was inflicting;
Persuasion	Russell in a discussion of her merits; and Anne preventing him. This was almost cruel. But she	could not be could not be	given to understand so much by her long ungrateful; he was sacrificing an
Emma	to refrain from making any answer at all. She aware, it was not worth bringing forward again it	could not be could not be	complying, she dreaded being done without a reproof to him, which
Pride and Prejudice	were of everything concerning either. Detection must be decisive, and whose behaviour at least	could not be could not be	in your power, and suspicion certainly mistaken for the affectation and

Austen frequently uses “could not be” to portray the limitations in which her characters must operate, whether they are related to emotions, societal norms, or moral principles. For example, in *Sense and Sensibility*, Elinor’s happiness is described as “could not be cheerful,” reflecting her inner turmoil and her struggle to maintain her effort to look happy. This limitation demonstrates how the characters encounter challenges in reconciling their emotions vis-a-vis societal expectations of how they should perform. Similarly, in *Mansfield Park*, Fanny found herself ‘silenced’ when her argument “could not be denied.” This suggests that the character encountered social constraints that limit her right to speak freely or assert herself in public.

The phrase “could not be” also portrays the conflict between a character’s desire and the reality of their situation. In *Emma*, the line “she could not be complying, she dreaded being quarrelsome” demonstrates Emma’s internal struggle to maintain peace and her true feelings. This line highlights the struggle between social conformity and personal integrity. In *Persuasion*, the mention of Anne “could not be given to given to understand so much by her friend, could not know herself” reflects not only her social and emotional limitations but also the challenges arising from a lack of self-awareness and external misunderstandings. In several instances, “could not be” is utilized to emphasize the permanent or unchangeable nature of specific circumstances. This adds a sense of sorrow to the story. For example, in *Pride or Prejudice*, the mention of “could not be in your power,” suggests the limitations of the character’s abilities.

The phrase “could not be” in Jane Austen’s novels signifies impossibility and prohibitions. The phrase frequently appears to highlight the realities or societal expectations that the characters themselves deem unable to change. The use of this modality expression suggests Austen’s critique of women’s vulnerability amidst patriarchal society and yet her being aware of the social injustice encountered by women by using this hedging feature to demonstrate her awareness of “modesty and politeness in the language [prescribed by] social and

communicative expectations” (Fischer-Starcke, 2009, p. 515). Similar to the frequent use of the words “to be” and “I would not,” the phrase “could not be” reflects how the author struggles internally with her own desires and externally with societal rules and expectations placed upon them.

CONCLUSIONS

This corpus stylistics study has explored the thematic and stylistic elements in Jane Austen’s novels by focusing on the function of specific word clusters such as “to be,” “it would be,” and “could not be.” Our findings demonstrate how these clusters are integral to Austen’s narrative style and the portrayal of societal constraints and expectations within her works. The analysis highlights the use of corpus stylistics in revealing subtle but significant patterns of language use that deepen our understanding of literary texts.

The frequent occurrence of the phrase “to be” throughout Austen’s novels highlights its role in constructing identities and social dynamics, especially in the portrayal of passive roles often assigned to women during the Regency period. These words subtly challenge and critique societal norms and expectations placed on women, emphasizing their often passive roles within societal structures. Similarly, “it would be” has shown to be crucial for exploring hypothetical scenarios within Austen’s texts that reflect the characters’ observations and the societal pressures they encounter. This modal expression enriches the narrative by allowing characters to consider potential outcomes and express desires or doubts in a socially acceptable manner, thus reflecting the constrained agency of individuals in Austen’s era. Lastly, the cluster “could not be” highlights the limitations and impossibilities faced by characters, especially women, in Austen’s novels. This phrase often marks the boundaries of societal and personal constraints, illustrating the conflicts between individual desires and societal expectations.

Overall, this corpus stylistic analysis not only affirms the literary and thematic depth of Austen’s works but also provides quantitative reasoning that supports traditional literary interpretations. By quantifying stylistic and thematic elements, we gain more understanding of Austen’s critique of her society and her literary sophistication. Future research could extend similar analyses to other authors and time periods, further exploring the intersections between linguistic structure, societal context, and literary expression.

REFERENCES

- Barchas, J. (2012). *Matters of fact in Jane Austen: History, location, and celebrity*. JHU Press.
- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Longman grammar of spoken and written English*. John Benjamins Publishing Company.
- Burrows, J. (2002). ‘Delta’: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267–287.

<https://doi.org/10.1093/LLC/17.3.267>

- Demmen, J. E. J. (2012). *A corpus stylistic investigation of the language style of Shakespeare's plays in the context of other contemporaneous plays* (Issue September). Lancaster University.
- Fischer-Starcke, B. (2009). Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*. *International Journal of Corpus Linguistics*, 14(4), 492–523. <https://doi.org/10.1075/ijcl.14.4.03fis>
- Flavin, L. (2004). *Jane Austen in the classroom: Viewing the novel/reading the film*. Peter Lang Publishing, Inc.
- Hoover, D. L. (2007). Corpus stylistics, stylometry, and the styles of Henry James. *Style*, 41(2), 174–203. <https://www.jstor.org/stable/10.5325/style.41.2.174>
- Hung, M., Chen, H.-Y., Chang, Y.-Y., Hung, W.-T., & Tsoi, C.-H. (2020). A corpus-based analysis of William Shakespeare's *The Winter's Tale*. *International Journal of Linguistics, Literature and Translation (IJLLT)*, 3(11), 117–123. <https://doi.org/10.32996/ijllt>
- Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.
- Johnson, C. L. (1995). *Jane Austen: Women, politics, and the novel*. University of Chicago Press.
- Johnson, C. L. (2014). *Jane Austen's cults and cultures*. University of Chicago Press.
- Kaplan, L. (2005). Persuasions 27. *The Jane Austen Journal*.
- Mahlberg, M. (2013). *Corpus stylistics and Dickens's fiction*. Routledge.
- Makhloof, S. (2020). Leech and short's checklist of lexical features in style in fiction: A theoretical analysis. *IJOHMN (International Journal Online of Humanities)*, 6(4), 19–43. <https://doi.org/10.24113/ijohmn.v6i4.189>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McIntyre, D., & Walker, B. (2019). *Corpus stylistics theory ad practic*. Edinburgh University Press Ltd. <https://doi.org/10.4324/9780367568887-27>
- Moretti, F. (2013). *Distant reading*. Verso.
- Murphy, S. (2006). Now I am alone: A corpus stylistic approach to Shakespearian soliloquies. *Lancaster University Postgraduate Conference in Linguistics & Language Teaching*, 1, 66–85. <http://www.lancaster.ac.uk/fass/events/laelpgconference/papers/v01/Murphy.pdf>
- O'Halloran, K. (2007). Critical Discourse Analysis and the Corpus-informed Interpretation of Metaphor at the Register Level. *Applied Linguistics*, 28(1), 1–14. <https://doi.org/10.1093/applin/aml046>

- Oakes, M. P., & Ji, M. (Eds.). (2012). *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research* (Vol. 51). John Benjamins Publishing Company. <https://doi.org/10.1075/SCL.51>
- Rybicki, J., Eder, M., & Hoover, D. L. (2016). Computational stylistics and text analysis. In C. Crompton, R. Lane, & R. Siemens (Eds.), *Doing digital humanities* (p. 22). Routledge.
- Scott, M. (2008). *WordSmith tools version 5.0: User manual*. Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus, and discourse*. Routledge.
- Stubbs, M. (2005). Conrad in the computer: examples of quantitative stylistics methods. *Language and Literature*, 14(1), 5–24. <https://doi.org/10.1177/0963947005048873>
- Tannen, D. (1996). *Gender and discourse*. Oxford University Press. <https://www.deborahannen.com/gender-and-discourse>
- Underwood, T. (2019). *Distant horizons digital evidence and literary change*. The University of Chicago Press.
- Wiltshire, J. (1992). *Jane Austen and the body: "The picture of health."* Cambridge University Press.