

DOI: 10.5281/zenodo.12426550

LINGUISTIC VARIATION AND CULTURAL REPRESENTATION IN CLASSICAL AND MODERN TEXTS: A COMPARATIVE CORPUS-BASED STUDY

Siddikova Iroda Abduzuhurovna^{1*}, Sadriddinzoda Safiya Shaxobiddinovna²
Muzafarova Luiza Umarovna³ Abdullaeva Sanobar Bakhrillaevna⁴, Muhitdinov Shavkat
Muhammedjanovich⁵ Ibatova Amira⁶ Davlatova Mukhayyo Hasanovna⁷

¹Department of Theory of Translation and Comparative Linguistics National University of Uzbekistan named after MirzaUlugbek 100174, Uzbekistan, Tashkent, Almazar district, st.Universitetskaya street, 4

²Samarkand State Institute of Foreign Languages Samarkand, Uzbekistan

³Samarkand State Institute of Foreign Languages PhD in Philology, Associate Professor at the department of the English theory and practice, Samarkand, Uzbekistan

⁴Philological Sciences. English Philology. Samarkand State Institute of Foreign Languages, Samarkand, Uzbekistan

⁵Biological Sciences, Associate Professor, Department of Medical Biology Samarkand State Medical University

⁶Samarkand Branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi

⁷Department of Uzbek Language and Literature, Russian and English, Bukhara State Medical Institute

Received: 17/10/2025

Accepted: 15/01/2026

Corresponding Author: Siddikova Iroda Abduzuhurovna

(ravsidd@mail.ru)

ABSTRACT

This study presents a comprehensive review of linguistic variation and cultural representation in classical and modern texts through a corpus-informed analytical perspective. Drawing on interdisciplinary insights from corpus linguistics, literary studies, and cultural theory, the study examines how language evolves across historical periods while simultaneously reflecting shifting socio-cultural and ideological contexts. The review synthesizes existing research to explore patterns of lexical, syntactic, and stylistic variation, highlighting key differences between classical texts, which emphasize formality, tradition, and collective identity, and modern texts, which foreground diversity, subjectivity, and social change. In addition, the study investigates how cultural representation is constructed and transformed in literary discourse, focusing on themes such as identity, power, marginalization, and globalization. A corpus-based analytical framework is employed to evaluate methods such as keyword analysis, collocation, and semantic prosody, demonstrating their effectiveness in uncovering linguistic patterns and deeper cultural meanings. The findings reveal that while corpus approaches provide valuable empirical insights, their integration with qualitative interpretation is essential for capturing the complexity of cultural representation. The study also identifies key methodological gaps in existing literature, particularly the lack of comprehensive comparative reviews bridging classical and modern texts. Overall, this review contributes to advancing corpus stylistics and interdisciplinary research by offering a systematic understanding of how linguistic variation and cultural representation interact across time, while also providing directions for future research in corpus-based literary analysis.

KEYWORDS: Linguistic variation; Cultural representation; Corpus linguistics; Corpus stylistics; Diachronic analysis

1. INTRODUCTION

Language is dynamic in nature and it keeps on changing with changing social, cultural and historical circumstances. The linguistic forms also change over time in the vocabularies, syntax, and style of expression, and they represent the future changes in communication and culture values patterns. Literary texts, being artefacts that exist in a culture, offer an especially fertile point of contribution to exploring such variation. They are not only formalising the linguistic techniques of their epochs, but also formalising the cultural, ideological, and social realities of their communities of production. Classical texts are usually characterised by the collective traditions, religious attitudes, hierarchic social order; the modern texts are usually characterised by the anticipation of personal identity, evolution of social life, and various ideological attitudes. In this respect, linguistic variation and cultural representation are closely related to each other because language is a tool that helps to create and pass on cultural meanings.

Over the last few years, corpus linguistics has made enormous steps in diagnosing language and literature since it allows analysing extensive textual data systematically and empirically. The methods that rely on the corpus enable a researcher to define similarities between linguistic patterns and track the semantic change through time, which provides the information about both linguistic change and the cultural one. As an example, semantic change computational analysis has been able to show how language can be used to reflect cultural change in various historical situations (Hamilton et al., 2016). Corpus stylistics has also been proposed as a helpful interdisciplinary tool, techniques of which are integrated with a qualitative interpretation of the intended meaning of stylistic characteristics (Walker and McIntyre, 2019). The combination of concordance-based tools and corpus methods has also enabled the emergence of new methods in the analysis of the literary work, with such studies investigating how linguistic patterns and cognitive processes interact during reading (Mahlberg et al., 2016).

Furthermore, the provision of the massive digital corpora has widened the horizons of the linguistic studies, as researchers got the opportunity to examine a variety of genres and eras with more specificity. Large-scale studies of language use and variation have been facilitated by novel access to organised textual information, including those in the Semantic Scholar Open Research Corpus (S2ORC) (Lo et al., 2020). Corpus-based approaches have been found to be useful in the study of variation between

language varieties and dialects, with computational measures being shown to be reliable in reflecting linguistic differences (Dunn, 2021). These developments highlight the increased relevance of corpus linguistics as a methodology of the study of both linguistic and cultural phenomena.

Irrespective of these improvements, current studies are rather disorganised both in terms of field and time. Linguistic variation is a subject of many studies where most of the investigations are conducted on the structural issues of language without correspondingly emphasising the cultural implications of it. On the other hand, studies of cultural and literary research usually focus on thematic and ideological interpretation and do not have an empirical basis on the basis of linguistic data. Corpus-assisted discourse studies (CADS) such as have played a significant role in discourse comprehension of meaning patterns but little has been done to make use of them in diachronic comparison of the literature (Walker, 2015). Correspondingly, media discourse research has shown how language can be used to form the cultural values and ideologies, especially concerning how perceptions of newsworthiness are made (Bednarek and Caple, 2017), but they are not always applied to comparative literary domains.

In addition, the intersection of corpus linguistics and cultural representation (especially in translation and cross-cultural studies) is a topic that has recently started to be studied. The results of corpus-based studies of bilingual literary works have demonstrated the process of mediating cultural significances and cross-cultural communication through linguistic choice (Mahmood et al., 2025). Similarly, studies on translator style also have emphasized the influence of linguistic patterns on the development of interpretive and functional features of translated texts (Wu, 2025). Although these studies do not deny the possibility of corpus methods in studying cultural representation, there is still a necessity to carry out thorough reviews that would incorporate such findings into a wider comparative context between classical and modern texts.

Against the backdrop of these gaps, the current paper intends to offer a critical and integrative review of linguistic variation and cultural representation in classical versus modern texts in a perspective of corpus. This work is based on the synthesis of the existing research results instead of an empirical study of a single datum in order to find out the major tendencies, methodologies, and theoretical input in the sphere. Using the lens of comparison, the paper analyses the differences in linguistic aspects over time, how such differences are indicative of the

changing cultural situation and how the use of a corpus based approach can help shed more light on the processes involved.

In particular, the aims of the review are three-fold. First, it attempts to examine the trends of linguistic variation between classical and contemporary works, such as the shift in the lexical means, syntactic forms, and stylistic characteristics. Second, it will seek to explore the production and translation of cultural representations in literature and will particularly focus on the issues of identity, power, tradition and social change. Third, it also examines the contribution of corpus linguistics to the process of analysis of the linguistic and cultural variation, both its benefits and drawbacks.

Based on these aims, the research questions of the study include: How linguistic difference is different in classical and modern texts? What are the cultural trends and themes in each era and how do they depict the changing socio-cultural environments? In what ways can corpus-based methods of analysis contribute to our knowledge about linguistic and cultural variation in the field of literature? Through the response to these questions, the paper aims to add to the current debate surrounding how to combine quantitative with qualitative methodology in the humanities and to emphasise the usefulness of corpus-informed methodologies in the development of comparative literature studies.

To sum up, this review is placed in the interdependence of language, literature, and culture and it is profoundly significant to consider the use of linguistic variation as well as cultural representation as two phenomena. Combining the existing research and emphasising the methodological advances, the study will be helpful in the future work in corpus-based literary analysis, as well as in emphasising the fact that corpus linguistics is a powerful tool in the investigation of dynamic interaction of language and culture over time.

2. THEORETICAL FRAMEWORK

2.1 Linguistic Variation

The linguistic variation is one of the key concepts of the way in which language changes through time, space and social condition. The discipline of diachronic linguistics, the study of the historical evolution of language, is one of the most important methods of analysing such variation. Diachronic studies examine the lexical changes and grammatical structures and stylistic patterns of languages that help scholars to trace how linguistic forms are changing in tandem with greater socio-cultural changes. This deviation can be used in a literary

situation, where it is a representation of evolving rules of communication and cultural priorities. The recent studies on corpus evidence prove further that variation is not arbitrary but rather systematic, and register differences do not change dramatically across languages as the situations change (Li et al., 2023).

Besides historical change, the linguistic variation is influenced by the register and stylistic difference, which is depending on the context, purpose, and audience. Register theory places more emphasis on language variation as a result of communicative circumstances whereas stylistic variation places more emphasis on authorial decisions that create aesthetic and functional outcomes. Corpus-controlled research has progressively been applied in the analysis of stylistic variation in both literary and translated works to show the variation in linguistic patterns among authors, genres, and even human and machine-generated texts (Zhou and Cheng, 2025). Such results demonstrate the significance of analysing stylistic elements as important markers of the linguistic variation between classical and modern writings.

Moreover, sociolinguistic views offer an understanding of the social aspects of the variation in language which are caused by the social factors including identity, culture and power relations. Language is not a neutral mode of communication since it is a socially situated activity that is guided by cultural practises and ideological frames. The modern studies highlight social media and digital communication in the change of the linguistic practises with the introduction of new forms, like the so-called *algospeak* that represent the alteration of the cultural and technological context (Aleksic, 2025). These changes show that the linguistic variation can be considered not only as the aspect of the past but also when considering the current socio-cultural phenomena.

2.2 Cultural Representation

The notion of cultural representation is central to the realisation of the way meaning is created with the use of language. Based on the representation theory, culture is considered to be an active process of meaning-making where the language is central. Literary texts are viewed as places, where cultural meanings are created and negotiated in both powerful ideologies and other ways of thinking. Here, the cultural representation is closely associated with the discourse whereby, language determines the way the reality is perceived and interpreted.

The connexions between the language and the culture have been extended to the posthumanist and

sociocultural approaches which propose a challenge to the traditional boundaries of language, identity and context. Such applied linguistics as posthumanist focuses on the unity of human and non-human aspects of communication and the way in which language is conditioned by technological, material and social influences (Pennycook, 2017). On the same note, the research on popular culture and linguistic diversity illustrates how voices and identities are created in both online and offline environments in the complexity of the modern cultural representation (Dovchin et al., 2018).

The role played by ideology and discourse as another significant dimension of cultural representation in the creation of meaning. Corpus-assisted discourse studies (CADS) offers a methodology of analysing the ways in which language finds a reflection and reproduction of ideological patterns within text. Through the analysis of the repetitive language elements and discourse patterns, it is possible to identify the values and assumptions formed in the cultural narratives (Gillings et al., 2023). This method is especially applicable in comparative literary analysis because it will allow to determine the changes in cultural representation of the classical and modern texts.

2.3 Corpus Linguistics and Corpus Stylistics

Corpus linguistics involves a technique of empirical and systematic analysis of language in

terms of large textual bodies. It allows the investigator to establish the patterns of linguistic variation and the ways those patterns are connected to the larger cultural and social settings. The modern literature shows the adaptability of corpus linguistics and indicates that it can be implemented in the following fields: qualitative analysis, language learning, and discourse studies (Misnawati et al., 2024; McEnery et al., 2019). All these shows that corpus-based approaches play a significant role in the research of language and literature.

One of the major differences in the field of corpus linguistics is the corpus-based and corpus-driven approach as each of them provides an important input into the usage of a language. Corpus-based methods are used to verify predetermined hypotheses and corpus-driven methods give patterns the opportunity to emerge inductively out of the information. In literary research, these two methods are usually intertwined to examine the stylistic elements and to derive their meaning within the context. The methods of analysing keywords, collocating words, and conducting a concordance allow the researcher to identify a pattern of meaning and how language creates a cultural representation. Figure 1 depicts the conceptual framework that informs this study as it shows the relationship between linguistic variation, cultural representation and the corpus-based approaches.

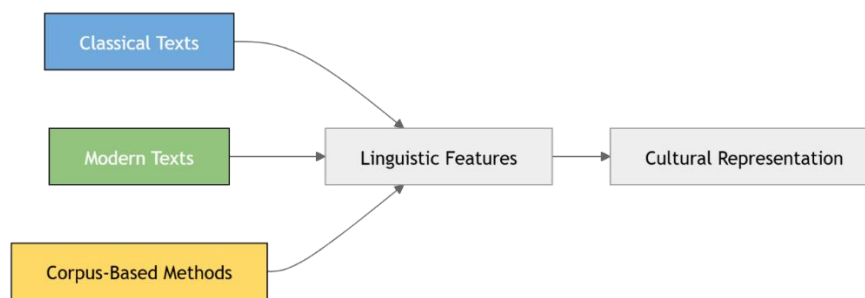


Figure 1. Conceptual Framework of the Study

A summary of the most significant linguistic differences between the classics and modern texts is displayed in Table 1. It identifies the differences in the lexical preference, syntax, and the stylistic

elements, and how language changes over time. These differences give a source of appreciation of larger trends of linguistic development and cultural transformation.

Table 1. Linguistic Differences Across Periods

Aspect	Classical Texts	Modern Texts
Lexicon	Archaic, formal vocabulary (Hilpert & Gries, 2016)	Dynamic, colloquial, evolving (Römer & Garner, 2022)
Syntax	Complex, embedded structures (Biber & Gray, 2016)	Flexible, simplified patterns (Love et al., 2017)
Style	Formal, standardized (Nini, 2019)	Diverse, experimental styles (Zhou & Cheng, 2025)

The main advantage of corpus linguistics is that it allows quantitative and qualitative analysis to be

combined. However, quantitative analysis gives statistical data of linguistic patterns whereas the

qualitative interpretation gives a deeper understanding of the cultural and contextual meaning of the same. This correlation is the key feature of corpus stylistics that involves linguistic examination alongside literary analysis to explore the role of stylistic decisions in the creation of meaning. The corpus linguistics provides a well-developed framework to study the linguistic variation and the cultural representation in the literary works by bridging the gap between numbers and holistic analysis (interpreting).

3. CORPUS-BASED ANALYTICAL FRAMEWORK

3.1 Literature Selection Strategy

In this study, the choice of literature is made in a systematic and clear manner to provide reliability and fullness of the review. The major academic databases such as Scopus, Web of Science, JSTOR and Google Scholar were used to identify relevant studies. Relevant publications were obtained by means of a mix of search terms, including but not limited to linguistic variation, cultural representation, corpus linguistics, corpus stylistics,

and literary analysis. Peer-reviewed journal articles, books, and edited volumes published since 2016 and previous foundational methodological works in corpus linguistics were prioritised by the selection process to guarantee the analytical rigour of the study (Paquot and Gries, 2021).

Inclusion criteria were created to make sure that the articles sampled were closely related to the purpose of this review. Only those studies that dealt with the linguistic variation, cultural representation, or corpus-based method of textual analysis were considered. The preference was given to studies that used corpus techniques or involved the discussion of the literary and discourse analysis in terms of linguistic orientation. Articles that were not relevant or clear methodologically were eliminated. This systematic selection plan is in line with the best practises in corpus-based research, which focuses on transparency, replicability, and planned data management (Egbert and Baker, 2019). Table 2 gives a summary of the main corpus-based studies that were considered in this paper.

Table 2. Summary of Reviewed Studies

Author(s)	Year	Focus Area	Method Used	Key Contribution
Hamilton et al.	2016	Semantic change	Computational corpus	Cultural shift detection (Hamilton et al., 2016)
Mahlberg et al.	2016	Corpus stylistics	Concordance analysis	Integration of corpus and literary analysis (Mahlberg et al., 2016)
Walker & McIntyre	2019	Corpus stylistics	Corpus-based approach	Stylistic meaning in texts (Walker & McIntyre, 2019)
Dunn	2021	Language variation	Corpus similarity	Reliability of linguistic variation (Dunn, 2021)
Mahmood et al.	2025	Translation studies	Corpus analysis	Cross-cultural meaning (Mahmood et al., 2025)
Wu	2025	Translator style	Corpus stylistics	Cultural mediation in translation (Wu, 2025)

3.2 Corpus Scope and Textual Domains

Even though this research does not create a new empirical corpus, it follows a corpus-informed approach by referring to the existing corpora and corpus-based research, which compiles literary texts through various periods of time. The review includes two extensive textual domains, including classical texts and modern texts. Classical texts are canonical literature works like epics, drama and early prose whereas modern texts are contemporary fiction, postmodern literature and digitally mediated texts.

The difference between these fields allows one to explore the comparison of diachronic difference in language and culture. The analysis of the differences between the linguistic features and cultural themes in these categories reveals continuity and change of literary discourse. This method aligns with the practises of corpus linguistics that support the significance of representative and well-defined data sets in the comparative study (Paquot and Gries, 2021).

3.3 Analytical Tools and Techniques

The review is based on the corpus linguistics tools and techniques as an abstract concept of the linguistic and cultural patterns analysis. Tools like AntConc are widely used as examples to show how corpus data may be explored systematically by use of concordance, frequency, and collocation analysis (Anthony, 2019). These tools allow researchers to deal with a large amount of textual data and determine recurring linguistic features with accuracy.

Corpus-based analysis is based on several important analytical methods. The common patterns of lexical and structural features are identified through frequency analysis and they give information on the prevailing linguistic tendencies. The keyword analysis is used to identify statistically significant lexical items against a reference corpus, whereas the collocation analysis is a method used to analyse the occurrence of words in co-occurrence and identify meaning-building patterns. Also, concordance analysis enables one to study words in context giving more details of how they are used and interpreted.

Corpus findings are especially important in regard to the validity of statistical methods used. Quantitative methods such as statistical modeling and significance testing help a researcher to differentiate between meaningful patterns and random variation (Brezina, 2018; Gries, 2021). Such methods improve the stability of corpus-based analysis and help to make more impressive interpretations of linguistic information.

3.4 Analytical Approach

The analytical methodology of the present work is premised on a theme and comparative synthesis of the results of the used literature. Instead of performing original corpus analysis, the research incorporates the knowledge of various corpus-based and discourse-based researchers to discover general trends and patterns. A comparative approach is used to analyse distinctions and resemblances between works of classical and modern literature not only in terms of language peculiarities but also in cultural contexts.

The approach will be a mixture of quantifying information, obtained with the help of studies of the corpus, and qualitative interpretation through the theory of literature and culture. The combination of

statistical evaluation and interpretation is an essential idea of corpus linguistics, which aims to find the middle ground between the empirical rigour and the contextual interpretation (Gries, 2021). The synthesis of the research findings in the review gives a detailed view of the manifestation of linguistic variation and cultural representation through the literary discourse.

3.5 Rationale for a Corpus-Informed Review Approach

The implementation of a corpus-based approach on review research signifies the increased significance of corpus linguistics in the modern studies. Corpus methods present empirical data which adds credibility to analytical statements, and it is also possible to discover patterns that would not be evident in the conventional methods of qualitative research. Systematic and replicable analysis in the corpus tools and statistical methods enables the study to be sounder in terms of the methodological base (Brezina, 2018). Figure 2 indicates the analytical process based on corpus that was applied in this study and shows the steps that must be followed starting with text selection until the interpretation.

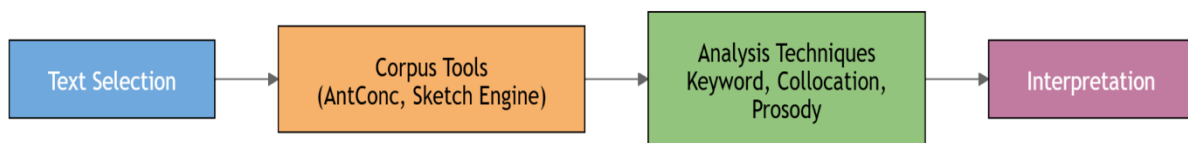


Figure 2. Corpus-Based Analytical Process

Moreover, corpus-based review allows critical review of the current research methods and its advantages and shortcomings. Although corpus methods present great opportunities in the analysis of the linguistic data, the interpretation of the results must be closely based on the context and theories. This research will combine the insights of corpus with the qualitative analysis in order to have a balanced and comprehensive picture of linguistic variation and cultural representation in literary texts.

4. LINGUISTIC VARIATION ACROSS CLASSICAL AND MODERN TEXTS

4.1 Lexical Variation

Lexical variation is one of the most evident evidence of language transformation over historical time. Classical texts tend to be defined by archaic and high diction and heavy dependence on traditional or culturally determined vocabulary. These lexical decisions are also influenced by the social-cultural and historical backgrounds within the textual

production period, and usually focus on the themes of religion, hierarchy, and communal identity. The use of archaic vocabulary also indicates temporal distance as well as adds to the stylistic and aesthetic features of classical literature and supports its formal and authoritative voice.

On the contrary, the contemporary text has a wider and moving lexical variety and consists of current words, colloquialisms, and neologisms. The usage of new words and expressions is often technologically-dependent, which is connected with the globalisation process and the change in cultural traditions. Corpus-based studies have shown that lexical variation may be actively studied in various contexts and registers, which can give us ideas of how language changes with new communicative situations (Römer & Garner, 2022). This is a lexical flexibility that helps contemporary texts to be more in touch with various social realities and touch a broader audience.

The other critical factor of lexical variation is that of semantic shift which means changes of meaning of words over time. Words used in both the classical and the modern writings might have varying meanings in a historical and cultural context. Such semantic change has been demonstrated using quantitative methods of diachronic corpus linguistics to demonstrate patterns of usage and frequency, providing empirical evidence of linguistic change (Hilpert and Gries, 2016). Contemplations of how the linguistic transformation can be made and researched with the assistance of sophisticated strategies further underline the dynamic character of the lexical evolution (Huang and Joo, 2025). Figure 3 provides a comparative description of linguistic characteristics of classical and modern texts, which shows the main distinctions between lexical, syntactic, and stylistic patterns.

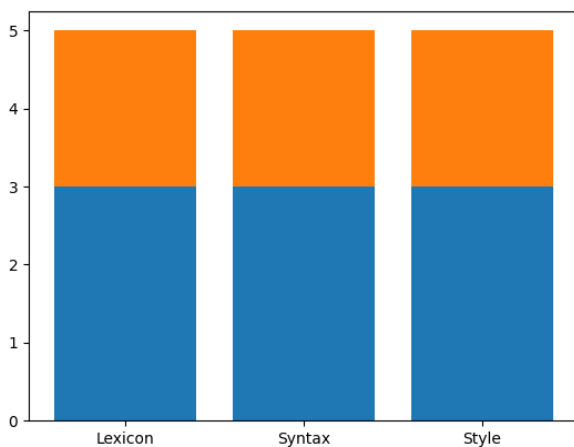


Figure 3. Comparative Linguistic Features Across Periods

4.2 Syntactic Variation

Another important dimension of linguistic change is syntactic variation that involves the differences in sentence structure, grammatical patterns and the levels of complexity. The classical texts are usually marked by the use of complicated syntactic structures such as lengthy sentences, numerous subordinate clauses, and the use of complicated rhetorical structures. These characteristics are indicative of the formal habits of previous literary traditions, which adhered to elaboration and structural elaborateness.

Modern texts, on the contrary, are more inclined to the simpler and more adaptable syntactic forms, which add to its clarity and readability. This movement is in line with the general trends in communicative practises and expectations of readers. The research on grammatical complexity has proved that the current written language tends to have a

variety of complexity patterns that are not inherited by the older forms and have evolved due to changes in language conventions and style inclinations (Biber and Grey, 2016).

Simultaneously, syntactic change does not necessarily consist of its simplification; the modern texts often flourish with new forms, such as discontinuous sentences and non-linear syntax. Spoken and written language corpus-based studies have also demonstrated variations in syntactic structure between the two modes of communication and have shown that contextual factors are important in determining grammatical variation (Love et al., 2017). Such results highlight the use of syntax as a structural and stylistic tool in the expression of literature.

4.3 Stylistic Features

The wider range of linguistic options that adds tone, voice, and aesthetic properties to a text includes stylistic variation. The use of formal or informal language is one of the greatest differences between classical and modern texts. Classical literature generally uses a very formal register that is typified by standardised grammar, high vocabulary, and regular patterns of narration. This is a representation of cultural and institutional contexts in which such texts were created.

The contemporary texts are, on the contrary, more stylistically varied, and tend to include conversational language, local dialects and non-standard forms. This change is a sign of democratisation of language and the rising number of voices represented in literature. It has been demonstrated that stylistic analysis based on corpus allows one to quantify the variation in linguistic features presented in the text in a systematic way, which allows identifying the specific stylistic pattern (Nini, 2019).

Narrative techniques also are another significant element of stylistic variation. Classical literature commonly utilises linear plots and omniscient point of view whereas modern literature commonly employs experimental writing styles like stream of consciousness, fragmented narrative, and multi-point of view. These stylistic inventions are a mirror of other cultural and literary trends, in which subjectivity is stressed and the storey is complicated.

Altogether, the variation in style provides an emphasis on how linguistic options can be used as a tool of artistic expression and culture presentation. Through the combination of the corpus-based approaches with the literary analysis, researchers will get a better idea of how the stylistic

characteristics change over time and how they help to shape the nature of the literary discourse.

5. CULTURAL REPRESENTATION IN CLASSICAL AND MODERN TEXTS

5.1 Cultural Themes in Classical Texts

Classical texts represent culture in the most embedded ways within their socio-religious and philosophical historical contexts. They tend to be more collective in their values, tradition, morality, and social order. Such themes as religion, mythology, heroism, and hierarchical organisation are evident which were used to support the supreme ideologies of the culture. Classical literature often creates a surrounding world where people are situated within social structures that are laid down, and cultural standards are introduced as being conservative and dominant.

In addition, classical literature is more inclined to collective identity than to individual subjectivity, and the characters are depicted as representatives of the culture as a whole and not as complex personalities. Archetypal characters like heroes, rulers and morally upright characters can be found in this. Examples of these types of cross-cultural studies of classical texts include the study of Shakespearean reception in other cultures, which show that these texts serve as vectors of long-lasting cultural codes and values (Joubin, 2021). In these texts, language is very important in continuation of this representation as the use of formal and high styles of language creates the impression of authority and tradition.

The focus on continuity and retention of cultural heritage is another aspect that is important in the cultural representation of classical texts. These writings tend to serve as storage of communal memory and transfer cultural information and values between generations. Consequently, they are less flexible to change and more interested in stability, which are characteristic of societies in which tradition is the main element. Simultaneously, the terms representation and agency emphasise the ways in which the classical works can support the dominant ideas and suppress the voices of opposing perspectives (Kian et al., 2020).

5.2 Cultural Themes in Modern Texts

Unlike the literary works of the classical period, the contemporary ones are characterised by a gradual transition to the individualism, diversity, and social transformation. The modern literature also tends to address the issues of identity, globalisation, migration, gender, and social inequality. These motifs raise the issues of the complication of the

contemporary life and how cultural identities are bargained in the dynamically shifting social scenes.

Contemporary literature is inclined to defy old traditions and criticise existing systems of power presenting other opinions that were usually pushed to the sidelines in literature. Examples of cultural representation in educational and literature sources prove that modern writings are becoming more and more interlaced with cultural views and are not limited to stereotypical images (Zhang et al., 2024; Taqdir, 2025). The language has been instrumental in this process because the contemporary writers have a broad variety of linguistic styles in an effort to reflect the breadth of modern experience.

Moreover, the subjectivity and psychological depth of cultural representation in the contemporary texts are more characteristic of modern texts. Characters are usually presented as multifaceted people who are experiencing several and even contrary identities. The multimodal and corpus-based text analyses of modern texts also inform about the way that cultural meanings are created by using both linguistic and visual means of representation, which is related to the multidimensionality of the contemporary discourse (Suwandani and Hidayati, 2025). Such transformation highlights the changing theme of cultural representation in contemporary literature.

5.3 Ideological and Discursive Shifts

The classical-modern changes are characterised by immense ideological and discursive changes in the depiction of power and identity, social relations, etc. Classical texts tend to strengthen the status quo, and hierarchical relations are evident, being accepted as natural. Modern texts, on the contrary, often question such constructions uncovering disparities and promoting social transformation.

The role of the discourse in creating these representations is paramount because the linguistic decisions are an indicator of the ideological standpoints. Corpus-based research has revealed the way cultural and religious ideologies are encoded in the language that is used, especially in educational and literary texts (Liu et al., 2022). Computational techniques are becoming increasingly important in modern discourse analysis, to understand the processes of the production and reproduction of cultural meanings over large volumes of data.

Recent studies also emphasise the power of digital technologies and artificial intelligence in the cultural representation formation. Research on large language models (LLMs) shows how cultural trends and biases could be encoded in textual data, which

are applicable to the rest of the society (Adilazuarda et al., 2024). In a similar manner, the analyses of the AI-generated content suggest the significance of considering cultural sensitivity and representation in the new discourse forms (Kiden et al., 2025). These changes denote that the aspect of cultural representation is not merely one of the characteristics of the traditional literature works, but it is also one of the major issues of the digital communication nowadays.

5.4 Representation of Marginalized Voices

Among the most important trends in the contemporary literary discourse, the greater presence of the marginalised voicing must be mentioned. Whereas in many classical literature, the dominant cultural viewpoints are usually favoured, modern literature aims at incorporating the voices that have been excluded in the past or are underrepresented. This encompasses views founded on gender, ethnicity, and class among other social groups.

The figure of the marginalised voices is closely connected with larger cultural and political movements promoting the ideas of equality and social justice. It can be proven by the contemporary research that the modern texts and digital corpora are becoming more and more representative of various opinions, but the issues related to bias and representation still persist (Madhusudan et al., 2025). Modern texts disrupt the rule of the standard language and emphasise the variety of linguistic variation by using various forms of languages, dialects, non-standard forms of languages.

Simultaneously, the idea of the inclusion of the marginalised voices brings up several significant issues of authenticity, representation, and power. Writers and researchers have to carefully analyse the process of creating cultural narratives and who is given voice or is left without a voice. Therefore, cultural representation in contemporary works is both a chance and a challenge, and one should take into consideration the ethical and ideological aspects.

5.5 Comparative Insights on Cultural Representation

A comparison of classical texts and the modern ones shows continuity and change in terms of cultural representation. Whereas the classical literature is concerned with stability, tradition, and collective identity, the modern literature foreshadows change, difference and individual experience. These variations represent the larger socio-cultural changes, such as globalisation,

technological change, and social norm transformation.

However, some elements remain relevant in both periods, including the quest of morality, identity and relationships between people. Their representation, however, is quite different, depending on the practises of language change and cultural differences. The methods of corpus and computation can be useful in detecting these patterns and provide researchers with opportunity to observe the transformation of cultural meanings throughout time.

6. CORPUS-INFORMED COMPARATIVE INSIGHTS

6.1 Keyword Trends Across Periods

The analysis of keywords based on corpus is also an excellent tool of determining the predominant lexical patterns between classical and contemporary writings. Through the frequencies of words in corpora, the researcher is able to identify the lexical items that are statistically salient within a given period of time. Keywords in classical literature typically fall into the theme of religion, power, morality, and communal identity due to the socio-cultural values of older times. These forms of lexis support the institutional and formal quality of the classical speech, in which language is used to preserve cultural continuity and ideological stability.

Conversely, the trends on key words in contemporary texts would emphasise on themes that concern individuality, identity and social change. Personal experience and other identities related words are more represented which means transition to more subjective and dynamic forms of representation. The computational analysis of semantics also contributes to the detection of such patterns, as it allows making more narrow comparisons even with large textual volumes (Mao et al., 2024). Also, the cross-linguistic and semantic similarity models have increased the possibility to make comparisons of the textual patterns in corpora and helped in strengthening the key-based analysis (Cer et al., 2017).

6.2 Collocation and Meaning Construction

The collocation analysis provides more information about the meaning construction of words associations patterns. The collocations in the classical works tend to be more representative of the fixated and conventionalized forms, supporting the proven cultural images and ideological frameworks. The combination of words that are associated with authority, morality, and tradition are often

predictable, as the classical discourse is based on the hierarchical and value systems.

On the other hand, there are more mobile collocational patterns in modern texts. Words are mixed up in new forms, which indicates the elasticity of the modern language and the intricacy of the modern culture life. The research of corpus-assisted discourse studies has demonstrated that these patterns have the potential to demonstrate the way in which cultural meanings are produced in situations, such as in the media and political discourse (Albota, 2024). This brings out the significance of collocation analysis in the discovery of explicit and implicit cultural sense that are contained in the use of language.

6.3 Semantic Prosody and Evaluative Meaning

Semantic prosody is an important factor in determining evaluative meaning because it brings out the way words obtain either positive or negative connotation after repetition in contextual association. Moral and ideological frameworks tend to be more in line with semantic prosody in classical writings, and some words in the lexicon always have a positive or negative connotation. These trends support the mainstream cultural discourses and make classical literature didactic.

Semantic prosody in modern texts is more heterogeneous and contextualised, which is characteristic of the discourse of contemporary texts. Corpus-based works have shown how semantic prosody can be effectively studied in order to discern delicate evaluative meanings of language (Sarhad and Mahmood, 2023; Mohammed, 2022). Also, recent studies indicate the need to study semantic prosody in terms of registers and contexts, especially in discerning how meaning can change in various communicative situations (Rutkowski, 2025).

The complexity of semantic prosody has also been highlighted by theoretical questions and statements that evaluative meaning is determined by several interdependent factors such as context, collocation, and discourse structure (Grabowski and Trklja, 2024). Moreover, the construction-based methods shed more light on the functioning of semantic

prosody grammatical and lexical forms (Ahmad and Saleem, 2026). Such views affirm the significance of semantic prosody as a major tool of analysis of corpus-based research on language and culture.

6.4 Quantitative and Qualitative Integration

Corpus linguistics is one of the most significant strength as it combines both quantitative and qualitative methods. Quantitative techniques (frequency analysis and statistical modelling) give objective cues of linguistic pattern on large sets of data. The approaches help researchers draw patterns and provide empirical bases of their inferences. The semantic processing methods have also improved the ability of corpus linguistics to analyse more complicated linguistic data and permit more complex models of meaning and variation (Mao et al., 2024).

Nevertheless, the quantitative results should be added with the qualitative interpretation to get a complete understanding of its cultural and contextual meaning. Through qualitative analysis, researchers are able to investigate the meaning and implications of patterns in language in a particular socio-cultural context. Such an integration is especially useful in comparative studies, during which the differences between classical and modern texts are to be explained in terms of historical and cultural issues.

6.5 Strengths and Limitations of Corpus-Based Approaches

The benefits of corpus-based approaches to the study of linguistic and cultural variation include the capacity to handle large amounts of textual data, and to reveal systematic trends. Such techniques improve the accuracy and validity of research results and hence can be used as a means of comparative analysis. Semantic similarity and corpus-based modelling allow a researcher to investigate the use of language in various contexts and time frames with more accuracy (Cer et al., 2017). Table 3 describes the key corpus-based methods of analysis and their purposes.

Table 3. Corpus Techniques and Their Functions

Technique	Function	Application in Study	Supporting Source
Frequency Analysis	Identifies common patterns	Lexical trends	(Brezina, 2018)
Keyword Analysis	Finds statistically significant words	Cultural themes	(Anthony, 2019)
Collocation	Reveals word associations	Meaning construction	(Egbert & Baker, 2019)
Semantic Prosody	Shows evaluative meaning	Ideological analysis	(Grabowski & Trklja, 2024; Sarhad & Mahmood, 2023)

Nevertheless, corpus-based methods have drawbacks, as well. The quantitative data might not

be quite sufficient to describe the cultural representation that is usually depth and relative in

nature. Also, the structure and the representativeness of corpora may impact the analysis results and introduce the bias. Nevertheless, the combination of the corpus approach and the qualitative analysis can be used to overcome these limitations and gain a more detailed picture of the language and culture.

7. IMPLICATIONS AND LIMITATIONS

This paper can contribute significantly to corpus linguistics and the study of literature. In theory, it shows how the corpus-based methods can be combined with the qualitative analysis method to analyse the linguistic variation and cultural representation in different periods. The study connects linguistic and cultural studies to bring home the importance of language as a life-force of cultural meaning and ideological production. The comparative framework is also beneficial in terms of the way the diachronic analysis is strengthened as it allows a systematic approach to the linguistic and cultural changes.

In real life, the research has placed importance on the importance of corpus tools to both education and digital humanities so that a more empirical and interactive method of analysing literature can be used. It can also be applied to computational linguistics and especially to enhance culturally sensitive language technologies.

There are, however, limitations of the study. The dependence on the chosen literature might not be able to reflect the diversity of the research in the area, and the lack of direct corpus analysis restricts the level of the empirical research. Moreover, the complexity of cultural meaning, which is usually a contextual interpretation, cannot be fully taken into consideration by the corpus methods. These restrictions emphasise the necessity of future studies to be more comprehensive in terms of data and mixed-method methods.

REFERENCES

- Adilazuarda, M. F., Mukherjee, S., Lavania, P., Singh, S. S., Aji, A. F., O'Neill, J., ... & Choudhury, M. (2024, November). Towards measuring and modeling "culture" in LLMs: A survey. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 15763-15784).
- Ahmad, S., & Saleem, T. (2026). Construction Approach to Semantic Prosody of the Verb study: A Corpus-based Study. *Corpus Pragmatics*, 10(1), 5.
- Albota, S. (2024). Linguistic representation of Ukraine in English-language media discourse: a corpus-assisted approach. In *COLINS* (4) (pp. 208-235).
- Aleksic, A. (2025). *Algospeak: How social media is transforming the future of language*. Knopf.
- Anthony, L. (2019). AntConc (version 3.5. 8)[computer software]. Tokyo, Japan: Waseda University.
- Bednarek, M., & Caple, H. (2017). *The discourse of news values: How news organizations create newsworthiness*. Oxford University Press.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

8. CONCLUSION

This study has provided a comprehensive review of linguistic variation and cultural representation across classical and modern texts through a corpus-informed perspective. By synthesizing insights from corpus linguistics, literary studies, and cultural theory, the research highlights how language evolves not only as a structural system but also as a reflection of changing cultural, social, and ideological contexts. The analysis demonstrates that classical texts tend to emphasize stability, tradition, and collective identity through formal and structured linguistic patterns, whereas modern texts exhibit greater diversity, flexibility, and subjectivity, reflecting contemporary concerns such as identity, globalization, and social transformation. Furthermore, the study underscores the value of corpus-based approaches in identifying patterns of lexical, syntactic, and stylistic variation, as well as in uncovering deeper cultural meanings through techniques such as keyword analysis, collocation, and semantic prosody. At the same time, it acknowledges the limitations of purely quantitative methods and emphasizes the importance of integrating qualitative interpretation to fully capture the complexity of cultural representation. Overall, the findings reinforce the significance of interdisciplinary approaches in advancing the study of language and literature. By bridging empirical analysis with cultural inquiry, this research contributes to a more nuanced understanding of how linguistic and cultural dynamics interact across time. It also provides a foundation for future studies to further explore these relationships using more diverse corpora and advanced computational methods, thereby expanding the scope and impact of corpus-based literary analysis.

- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., & Specia, L. (2017, August). SemEval-2017 task 1: Semantic textual similarity multilingual and crosslingual focused evaluation. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)* (pp. 1-14).
- Dovchin, S., Pennycook, A., & Sultana, S. (2018). *Popular culture, voice and linguistic diversity*. Cham: Palgrave Macmillan.
- Dunn, J. (2021, April). Representations of language varieties are reliable given corpus similarity measures. In *Proceedings of the eighth workshop on NLP for similar languages, varieties and dialects* (pp. 28-38).
- Egbert, J., & Baker, P. (Eds.). (2019). *Using corpus methods to triangulate linguistic analysis* (pp. 268-282). London: Routledge.
- Gillings, M., Mautner, G., & Baker, P. (2023). *Corpus-assisted discourse studies*. Cambridge University Press.
- Grabowski, Ł., & Trklja, A. (2024). Unravelling the complexity of semantic prosody: A theoretical inquiry. *Journal of Pragmatics*, 226, 89-102.
- Gries, S. T. (2021). *Statistics for linguistics with R: A practical introduction*. Walter de Gruyter GmbH & Co KG.
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016, November). Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 2116-2121).
- Hilpert, M., & Gries, S. T. (2016). Quantitative approaches to diachronic corpus linguistics. *The Cambridge handbook of English historical linguistics*, 36-53.
- Huang, D., & Joo, H. A. (2025). A Computational Approach to Analyzing Language Change and Variation in the Constructed Language Toki Pona. *arXiv preprint arXiv:2508.10246*.
- Joubin, A. A. (2021). *Shakespeare and East Asia*. Oxford University Press.
- Kian, A., Peccia, T., & Porret, A. A. 24 | 2020 Genre et transgressions: Représentation, agentivité, autodétermination.
- Kiden, S., Peter, O., Reyes-Cruz, G., Klyshbekova, M., Choi, S., Bergin, A. G., ... & Fischer, J. E. (2025). Back to the Communities: A Mixed-Methods and Community-Driven Evaluation of Cultural Sensitivity in Text-to-Image Models. *arXiv preprint arXiv:2510.27361*.
- Li, H., Dunn, J., & Nini, A. (2023). Register variation remains stable across 60 languages. *Corpus Linguistics and Linguistic Theory*, 19(3), 397-426.
- Liu, Y., Zhang, L. J., & Yang, L. (2022). A corpus linguistics approach to the representation of western religious beliefs in ten series of Chinese university english language teaching textbooks. *Frontiers in Psychology*, 12, 789660.
- Lo, K., Wang, L. L., Neumann, M., Kinney, R., & Weld, D. S. (2020, July). S2ORC: The semantic scholar open research corpus. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 4969-4983).
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319-344.
- Madhusudan, S., Morabito, R., Reid, S., Sadr, N. G., & Emami, A. (2025, April). Fine-tuned LLMs are "time capsules" for tracking societal bias through books. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 2329-2358).
- Mahlberg, M., Stockwell, P., Joode, J. D., Smith, C., & O'Donnell, M. B. (2016). CLiC Dickens: Novel uses of concordances for the integration of corpus stylistics and cognitive poetics. *Corpora*, 11(3), 433-463.
- Mahmood, M. A., Sattar, A., Imran, M., & Almusharraf, N. (2025). Bridging cultures through explicitation: a corpus-based analysis of bilingual literary translations. *Cogent Arts & Humanities*, 12(1), 2508210.
- Mao, R., He, K., Zhang, X., Chen, G., Ni, J., Yang, Z., & Cambria, E. (2024). A survey on semantic processing techniques. *Information Fusion*, 101, 101988.
- McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, 39, 74-92.
- MISNAWATI, M., Sahril, N. U. R., & TAHIR, S. Z. B. (2024). Corpus linguistics today: A qualitative approach. *Research and Innovation in Applied Linguistics*, 2(1), 45-62.
- Mohammed, A. O. G. (2022). Semantic prosody of research verbs: A corpus-informed study. *Journal of Language and Education*, 8(2 (30)), 48-65.

- Nini, A. (2019). The multi-dimensional analysis tagger. *Multi-dimensional analysis: Research methods and current issues*, 67-94.
- Paquot, M., & Gries, S. T. (Eds.). (2021). *A practical handbook of corpus linguistics*. Springer Nature.
- Pennycook, A. (2017). *Posthumanist applied linguistics*. Routledge.
- Römer, U., & Garner, J. (2022). What can corpus linguistics tell us about second language acquisition?. In *The Routledge handbook of corpus linguistics* (pp. 328-340). Routledge.
- RUTKOWSKI, M. S. (2025). The pedagogical value of register-specific semantic prosody in L2 teaching.
- Sarhad, J., & Mahmood, R. (2023). A Corpus-based study of semantic prosody across a native corpus. *Journal of Garmian University*, 10(3), 902-909.
- Suwandani, F., & Hidayati, S. (2025). Conceptual Representation of Culture in EFL Textbook Images: A Multimodal Analysis. *Dinamika Ilmu*, 25(1), 99-122.
- Taqdir, S. P. (2025). Cultural Stereotypes in Foreign Language Textbooks: A Systematic Review of Visual Representation. *Theory and Practice in Language Studies*.
- Walker, B., & McIntyre, D. (2019). *Corpus stylistics: Theory and practice*. Edinburgh University Press.
- Walker, R. J. (2015). Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS). *麗澤大学紀要*, 98, 129-130.
- Wu, K. (2025). Uncovering the functional aspect of translator style: corpus stylistic insights. *Humanities and Social Sciences Communications*, 12(1), 1-13.
- Zhang, H., Li, R., Chen, X., & Yan, F. (2024). Cultural representation in foreign language textbooks: A scoping review from 2012 to 2022. *Linguistics and Education*, 83, 101331.
- Zhou, P., & Cheng, J. (2025). Stylistic variation across English translations of Chinese science fiction: Ken Liu versus ChatGPT. *Frontiers in Artificial Intelligence*, 8, 1576750.