

ANALYZING *MACBETH* THROUGH VOYANT TOOLS: A CORPUS-BASED STUDY OF THEMES, WORD FREQUENCY, AND CHARACTER REPRESENTATION

Om Kumar
omk7341@gmail.com

MS. Scholar, Quaid-e-Awam University of Engineering,
Science & Technology (QUEST), Nawabshah, Sindh,
Pakistan.

Dr. Mahmood Dahri
dahrimahmood7@gmail.com

Assistant Professor, Government Sachal Sarmast
Postgraduate College, Nawabshah, Sindh, Pakistan.

Dr. Qurra-tul-ain Ali Sheikh
anne3_libra@yahoo.com

Associate Professor and Head of Economics Department,
Government Girls Degree College, Nawabshah, Sindh,
Pakistan.

Abstract

The present study applies Voyant Tools, which is a web-based text-analyzing software, to analyze the themes and characters in the play *Macbeth* by William Shakespeare. This work will enhance conventional approaches to the analysis of literature by integrating both quantitative and qualitative analysis with the help of digital humanities. The research discovered a vocabulary density of 0.185, which means that there is an abundant but easy-to-use language and a readability index of 8.126, which means that the text is average and difficult to read. The mean sentence length of 14.7 words indicates the style of Shakespeare, who wrote short sentences and had a powerful effect. The frequency analysis reveals the word occurrences including Macbeth (282), Macduff (106), and Lady (95) are analyzed as central words in the story. Thematic analysis shows that there are major themes in this analysis, which include ambition, guilt and conscience, and the supernatural, with recurring themes of power (11 times), blood (24 times), and witches (8 times) being utilized frequently. These findings show how digital tools can reveal patterns and trends in literary texts, offering fresh perspectives on classic works. The study emphasizes the ability of digital humanities to supplement and enrich traditional literary scholarship by providing deeper insights into the complex interplay of themes and characters in *Macbeth*.

Keywords: *Corpus Analysis, Corpus Linguistics, Digital Tools, Macbeth Themes, Macbeth, Voyant Tools, William Shakespeare.*

Corresponding Author: Om Kumar (Ms. Scholar, Quaid-e-Awam University of Engineering, Science & Technology (QUEST), Nawabshah, Sindh, Pakistan)
Email: omk7341@gmail.com

1. Introduction

This study applies corpus-based digital analysis to the play *Macbeth* by William Shakespeare. The purpose of the study is to examine the frequency of words, themes, and character representation using a corpus method to understand the meaning of the play better (Naqvi and Bhatti, 2025). Voyant Tools is the main analytical tool used to examine themes and characters in *Macbeth* and is a powerful and flexible web-based text analysis tool. The play also develops the ideas of time, fear, and moral decline, which is the psychological turmoil of the protagonists (Bondi & Sezzi, 2018). In addition, the chaotic arrangement of the story emphasizes how unregulated ambition impacts people and society (Soenarto & Wibawarta, 2023; Roy, 2025). This method improves the process of literary interpretation and encourages the application of digital tools when studying Shakespearean works.

Digital Humanities (DH) is a revolutionary convergence of computational approaches and literary studies, which substantially improve the study of literature. Using the methods of text mining, natural language processing (NLP), and machine learning, researchers can discover the latent patterns, themes, and emotional colorings in literary works and thus enrich the traditional qualitative methods (Preeti et al., 2024; Vinodini and Raj, 2024). For example, sentiment analysis and topic modeling can be used to learn more about the character dynamics and the thematic framework (Vinodini and Raj, 2024). Moreover, digital tools support new approaches such as data visualization and geographic mapping that bring new information about narrative structures and interrelations between literary elements (Puspitasari et al., 2025). Combining these computational methods not only encourage interdisciplinary cooperation but also creates new meanings of cultural objects, which ultimately transform the way literature is read and interpreted in the digital era (Thakre, 2024; Vanathi, 2023).

Corpus linguistics is important in interpreting the text, as it uses systematic approaches to the analysis of large amounts of text corpora, which allows an objective study of linguistic properties, including frequency, distribution, and collocation. This field focuses on the significance of language in practice, and the meaning is formed based on the relations formed during textual interactions (Li, 2010). Corpus linguistics involves the use of computer software and electronic databases to integrate empirical processes with quantitative and qualitative analyses that improve the interpretative process (Skichko, 2023). Technology has shifted the focus of textual analysis beyond the stagnant study of printed materials to active analysis of language, enabling researchers to study spoken and written language in different contexts and time periods (Meyer, 2012). Besides, word frequency and lexical dispersion research in corpora helps to understand the patterns and use of language, which contributes to the comprehension of linguistic phenomena (Burch & Egbert, 2020).

Macbeth is a good example of a corpus that can be analyzed using digital and thematic analysis because it contains a complex language, multiple themes, and the plot that is character-oriented. The ambition that the play explores, which is triggered by the prophecies of the witches and exacerbated by the manipulation of Lady Macbeth, shows the psychological effects of uncontrolled desire, which results in moral corruption and social disorder (Roy, 2025). Also, the character of Ross demonstrates the intricacies of communication and information flow in early modern society and raises more general issues connected with surveillance and social life (Mazzola, 2023). The fact that the play is interdisciplinary, i.e., the past can be related to the present, only increases the relevance of the play and enables a wide range of analytical approaches (Nostbakken, 1997). Furthermore, both Macbeth and Lady Macbeth experience a psychological breakdown which highlights the dramatic influence of ambition on human nature, and the play is a good theme for thematic analysis (Roy, 2025; Lutz, 2019).

this paper emphasizes the significance of integrating corpus linguistics and digital humanities tools to learn more about literary works like Macbeth. Through the utilization of Voyant Tools, the study shows how the data analysis can be used to supplement the traditional interpretation of literature through the systematic analysis of themes, language use, and characterization. This method not only contributes to better knowledge about Shakespeare's plays but also demonstrates the increased importance of digital instruments in contemporary literary studies and education.

1.1. Research Objectives

1. To Analyze the Frequency and Distribution of Words in Macbeth.
2. To Visualize Thematic Structures and Patterns in Macbeth.

2. Literature Review

2.1. Corpus Linguistics as a Method in Literary Analysis

Corpus linguistics can play an important role in literary analysis, especially the analysis of Shakespearean plays, by offering empirical data regarding the use of language and features of style. Corpus stylistics has been applied to soliloquies and has demonstrated a unique set of linguistic features: the high use of vocabulary of mental states and first-person pronouns, which deepens the characterization and increases the emotional intensity (Murphy, 2015). Also, the comparative studies with the playwrights of the time reveal the presence of both distinctive and generalized language features, which contribute to the generalization of the stylistic register of Early Modern English drama (Demmen, 2012). In addition, the analysis of the keywords in particular texts such as the Romeo and Juliet reveals the power of corpus analysis to identify the important thematic and functional elements of the text that can be easily missed when the literary piece is considered in conventional literary criticism (Ha, 2013).

The investigation of lexical patterning within the tragedies, including the investigation of such significant words as love and grief, also helps to see how the subject matter development and the character formation are influenced by repeated lexical exposure (Naqvi & Bhatti, 2025). On the whole, corpus linguistic techniques also contribute to the study of Shakespearean works by unveiling subtle language processes and contextual interactions in the texts (Keshabyan, 2009).

2.2. Applications of Corpus-Based Analysis in Shakespearean Studies

The corpus analysis has made a great contribution to the study of Shakespeare by using computational approaches to study the language, style, and themes in his works. As an example, Naqvi and Bhatti (2025) use Sketch Engine to examine lexical patterns in Shakespearean tragedies and find out how certain words like love and grief help to develop characters and add depth to the themes used by them in terms of frequency and collocational patterns. On the same note, Craig and Greatley-Hirsch (2017) do not limit computational methods to the authorship attribution, but instead, explore linguistic variation among the plays of Shakespeare and, in so doing, identify differences in verse and prose styles, which add to the history of literature.

According to the analysis conducted by Choi (2024) about Richard III with Voyant Tools, it is possible to observe that there is a complex character of love in the play in terms of thematic importance and the lack of charity between the characters. Moreover, Ha (2013) shows how the analysis of keywords in Romeo and Juliet can be used to identify additional meanings and structural characteristics by applying the tool of corpus linguistics. Taken together, these works demonstrate the radical nature of digital tools in Shakespeare studies.

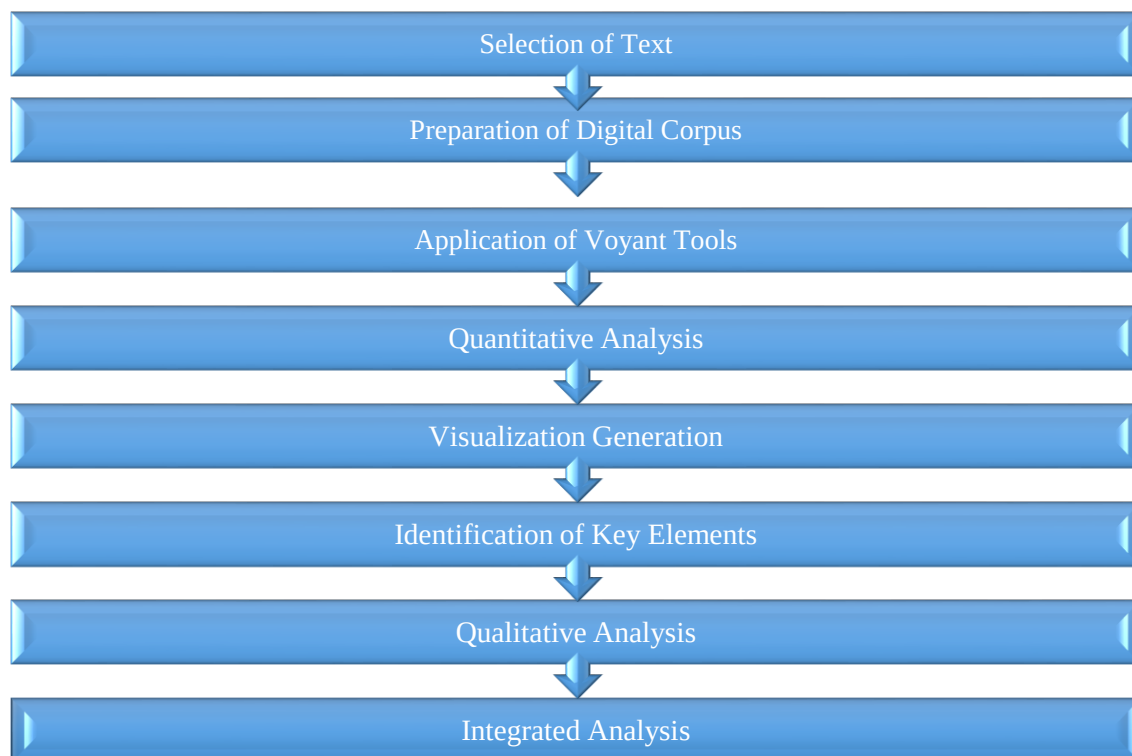
3. Research Methodology

The research method employed in this study is a mixed-method approach as the themes and characters of Macbeth are analyzed through a combination of quantitative and qualitative analysis. The qualitative section will include the interpretation of these visualizations as related to the themes and the story of the play. The quantitative part includes analyzing the frequencies of words and producing visualizations with the help of Voyant Tools. The primary goals of the quantitative analysis are to determine which words are the most widely used in Macbeth and plot their appearance throughout the text. By building word clouds and frequency lists, we are able to identify the primary themes and characters in the play.

Voyant Tools is the primary research instrument of this project; it is a powerful text analysis online application designed with regard to the digital humanities. With a range of tools, such as word clouds, trend graphs, contextual concordances and more complex visualisations than mere word frequency counts, Voyant Tools can help researchers to perform a wide range of different text analysis tasks. It is possible to study

thematic structures and patterns in texts interactively and in a user-friendly manner using this tool.

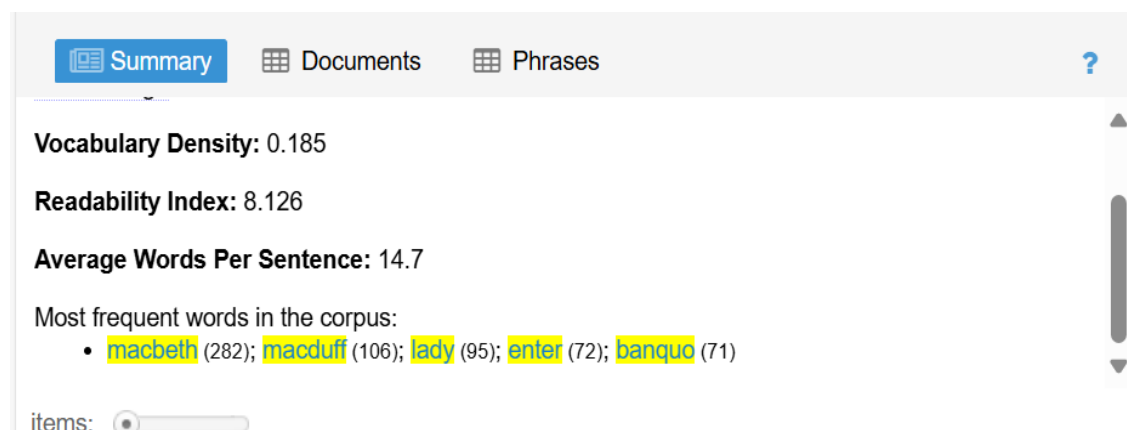
Figure 1: Flow Chart of Research Methodology



4. Results

The vocabulary density, readability, sentence structure, and frequency of words in Macbeth were the key results of the analysis of this play with the help of Voyant Tools. The vocabulary density of 0.185 shows that there is a moderate lexical variety in the work, and it is not very complex but rather rich, characteristic of the Shakespearean works. This equilibrium gives an opportunity to have a wide range of expression of themes and characters, and at the same time to be easily accessible to the reader.

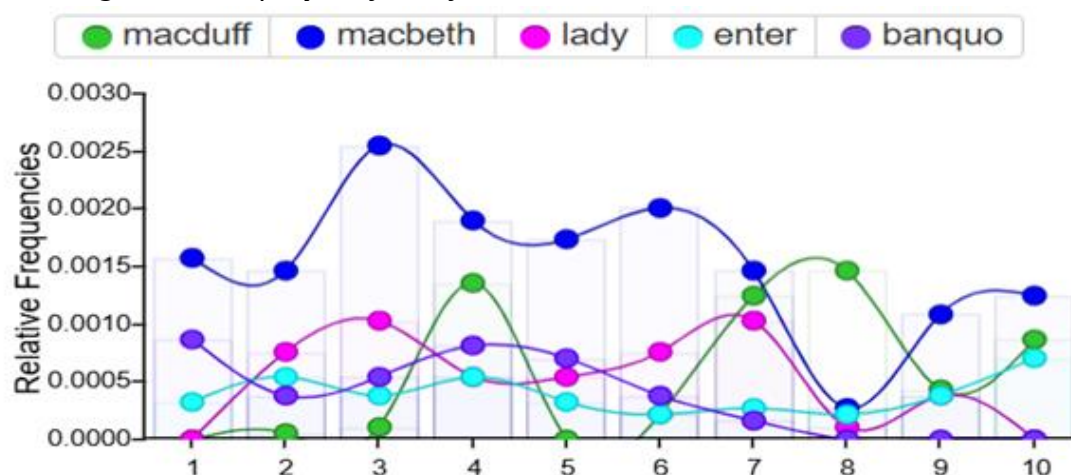
Figure 2: Summary



The difficulty of the text, as indicated by the readability index of 8.126, is in line with the expectation that Shakespearean plays containing the texts in the Early Modern English language are moderate to high in difficulty to read by modern readers. According to this index, the language of Macbeth is complex, but it can be easily understood, which is consistent with the dramatic and thematic nature of the play.

The mean number of words in each sentence in Macbeth is 14.7, and it shows that Shakespeare managed to express deep meaning and emotion. This brevity also helps to sustain the dramatic flow and intensity of the play so that the dialogue will be interesting and effective.

Figure 3: The frequency analysis



The frequency analysis showed that the most common words in the corpus are; macbeth (282 occurrences), Macduff (106 occurrences), lady (95 occurrences), enter (72 occurrences) and Banquo (71 occurrences). It is emphasized that the most frequent word was macbeth and this highlights the fact that the titular character is central to the play. His huge dominance in the text is a manifestation of his dominant position and the immense attention paid to his psychological and moral path. The second word that appears the most

is the word, Macduff, which suggests the important role played by this character as the main antagonist and subsequent foe of Macbeth. The repetition of the phrase by Macduff highlights his significance in the plotting in the story and his significant role at the climax and denouement of the play. The recurrence of the name lady is indicative of how important Lady Macbeth is in the play. Her effect on Macbeth and her madness also play a pivotal role in the plot and her mentioning so many times emphasizes her multifaceted personality and the key role in the drama. The presence of the word “enter” as one of the most repeated words is an indicator of the dynamic staging of the play and constant changes of scenes and characters. This repetitive stage direction highlights the constant motion and action that propel the story. Finally, the word of Banquo that is used repeatedly emphasizes the importance of Banquo.

**Table 1: Most repeated 50 terms**

Come	54
Ross	53
Witch	52
Good	52
Hath	50
Time	45
Like	42
Say	40
Make	39
Lord	38
Tis	37
Let	36
Night	35
Fear	35
Murderer	34
Man	34
I'll	34
Duncan	34
Lennox	33
Know	33
King	33
Exeunt	33
Great	31
Speak	30
Sir	29
Doctor	29
Second	28
Sleep	26
Siward	26
Thane	25

Son	24
Look	24
Castle	24
Blood	24
Things	23
Hand	23
Exit	23
Heart	22
Cawdor	21
Till	20
Nature	20
Death	20

Figure 4 and Table 1, together, provide the most common words of Macbeth by giving a quantitative description of the words lexical patterns of the play. A visual representation of the word cloud allows highlighting the most common words, including Macbeth, Macduff, lady, and Banquo, which were the protagonists and structure of the play. In line with this, Table 1 gives definite numbers of the most repeated words (50) that validate the priority of narration of major characters, thematic ideas such as blood, fear, and death, stage directions, like enter, and exeunt. Collectively, these results show that the computational tools are useful in bringing out character dominance, theme as well as dramatic structure in the text.

Table 2. *Character representation in frequency*

Terms	Count	Terms	Count
Macbeth	282	Fleance	15
Macduff	106	Donalbain	15
Banquo	71	Gentlewoman	12
Lady macbeth	71	Angus	10
Malcolm	58	Porter	8
Ross	53	Menteith	7
Duncan	34	Young Siward"	7

Siward	26	Hecate	7
Lady macduff	21	The three witches	3

Figure 5: Character representation in frequency

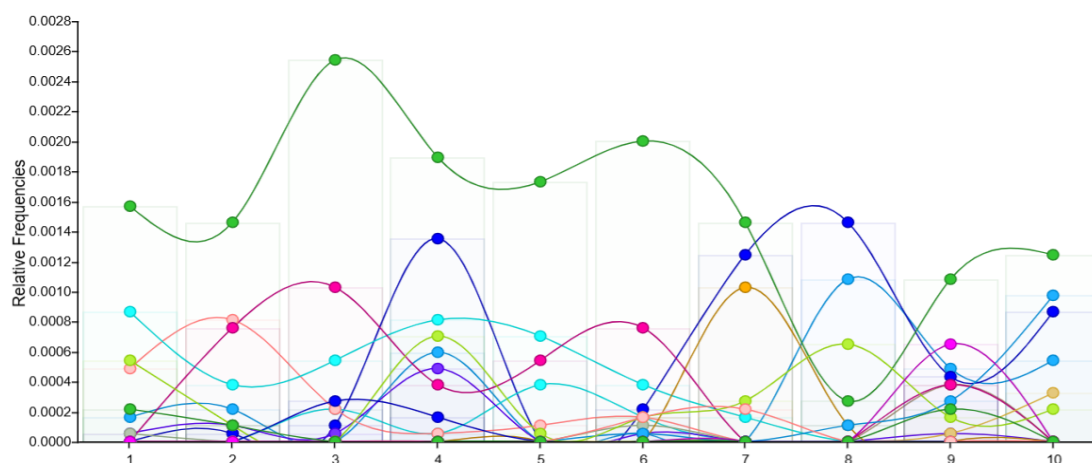


Table 2 and Figure 5, presents the frequency distribution of characters in Macbeth with evident difference in the prominence of characters in the play. There are a few characters that have very high frequency, which means that they play a central role in advancing the storyline, whereas the number of other characters, and they appear less frequently, which means that they have a supporting or a symbolic role in the play.

Table 3. Themes representation in frequency

Theme	Related Word	Frequency	Theme	Related Word	Frequency
Ambition	Ambition	3	Supernatural	Witches	8
	Power	11		Ghosts	6
	King	33		Spirits	5
	Throne	4		Prophecy	1
	Success	4		Magic	1
	Crown	6		Weird	6
	Rise	6		Unnatural	3
Guilt and Conscience	Guilt	2	Manhood and Gender	Man	34
	Blood	24		Woman	10
	Wash	3		Lady	95
	Sleep	26		Wife	13
	Fear	35		Husband	4
	Dagger	3		Courage	3
	Murder	9		Brave	2

Fate and Free Will	Fate	6	Violence and Death	Kill	3
	Destiny	1		Murder	9
	Prophecy	1		Death	20
	Witches	8		Blood	24
	Stars	2		Sword	13
				Fight	12
				Battle	1
				Slain	4

The frequency of the key themes in Macbeth is given in Table 3 in terms of the attached lexical items. Recurring words related to power, morality, destiny, and conflict represent themes such as ambition, guilt and conscience, fate and free will, violence and death, the supernatural, and manhood and gender. The difference in word frequencies outlines the prevalence of some thematic issues over the others. In general, the table shows that the thematic structure of the play

5. Discussion

The findings of this study demonstrate that Voyant Tools is very helpful in determining the central characters, themes and language features of Macbeth. The frequency results clearly demonstrate that Macbeth is the main character in the play, appearing 282 times. This frequency shows that the play is primarily based on Macbeth's ambition, rise, guilt, fear, and fall. Macduff also has a high frequency, which demonstrates his role in the play as Macbeth's chief antagonist, and the one who brings about justice.

The high-frequency words "king," "power," "crown" and "thane" are evidence of ambition. These words indicate that power is a major topic of the play. Macbeth wants to be king, and this ambition takes him to kill Duncan, but it also consumes his sanity, friendships and country. The corpus results support the traditional view that Macbeth's tragic flaw is his uncontrollable ambition.

Guilt and conscience is also a prominent theme in the results. The words "blood," "sleep", "fear", "murder" and "wash" are repeated frequently. This is related to the psychological struggle of Macbeth and Lady Macbeth in the aftermath of the murder. The word "blood" is significant for it represents guilt, violence and corruption. And the word "sleep" symbolizes the loss of innocence. This demonstrates the Shakespeare's use of imagery to show psychological consequences of crime.

Supernatural is another key theme of the play. The words "witches," "ghosts," "spirits," "weird" and "unnatural" make us aware of the supernatural influence in Macbeth's behavior. Macbeth is influenced by the witches' prophecies, but on the other hand, the consequences also reveal that Macbeth chooses to pursue his ambition. Thus the play is about the tension between fate and choice. Macbeth is tempted by prophecy, but he is responsible for his own actions.

The frequency of characters also demonstrates the role of Lady Macbeth. Her name is not used as frequently, but she has a significant role. She manipulates Macbeth to do what she wants him to do in the beginning of the play, and tells him to kill Duncan. But in the end she is weak and mad with guilt. This enhances the theme of gender and manhood, as the play explores definitions of bravery, courage, masculinity and weakness.

The research also demonstrates digital tools can aid the process of traditional literary analysis. Voyant Tools makes it easier to see patterns, words and themes that might be missed when reading a text. However, the tool cannot explain meaning by itself. The researcher still needs to interpret the results carefully. For instance, the frequency of the word "blood" makes sense when it is associated with guilt, murder, violence in the play.

In conclusion, the results demonstrate that ambition, guilt, violence, fate and the supernatural are prominent themes in Macbeth. The corpus analysis provides confirmation for many of the conventional interpretations of the play, but it also offers quantitative evidence. Hence, this research demonstrates that digital humanities methods, such as Voyant Tools, can complement traditional literary analysis by offering statistical data.

6. Conclusion

This article explores Macbeth through a corpus-based analysis with Voyant Tools. The findings demonstrate that new technologies can aid us in understanding the language, characters and themes of the play. The frequency analysis showed that the play is focused on Macbeth, other important characters include Macduff and Lady Macbeth. This confirms that the plot mainly centers on Macbeth's ascent and downfall.

The thematic analysis also reveals other themes like ambition, guilt, supernatural and violence. Words like "blood," "fear" and "murder" reflect the psychological turmoil of characters, while words like "king" and "power" reflect the theme of ambition. The supernatural characters such as witches and ghosts highlight the role of fate and prophecy.

This research also reveals that Shakespeare's language is simple but effective. The relatively low vocabulary and sentence density makes the play rich, but not too difficult. This contributes to the drama of the play. In conclusion, this study shows that this kind of research using digital tools and traditional literary analysis offers new knowledge about literary works. Voyant Tools helps researchers discover patterns and trends, and humans interpret their significance. So, digital humanities is a valuable way of learning about literature, particularly the classics such as Macbeth.

7. Recommendations

1. The same approach can be used to study other plays such as Hamlet, Othello and King Lear to analyze themes, language and characters.
2. Future studies can involve sentiment analysis, topic modeling and machine learning to identify tone, emotions and patterns in the text.

3. Further research can look at character interactions using network analysis and collocation data.
4. Educators can employ digital methods such as Voyant Tools in the classroom to enrich students' reading experience.
5. Future work can involve multiple texts or historical corpora to analyze language variation and thematic connections in different works.

References

- Bondi, M., & Sezzi, A. (2018). "Come what come may, Time, and the Houre, runs through the roughest Day" Temporal phraseology and the conceptual space of futurity in Macbeth. *English Text Construction*, 11(1), 81–104. <https://doi.org/10.1075/ETC.00005.BON>
- Burch, B. D., & Egbert, J. (2020). Zero-inflated beta distribution applied to word frequency and lexical dispersion in corpus linguistics. *Journal of Applied Statistics*, 47(2), 337–353. <https://doi.org/10.1080/02664763.2019.1636941>
- Choi, S. (2024). “그대는 자선의 원칙을 모르는구려”: 보이언트 도구를 활용한 셰익스피어 『리처드 3세』에 나타난 ‘사랑’의 코퍼스 분석. 60, 89–125. <https://doi.org/10.25093/ibas.2024.60.89>
- Craig, H., & Greatley-Hirsch, B. (2017). *Style, Computers, and Early Modern Drama: Beyond Authorship*. <https://www.cambridge.org/core/books/style-computers-and-early-modern-drama/D69296E35A5E3C510B1AA0E46F46E4A1>
- Demmen, J. (2012). *A corpus stylistic investigation of the language style of Shakespeare's plays in the context of other contemporaneous plays*. <https://eprints.lancs.ac.uk/id/eprint/133553/>
- Ha, M.-J. (2013). Corpus-Based Literary Analysis. *The Journal of the Korea Contents Association*, 13(9), 440–447. <https://doi.org/10.5392/JKCA.2013.13.09.440>
- Keshabyan, I. (2009). Analysing the Concepts of Vengeance and Hono(u)r in Shakespeare's Hamlet and Sumarokov's Gamlet: A Corpus-based Approach to Literature. *International Journal of English Studies*, 9(3), 235–258. <https://doi.org/10.6018/IJES.1.1.99621>
- LI, W. (2010). The Research Perspectives of Corpus Linguistics. *Journal of PLA University of Foreign Languages*. https://en.cnki.com.cn/Article_en/CJFDTotal-JFJW201002009.htm
- Lutz, M. (2019). "Strange Intelligence": Humans, Nonhumans, and History in Macbeth. *Modern Philology*, 117(2), 194–213. <https://doi.org/10.1086/705438>
- Mazzola, E. (2023). Noise, Sound and Fury in Macbeth. *Critical Survey*, 35(1), 1–18. <https://doi.org/10.3167/cs.2023.350101>

- Meyer, C. F. (2012). *Textual analysis: from philology to corpus linguistics* (pp. 23–42). Brill. https://doi.org/10.1163/9789401207935_004
- Murphy, S. E. (2015). I will proclaim myself what I am: Corpus stylistics and the language of Shakespeare's soliloquies. *Language and Literature*, 24(4), 338–354. <https://doi.org/10.1177/0963947015598183>
- Naqvi, U.-L., & Bhatti, Z. I. (2025). *Lexical patterning in Shakespearean tragedies: a corpus-based analysis using sketch engine through the lens of lexical priming theory*. 3(3), 1360–1375. <https://doi.org/10.63878/cjssr.v3i3.1104>
- Nostbakken, F. (1997). *Understanding Macbeth*. Greenwood. <https://doi.org/10.5040/9798216029588>
- Preeti, Dr., Verma, Dr. J., R, L., J, D., & Bheemraj. (2024). *Quantitative Analysis of Literary Texts: Computational Approaches in Digital Humanities Research*. <https://doi.org/10.53555/kuey.v30i5.3770>
- Puspitasari, F., Efendi, A., Pratama, R. A., & Chairu, T. (2025). From close to distant reading: digital humanities in literary studies. *Literature & Literacy*, 3(1), 12–20. <https://doi.org/10.21831/litlit.v3i1.1949>
- Roy, A. K. (2025). Ambition and Its Consequences: A Study of Macbeth. *International Journal of Language, Literature and Culture*, 5(3), 63–68. <https://doi.org/10.22161/ijllc.5.3.10>
- Skichko, A. S. (2023). Corpus driven analysis of motivational speeches with antconc. *Scientific practice: modern and classical research methods*. <https://doi.org/10.36074/logos-26.05.2023.050>
- Soenarto, I., & Wibawarta, B. (2023). Shakespeare's tragedy of macbeth as a strange attractor for theater performances and cinematic adaptations. *International Review of Humanities Studies*, 8(2). <https://doi.org/10.7454/irhs.v8i2.1124>
- Thakre, S. S. (2024). *echoes of the future: bridging tradition and innovation in digital humanities and english literature*. <https://doi.org/10.29121/shodhgyan.v2.i1.2024.21>
- Vinodini, S., & Raj, A. P. (2024). *Enhancing Literary Analysis through Artificial Intelligence and Machine Learning: Insights from "The Alchemist."* 136–141. <https://doi.org/10.1109/icrais62903.2024.10811737>
- Vanathi, A. (2023). Exploring Literary Studies through Digital Humanities. *Shanlax International Journal of English*. <https://doi.org/10.34293/rtdh.v12is1-dec.61>