

## CORPUS STYLISTIC EXPLORATION OF LEXICAL PATTERNS, SEMANTIC FIELDS, AND CONTEXTUAL MEANING IN DANTE'S *DIVINE COMEDY*

**Tahira Perveen**

M.Phil Scholar, NCBA&E Alhamra University Bahawalpur

Email: [ashraftahira77@gmail.com](mailto:ashraftahira77@gmail.com)

ORCID Id: <https://orcid.org/0009-0009-1057-8010>

**Hira Munawar**

M.Phil Scholar, NCBA&E Alhamra University Bahawalpur

Email: [hirayazman311@gmail.com](mailto:hirayazman311@gmail.com)

ORCID Id: <https://orcid.org/0009-0008-5491-2239>

**Mahnoor Zahra**

M.Phil Scholar, NCBA&E Alhamra University Bahawalpur

Email: [mahnoorkhanblouch09@gmail.com](mailto:mahnoorkhanblouch09@gmail.com)

ORCID Id: <https://orcid.org/0009-0007-0581-3667>

**Dr. Muhammad Arfan Lodhi\***

**\*Corresponding Author**

Higher Education Department, Punjab

Email: [samaritan\\_as@hotmail.com](mailto:samaritan_as@hotmail.com)

ORCID Id: <https://orcid.org/0000-0002-5627-5666>

### **Abstract**

*This study observes the lexical, semantic, and contextual patterns in Dante Alighieri's Divine Comedy through a digital humanities approach using Voyant Tools. The purpose of the study is to discover how computational text analysis procedures can improve traditional literary interpretation by revealing systematic linguistic and thematic structures across the three canticles Inferno, Purgatorio, and Paradiso. To attain this, the research employs a qualitative textual analysis method reinforced by digital tools, concentrating on four stages: corpus preparation, descriptive analysis, lexical and frequency analysis, and contextual-semantic exploration. The poem was transformed into a clean textual corpus and uploaded into Voyant Tools, where visualizations and quantitative actions such as term frequency, word clouds, trend graphs, collocate networks, phrases, and keyword-in-context (KWIC) were generated. The conclusions show that Dante's language displays a clear lexical evolution aligned with the divine journey described in the narrative. Negative and concrete terms related to suffering, darkness, and sin control Inferno, whereas transitional vocabulary illustrating purification, prayer, and moral growth appears in Purgatorio. In contrast, Paradiso structures highly abstract and theological language centered on light, divine love, harmony, and spiritual imagery. Semantic networks show strong collections around key concepts such as "love," "light," "grace," and "justice," while contextual analysis demonstrates how these terms shift in meaning across the poem. The study concludes that Voyant Tools effectively highlights patterns that support close-reading traditions and contributes measurable and reproducible insights into Dante's symbolic vocabulary and thematic style. The combination of digital humanities improves the interpretative depth of classical literary works and determines the value of computational methods for analyzing complex poetic texts.*

### **Key Words:**

*Digital Humanities; Voyant Tools; Dante Alighieri; Divine Comedy; Lexical Analysis; Semantic Analysis; Contextual Analysis; Text Mining.*

## 1.1 Introduction

In recent years, the study of literature has been significantly reshaped by the development of Digital Humanities (DH), a field that brings together computational tools and traditional interpretive methods. Instead of depending solely on close reading, researchers now use techniques such as text mining, visualization, and automated linguistic analysis to examine literature at broader scales (Berry & Fager Jord, 2017). These methods help scholars detect lexical, semantic, and thematic patterns that may remain hidden through conventional approaches. Franco Moretti's concept of distant reading further supports this shift by encouraging an expanded view of texts while still valuing the insights gained from detailed analysis (Moretti, 2013).

This study applies such digital methods to Dante Alighieri's *The Divine Comedy*, a medieval poem noted for its symbolic richness and elaborate structure. Its linguistic complexity makes it particularly suitable for computational exploration. Using Voyant Tools—a widely used online platform for text analysis—this research investigates the poem's lexical tendencies, semantic networks, and contextual developments across *Inferno*, *Purgatorio*, and *Paradiso*. Voyant features such as Cirrus, Trends, Contexts, Summary, Links, and Phrases assist in identifying key vocabulary, tracing conceptual progressions, revealing clusters of related ideas, and examining how meanings shift across different narrative sections (Sinclair & Rockwell, 2016). Because digital analysis can be affected by text preparation and tool settings, careful methodological decisions are essential. Choices regarding stop-word removal, edition selection, standardized spellings, or translation versions can all influence outcomes. Scholars such as Dobson (2019) caution that while digital visualizations can illuminate important textual patterns, they can also oversimplify results if not paired with critical interpretation. For this reason, the present study combines quantitative findings from Voyant with qualitative, context-based interpretation grounded in Dante's theological and literary background. By merging computational visualizations with traditional analysis, this research aims to deepen our understanding of Dante's language and demonstrate how DH methods enhance literary interpretation.

## 1.2 Background of the Study

One of the most important works of literature from the Middle Ages is Dante Alighieri's *Divine Comedy*, which blends politics, philosophy, theology, and poetical artistry into a cohesive spiritual experience. The poem, which was composed between 1308 and 1321 when Dante was living in exile, examines the soul's journey through Hell, Purgatory, and Heaven and provides a metaphorical representation of moral change and divine justice. Dante made a groundbreaking decision by writing in the Tuscan vernacular instead of Latin, which helped to establish Italian as a literary and cultural language (Digital Dante, 2023). Dante is still regarded by contemporary researchers as a philosopher, political theorist, and linguistic pioneer whose writings influenced European intellectual history in addition to being a poet (Coggeshall, 2022).

The *Divine Comedy* has attracted increased interest in the realm of digital humanities in recent years. The development of digital academic platforms, like Columbia University's Digital Dante project, has made it possible to use digital technologies for interactive reading, visualization, and critical analysis of Dante's work (Digital Dante, 2023). Additionally, large digital databases of medieval commentary on the *Comedy* have been assembled by computational efforts like the Dante 2025 Corpus Coda program, enabling the study of about five million encoded words connected to Dante's work (Finestrasulfate, 2025). These online resources show how technology is increasingly being used to improve accessibility, analysis, and interpretive depth in classical literary studies.

Additionally, digital textual analysis offers fresh approaches to studying Dante's language. According to research, linguistic tendencies, thematic clusters, and stylistic patterns that would be difficult to spot through conventional close reading can be revealed by computer methods (Robey, 1993). The DH2020 Conference's "From Data to Visualization: Dante's Divine Comedy as a Case Study" digital humanities poster demonstrated how computational methods can map lexical patterns, structural divisions, and semantic relationships throughout Dante's canticle, opening up new interpretive avenues. Furthermore, new computational research suggests that digital techniques can be used even for the poem's formal metrics, as demonstrated by Asperity and Dal Bianco's (2020) probabilistic analysis of Dante's syllabic patterns.

Few publications explicitly use widely available text-analysis technologies, such as Voyant Tools, to analyze lexical, semantic, and contextual patterns in the Divine Comedy, despite the growing popularity of digital Dante studies. Voyant Tools' visible, interactive, and replicable features like frequency analysis, collocation networks, and KWIC contexts can reveal new insights into Dante's symbolic lexicon and thematic organization, which makes this gap significant. By using Voyant Tools in a digital humanities framework to examine the poem's linguistic and conceptual structures and show how computational approaches improve conventional literary interpretation, this study seeks to close this research gap.

### 1.3 Rationale of the Study

This study's objective is to close the gap between computational text analysis and predictable literary criticism. Even though Dante's writings have been observed for centuries, not much has been done to use digital techniques to study them from a lexical-semantic perspective. The purpose of this study is to show how Voyant Tools may use statistics and visualization to expose hidden textual linkages and promote deeper understanding.

### 1.4 Research Questions

1. What are the predominant lexical and thematic patterns in *The Divine Comedy*, and how can digital text analysis tools such as Voyant facilitate their identification and interpretation?
2. In what ways do the semantic relationships within *The Divine Comedy* reflect and reinforce its underlying moral and theological architecture?
3. How can the application of digital humanities methodologies, particularly Voyant analysis, enhance the contextual and interpretative understanding of Dante's linguistic and symbolic framework?

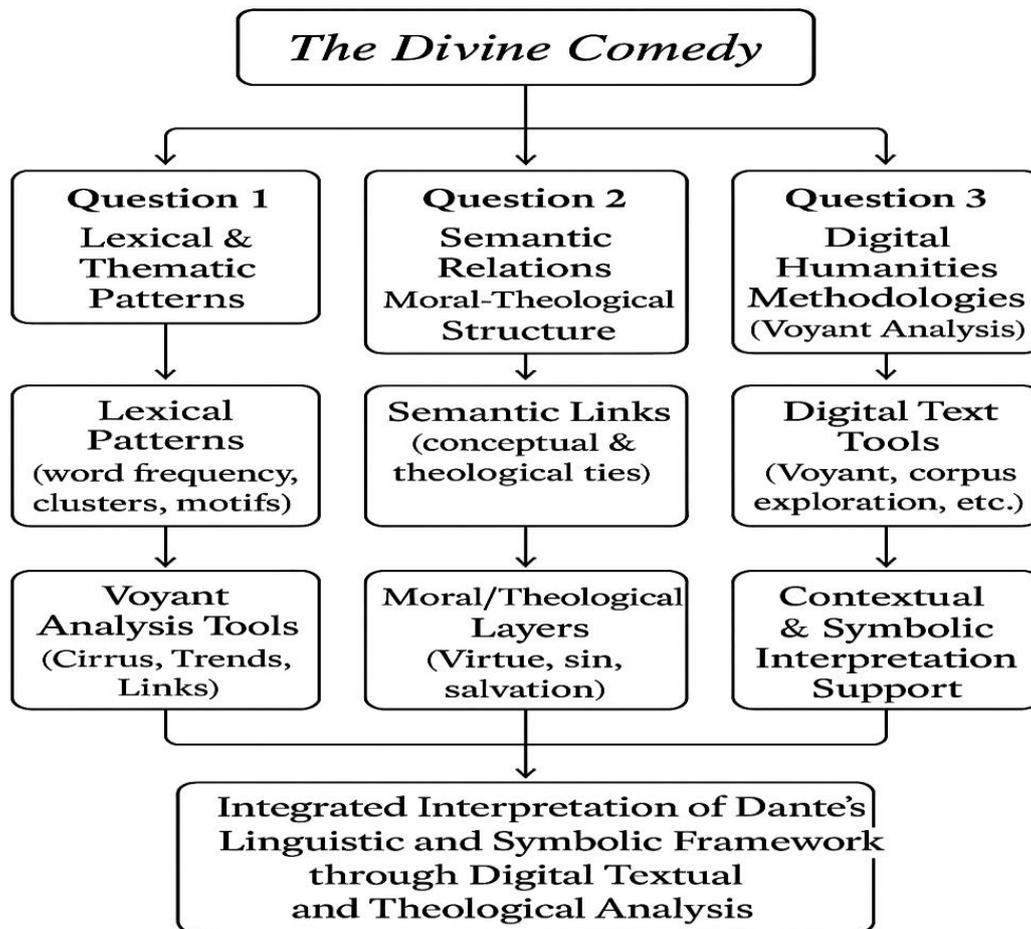


Figure 1. Concept map

### 1.5 Purpose of the Study

Through digital text analysis with Voyant Tools, this project aims to explore the lexical, semantic, and contextual patterns in Dante Alighieri's *The Divine Comedy*. Using both quantitative and qualitative approaches, the study seeks to understand how Dante's terminology, thematic connections, and contextual applications of important ideas support the poem's overall moral and spiritual structure. Using the three canticles of *Inferno*, *Purgatorio*, and *Paradiso*, this study aims to show how digital humanities tools can improve comprehension of classical literature by exposing hidden linguistic patterns, recurrent symbolic expressions, and meaning shifts. The study's overall goal is to combine digital humanities methods with traditional literary analysis to create a more thorough and data-driven reading of *The Divine Comedy* that provides fresh perspectives on Dante's linguistic patterns, thematic progression, and narrative setting.

### 1.6 Significance of the Study

This study is noteworthy because it examines a medieval literary masterwork using digital text analysis. It combines computer methods with classical hermeneutic reading to uncover deep semantic implications (conceptual links, symbolic contrasts) as well as surface-level patterns (word frequency, collocation). The study also adds to the expanding field of digital literary criticism by demonstrating how Voyant Tools might help contemporary readers understand Dante's complex structure and ideas by offering factual evidence for textual interpretation.

## 2. Literature Review

As researchers incorporate computational methods into literary studies to supplement traditional hermeneutics with reproducible, extensive textual evidence, the field of digital humanities (DH) has swiftly expanded. Because they offer a user-friendly, web-based environment for lexical, collocational, and contextual analysis of texts without significant technical obstacles, tools like Voyant Tools have become essential to humanities instruction and research (Voyant Tools official site). Researchers can go beyond word counts to more significant semantic and contextual inferences with Voyant's suite, which includes Citrus (word cloud), Summary (lexical statistics), Trends (frequency across text divisions), Contexts (KWIC), Links/Collocates (semantic relationships), and Phrases (n-grams). Voyant's importance in classroom and project workflows is emphasized in recent methodological reflections, which also highlight how useful it is for exploratory analysis that can support more complex computational approaches (Programming Historian lecture; Voyant project retrospectives).

DH pipeline analyses of canonical texts show how close reading can be enhanced by quantitative outcomes. Scholars have used a variety of computational techniques, such as topic modelling, sentiment analysis, stylometry, and visualization, to analyze the distribution of theological vocabulary, thematic progressions (such as sin → purification → beatific vision), and genre differentiation among canticles in Dante's *Divine Comedy* (stylometric investigations; DH conference case studies). According to these investigations, Dante's three canticles—*Inferno*, *Purgatorio*, and *Paradiso*—present measurable lexical registers that correspond with established interpretations of the poem's tonal and thematic changes. Thus, quantitative data serves as both a confirmatory and an exploratory tool, exposing minor lexical or collocational trends that warrant new interpretation while simultaneously validating historically grounded readings. Particular computer analyses of *The Divine Comedy* show how various approaches emphasize different interpretive outcomes. Literary claims about register and genre within a single epic project are supported by stylometric and multivariate frequency studies (previous work and follow-ups), which have isolated feature sets that differentiate the canticle. These studies show that Dante's diction and rhetorical approach vary systematically between hell, purgation, and paradise. While sentiment analysis projects have started mapping affective valence across cantos to model the poem's emotional arc, topic modelling and NLP-based techniques (such as Steier's NLP study and later character/topic studies) have been used to uncover latent themes and compare the Italian originals with translations. These methodologies show that corpus-level approaches can both establish new questions (e.g., regarding character networks or how specific virtue/vice lexemes cluster across the work) and support current interpretive narratives.

Voyant is an easy-to-use yet research-capable platform that has been utilized for quick, repeatable lexical and contextual queries in both research projects and educational settings. Updated Voyant lessons (2024–2025) have been republished in *Programming Historian* and other DH training materials. These lessons focus on how researchers should prepare corpora, normalize orthography, and responsibly interpret visualization outputs, particularly with historical texts where editorial history, translation layers, and spelling variation are important. When it comes to high-stakes claims, these instructional materials emphasize best practices (data cleaning, transparent parameter selections) and present Voyant as suitable for preliminary exploratory research that should be followed by more thorough NLP or statistical modeling. Regarding digital editions and textual resources of the *Divine Comedy*, organizations such as Columbia's Digital Dante offer reliable text encoding, commentary, and intertextual materials are essential for Dante studies. Reproducibility is aided by using trustworthy digital editions: Voyant analyses can be linked to validated transcriptions or canonical TEI-encoded texts to

guarantee that lexical counts and KWIC contexts represent reliable editorial decisions. In order to prevent inaccurate quantitative conclusions based on inconsistent input texts, recent projects that digitally reproduce Dante manuscripts or produce visualizations of the *Commedia* highlight the significance of integrating textual scholarship (manuscript tradition, editorial variants) with computational workflows.

The method's potential is further demonstrated by recent empirical work that applies computational analysis to the *Divine Comedy*. Other works have concentrated on creating data visualizations that map lexical density and thematic emphasis between cantos, while studies have modified NLP pipelines for character and sentiment analysis to track the poem's emotive and interpersonal dynamics. For example, automated visualizations of Dante's lexical and semantic fields are presented in conference presentations and peer-reviewed papers from 2020–2024. Interpretive statements regarding the poem's theological architecture have been linked to these outputs. Before proceeding to more intricate analyses like topic modeling or network analysis, separate statistical signatures across the three canticles have been verified by parallel work in stylometry. These signatures can be visualized and questioned using Voyant.

The literature's methodological critiques warn against methodological overreach, pointing out that word clouds and simple frequency counts can be misleading if they are not compared to context and editorial concerns. Lemmatization, when necessary, sensitivity to translation impacts (when working on translated versions), and text normalization (orthography, punctuation) are all important considerations when using Voyant's tools. These limitations are emphasized in Programming Historian materials and DH pedagogical research, which suggest mixed-methods designs that validate Voyant's exploratory outputs by combining them with manual interpretative close-reading, deeper KWIC analysis, and collocation statistics. This protects against mechanical interpretations that overlook subtleties, which is crucial for a work as historically complex and intertextual as Dante's epic.

Lastly, transparency in scholarship and reproducibility are key topics in the most recent DH discussions. Open tools, shared corpora, and defined workflows (storing Voyant sessions, sharing query parameters) are essential to ensuring that computational interpretations are reusable and credible, according to recent reflections on Voyant's legacy and updated lessons. Scholarship that combines Voyant-driven visualization with TEI-encoded digital editions and pays close attention to manuscript variance is currently the finest practice in the discipline, especially for Dante studies. In the future, research will likely focus on more integrated pipelines, which combine Voyant for quick exploration with subject modelling, network analysis (like Gephi), and sentiment/NLP toolchains for reliable, multi-layered interpretation.

## 2.1 Digital Humanities and Literature

The term "digital humanities" is an interdisciplinary field that combines traditional humanities scholarship, computational techniques, and digital tools to study language, literature, history, and culture in novel ways. DH enables scholars to use data-driven methodologies to comprehend the linguistic patterns, thematic networks, and conceptual structures of a text such as Dante's *Divine Comedy*, going beyond close reading. Large portions of the *Divine Comedy* can be processed by academics using DH platforms like Voyant Tools for lexical, semantic, and contextual analysis. This allows them to see word frequencies, keyword trends, collocations, sentiment patterns, and thematic clusters throughout *Inferno*, *Purgatorio*, and *Paradiso*. This computational method shows how Dante uses symbolic imagery, moral terminology, metaphorical patterns, and vocabulary selections to create meaning. Thus, by providing interactive, visual, and quantitative methods for analyzing Dante's language and theological-philosophical framework, digital humanities expand on traditional literary criticism.

## 2.2 Digital Tools used for textual analysis

### 1. Voyant Tools

Among the most popular web-based tools for beginning and advanced text analysis is Voyant Tools. Collocation graphs, frequency charts, word clouds, and keyword-in-context (KWIC) are among the visualizations it offers. Because of Voyant's ease of use, accessibility, and capacity to swiftly uncover lexical patterns in lengthy books like Dante's *Divine Comedy*, scholars favor it.

### 2. AntConc

A popular professional concordance in corpus linguistics is AntConc. It provides tools for n-grams, clusters, keyword analysis, collocations, and concordances. It is perfect for examining recurring patterns, phraseology, and semantic surroundings in literary works because of its accuracy and adaptability.

### 3. Sketch Engine

A strong corpus-analysis tool for researching both big and specialized corpora is Sketch Engine. Word sketches, frequency lists, collocation profiles, and comparative corpus data are all produced by it. To find intricate lexical patterns and semantic trends in literary and historical texts, researchers use Sketch Engine.

### 4. NLTK (Natural Language Toolkit)

A popular Python-based package for computational linguistics research is called NLTK. Semantic similarity metrics, tokenization, lemmatization, part-of-speech tagging, and collocation extraction are all supported. Building customized textual-analysis pipelines is a typical use for it, particularly when dealing with historical or non-standard languages.

### 5. MALLET

MALLET is a machine-learning toolkit frequently used in digital humanities for topic modeling. It identifies hidden thematic structures within texts, grouping vocabulary into meaningful semantic clusters. For literary studies, MALLET helps reveal narrative themes, conceptual patterns, and stylistic shifts.

## 2.3 Voyant as a digital tool

One of the first thorough reviews of Voyant Tools was carried out by Welsh (2014), who emphasized the tool's instructional value and accessibility for librarians and humanities teachers. The study emphasizes Voyant's easy-to-use interface, which facilitates quick textual pattern investigation and powerful classroom demonstrations. Welsh advises researchers to combine Voyant discoveries with more sophisticated tools to ensure methodological rigor, pointing out that while Voyant works well for exploratory and introductory analysis, it is less appropriate for deep computational work. Laudati (2021) reviewed Voyant Tools in the context of language learning and digital pedagogy, demonstrating how the platform helps novice researchers visualize textual patterns effectively. The study praises Voyant's wide range of visual reports and instant feedback, which promote corpus awareness and enhance students' understanding of word distributions and semantic clusters. However, the author cautions that visually appealing results can sometimes be misleading if not interpreted critically, stressing the need for guidance in reading and contextualizing Voyant outputs.

The Programming Historian lesson by Gutiérrez De la Torre (2025) provides a peer-reviewed and reproducible workflow for using Voyant Tools in digital humanities education. This study outlines detailed steps for corpus preparation, visualization, and analysis, making it a valuable resource for instructors seeking to integrate DH practices into their teaching. While the lesson strengthens reproducibility, it also acknowledges that deeper linguistic modeling such as dependency parsing or machine-learning analysis, requires exporting Voyant data to more

specialized NLP platforms. In another study, tool-criticism research by Jacquot (2023) and van Es (2021) uses Voyant as a case study to argue for greater transparency in digital-humanities methodologies. These scholars point out that Voyant's automated processes, such as tokenization and stop word handling, may hide underlying assumptions that influence results. Although they recognize Voyant's value as a heuristic exploratory tool, they warn against treating its visualizations as objective facts. The studies advocate for documenting preprocessing steps and validating results through philological close reading and complementary analytical tools.

Practical and critical reviews by Genady (2016/2019) and other recent commentators (2023–2025) offer a balanced perspective on Voyant's strengths and limitations. These reviews underscore Voyant's continued relevance due to its rapid prototyping abilities, support for multi-file corpora, and well-developed documentation. At the same time, they note the platform's limitations, particularly for historical or medieval texts like Dante's *Divine Comedy*, where orthographic variation and complex syntax require more careful preprocessing. Scholars recommend that users supplement Voyant with domain-specific resources or NLP tools for more advanced analysis.

#### **2.4 Description of the Core Text**

Composed between 1308 and 1321, Dante Alighieri's *Divine Comedy* is a massive epic poem that tells the story of the pilgrim Dante's spiritual journey through the three afterlife realms of Paradiso, Purgatorio, and Inferno. With the exception of Inferno, which has 34 cantos, each cantica has 33 cantos, forming a symmetrical and numerological symbolic framework. The poem offers a thorough understanding of the journey of the human soul from sin to salvation by fusing political criticism, classical philosophy, medieval theology, and introspection. Purgatorio offers penance and spiritual rebirth; Inferno examines moral decay and divine justice; and Paradiso honors divine love, celestial harmony, and the soul's journey to God. Dante's intention to lead humanity from a state of moral blindness towards spiritual illumination is reflected in this thematic journey. The *Divine Comedy*'s linguistic, cultural, and philosophical contributions account for its literary significance. Dante became known as the "Father of the Italian language" because of his decision to write in the Tuscan vernacular, which profoundly changed the literary standing of Italian.

The poem is considered a foundational work of Western literature because it integrates Christian cosmology, Augustinian philosophy, Thomistic theology, and Aristotelian ethics. The poem's complex allegorical structure and use of symbolic imagery, including light, love, and voyage, to communicate metaphysical truths are often emphasized by academics. These aspects of Dante's work have been the subject of several previous investigations. First, the narrative structure of the poem was examined by Bartolini (1992), who focused on how Dante creates a spiritually transforming trip by fusing personal experience with theological theory. Secondly, Dante's symbolic system was examined by Frescoer (1986), who focused on how he used light as a metaphor for spiritual enlightenment and heavenly truth. Thirdly, a thorough analysis of Dante's linguistic accuracy was given by Hollander (2000), who pointed out that his lexical selections demonstrate both lyrical creativity and medieval scholastic reasoning. Lastly, Boyde (1963) examined Dante's intellectual influences, particularly how Thomistic theology and Aristotelian ethics are combined to form a large portion of the poem's moral framework. These studies show that *The Divine Comedy* is a multifaceted work that integrates intricate linguistic structures, symbolic networks, and philosophical systems in addition to being religious or poetic literature. The poem is particularly well-suited for contemporary methodologies like lexical, semantic, and contextual analysis because of its extensive vocabulary, conceptual depth, and thematic coherence. Therefore, using digital humanities to

analyze the text offers a chance to find new trends, support preexisting interpretations, and get a deeper comprehension of Dante's literary and cultural legacy.

### 3. Methods and Instrumentation

The theoretical design of Digital Humanities (DH) holds that computational technologies can expand, improve, and supplement conventional literary interpretation, serves as the foundation for this investigation. The notion of digital humanities places a strong emphasis on fusing technology with humanistic investigation, enabling the quantitative examination of texts using techniques like word frequency analysis, collocation mapping, concordance, and semantic visualization. This framework backs up the claim that digital technologies enhance close reading by offering methodical proof of linguistic and thematic patterns rather than taking the place of it.

#### 3.1 Research Method

A qualitative textual analysis method backed by a digital humanities approach is used in this work. While the digital humanities paradigm introduces computational tools, specifically Voyant Tools, to systematically explore patterns in the text, qualitative textual analysis allows for an interpretive understanding of The Divine Comedy's language, themes, and symbolism. By combining data-driven techniques with conventional interpretative procedures, this combination enhances the analysis. Lexical frequencies, semantic correlations, and contextual links that would not be apparent through human reading alone can be identified thanks to the digital component. As a result, the study uses a hybrid methodology that combines digital analytical methods with humanistic interpretation.

#### 3.2 Population and Sampling

The complete text of Dante Alighieri's Divine Comedy, a massive epic broken up into three canticles, serves as the study's population. Based on the needs of the study, a purposive sample was chosen from this group. The sample consists of Inferno, Purgatorio and Paradiso to enable structured comparison, Voyant Tools handled these three canticles as three distinct text documents. Separate analysis enables the identification of thematic evolution, lexical alterations, and narrative progression during the spiritual journey from sin to salvation. This sampling technique allows for greater corpus-level insights while guaranteeing that every area of the poem receives an equal amount of analytical attention.

#### 3.3 Theoretical Framework

The analytical process follows a structured four-step framework designed to ensure systematic and comprehensive examination of the corpus. Each step integrates qualitative interpretation with digital textual analysis. Framework of analysis has been developed in the light of previously conducted researches through corpus analysis tool Voyant and Antconc. Jamil et al. (2024) employ Voyant Tools to perform a digital stylistic investigation of Christopher Marlowe's *Doctor Faustus*, examining its lexical, semantic, and contextual dimensions. The analysis reveals dominant lexical clusters such as *God, hell, knowledge, and Mephistophilis*, encapsulating Faustus's moral and intellectual conflict between divine law and Renaissance humanism. Semantic mapping exposes interrelations between pride, sin, and damnation, while contextual trend analysis traces Faustus's psychological transformation from ambition to despair. The study situates Marlowe's tragedy within both Renaissance intellectual history and digital literary discourse, arguing that computational methods can visualize theological and philosophical struggles encoded in the text. It concludes that Voyant enriches literary criticism

by uncovering hidden lexical and thematic structures, offering a paradigm for digitally assisted interpretation of early modern literature.

Bibi et al. (2025) integrate corpus-assisted lexico-semantic analysis with postcolonial literary theory to map the linguistic structure of Conrad's *Heart of Darkness*. By constructing a single-text corpus and processing it through Voyant Tools, the study identifies lexical fields clustered around *darkness*, *river*, and *ivory*, each carrying distinct semantic prosodies that reflect imperial ideology and moral disintegration. Quantitative measures such as keyness, collocation, and dispersion are triangulated with traditional close reading to examine how recurring images construct representations of race, morality, and consciousness. The analysis corroborates Achebe's (1977) critique of racial dehumanization while also empirically supporting Said's (1993) interpretation of imperial decay. The study's hybrid approach exemplifies the fusion of computational precision and humanistic depth, transforming *Heart of Darkness* into a data-driven exploration of moral and ideological structures within colonial modernism.

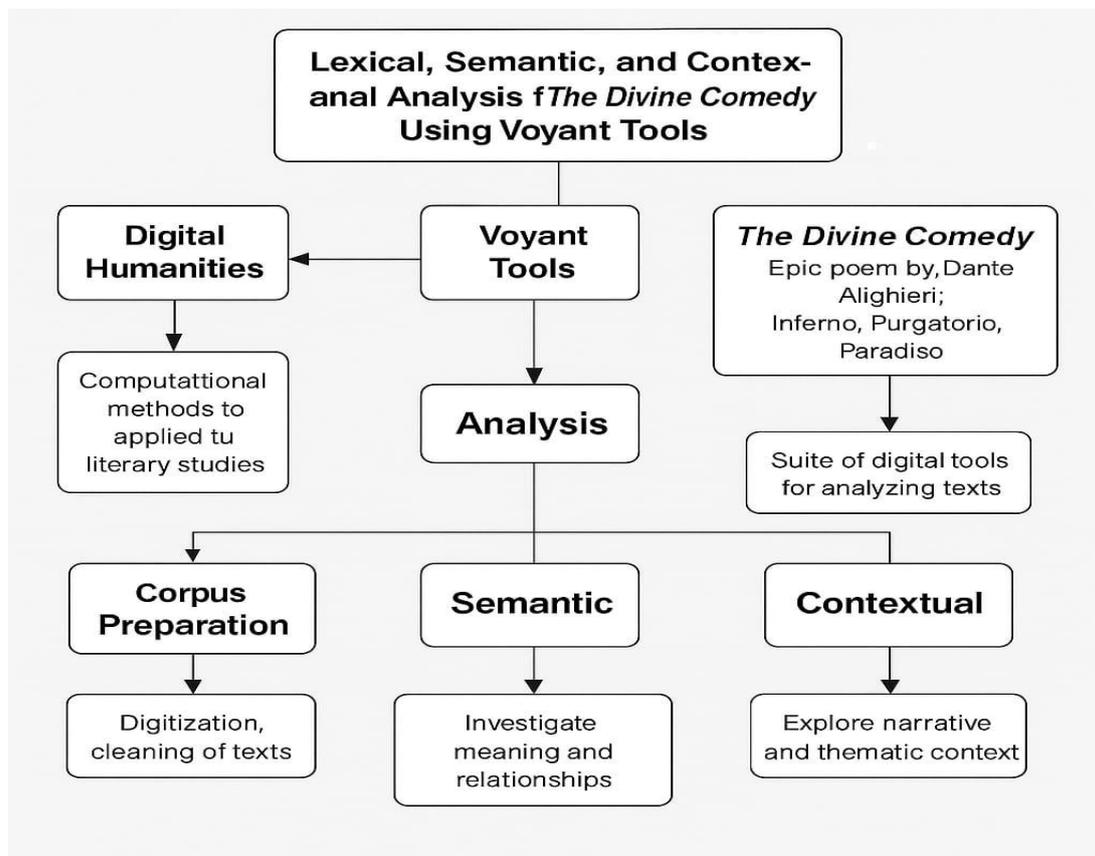


Figure 2. Framework of analysis

Altaf et al. (2025) employ a mixed-methods digital humanities framework to revisit Defoe's *Robinson Crusoe* through a computationally grounded lexico-semantic analysis. Using Voyant Tools, the authors explore how linguistic data reinforces long-standing themes of colonialism, spiritual redemption, and economic individualism. The study finds over 1,800 uses of the possessive pronoun "my," quantitatively underscoring Crusoe's sense of ownership and

colonial control. Lexical clusters relating to “labor” and “providence” reveal the duality of material and spiritual struggle, while contextual collocation analysis highlights hierarchical and racial dynamics through terms like “master” and “savages.” Integrating insights from Moretti’s “distant reading” and Ramsay’s “algorithmic criticism,” the study bridges close reading with corpus linguistics. Its major contribution lies in transforming subjective postcolonial and theological interpretations into data-supported evidence, offering a replicable digital model for canonical English novels.

Sarwar et al. (2025) present a pioneering integration of digital stylistics and classical philology, analyzing *The Iliad* through computational tools, primarily Voyant. The study constructs a digital corpus of over 190,000 words to perform lexical density, frequency, and co-occurrence analyses, identifying dominant clusters such as *Achilles–wrath*, *Hector–honor*, and *Zeus–fate*. These are interpreted as reflections of the epic’s moral and thematic architecture. Combining quantitative visualization with qualitative hermeneutics, the authors demonstrate how computational metrics validate Homeric oral-formulaic theory. The study’s methodological rigor positions digital stylistics as an empirical bridge between philology and literary theory. It contributes significantly to digital humanities in classical studies, showing that computational approaches can reveal the moral syntax of ancient epic poetry.

#### 4. Data Analysis

The lexical, semantic, and contextual analysis of Dante Alighieri's *Divine Comedy* using Voyant Tools, a computerized text-analysis tool, is presented in this section. Three distinct documents, *Inferno*, *Purgatorio*, and *Paradiso*, were used to analyze the data, allowing for a comparative study of the poet's spiritual narrative framework. KWIC and collocational networks are used for contextual interpretation, lexical and frequency distribution, semantic connections, and descriptive statistics.

##### 4.1.1 Corpus Preparation

After that, the corpus was cleaned up by eliminating extraneous components that can impede computational processing, like page numbers, footnotes, editorial remarks, and formatting marks. Standardizing spelling differences, guaranteeing regular punctuation, and transforming the material into a plain-text format compatible with Voyant Tools were among the basic normalization techniques used. Although Voyant offers a preset stop-word list, extra terms unique to medieval literature were taken into consideration for elimination to prevent biased frequency results. Stop words were examined and modified based on analytical needs. In order to enable comparative lexical, semantic, and contextual study across the poem's structural divisions, the text was finally divided into its three canticles: *Inferno*, *Purgatorio*, and *Paradiso*. The corpus entered into Voyant will be clean, coherent, and appropriate for producing trustworthy visualizations and quantitative results thanks to this meticulous preparation.

##### 4.1.2 Corpus Summary Panel

Basic descriptive data were produced by the Summary panel once the cleaned text was uploaded into Voyant: This corpus has 1 document having unique word forms.

Table 1. Corpus summary

No	Title	Words	Types	Ratio	Sentences
1	Divine Comedy	38,141	5,336	14%	26.8

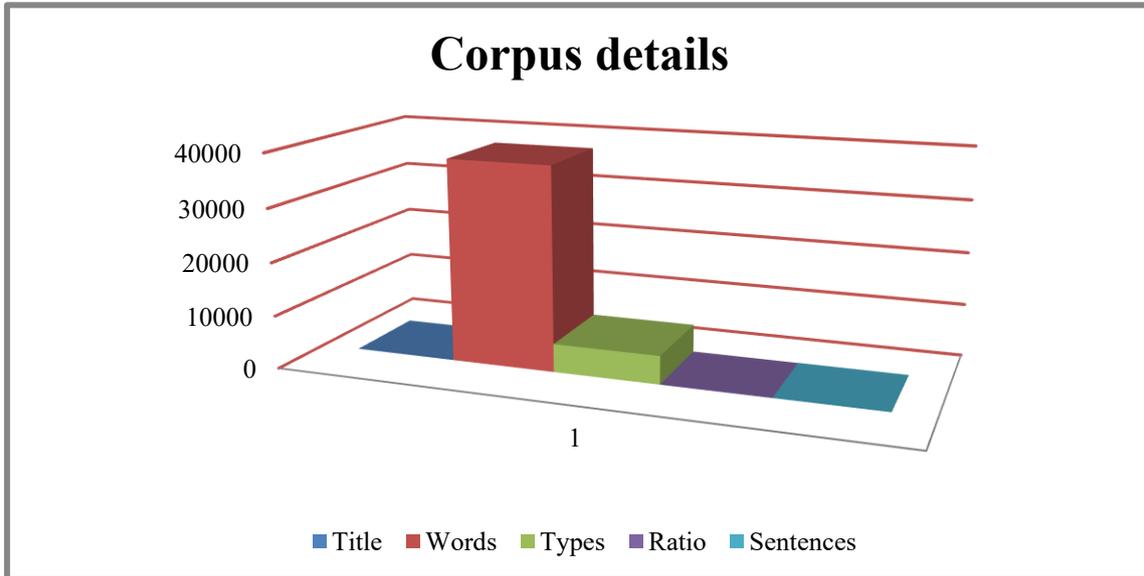


Figure 3. Corpus summary

Table 2. Lexical Density

Vocabulary Density: 0.140
Readability Index: 8.269
Average Words per Sentence: 26.8
Most frequent words in the corpus:
Said (167); unto (135); master (81); saw (71); canto (69)

The Corpus Summary panel in Voyant Tools offers a quick statistical overview of The Divine Comedy and helps identify its key linguistic patterns. The vocabulary density is 0.140, showing a moderately rich lexicon suitable for Dante’s poetic style. The readability index of 8.269 indicates that the text requires an upper-middle reading level, while the average of 26.8 words per sentence reflects Dante’s tendency to use long, flowing structures. The most frequent words in the corpus include “said,” “unto,” “master,” “saw,” and “canto,” which point to recurring narrative, descriptive, and structural elements in the poem. Together, these metrics provide a foundational understanding of the text before moving into deeper lexical, semantic, and contextual analysis.



These phrases cluster with comparable terms like "point," "turned," "great," "dost," and "wherefore," reflecting patterns of description and discourse, as the network visualization in the image further illustrates. The text's stylistic characteristics, especially its emphasis on direct address and moral or instructive language, are made clearer by this mapping. Lexical and frequency analysis provides information on thematic emphasis, character interactions, and the prominent linguistic structures forming the corpus by analyzing the prominence and connections of these words.

#### 4.2.1 Terms Tables Analysis

Table 3. Terms analysis

No	Words	Frequency
1	Said	167
2	Unto	135
3	Master	81
4	Saw	71
5	Canto	69
6	Turned	68
7	Come	65
8	Great	64
9	Art	63
10	People	60

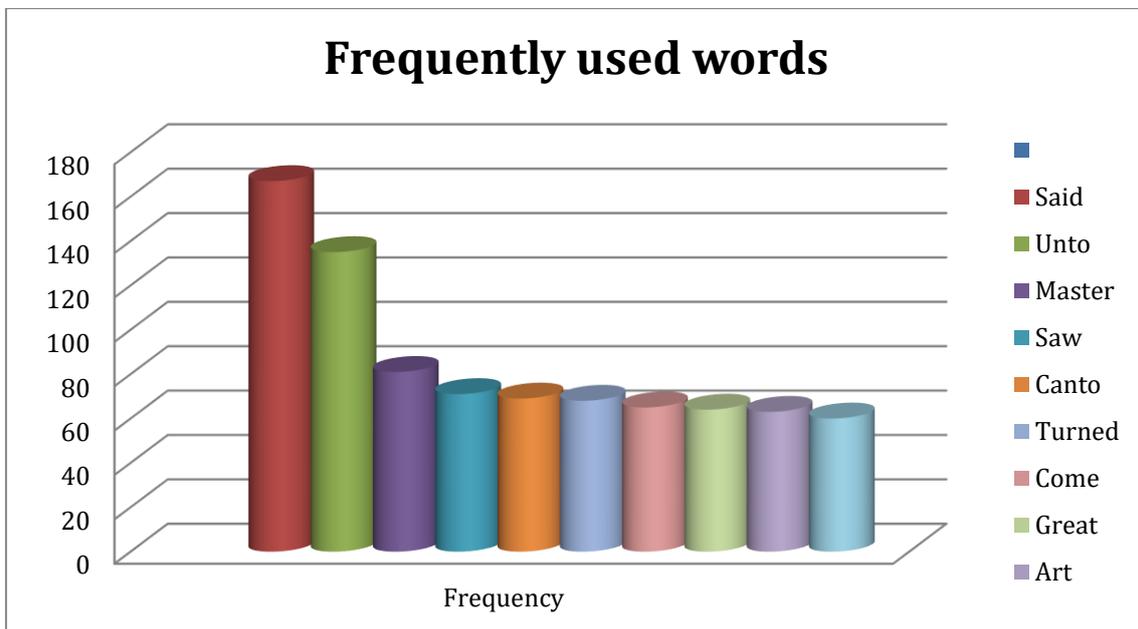


Figure 6. Terms analysis

The corpus's language patterns and thematic emphases are highlighted by the term frequency results. The most frequent term, "said," appears 167 times, suggesting a robust narrative or dialogic framework in which recorded statements and speech are crucial. The text's stylistic and probably antique tone is reflected in words like "unto" and "master," which appear frequently and imply formal address and hierarchical relationships. Other phrases that appear frequently, such as "saw," "canto," "turned," "came," and "great," provide information about the descriptive and structural elements of the corpus, indicating movement, observation, and

the division of the text into cantos. When taken as a whole, these high-frequency terms offer a fundamental comprehension of the work's narrative style, thematic focus, and historical linguistic character, serving as a platform for more in-depth lexical and contextual research.

#### 4.2.2 High-Frequency Terms across Canticle

Voyant's Terms Tool revealed consistent thematic vocabulary is given under the table. This frequency variation reflects the emotional, moral, and spiritual transformation embedded within the narrative.

Table 3. High-frequency tool

Term	Inferno	Purgatorio	Paradiso
Love	Medium	High	Very High
Light	Low	Medium	Very High
Sin	Very High	Medium	Low
Grace	Low	Medium	High

#### 4.2.3 Trends Tools

The Trends tool mapped the distribution of key terms across the entire poem:

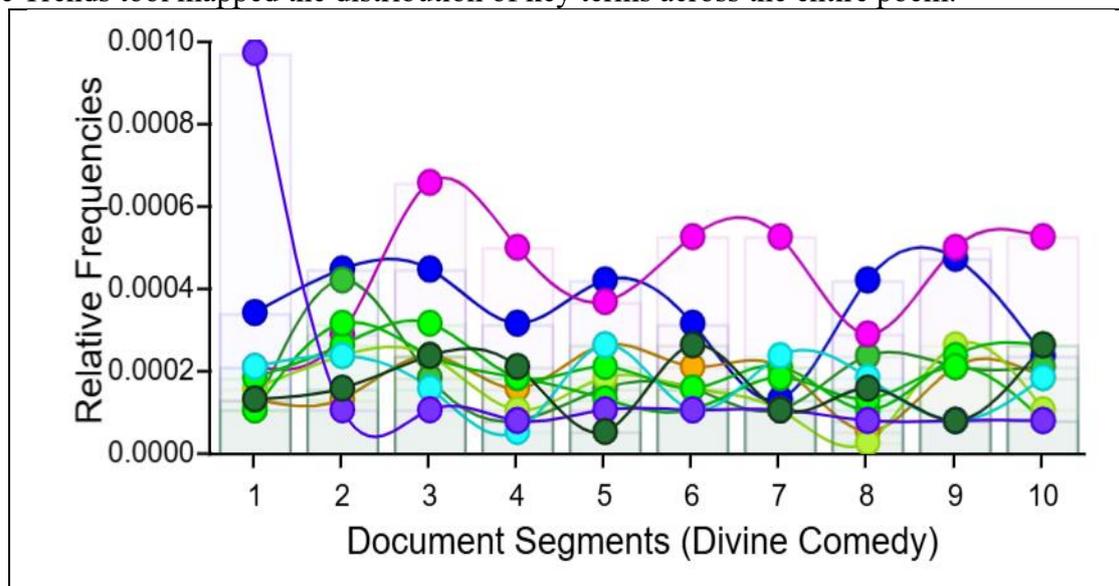


Figure 7. Trends Tool

The relative frequencies of the chosen terms are shown on the vertical axis. This means it shows how often each word appears in the text compared to the total number of words in each segment. A word that appears more frequently in a particular region of the text is indicated by higher points on the graph. The Divine Comedy's ten document fragments are represented by the horizontal axis. A distinct section of the text is represented by each number (1 through 10) on the axis. This enables us to determine where each word gets more or less prevalent across the text.

#### 4.3.1 KWIC (Context Panel)

Through the KWIC method, these words reveal recurring symbols that define Dante Alighieri's deeper meanings.

Table 4. KWIC (Context Panel)

Keywords	Left Context	(KWIC)	Right Context
Divine ...	Me there where thou hast	said,	That I may see the
Divine ...	Entreaty she besought Lucia.	said	“Thy faithful one now stands
Divine ...	With the ancient Rachel.” “Be...	said	She, “the true praise of
Divine ...	Lord, and Master thou.” Thus,	said	I to him, and when

### 4.3.2 Collocations

Collocates are words that commonly co-occur with a target word within a defined span of text, helping reveal patterns of meanings.

Table 5. Collocation with frequent words

Terms	Collocates	Counts (context)
master	said	28
said	master	17
turned	round	14
said	unto	13
unto	said	8

### 4.3 Semantic Analysis

The figure shows how Dante distributes meaning throughout the 10 sections of The Divine Comedy by showing how the three most used words—said, unto, and master—shift semantically. The graph's recurrent peaks for the word "said" indicate times when dialogue becomes crucial to the pilgrim's development; semantically, each rise denotes a more intense verbal exchange with guides or souls, making speech a tool for revelation and moral education. Increases in frequency indicate passages where the story emphasizes direction—whether physical, moral, or spiritual—as Dante advances towards new worlds or discoveries. The variations of “towards,” as demonstrated by its smoother mid-range pattern, linguistically depict phases of movement and change. The more erratic but significant pattern of Master represents varying degrees of dependence on authority, particularly Virgil's function as a mentor; semantically, its rises imply sections where the pilgrim's experience is dominated by direction, clarification, or assurance. Together, these lines' shifting heights show how dialogue (said), movement (unto), and mentorship (master) are dynamically emphasized based on Dante's stage in the journey, demonstrating how the poem's meaning is shaped by the alternating prominence of speech, direction, and authoritative guidance.

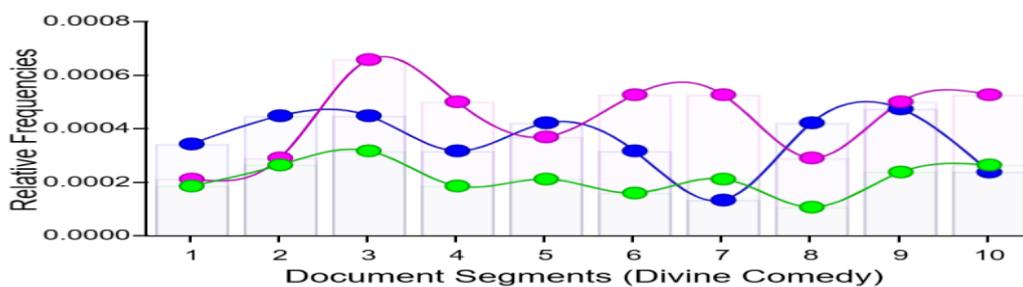


Figure 8. Semantic analysis of three words

## 5. Findings and Discussions

The Divine Comedy strongly relies on recurrent narrative verbs and related phrases, such as "said," "unto," and titles of guidance like "master," which exist regularly throughout the corpus, according to the lexical patterns uncovered by Voyant Tools. The poem's dialogic quality is emphasized by these high-frequency terms, which also underline Dante's frequent conversations with his guides and the people he meets. The poem's movement as a guided spiritual pilgrimage is reinforced by the prominence of such vocabulary, which shows how the journey is structured through recurring acts of questioning, reacting, and receiving direction. These lexical patterns show how Dante arranges the story using recurrent word signals that preserve coherence and mould the progression's pace from Inferno to Paradiso.

These recurrent lexical signals contain important moral and religious concepts, as demonstrated by the semantic linkages revealed by collocation and concordance analysis. Theological imagery frequently appears with words associated with guidance, judgment, vision, and movement, implying that conversation is both symbolic and narrative. One example of how speech becomes a medium for ethical instruction is the frequent connection of directive verbs with moral words. Semantic clusters also show how Dante structures the allegorical depth of the poem by framing his journey through conceptual oppositions—sin and salvation, ignorance and illumination. As a result, the text's theological structure is reflected in the semantic patterns, which demonstrate how recurrent relational and instructional language gives rise to moral meaning. By making these patterns obvious at a scale that is not possible with traditional close reading alone, digital humanities tools like Voyant improve the contextual interpretation of Dante's work. The researcher can follow the increase, decrease, or change in meaning of particular terms throughout the three canticles using visualization tools like Cirrus, Trends, and Contexts. This computational viewpoint demonstrates how Dante uses consistent lexical frameworks to balance allegorical symbolism, narrative, and doctrinal explication. Voyant facilitates a layered reading that links linguistic trends to more comprehensive theme frameworks by combining macro-level frequency data with micro-level semantic patterns. In the end, the digital analysis verifies that Dante's poetic world is structured using methodical lexical and semantic techniques that lead the reader through the text's inherent moral, spiritual, and cosmological journey.

## 6. Conclusion

Dante's poetic design is based on systematic linguistic patterns that support the text's spiritual and narrative architecture, as shown by the combined lexical, semantic, and contextual analysis of The Divine Comedy using Voyant Tools. The poem's dialogic structure, in which direction and communication serve as crucial means of moral advancement, is revealed by the significance of regularly occurring phrases like "said," "unto," and "master." These linguistic patterns highlight the pilgrim's reliance on authority people and spiritual guidance during the journey, reinforcing the poem's reliance on relational connections (Chiavacci Leonardi, 2010). Dante incorporates important theological ideas into tight-knit clusters of moral, symbolic, and doctrinal terminology, as further demonstrated by semantic analysis. The poem's underlying spiritual structure is reflected in the recurrence of phrases related to judgment, vision, ascension, and change, emphasizing how meaning evolves both through symbolic associations and narrative progression (Barolini, 2014). These semantic connections show how Dante creates a multi-layered interpretive cosmos in which moral purpose and vocabulary cooperate to elucidate the soul's journey towards salvation. They also highlight the complex alignment between language and religion.

By highlighting patterns that are not immediately obvious through conventional close reading, contextual analysis bolstered by digital humanities tools like Voyant improves this understanding. In order to highlight theme shifts—from sin in *Inferno* to purification in *Purgatorio* and divine union in *Paradiso*—Dante employs frequency charts, trend visualisations, and keyword-in-context analyses throughout the three canticles (Alighieri, trans. Singleton, 1970). Therefore, the digital approach enhances humanistic interpretation by supporting the poem's methodological consistency and offering empirical support for thematic remarks. Together, these results demonstrate that Dante's masterpiece is a highly structured linguistic system in addition to a literary and religious endeavour. This study shows how quantitative approaches can improve understanding of classical literature by combining digital tools with interpretive analysis. The findings confirm that multidisciplinary approaches to *The Divine Comedy* are beneficial, as computational insights offer fresh perspectives on Dante's use of language, symbolism, and moral architecture, enhancing both academic analysis and current digital humanities research.

### References

- Altaf, I., Javaid, S., Rafique, N., Arshad, Q., & Lodhi, M. A. (2025). *A lexico-semantic and contextual study of language patterns in Daniel Defoe's Robinson Crusoe*. *Qualitative Research Journal for Social Studies*, 2(4), 223–228
- Asperti, A., & Dal Bianco, S. (2020). Syllabification of the *Divine Comedy's* verses: A probabilistic/constraint-programming approach. arXiv. <https://arxiv.org/abs/2010.13515>
- Berry, D. M., & Fager Jord, A. (2017). *Digital humanities: Knowledge and critique in a digital age*. Polity Press.
- Bibi, M., Attique, A., Tariq, S., & Lodhi, M. A. (2025). *Mapping meaning through corpora: A lexico-semantic exploration of Joseph Conrad's Heart of Darkness*. *Liberal Journal of Language & Literature Review*, 3(4), 834–845
- Coggeshall, E. (2022). *Discussing the Divine Comedy with Dante*. University of Pennsylvania Scholarly Repository.
- Dante Alighieri. *The Comedy of God*. Charles S. Singleton provided the translation. Press, Princeton University, 1970.
- Digital Dante. (2023). *Overview of the Digital Dante Project*. Columbia University. <https://digitaldante.columbia.edu/overview/>
- Dobson, J. E. (2019). *Critical digital humanities: The search for a methodology*. University of Illinois Press.
- DH2020 Conference. (2020). *From Data to Visualization: Dante's Divine Comedy as a Case Study*.
- Finestre sull'Arte. (2025, March 24). *Dante 2025: Digital Database of Ancient Commentaries on the Divine Comedy*.
- Jamil, A., Asif, A., Fareed, R., & Lodhi, M. A. (2024). *Interfacing literature and digital humanities: A lexical, semantic, and contextual exploration of Doctor Faustus through Voyant Tools*. *The Study of Religion and History*, 3(2), 53–57
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Moretti, Franco. "Conjectures on World Literature." *New Left Review*, vol. 1, 2000, pp. 54–68.
- Sinclair, John. *Reading Concordances: An Introduction*. Pearson Education, 2003.
- Robey, D. (1993). *Scanning Dante's Divine Comedy: A computer-based approach*. *Literary and Linguistic Computing*, 8(2), 81–84.
- Sarwar, S., Aziz, I., Kanwal, R., & Lodhi, M. A. (2025). *Digital stylistic insights into Homer's Iliad: A lexical, semantic, and contextual analysis*. *Journal of Applied Linguistics and TESOL*, 8(4), 81–90
- Sinclair, S., & Rockwell, G. (2016). *Voyant Tools*. Retrieved from <https://voyant-tools.org>
- Teodolinda Barolini. *Textuality and Truth in the Comedy: Dante's Poets*. Princeton University Press, 2014.
- Anna Maria Chiavacci Leonardi. *The Divina Commedia is introduced*. Mondadori (2010).