

TRACING TENNYSON’S STYLISTIC SHIFTS: STYLOMETRIC INSIGHTS FROM EARLY LYRICS TO LATE EPICS

SUNGHYUN JANG *

Abstract This study employs computational stylometry to empirically trace the evolution of Alfred, Lord Tennyson’s poetic style across three chronological phases: his early lyrics, his middle period centered on *In Memoriam A.H.H.*, and his late epic cycle, *Idylls of the King*. To investigate these stylistic shifts, the research adopts a hybrid methodology that combines macroscopic distant-reading techniques (using the *stylo* package in R for Cluster Analysis and PCA) with granular linguistic profiling (using a transformer-based NLP pipeline in Python). The macroscopic analyses reveal a stylistic rupture in the late period; this demonstrates that Tennyson’s epic voice represents a structural departure from the stylistic foundation shared by his early and middle-period works. The Python-based analysis further shows that this late epic style was achieved through an active syntactic reconstruction characterized by increased sentence length, extended dependency distances, a shift from adjectival description to action-oriented verbs, and a heightened density of archaic diction. In contrast, the analyses identify the middle period as a peak in lexical diversity. It is marked by an objective, philosophical tone operating within a regularized syntactic framework. In conclusion, by translating literary concepts such as “lyric” and “epic” into quantifiable linguistic metrics, this digital-humanities approach provides robust empirical evidence that supports and refines traditional critical understandings of Tennyson’s dynamic artistic development.

Keywords: Tennyson, stylometry, *In Memoriam A.H.H.*, *Idylls of the King*

1. INTRODUCTION

Alfred, Lord Tennyson (1809–1892) stands not merely as a preeminent figure of nineteenth-century literature, but as the quintessential voice of the Victorian age itself. Spanning a creative career of over six decades, his oeuvre serves as a mirror to the intellectual and aesthetic transformations of his century—from the lingering shadows of Romanticism in the 1830s to the grand, imperial anxieties of the *fin de siècle* (Armstrong 1993). Literary critics have long recognized the complex duality of his artistic identity: the ‘private’ poet of melancholic introspection, best exemplified by the sensory richness of his early lyrics, and the ‘public’ Poet Laureate who constructed the monumental national mythologies of the *Idylls of the King*. While traditional scholarship has exhaustively charted these thematic currents, the precise linguistic mechanisms underlying his transition—from a lyricist of sensation to an architect of epic narrative—remain a fertile ground for empirical investigation. To explore this uncharted terrain,

* Korea University, Department of English Language and Literature

this study proposes a methodological pivot to address these long-standing critical questions: What if the evolution of Tennyson's style could be measured empirically? By treating Tennyson's poetic corpus not merely as an object of aesthetic appreciation but as a dataset amenable to computational analysis, this research employs stylometry to track the specific mutations in his language, sentence structure, and vocabulary over time (Moretti 2013; Jockers 2013).

The primary objective of this study is twofold, addressing both the 'what' and the 'why' of Tennyson's development. First, we aim to determine whether Tennyson's poetic style exhibits statistically significant changes across his career. To investigate this, his body of work has been categorized into three creative periods: Early, Middle, and Late. Second, the study seeks to interpret the nature and causes of these quantitative shifts. If statistical distinctiveness is observed, we must then transition from data to interpretation, situating these stylistic metrics within their broader literary-historical context. To achieve this, we adopt a hybrid methodology that synthesizes quantitative stylometric techniques with qualitative literary interpretation. This digital humanities approach allows for a granular analysis of how Tennyson's concrete choices in diction, syntax, and overall style evolved in tandem with the massive intellectual and cultural transformations that defined the Victorian age.

2. METHODOLOGY: COMPUTATIONAL FRAMEWORK

2.1 Corpus Construction and Periodization

To operationalize the research objectives, a comprehensive corpus of Tennyson's major works was curated and stratified into three chronological phases. This periodization strategy is not merely chronological but follows the generally accepted trajectory of Tennyson's poetic development, allowing for a comparative analysis of his stylistic evolution from private lyricism to public epic (Ricks 1989; Perry 2005; Douglas-Fairhurst 2010). The corpus is divided as follows. 1) The Early period (1830s–1840s): This phase encompasses Tennyson's major lyrical poems, including "The Lady of Shalott," "Mariana," "Ulysses," and "The Lotos-Eaters." These works are characterized primarily by rich Romantic imagery and deep psychological introspection (Tucker 1988; Perry 2005). By aggregating these texts, we aim to capture the baseline 'fingerprint' of Tennyson's younger, more sensory-oriented style; 2) The Middle period (1850s): Uniquely, the Middle period is defined by a single, monumental text: *In Memoriam A.H.H.* (1850). Although published in 1850, this elegy was composed over a span of seventeen years. Treating this text as a standalone period is crucial; it serves as a bridge between his early lyrics and later narratives, encapsulating a journey of introspective maturity (Ricks 1989; Perry 2005); and 3) The Late period (1860s–1880s): The final phase focuses on the epic cycle *Idylls of the King*, written and published over more than two decades. This corpus represents the culmination of Tennyson's mature style and reflects his evolved role as Poet Laureate. The language here engages with grand national and moral themes, distinct from the solitary introspection of his earlier years (Tucker 1991; Douglas-Fairhurst 2010).

To facilitate a macro-analytical comparison, individual poems within each period were not analyzed in isolation. Instead, the texts were aggregated into three consolidated plain-text files: 'tennyson_early.txt', 'tennyson_middle.txt', and 'tennyson_late.txt'. For the computational tests, the analysis used machine-readable texts obtained from Project Gutenberg, and

Christopher Ricks's edited edition of Tennyson's collected poems was consulted as a scholarly reference to verify corpus selection and textual organization (Ricks 1987). This consolidation enables a period-level analysis, and this in turn allows us to compare the collective stylistic signatures of Tennyson's career phases directly.

2.2 Critical Framework: From Lyric to Epic

Before examining the stylometric data, it is essential to characterize the intuitive stylistic trajectory that literary criticism has long attributed to Tennyson. This study postulates that the statistical shifts in the corpus will correspond to the distinct thematic and aesthetic pillars of his career: the movement from private lyricism to philosophical inquiry, and finally to public epic narrative.

The Early period is anchored by poems such as "The Lady of Shalott" and "Mariana." As illustrated in the corpus selection, these works are defined by an intense pictorialism. Critics have long noted that Tennyson's early style is characterized by dreamlike atmospheres and melancholic introspection (Tucker 1988; Perry 2005). The visual potency of these lyrics was such that they profoundly influenced late nineteenth-century visual artists, particularly the Pre-Raphaelites, who sought to capture the same vivid, detailed imagery on canvas (Kooistra 2002; Carretero González 2012).

The Middle period is represented exclusively by *In Memoriam A.H.H.* Triggered by the sudden death of Arthur Henry Hallam in 1833, the poem is a massive structure of 133 cantos that blends personal grief with the Victorian era's most pressing intellectual anxieties (Eliot 1936; Ricks 1989; Shaw 1976; Perry 2005). Unlike the visual stasis of the early lyrics, *In Memoriam* is discursive and argumentative. It functions as a stylistic bridge and carries both philosophical weight and emotional range, ultimately demanding a more complex—and at times more abstract—lexical field to articulate its full effect (Shaw 1976; Tucker 1988; Kincaid 1975).

Finally, the Late period, embodied by *Idylls of the King*, showcases Tennyson in his mature role as Poet Laureate. This cycle of twelve narrative poems, written in blank verse, retells the Arthurian legend not merely as myth, but as a mirror for nineteenth-century British society. Here, Tennyson's style transforms into a vehicle for grand national and moral themes. The shift to blank verse and narrative momentum suggests a move away from the dense lyricism of his youth toward a more direct, didactic, and perhaps public rhetorical style suited for addressing the social anxieties of the age (Tucker 1991; Douglas-Fairhurst 2010).

2.3 Stylo and Multivariate Analysis

To empirically measure the stylistic distances between Tennyson's three creative periods, this study employs the stylo package: a comprehensive library within the R programming environment widely recognized as a standard tool in computational stylistics (Eder, Rybicki, and Kestemont 2016). Stylo provides a robust, semi-graphical interface for performing multidimensional statistical analyses based on word frequency distributions. The core premise of this analysis is that a writer's style can be quantified through the frequency of function words (most frequent words, or MFWs) in the corpus (Burrows 2002; Kestemont 2014). By calculating the relative frequencies of these words, the software generates a high-level, visual representation of the textual relationships across the corpus.

Two primary exploratory techniques were utilized to investigate the dataset. First, we performed Cluster Analysis to assess the hierarchical relationships between the texts. This method groups texts based on their stylistic proximity, visualizing the results as a dendrogram, or a tree diagram (Eder, Rybicki, and Kestemont 2016). In this study, Cluster Analysis is used to determine whether Tennyson's works naturally cluster according to the proposed chronological divisions—Early, Middle, Late—without prior labeling. Second, the study utilized Principal Components Analysis (PCA) to reduce the multidimensionality of the frequency data into interpretable two-dimensional visualizations (Binongo and Smith 1999; Eder, Rybicki, and Kestemont 2016). Whereas Cluster Analysis shows hierarchical grouping, PCA maps the texts onto a coordinate system, thus revealing the magnitude and direction of stylistic variance. This allows us to observe not just that the periods differ, but how they are spatially related.

2.4 NLP Pipeline and Feature Extraction

The macro-stylometric analysis using R successfully mapped the topological relationships between Tennyson's periods, but it primarily measures relative distances based on high-frequency function words. To understand the nature of these distances, we must transition from a macroscopic view to a granular linguistic perspective. For this phase of the research, Python was selected as the primary computational environment (van Rossum and Drake 2009). Python was chosen specifically for its robust ecosystem in Natural Language Processing (NLP), statistical computing, and data visualization, which allows for the extraction, visualization, and statistical testing of specific linguistic features beyond simple word counts (Hunter 2007; Honnibal et al. 2020; Virtanen et al. 2020).

To construct a comprehensive picture of Tennyson's stylistic evolution, we designed a multi-dimensional suite of analyses. This study investigates the following specific linguistic variables. 1) Syntactic and Structural Complexity: We examine sentence length and syntactic complexity to determine if Tennyson's sentence structures became more elaborate or concise over time (Lu 2010). 2) Lexical Composition: The study measures lexical richness to assess vocabulary diversity and tracks the frequency of archaism to see if his engagement with the Arthurian legend precipitated a rise in antiquated diction (McCarthy and Jarvis 2010). 3) Grammatical and Tonal Shifts: By analyzing POS (Part-of-Speech) distribution, we aim to identify shifts in grammatical preference. This follows corpus-linguistic approaches that treat grammatical feature distributions as indicators of register and stylistic variation (Biber 1988). Additionally, Sentiment Analysis is employed to quantify the shifting emotional landscapes from the early lyrics to the late epics (Pang and Lee 2008). Furthermore, the Jensen-Shannon distance is calculated to measure distributional divergence, and results are validated using the Kruskal-Wallis H-test and Dunn's post-hoc test to confirm statistical significance across the three periods (Lin 1991; Kruskal and Wallis 1952; Dunn 1964).

To execute the linguistic analyses outlined above, we constructed a specialized Python environment, which centers on high-precision NLP tools. The primary analytical engine selected for this study is the spaCy library (v3.7+), a widely used framework for syntactic parsing, POS tagging, lemmatization, and dependency annotation (Honnibal et al. 2020). Crucially, this study deviates from standard, lightweight statistical models, such as `en_core_web_sm`, in favor of the English Transformer Pipeline, `en_core_web_trf`. This model leverages transformer-based architectures, specifically RoBERTa-based representations, to generate contextual word embeddings rather than relying on static vectors (Vaswani et al. 2017;

Liu et al. 2019; Honnibal et al. 2020).

The deployment of a transformer model is fundamental to the study's validity for the following reasons. 1) The transformer model analyzes words within their specific sentence context. This allows for a more accurate resolution of polysemy and ambiguous grammatical structures, which is essential when analyzing poetic syntax (Vaswani et al. 2017; Liu et al. 2019). 2) The model provides context-sensitive dependency parsing, enabling a more precise measurement of complex sentence structures (Honnibal et al. 2020; Lu 2010).

3. RESULTS: EMPIRICAL EVIDENCE OF STYLISTIC EVOLUTION

3.1 Macroscopic Topography: Visualizing the Stylistic Rupture

3.1.1 Hierarchical Clustering

In order to test the hypothesis of stylistic progression, we first performed a Cluster Analysis to visualize the hierarchical relationships between the chronological periods. To ensure that the analysis was not skewed by the varying lengths of individual poems, the consolidated files were segmented into equal-sized samples. Specifically, the corpus was divided into 3,000-word chunks to standardize the text density for comparison. The analysis relied on the frequencies of the 100 MFWs, and Euclidean distance was used as the metric to measure stylistic proximity.

Figure 1, the resulting dendrogram, reveals a striking primary bifurcation, separating the corpus into two massive branches.

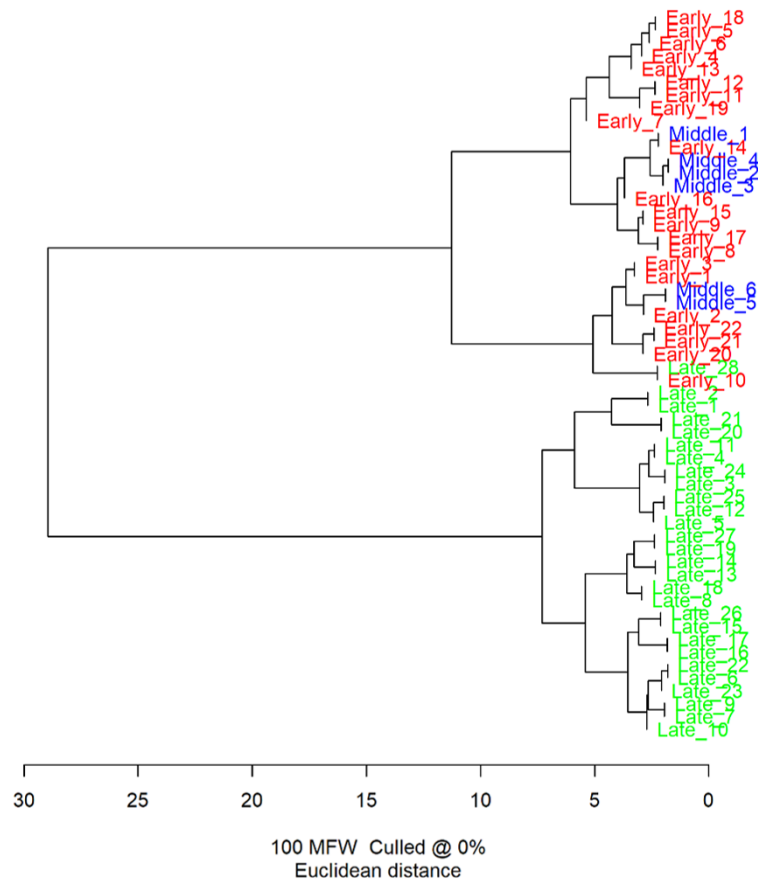


Figure 1 Cluster Analysis #1

The bottom branch of the dendrogram is composed exclusively of chunks from the Late period (labeled in green). This homogeneity indicates that the stylistic signature of the *Idylls of the King* is consistently different from Tennyson's earlier output. The fact that no Early or Middle chunks appear in this cluster suggests that his transition into the epic mode involved a fundamental shift in his usage of high-frequency function words—the basic building blocks of syntax. By comparison, the top branch presents a more complex relationship. It contains almost all chunks from the Early period (red) intermingled with the Middle period (blue). Crucially, the Middle chunks are not isolated in a separate major branch but are nested within the Early cluster.

This nesting offers a significant interpretive insight: While *In Memoriam* (Middle period) represents a thematic and philosophical maturity, its underlying lexical architecture retains a strong affinity with the early lyrics. The dendrogram confirms that the 'stylistic DNA' of the Middle period is still rooted in his earlier phase. However, it is noteworthy that the Middle chunks are not entirely scattered but instead form small sub-clusters, suggesting internal consistency even while being embedded within the broader early style. In summary, the Cluster Analysis demonstrates a clear stylistic rupture, but it places that rupture after the Middle period and distinguishes the late epics from the lyrical foundations of both the Early and Middle phases.

To ensure that the initial clustering was not an artifact of specific parameter selection (i.e., word frequencies and Euclidean distance), a secondary analysis was conducted to test the

methodological robustness of the findings. In quantitative literary studies, verifying stability across different feature sets is crucial for validating stylistic signatures. For this validation phase, the feature set was shifted from whole words to the 100 Most Frequent Character 3-grams (MFC 3-grams). Unlike word frequencies, character n-grams capture morphological patterns (such as suffixes, prefixes, and common letter sequences) rather than purely lexical choices, often revealing deeper, subconscious stylistic habits. Furthermore, the distance metric was adjusted to Classic Delta, a measure widely accepted as the standard for authorship attribution.

Figure 2 confirms the structural stability of the corpus. The primary bifurcation remains intact.

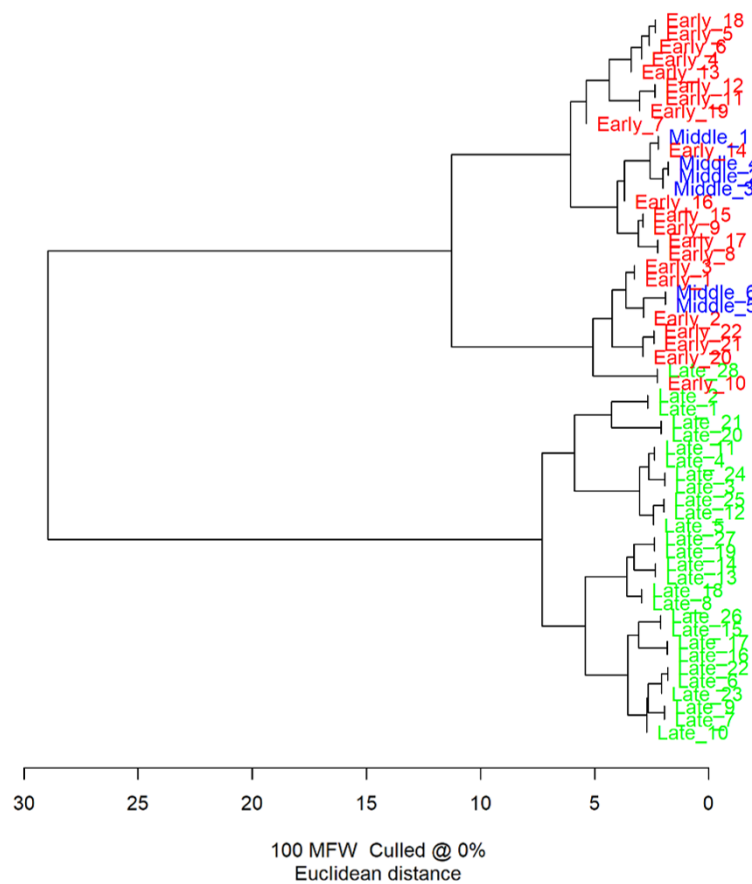


Figure 2 Cluster Analysis #2

Once again, the Late chunks are isolated in a discrete branch, which reinforces the conclusion that Tennyson’s epic style represents a departure from his previous work. Meanwhile, the Early (Red) and Middle (Blue) periods continue to inhabit the same primary branch, maintaining their stylistic affinity.

To further interrogate the dataset and ensure that the clustering was not limited to the highest-frequency function words, a tertiary analysis was conducted by expanding the feature set to the 300 MFWs. The distance metric was maintained as Classic Delta to allow for a direct comparison with the secondary analysis. The results of this expanded analysis corroborate the fundamental bifurcation observed in the previous trials.

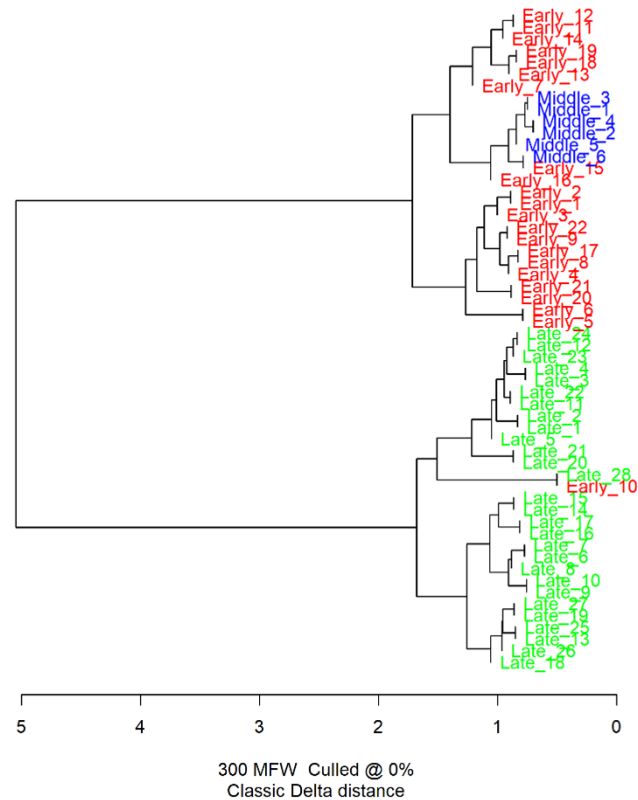


Figure 3 Cluster Analysis #3

Figure 3 once again displays a stark primary split, with the Late period (green) forming a super-cluster, separate from the Early and Middle periods. This consistency across varying parameters—from 100 words to 3-grams, and now to 300 words—establishes that Tennyson’s late epic style is substantially differentiated from the earlier poetic modes.

The most significant insight from this iteration, however, concerns the internal structure of the Middle period. In the initial analysis (100 MFWs), the Middle chunks were nested within the Early branch but were slightly fragmented. In this expanded analysis (300 MFWs), all chunks corresponding to *In Memoriam* (blue) are consolidated into a singular, contiguous sub-cluster. This visual unification suggests that although *In Memoriam* shares the genetic markers of the early style, it nonetheless operates as a cohesive stylistic unit. The text is not merely a collection of lyrics written over seventeen years, but a unified work with a consistent lexical density that distinguishes it from the shorter, more varied lyrics of the 1830s and 40s.

A Bootstrap Consensus Tree analysis was performed to provide a definitive statistical validation of the observed clusters. Unlike single-run cluster analyses, which can be sensitive to specific parameter settings, this technique aggregates results from multiple iterations to identify the most stable and recurrent stylistic groupings. The analysis was configured to iterate across a range of features, from the Top 100 to the Top 300 MFWs, with the Classic Delta distance metric applied consistently. By synthesizing these iterations, the consensus tree filters out noise and visualizes only those relationships that appear most frequently across the experimental spectrum.

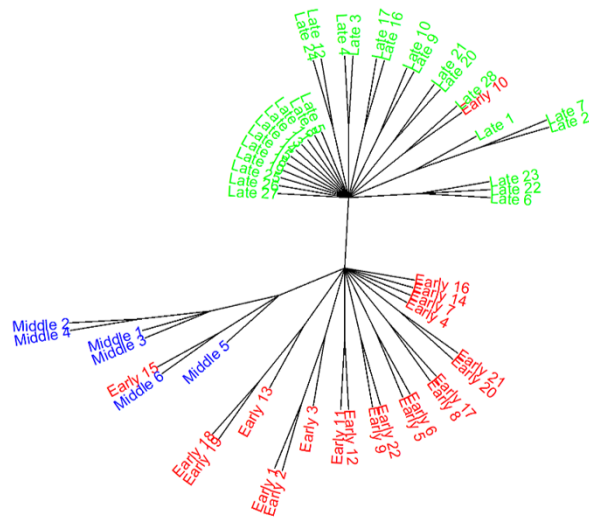


Figure 4 Bootstrap Consensus Tree

Figure 4, the resulting radial diagram, offers robust corroboration of the findings from the individual cluster analyses. The visualization displays a deep, fundamental split in the data. The chunks from the Late period (green) diverge into a discrete lineage, forming a unified cluster separate from the rest of the corpus. This reinforces the hypothesis that Tennyson's mature epic voice constitutes a departure from his earlier aesthetics.

Importantly, the consensus tree resolves the ambiguity regarding the Middle period. Within the primary branch shared with the early works (red), the middle-period texts (blue) consolidate to form their own sub-cluster. This structural positioning confirms a nuanced dual identity for *In Memoriam*: (1) By sharing the main branch with the early works, it retains a foundational stylistic similarity to the lyrics of the 1830s and 40s. (2) By forming a separate sub-cluster, it proves to be distinguishable from those earlier works.

3.1.2 Spatial Analysis: PCA

Although Cluster Analysis provides a hierarchical view of textual relationships, it forces data into a branching structure. To verify these findings in a continuous spatial context, we employed PCA. This multivariate technique reduces the high-dimensional data—specifically, the frequencies of the 100 MFWs—into a two-dimensional scatterplot, thereby enabling the observation of geometric distances between texts. By utilizing a Correlation matrix to normalize the data, the analysis ensured that high-frequency function words did not disproportionately skew the results. The resulting plot (Figure 5) maps the texts onto two axes, PC1 and PC2, which represent the two strongest dimensions of stylistic variance within the corpus.

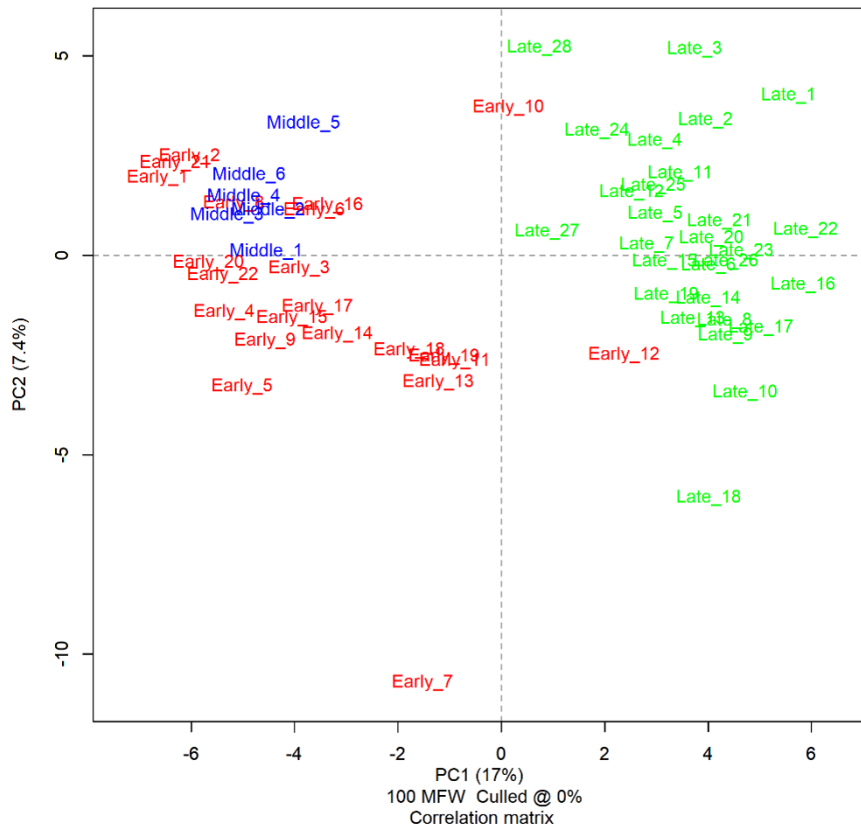


Figure 5 PCA #1

The PCA results provide a powerful spatial corroboration of the hierarchical clustering. PC1, represented by the horizontal axis, creates a nearly perfect stylistic division. Situated to the right of the central axis, the Late chunks form an isolated cloud. This spatial segregation confirms that the *Idylls of the King* occupies a distinct stylistic territory, remote from Tennyson's earlier output.

The PCA elucidates the relationship between the Early and Middle periods. As observed in the scatterplot, the Middle chunks (blue) are positioned within the same coordinate space as the Early chunks (red), to the left of the central division. They share the same lexical territory as the early works. This visual evidence reinforces the conclusion that *In Memoriam* (Middle) does not represent a radical break from Tennyson's foundational voice. Instead, it operates within the established parameters of his early lyrical style, whereas the Late period marks a definitive migration to a new stylistic paradigm.

To verify the stability of these spatial configurations, we conducted a secondary PCA using the 100 MFC 3-grams. This shift from a lexical feature set (whole words) to a morphological one (character sequences) serves as a stress test for the observed stylistic patterns. As anticipated, the alteration of the feature set resulted in a reconfiguration of the specific coordinate topology. However, as Figure 6 illustrates, the underlying structural narrative of the plot remains consistent.

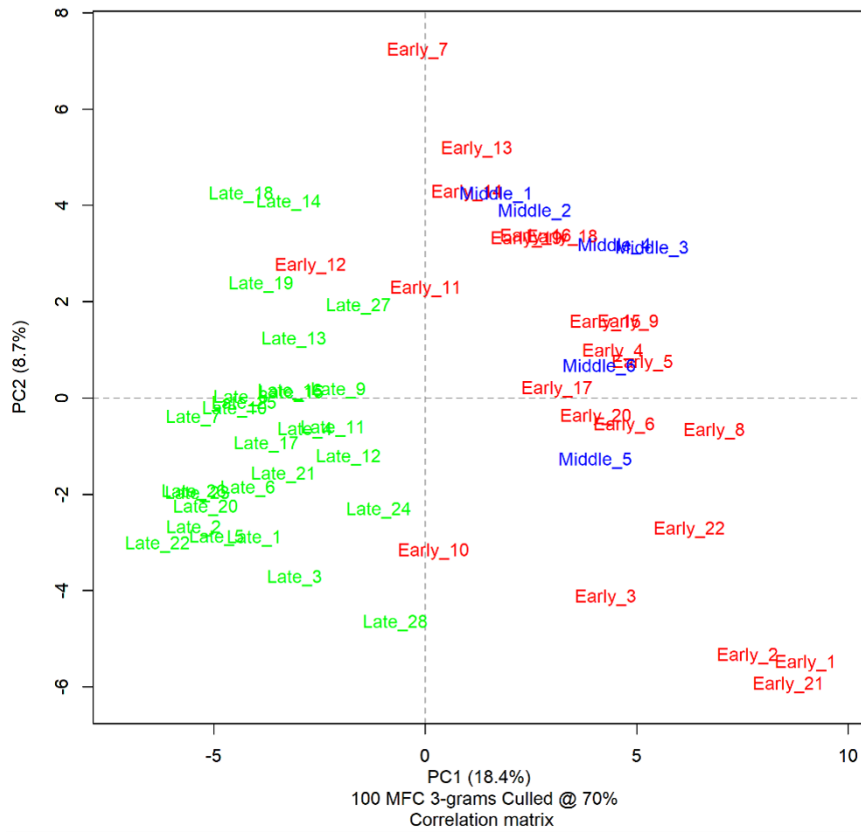


Figure 6 PCA #2

The Late period chunks (green) continue to form a distinct cluster. This confirms that the distinctiveness of the *Idylls* is not merely a matter of vocabulary selection but is deeply encoded in the morphological structure of the text. Similarly, the Early (red) and Middle (blue) chunks remain co-located within the same region of the vector space. This reiterates the finding that *In Memoriam*, despite its thematic gravity, does not morphologically diverge from the stylistic habits established in the early lyrics.

To conclude the spatial verification of the corpus, we executed a final PCA iteration utilizing the 100 MFW 2-grams (bigrams).

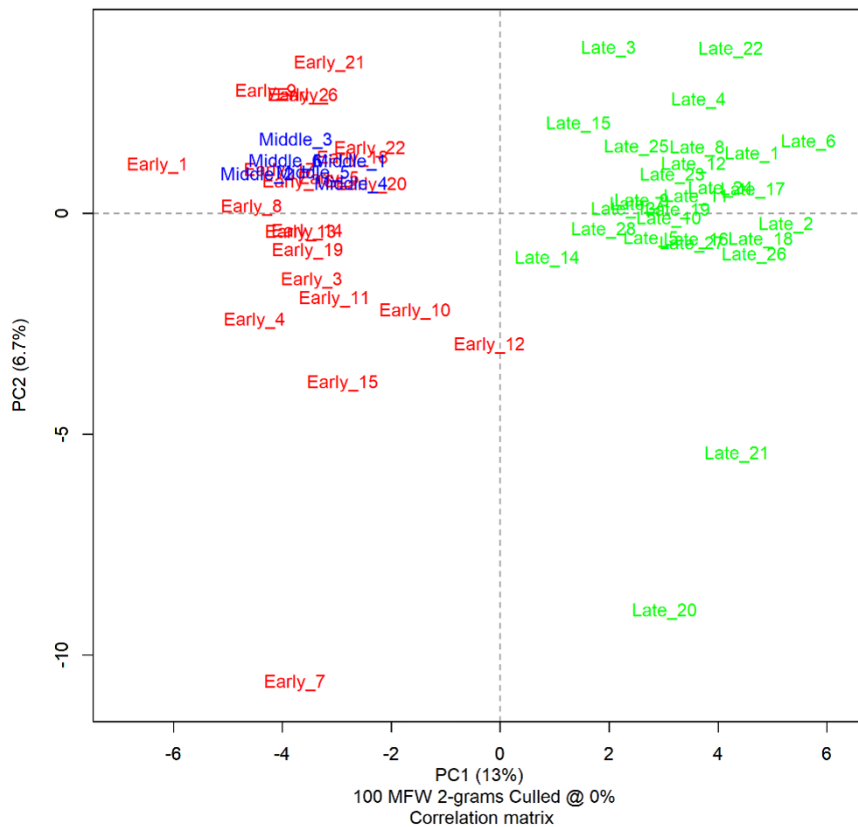


Figure 7 PCA #3

Despite the radical shift in feature selection, as shown in Figure 7, the fundamental structure of Tennyson's stylistic evolution remains largely unchanged. PC1 continues to function as the primary discriminator; it enforces a delineation between the mature epic style (Late) and the earlier lyrical modes. This final analysis also crystallizes the status of the Middle period. As illustrated in the scatterplot, the Middle period texts (blue) form a cohesive cluster within the Early stylistic territory. This high degree of vector clustering indicates that *In Memoriam* possesses a unique internal consistency; it is stylistically homologous to the early works yet distinguished by a disciplined uniformity that the more scattered early lyrics lack.

Across all computational trials, the data unequivocally support two conclusions: 1) There is a statistically significant stylistic break in Tennyson's career: the Late period stands on one side, and the Early and Middle periods on the other. 2) The Middle period (*In Memoriam*) has a dual identity. It is stylistically closer to his early works, yet it also displays a unique and consistent stylistic signature of its own.

3.2 Linguistic Architecture: The Mechanisms of Change

3.2.1 Corpus Parsing and Statistical Overview

The analysis now shifts from the high-level clustering approach used in R to a more granular feature analysis implemented in Python. Following the initialization of the NLP pipeline described in the Methodology section, the consolidated text files were processed through spaCy's `en_core_web_trf` model. This parsing stage transformed the raw string data into structured linguistic objects, enabling the precise extraction of syntactic boundaries and morphological features. The parsing process yielded a fundamental quantitative baseline for

the study. The total counts of sentences and tokens (defined here as individual words or punctuation marks) identified for each period are summarized in Table 1:

Table 1 Number of Sentences and Tokens per Period

	Early	Middle	Late
Sentence	3259	766	3165
Token	82157	22713	103582

As the data indicate, there is a marked variance in the volume of text across the three periods. The Late period, comprising the expansive epic cycle *Idylls of the King*, contains over 100,000 tokens, whereas the Middle period, represented solely by *In Memoriam*, contains approximately 22,700 tokens. This significant disparity in corpus size underscores a critical methodological imperative: Raw frequency counts cannot be used for direct comparison. Consequently, in order to ensure statistical validity and comparability across corpora of unequal size, all subsequent analyses in this study apply different normalization procedures depending on the feature being measured. With this foundational framework established, we proceed to the first granular metric: the analysis of sentence length and syntactic complexity.

3.2.2 Syntactic and Structural Evolution

The first dimension of the investigation focuses on the structural architecture of Tennyson's verse. By leveraging the dependency parsing capabilities of the spaCy transformer model, we extracted four key metrics to quantify the complexity of Tennyson's syntax, as shown in Table 2: Mean Sentence Length, Median Sentence Length, Subordinate Clauses per Sentence, and Mean Dependency Distance (MDD).

Table 2 Syntactic Complexity Metrics

Period	Mean Sentence Length	Median Sentence Length	Subordinate per Sentence	MDD per Sentence	Number of Sentences
Early	25.209267	21	1.783983	2.864744	3259
Middle	29.651436	28	2.415144	2.771469	766
Late	32.727330	24	2.567773	3.308482	3165

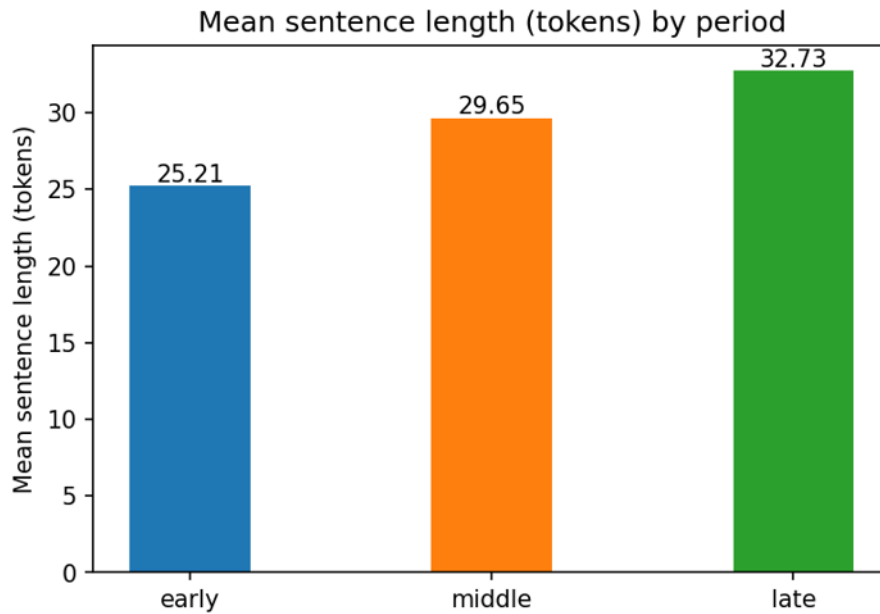


Figure 8 Mean Sentence Length

As illustrated in Figure 8, the progression from Early to Middle and finally to Late is visually linear. In the context of Victorian stylistics, this steady rise points to an increasing reliance on periodic phrasing—a rhetorical style where the syntactic completion of a sentence is delayed until the end. However, the more revealing metric regarding Tennyson’s late style lies in the divergence between the mean and the median values. This differential serves as a proxy for distributional skewness, indicating the presence of outliers within the text. The analysis yielded the following Mean minus Median values (Table 3):

Table 3 Mean vs. Median

	Mean – Median
Early	4.2
Middle	1.65
Late	8.73

The dramatic spike in the Late period is statistically diagnostic. It suggests that the increase in average sentence length in the *Idylls of the King* is driven by the inclusion of exceptionally long, complex sentences that pull the mean upward, rather than a uniform increase across all sentences. Conversely, the Middle period exhibits the lowest skew (1.65), despite having a high average length. The stanzas of *In Memoriam* are structurally homogenous; the sentences are long and weighty, but they are consistently so. In conclusion, the *Idylls* are defined not merely by length, but by syntactic elasticity—the capacity to stretch language to extreme limits to accommodate the grand, serious subjects of national myth.

While Mean Sentence Length indicates the volume of Tennyson’s utterances, the Subordinate per Sentence metric reveals their architectural quality. A mere increase in length could theoretically be achieved through parataxis—stringing together simple clauses with “and” or “but.” However, the data confirm that Tennyson’s evolution was driven by a fundamental

shift toward hypotaxis, a style characterized by the hierarchical nesting of ideas. Figure 9 demonstrates an upward trajectory that mirrors the elongation of his sentences:

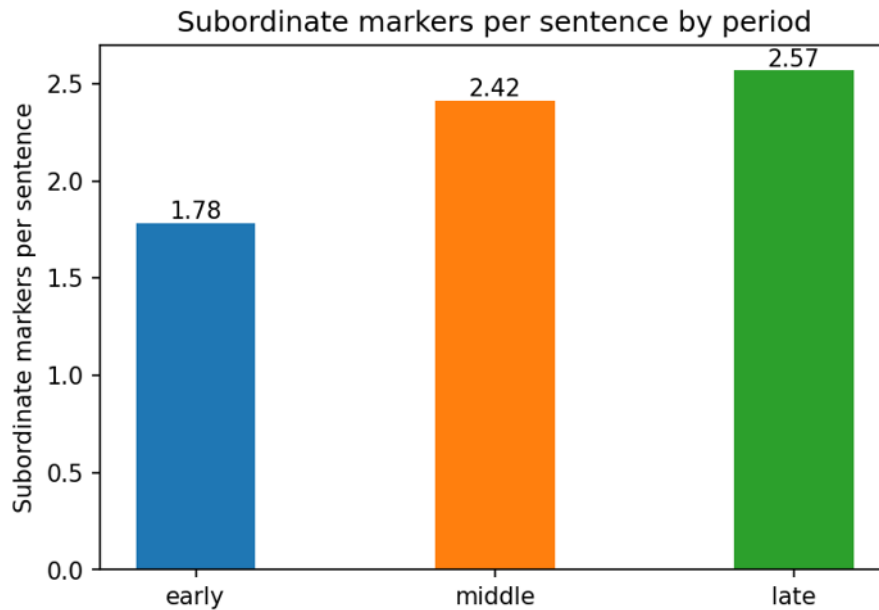


Figure 9 Subordinate per Sentence

This metric, derived from spaCy's dependency parsing, counts specific grammatical dependencies such as relative clauses ("which"), adverbial clauses ("because," "when"), and complement clauses. Both *In Memoriam* and *Idylls of the King* required a grammar capable of subordination and embedding—placing one thought in a conditional or otherwise dependent relationship to another. The data confirm that the late style is defined by a dense, intricate syntax typical of Victorian formality, where narrative and logic are tightly interwoven.

To complement the analysis of subordination, the study employed a more advanced computational metric: MDD. Whereas subordination counts the number of grammatical layers, MDD measures the linear spread of those layers. Specifically, it calculates the average distance (in tokens) between a word and the word it syntactically depends on (Liu 2008). A higher MDD signifies that grammatically related elements are widely separated by intervening phrases or clauses. This is a hallmark of highly complex, nested sentence structures.

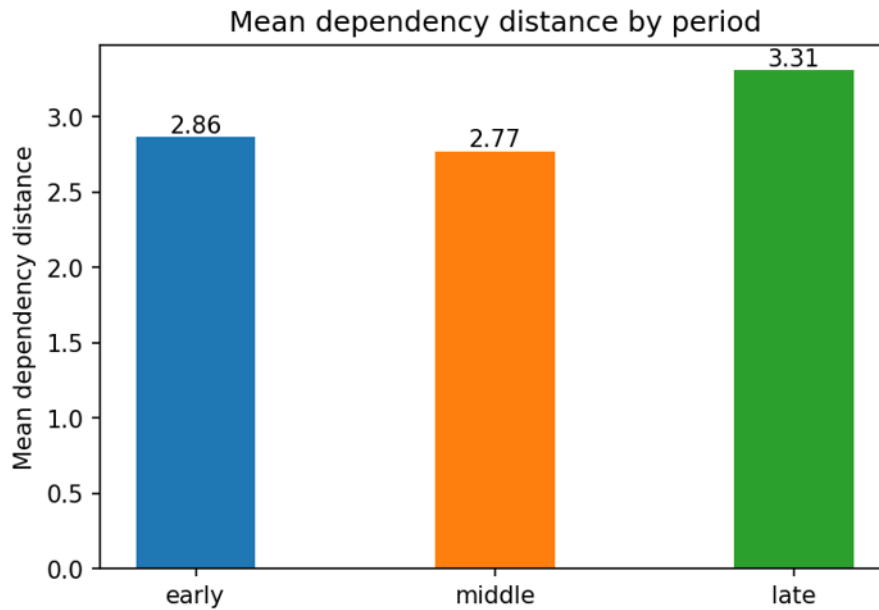


Figure 10 MDD per Sentence

The substantial spike to 3.31 in the Late period, illustrated in Figure 10, suggests that the syntax of *Idylls of the King* is characterized by syntactic suspension—a stylistic device in which the resolution of meaning is deliberately delayed. The Middle period (*In Memoriam*) actually shows a slight contraction in dependency distance, likely due to the constraints of the tetrameter quatrain form. By contrast, the unrhymed blank verse of the *Idylls* allows Tennyson to stretch his syntax across line endings. These data quantitatively capture an epic style where the architecture of the sentence demands sustained intellectual engagement from the reader.

While the metrics for sentence length and subordination display a linear progression, the MDD reveals a significant structural nuance. As visualized in Figure 10, the data do not follow a simple upward trend; rather, the Middle period (*In Memoriam*) dips below the Early period's 2.86. This statistical anomaly offers an insight into the relationship between form and syntax. Although *In Memoriam* contains longer and more subordinated sentences than the early lyrics, its internal grammatical structure appears more regular. This regularization can be attributed to the poem's strict and uniform stanzaic form (the *abba* tetrameter quatrain). The constraints of this *In Memoriam* stanza likely served to channel the syntax into more predictable patterns. In contrast, the Late period registers a leap in MDD to 3.31, the highest value in the corpus. This surge confirms that the complexity of the *Idylls of the King* is multi-dimensional. Tennyson is not merely writing longer sentences; he is arranging his clauses in significantly more intricate and spread-out ways. This structural stretching corresponds to the demands of the grand, high epic style. Freed from the stanzaic boundaries of *In Memoriam* and utilizing the flexibility of blank verse, Tennyson constructs sentences where the syntactic resolution is delayed over multiple lines. The high MDD score thus quantitatively captures the epic quality of the *Idylls*, which requires the reader to traverse a greater cognitive distance to connect subject to verb and antecedent to relative clause.

3.2.3 Lexical Composition and Diversity

Having established the structural evolution of Tennyson's syntax, the study proceeds to examine the lexical dimension of his style. To ensure accurate lexical profiling, the corpus

underwent a specific filtration process: Only alphabetic tokens were extracted and converted to lowercase, effectively removing punctuation and numerical data. Initial calculations using the standard Type-Token Ratio (TTR)—the ratio of unique words (Types) to total words (Tokens)—yielded misleading results. As a text grows longer, the rate of new word introduction naturally slows, depressing the TTR; this length sensitivity has long been recognized as a central limitation of traditional lexical richness measures (Tweedie and Baayen 1998; Covington and McFall 2010). To correct for textual length, two sophisticated, length-independent metrics were employed: 1) MATTR (Moving-Average Type-Token Ratio): This algorithm divides the text into overlapping windows of 500 words, calculates the TTR for each window, and averages the results (Covington and McFall 2010). 2) MTL D (Measure of Textual Lexical Diversity): This metric calculates the average sequence length required for the TTR to drop below a fixed threshold, set here at 0.72 (McCarthy and Jarvis 2010). For both metrics, a higher score indicates greater lexical richness.

The visualization of these lexical metrics—specifically the length-independent MATTR and MTL D—reveals a non-linear trajectory in Tennyson's career. As illustrated in the comparative line charts, the Middle period represents the peak of lexical creativity.

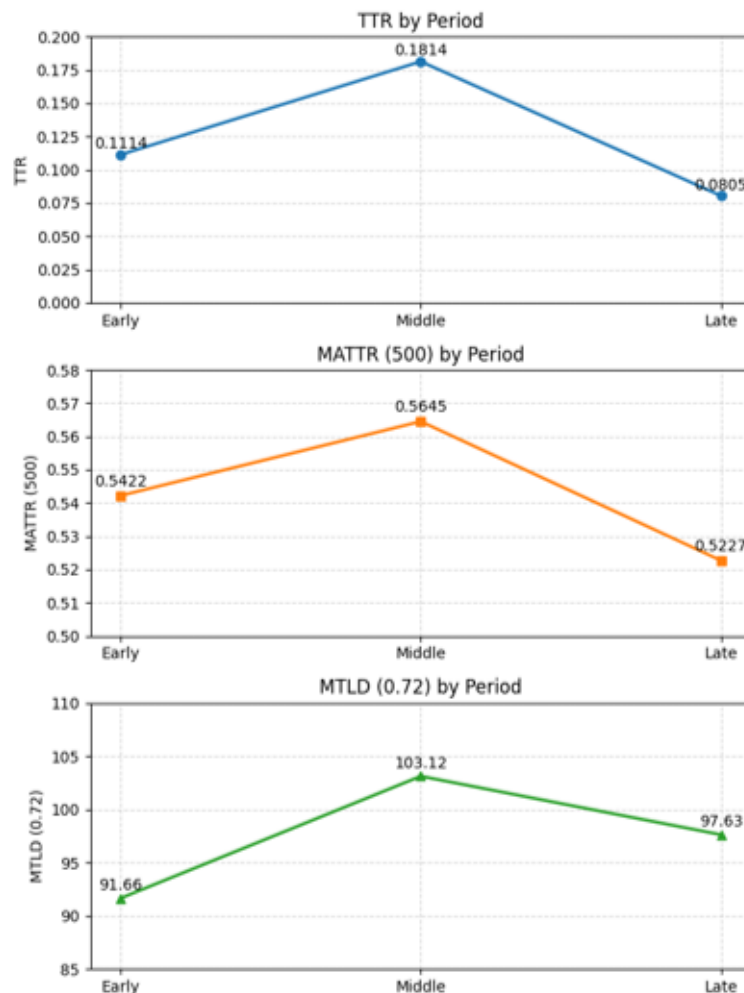


Figure 11 TTR, MATTR, and MTL D by Period

Both the MATTR and MTL D graphs in Figure 11 display an inverted V-shape, with their

peaks occurring in the Middle period. This quantitative peak aligns with the unique compositional history of *In Memoriam*. Written over a span of seventeen years, the elegy is a comprehensive philosophical inquiry grappling with evolution, theology, and the psychology of grief. Tennyson was therefore compelled to deploy an immense and varied vocabulary, which results in a text that is lexically denser than anything else in his corpus.

The data indicate a slight but notable lexical contraction in the Late period. The *Idylls of the King* exhibits a lower lexical diversity score compared to the Middle period. Although the *Idylls* are syntactically the most complex (as seen in the earlier MDD analysis), they are lexically more restrained. This contraction likely reflects the demands of the epic genre, which favors consistency over variety. In order to maintain the cohesion of the Arthurian world, Tennyson relies on a stabilized high diction and recurring formulaic language.

While the MATTR and MTLTD metrics provided robust point estimates, a single average score can obscure the internal variance of a text. To test the stability of these lexical patterns, a bootstrapping analysis was performed. This technique moves beyond simple averages by generating a full sampling distribution. Instead of calculating a single TTR for the entire corpus, the code extracted multiple random samples of 5,000 tokens from each period. By standardizing the sample size, this method neutralizes the text-length bias that plagues traditional TTR calculations. The resulting boxplots reveal insights into the consistency of Tennyson's style:

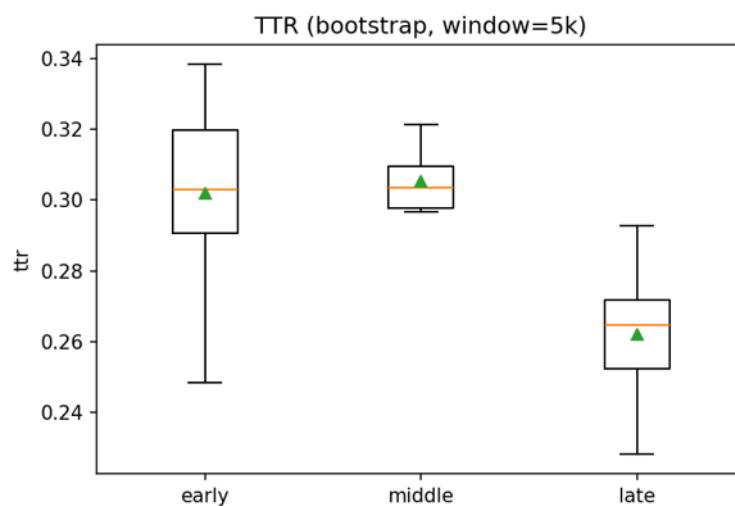


Figure 12 TTR (bootstrap, window=5k)

In Figure 12, the distribution for the Late period appears lower on the y-axis, with a median TTR markedly below that of the Early and Middle periods. This provides statistical confirmation of the lexical contraction hypothesis: the *Idylls of the King* employs a more restricted vocabulary across all its constituent parts. Moreover, a comparison of the Early and Middle plots reveals a fascinating structural difference regarding variance. The Early boxplot exhibits a tall box and long whiskers, indicating high variance. This reflects the nature of the corpus: The Early period is a collection of disparate lyrics, each with its own thematic vocabulary, which leads to fluctuations in lexical density. In contrast, the Middle period boxplot is compressed and narrow. This statistical tightness confirms that *In Memoriam*, despite being the most lexically rich (highest median), is also remarkably homogeneous. Overall, the bootstrapping analysis validates the previous findings: Tennyson's career trajectory involves a

peak of rich, consistent vocabulary in his middle years, followed by a disciplined contraction in his late epic phase.

The bootstrapping protocol was also applied to the MATTR. This approach combines a length-independent metric with a distributional sampling method, thereby neutralizing both text-length bias and local anomalies.

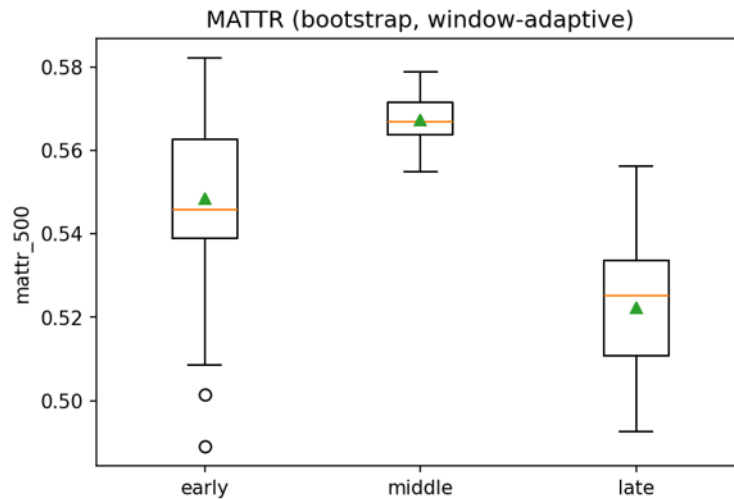


Figure 13 MATTR (bootstrap, window-adaptive)

As shown in Figure 13, the Middle period occupies the highest position on the y-axis. The box itself—representing the interquartile range—is exceptionally compact. This indicates that the high lexical diversity of *In Memoriam* is not sporadic but systemic. The Early period displays a median lower than the Middle but significantly higher than the Late. However, the most striking feature is its variance. The box is taller, and the whiskers extend widely, with visible outliers (circles) at the bottom. This high dispersion confirms the anthological nature of the Early corpus. On the other hand, the Late period is situated lowest on the scale. The distribution lacks the extreme variance of the early works. This reinforces the conclusion that the *Idylls of the King* relies on a relatively restricted lexicon.

To conclude the investigation into lexical diversity, the study applied the bootstrapped MTLT. As the most sophisticated metric in the current stylometric toolkit, MTLT offers a unique perspective: Rather than simply counting unique words, it measures lexical endurance—specifically, how long a text can sustain its narrative before succumbing to repetition. This metric is impervious to the distorting effects of text length. A higher score, therefore, indicates a text that resists repetition for longer stretches. Figure 14 provides the final corroboration of the Middle Peak hypothesis:

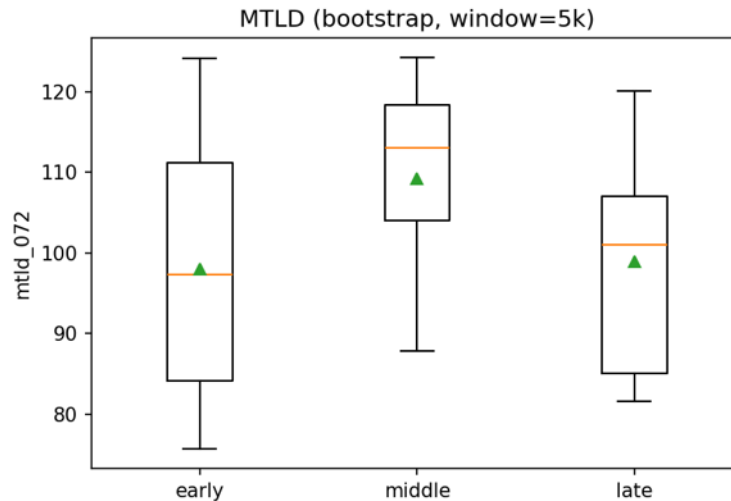


Figure 14 MTLD (bootstrap, window=5k)

The distribution for the Middle period is superior, with the highest median score exceeding 110. This signifies that *In Memoriam* possesses the highest repetition threshold in the corpus. This quantitative resilience mirrors the poem's intellectual complexity. The Late period falls short of the Middle period's diversity, which indicates that the epic voice is structurally more repetitive.

The most consistent finding across the three bootstrapped metrics is the prominence of the Middle period. This provides strong evidence that the elegy marks the peak of Tennyson's lexical richness and vocabulary dynamism. The philosophical demands of the poem appear to have compelled him to operate at the height of his verbal resources.

3.2.4 Statistical Verification of Lexical Variance

For rigorous hypothesis testing, a Kruskal-Wallis H-test was performed on the bootstrapped lexical data. This non-parametric method was selected as the appropriate alternative to One-Way ANOVA, as it allows for the comparison of three independent groups (Early, Middle, Late) without assuming a normal distribution of the underlying data. The objective was to test the null hypothesis that the lexical diversity distributions of the three creative periods are identical and that any observed differences are the result of random chance.

Table 4 Kruskal-Wallis H-test

Measure	H	p-value	Epsilon Squared
TTR	83.65355503	6.83693E-19	0.555466361
MATTR_500	90.33703841	2.41858E-20	0.600932234
MTLD_072	23.43699073	8.14183E-06	0.145829869

The results, summarized in Table 4, provide evidence against the null hypothesis. The p-values for all three metrics are infinitesimally small, falling orders of magnitude below the standard scientific threshold of 0.05. These values confirm that the probability of these stylistic differences occurring by chance is virtually zero. Consequently, Tennyson's lexical evolution is a genuine, measurable phenomenon.

While the p-value confirms the existence of a difference, the Epsilon Squared (ϵ^2) statistic measures the magnitude of that difference. Both TTR and MATTR demonstrate a large effect size. This implies that the Period variable explains a substantial portion (approx. 60%) of the variance in lexical diversity. Interestingly, MTLT shows a medium effect size. This suggests that although lexical endurance (the ability to avoid repetition) does change across periods, it is comparatively more stable than the diversity measured by MATTR.

3.2.5 Pairwise Distinction: Dunn's Post-Hoc Analysis

The Kruskal-Wallis test established that statistically significant differences exist somewhere within the corpus, but it did not identify the sources of variance. To isolate the location of these shifts, we conducted a Dunn's post-hoc test. This non-parametric method performs pairwise comparisons between every possible combination of periods (Early vs. Middle, Middle vs. Late, and Early vs. Late) to determine which relationships drive the global divergence.

Table 5 Dunn's post-hoc test

	Early	Middle	Late
Early	1	0.000044	1
Middle	0.000044	1	0.000164
Late	1	0.000164	1

The results for the MTLT metric (the most robust measure) presented in Table 5 reveal a striking structural pattern. The test confirms that the Middle period is the stylistic outlier. In both cases (Early vs. Middle, Middle vs. Late), the p-values are orders of magnitude below the $\alpha = 0.05$ threshold. This proves that the lexical richness of *In Memoriam* is fundamentally distinct from both the poetry that preceded it and the epics that followed. Another revealing finding appears in the comparison between the start and end of Tennyson's career (Early vs. Late: $p = 1.0$). This p-value indicates no statistically significant difference in lexical endurance between the Early and Late periods. This pairwise analysis refines our understanding of Tennyson's trajectory. It suggests that his career does not describe a linear evolution ($A \rightarrow B \rightarrow C$) but rather a parabolic arc ($A \rightarrow B \rightarrow \hat{A}$). While the styles of the early lyrics and late epics are qualitatively different (as seen in the Cluster Analysis), their lexical diversity profiles are statistically homologous.

3.2.6 Quantifying Archaic Diction

The investigation now turns to the specific historical texture of Tennyson's language. We conducted a Frequency of Archaism analysis using a two-pronged lexical retrieval method: 1) The Archaic Lexicon: A predefined list of Early Modern and poetic English function words (e.g., pronouns like *thee*, *thou*, *thine* and auxiliaries like *hath*, *doth*, *art*). 2) Morphological Suffixing: A regex-based search for historical verb inflections, specifically the third-person singular singular "-eth" (e.g., *goeth*, *cometh*) and the second-person singular "-est" (e.g., *lovest*, *speakest*).

Table 6 Archaic Lexicon

Period	Archaic Lexicon Hits	-eth/est Hits	Archaic Total	Per 10k	Tokens
--------	-------------------------	------------------	------------------	---------	--------

Early	1260	332	1592	239.94333	66349
Middle	438	93	531	280.83351	18908
Late	2153	436	2589	304.573902	85004

The results in Table 6, normalized per 10,000 tokens to account for text length, reveal a linear escalation in the density of archaic usage. The data confirm that the Late period represents the peak of Tennyson's archaism. With a normalized frequency of 304.57, the *Idylls of the King* is significantly more antique than the early lyrics. In composing the *Idylls*, Tennyson faced the task of resurrecting the medieval world of Arthurian legend for a Victorian audience. To achieve this, he made a conscious stylistic decision to saturate the poem with linguistic markers of the past. By weaving archaisms into the fabric of the verse, Tennyson establishes a grand, epic tone that separates the mythic time of Camelot from the modern industrial present.

3.2.7 Deep Grammatical Drift: JSD

Whereas the previous analyses focused on conscious stylistic choices, the Jensen-Shannon Distance (JSD) allows us to investigate a subconscious layer of style: the distribution of function words (e.g., *the, and, of, to, in*). Because these words constitute the non-lexical glue of the language, their usage patterns often represent a deep, unconscious authorial habit. As presented in Table 7, the study computed the JSD between the probability distributions of function words for each period. JSD is a metric in which a score of 0 denotes identical distributions, with higher scores indicating greater divergence.

Table 7 Jensen-Shannon distance (JSD)

	Mean – Median
Early to Middle	0.10462
Early to Late	0.12089
Middle to Late	0.15030

First, all the JSD scores fall below 0.2. In stylometric terms, this confirms that the corpus retains a consistent authorial fingerprint across periods. Despite the generic shifts from lyric to elegy to epic, the fundamental fingerprint of Tennyson's grammar remains recognizable across his career. However, the data indicate a progressive stylistic drift. The most significant finding is that the largest grammatical divergence occurs between the Middle and Late periods (0.150). This grammatical drift is sharper than the drift from his youth to his middle age. It suggests that the transition to the *Idylls of the King* involved a profound restructuring of Tennyson's syntactic rhythm. He did not merely add archaic words to his existing style; he altered the subconscious connective tissue of his verse to accommodate the weightier, more complex syntax of the epic mode.

3.2.8 Grammatical Architecture: POS Distribution

Next, the study calculated the Part-of-Speech (POS) Distribution for each period. By measuring the proportional frequency of major grammatical categories, we can trace the syntactic rhythm of Tennyson's evolving career.

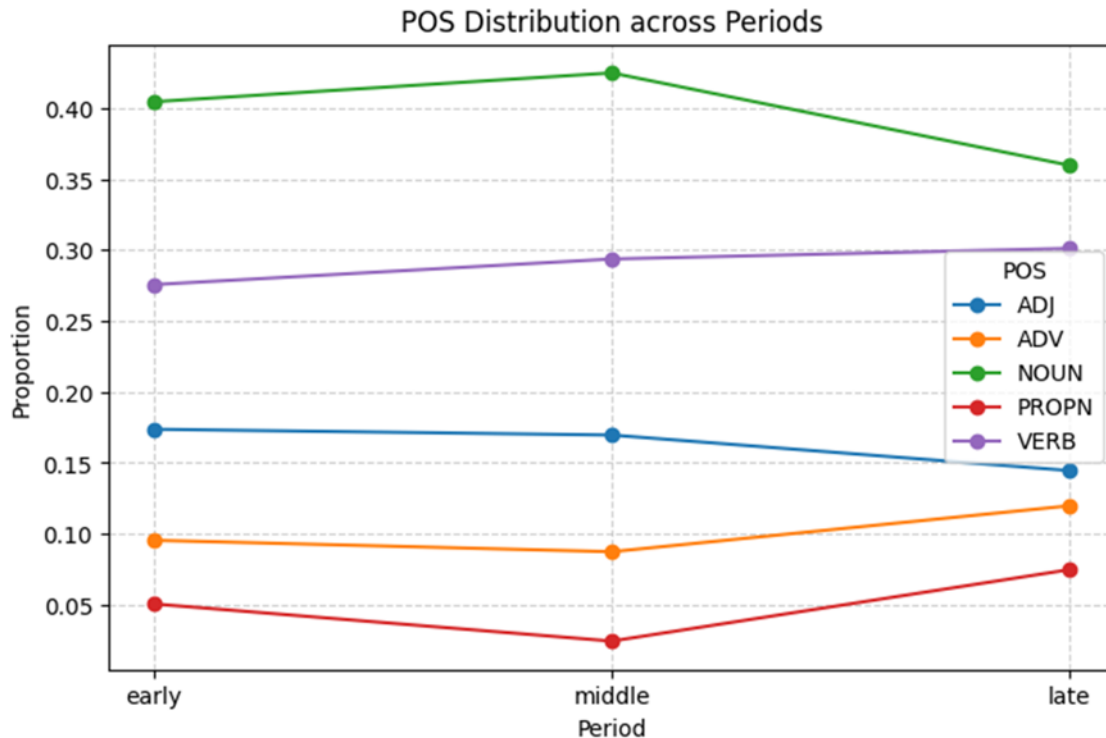


Figure 15 POS Distribution

In Figure 15, the Early period is defined by a peak in adjectival usage (Blue Line). The proportion of adjectives is at its highest in the early lyrics and steadily declines toward the Late period. This confirms the characterization of the early style as richly descriptive and lyrical. In contrast to the adjectives, the verbal frequency (Purple Line) exhibits a consistent upward trend. Verbs rise from approximately 27% in the Early period to over 30% in the Late period. The *Idylls of the King* is driven by plot—quests, battles, and betrayals. To sustain this forward momentum, Tennyson sheds the decorative weight of adjectives in favor of the muscularity of verbs. The most dramatic fluctuation occurs in proper nouns (Red Line). Usage dips to its lowest point in the Middle period but spikes sharply in the Late period. This validates the thematic distinction between *In Memoriam* and the *Idylls*. The Middle period is reflective and abstract; it concerns universal nouns rather than specific people, which results in the lowest count of proper names. Conversely, the Late spike reflects the storytelling of the epic, which necessitates a vast cast of named characters (Arthur, Lancelot, Guinevere) and geographic locations to populate its world. In short, the POS distribution provides strong, independent evidence that Tennyson dismantled the adjectival, descriptive grammar of his youth and reconstructed a verbal, action-oriented syntax to support the narrative weight of his late epics.

3.2.9 Sentiment Valence and Emotional Landscapes

To conclude the investigation, the study examined the emotional trajectory of Tennyson's career. Two different NLP tools were employed to capture different dimensions of the poet's affective landscape: VADER, a rule-based model designed to measure sentiment intensity and particularly well suited to short, context-sensitive expressions of valence, and TextBlob, used here specifically for measuring subjectivity versus objectivity (Hutto and Gilbert 2014; Loria 2020; De Smedt and Daelemans 2012).

Table 8 Sentiment Analysis: VADER

Period	vader_neg	vader_neu	vader_pos	vader_compound
Early	0.065246	0.824993	0.097722	0.046102
Middle	0.078861	0.806498	0.114312	0.046648
Late	0.073644	0.828789	0.097566	0.038754

The VADER analysis in Table 8, which calculates a compound sentiment score ranging from -1 (most negative) to +1 (most positive), reveals a stability in Tennyson's overall emotional tone. The slight dip in the Late period (0.038) reflects the tragic arc of the *Idylls of the King*, where the dissolution of the Round Table inevitably pulls the narrative toward a darker, more somber valence. However, the consistency of the scores suggests that Tennyson maintains a disciplined emotional baseline throughout his work, never succumbing to fluctuations of mood.

Table 9 Sentiment Analysis: TextBlob

Period	tb_polarity	tb_subjectivity
Early	0.119063	0.509933
Middle	0.111464	0.497202
Late	0.131354	0.521692

More nuanced insights emerge from the Subjectivity analysis in Table 9. The TextBlob metric (where 0 is objective and 1 is subjective) uncovers a U-shaped pattern that isolates the Middle period. *In Memoriam*, universally regarded as Tennyson's most intensely personal poem of grief, registers as the least subjective text in the corpus. This finding quantitatively supports the idea that the poem reads not simply as a cry of pain but as a philosophical inquiry. In grappling with the death of Arthur Hallam, Tennyson moves beyond the subjective self to interrogate universal laws of nature, theology, and evolution. The lower subjectivity score reflects this intellectual distancing.

4. CONCLUSION

This study set out to map the stylistic trajectory of Tennyson through the lens of computational stylistics. By integrating macroscopic distant-reading methods (stylo) with granular linguistic analyses (Python), we have quantified the stylistic fingerprint of Tennyson's evolution. The most definitive finding of this research is the statistical reality of the late style. As demonstrated by the PCA and Cluster Analyses, the *Idylls of the King* represents a rupture from the earlier corpus. The analysis confirms that this shift reflects an active syntactic reconstruction designed for the epic mode. To sustain the narrative weight of the Arthurian legend, Tennyson deliberately expanded his syntax—increasing sentence length and MDD; reoriented his grammar—moving from adjectival description to verbal action to drive the plot; and constructed a synthetic medievalism—systematically increasing the frequency of archaisms to establish a mythic distance. Moreover, the observed contraction in lexical diversity (lower TTR/MATTR) during this period is not a sign of decline but a functional necessity of the epic genre.


The Middle period emerges from these data as a distinct stylistic formation. Despite being structurally closer to the early works (clustering together in PCA), *In Memoriam* stands apart as the peak of lexical density. The characteristic *In Memoriam* stanza imposed a prosodic discipline that regularized Tennyson's syntax (lower MDD), whereas the philosophical demands of the elegy pushed his vocabulary to its upper limit.

Finally, even as Tennyson's syntax, vocabulary, and grammatical structures underwent marked transformation, the sentiment valence of his poetry remained stable across his career. Yet this emotional steadiness did not preclude deeper shifts in how he positioned himself within his verse. The subjectivity analysis reveals that his Middle period—paradoxically the era of his most personal grief—displays a turn toward objectivity, as *In Memoriam* reframes private sorrow into philosophical inquiry. For Tennyson, the mastery of grief required suppressing the lyric “I” in favor of a universalizing philosophical voice. In sum, computational stylistics reveals Tennyson not as a fixed monument of the Victorian age but as a dynamic architect of language who continually reshaped his style to meet the evolving demands of his art.

Methodologically, this study demonstrates the value of stylometry for understanding Victorian poetry. Computational analysis does not replace close reading; rather, it helps identify patterns that can sharpen literary interpretation. By measuring the linguistic features through which Tennyson's style changes, the study provides empirical support for long-standing critical claims while also refining them. It shows that terms such as “lyric,” “elegiac,” and “epic” function not only as generic or thematic categories but also as quantifiable linguistic configurations.

Several limitations remain. The present study aggregates texts into broad chronological periods, and future research could examine finer-grained stylistic changes within each period—especially within the long compositional history of *In Memoriam* and the multi-decade publication history of the *Idylls*. Further work might also compare Tennyson's stylistic trajectory with those of other Victorian poets, such as Browning, Arnold, or Swinburne, in order to determine which features are uniquely Tennysonian and which reflect broader nineteenth-century poetic developments. Even with these limitations, however, the present analysis demonstrates that Tennyson's movement from early lyrics to late epics is not only a literary-historical narrative but also a measurable transformation in the deep structure of poetic language.

ORCID

Sunghyun Jang  <https://orcid.org/0009-0009-0685-4772>

FUNDING

Funding support for this article was provided by the Humanities Utmost Sharing System (HUSS) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education.

REFERENCES

- Armstrong, Isobel. 1993. *Victorian Poetry: Poetry, Poetics and Politics*. London: Routledge.
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Binongo, José Nilo G., and M. W. A. Smith. 1999. "The Application of Principal Component Analysis to Stylometry." *Literary and Linguistic Computing* 14 (4): 445–66. <https://doi.org/10.1093/lc/14.4.445>.
- Burrows, John F. 2002. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17 (3): 267–87. <https://doi.org/10.1093/lc/17.3.267>.
- Carretero González, Margarita. 2012. "Floating down beyond Camelot: The Lady of Shalott and the Audio-Visual Imagination." *Odisea* 13: 237–54.
- Covington, Michael A., and Joe D. McFall. 2010. "Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR)." *Journal of Quantitative Linguistics* 17 (2): 94–100. <https://doi.org/10.1080/09296171003643098>.
- De Smedt, Tom, and Walter Daelemans. 2012. "Pattern for Python." *Journal of Machine Learning Research* 13 (66): 2063–67.
- Douglas-Fairhurst, Robert. 2010. "Tennyson." In *The Cambridge History of English Poetry*, edited by Michael O'Neill, 596–616. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CHOL9780521883061.034>.
- Dunn, Olive Jean. 1964. "Multiple Comparisons Using Rank Sums." *Technometrics* 6 (3): 241–52. <https://doi.org/10.1080/00401706.1964.10490181>.
- Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8 (1): 107–21. <https://journal.r-project.org/articles/RJ-2016-007/index.html>.
- Eliot, T. S. 1936. "In Memoriam." In *Essays Ancient and Modern*, 186–203. London: Faber and Faber.
- Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. "spaCy: Industrial-Strength Natural Language Processing in Python." Computer software. Zenodo. <https://doi.org/10.5281/zenodo.1212303>.
- Hunter, John D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Hutto, Clayton J., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 216–25. AAAI Press. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Kestemont, Mike. 2014. "Function Words in Authorship Attribution: From Black Magic to Theory?" In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, edited by Anna Feldman, Anna Kazantseva, and Stan Szpakowicz, 59–66. Gothenburg, Sweden: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-0908>.
- Kincaid, James R. 1975. *Tennyson's Major Poems: The Comic and Ironic Patterns*. New Haven, CT: Yale University Press.
- Kooistra, Lorraine Janzen. 2002. "Poetry and Illustration." In *A Companion to Victorian Poetry*, edited by Richard Cronin, Alison Chapman, and Antony H. Harrison, 392–418. Oxford: Blackwell.

- Kruskal, William H., and W. Allen Wallis. 1952. "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association* 47 (260): 583–621. <https://doi.org/10.1080/01621459.1952.10483441>.
- Lin, Jianhua. 1991. "Divergence Measures Based on the Shannon Entropy." *IEEE Transactions on Information Theory* 37 (1): 145–51. <https://doi.org/10.1109/18.61115>.
- Liu, Haitao. 2008. "Dependency Distance as a Metric of Language Comprehension Difficulty." *Journal of Cognitive Science* 9 (2): 159–91. <https://doi.org/10.17791/jcs.2008.9.2.159>.
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv. <https://doi.org/10.48550/arXiv.1907.11692>.
- Loria, Steven. 2020. *TextBlob: Simplified Text Processing*. Version 0.16.0. Computer software. <https://textblob.readthedocs.io/en/dev/>.
- Lu, Xiaofei. 2010. "Automatic Analysis of Syntactic Complexity in Second Language Writing." *International Journal of Corpus Linguistics* 15 (4): 474–96. <https://doi.org/10.1075/ijcl.15.4.02lu>.
- McCarthy, Philip M., and Scott Jarvis. 2010. "MTLD, vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment." *Behavior Research Methods* 42 (2): 381–92. <https://doi.org/10.3758/BRM.42.2.381>.
- Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- Pang, Bo, and Lillian Lee. 2008. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135. <https://doi.org/10.1561/1500000011>.
- Perry, Seamus. 2005. *Alfred Tennyson. Writers and Their Work*. Tavistock: Northcote House.
- Project Gutenberg. n.d. "Project Gutenberg." Accessed September 1, 2025. <https://www.gutenberg.org/>.
- Ricks, Christopher, ed. 1987. *The Poems of Tennyson*. 2nd ed. 3 vols. Harlow: Longman.
- Ricks, Christopher. 1989. *Tennyson*. 2nd ed. Berkeley: University of California Press.
- Shaw, W. David. 1976. *Tennyson's Style*. Ithaca, NY: Cornell University Press.
- Tucker, Herbert F. 1988. *Tennyson and the Doom of Romanticism*. Cambridge, MA: Harvard University Press.
- Tucker, Herbert F. 1991. "The Epic Plight of Troth in *Idylls of the King*." *ELH* 58 (3): 701–20. <https://doi.org/10.2307/2873461>.
- Tweddle, Fiona J., and R. Harald Baayen. 1998. "How Variable May a Constant Be? Measures of Lexical Richness in Perspective." *Computers and the Humanities* 32 (5): 323–52. <https://doi.org/10.1023/A:1001749303137>.
- van Rossum, Guido, and Fred L. Drake Jr. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems* 30, 5998–6008. Red Hook, NY: Curran Associates.
- Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, et al. 2020. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." *Nature Methods* 17: 261–72. <https://doi.org/10.1038/s41592-019-0686-2>.