A Study on Characteristic Sounds in Tennyson and Browning: Using Stylometric Approaches

Iku Fujita

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract This study investigates whether tabulated phoneme frequencies may reveal aspects of Alfred Tennyson's (1809–1892) and Robert Browning's (1812–1889) works. Tennyson and Browning are two of the most representative Victorian poets in the United Kingdom. Although studies on their topics, motifs, and styles, as well as comparative studies of Tennyson, Browning, and other poets, abound, only a few studies have divided the sounds in their works into phonemes and examined them. In addition, few studies have used quantitative methods to contrast the tones in Tennyson's and Browning's works. To perform a more in-depth analysis of Tennyson's and Browning's poetry collections, this study examines the top 50 phoneme frequency poems from each author while considering the limitations of earlier studies using cluster analysis and principal component analysis. This study finds that the tabulated international phonetic alphabet (IPA) data effectively identifies the differences between the two writers and captures the features present in their poetry using cluster and principal component analyses.

Keywords Alfred Tennyson, Poetry, Robert Browning, Stylometry, Sound

Tennyson と Browning の音声比較: 計量文体論手法を用いて

藤田　郁

大阪大学大学院言語文化研究科

〒 560-0043 豊中市待兼山町 1-8

E-mail: u256780k@ecs.osaka-u.ac.jp

あらまし 本稿は，19 世紀を代表する 2 詩人，アルフレッド・テニスン (1809–1892) とロバート・ブラウニング (1812–1889) の作品を国際音声字母 (IPA) に変換し，IPA 記号から得られる頻度表に計量文体論的アプローチ（クラスター分析及び主成分分析）を応用することにより，各詩人の作品の音声特徴を明らかにすることができるかを探索的に分析することを目的とする。テニスンとブラウニングは，イギリスを代表するヴィクトリア朝の詩人であり，彼らの詩に現れるテーマやモチーフ，各々の文体に関する研究や，テニスン，ブラウニングのそれと他の詩人のそれとを比較する質的研究は数多くあるものの，作品中の音を音素に分割し，分析した研究はごくわずかである。また，テニスンとブラウニング の作品に含まれる音の要素を定量的な手法で比較した研究も少ない。本研究では，テニスンとブラウニングの韻文についてより詳細な分析を行うため，先行研究に挙げられる制約を考慮しつ

つ，各作家の音素頻度上位 50 の詩を対象として分析した。クラスター分析及び主成分分析の結果により，2 人の作家の作品内で使用されている音の違いが効果的に識別され，各詩人の特徴を捉えることができることが明らかになった。

キーワード　アルフレッド・テニスン，韻文，音声，計量文体論，ロバート・ブラウニング

1. Introduction

Alfred Tennyson (1809–1892) and Robert Browning (1812–1889) are two of the most representative Victorian poets in the United Kingdom. While studies on their themes, motifs, and styles, as well as comparative studies of Tennyson, Browning, and other poets, abound, only a few have divided the sounds in their works into phonemes and analyzed them. Moreover, only a few studies have compared sounds in Tennyson's and Browning's works using quantitative approaches. Tawfiq (2020) indicated the number of phonological poetic devices (e.g., alliteration, onomatopoeia, consonance, assonance, and rhyme). While Tawfiq (2020) succeeded in highlighting the differences in the five poetic devices between Tennyson and Browning using decimal information, some issues remained because its target data were limited to five examples from ten randomly selected poems by Tennyson and Browning. Notably, detecting (phonological) poetic devices poses a significant challenge for analysts, as it fully depends on qualitative, close reading techniques. Therefore, the limitation of the data under investigation was inevitable. However, the question remains whether the data and results of Tawfiq could represent Tennyson's and Browning's entire works. Another question is whether the numerical differences between Tennyson and Browning were statistically significant. Tawfiq concluded that Browning's frequency of onomatopoeia was significantly higher than that of Tennyson. However, the difference in their frequencies of onomatopoeia was only two: Browning used onomatopoeia four times in his selected poems, while Tennyson used it just twice (Tawfiq, 2020: 32).

Plamondon (2005: 153) issued the problems with phonetic analysis, mentioning that the terminology, symbols, and rules, as well as the large amounts of phonetic data, made the phonetic analysis of literature uneasy for students. The latter issue can also overwhelm analysts and researchers because the difficulty in finding phonological devices increases with larger amounts of data. However, an automated model for detecting phonological poetic devices remains unestablished. Plamondon (2005) challenged the difficulties in analyzing phonetic elements in poetry using the program AnalysePoems, which is 'designed to assist students to enhance their interaction with poetry and assist scholars in accumulating data, phonetic and otherwise, about poems' (Plamondon, 2005: 154). Plamondon described some results of his explorative analyses of Tennyson's and Browning's poems, which were converted into phonetic alphabets, through AnalysePoems.

Based on previous research and criticisms on sounds in Tennysonian and Browning's poems, Plamondon's attempts were 'to determine what in Browning's poetry produces the 'clamorous chaos' and what in Tennyson's poetry produces the 'melodious tone' and 'to determine if the

phonetic makeup of (Tennyson's and Browning's) poetry differs in any substantial manner.' Plamondon hypothesized that the more 'musical' Tennyson's poetry is, the more pleasant-sounding vowels it should have; the more 'harsh' Browning's poems are, the more harsh-sounding consonants they should contain (Plamondon, 2005: 163–164). Notably, Plamondon obtained the results by counting the phonemes in several poems of Tennyson and Browning, tabulating them, and calculating the proportion of each phoneme in the poems. The outcome was that Tennyson uses more vowels than Browning while Browning employs a higher proportion of consonants in his writing. Furthermore, it suggested that Tennyson's poetry contained shorter vowel sounds, which were allegedly responsible for its melodic quality, whereas Browning's poetry had more plosive consonants, which are presumably responsible for its harshness. However, as Plamondon already argued, the proportions of phonemes in Tennyson's and Browning's works differed slightly. Plamondon mentioned that the restrictions on the number of works might have caused the insignificant divergences. Plamondon continued:

> It seems, therefore, that either the counting of individual phonemes and tabulating them by class is not sufficient to characterize the difference in the quality of their poetry, or the long held opinion about the relative 'musical' quality of their verse is flawed.
>
> (Plamondon, 2005: 166)

A mere tabulation of individual phoneme frequencies might be insufficient to explain the differences in poetry. However, other quantitative approaches exist for effectively finding characterizations in poems using tabulated information. Considering the limitations and issues of previous studies, this study aims to explain whether tabulated phoneme frequencies can demonstrate the characteristics of Tennyson's and Browning's poetry and, if so, which phonemes are features of each poet. To investigate the poems of Tennyson and Browning more widely, this study used the top 50 poems of phoneme frequencies from Browning and Tennyson. The 100 poems of Tennyson and Browning were further tabulated and analyzed via cluster analysis (CA) and principal component analysis (PCA). The results were visualized using `stylo` by `R` (Eder et al., 2016), `Python`, and `Gephi` (Bastian et al., 2009) to facilitate the reading of results.

2. Methodology and Data

2.1. Methodology: Quantitative approaches in this study

This study employed CA and PCA to detect the characteristics of Tennyson's and Browning's poems. CA and PCA are statistical methods used to classify text variables (elements) into several groups. Stylometric studies often use these methods to reveal the stylistic features of their target texts or corpora and to detect authorship distributions (Burrows, 1987; Hoover, 2012; Tabata,

2004; Tabata, 2012; to name but a few). CA is a statistical method that uses the information in the data to form clusters by grouping data that are close to each other (Ishikawa et al., 2010: 163). The two main types of CA are hierarchical and nonhierarchical CAs. However, this paper focuses on the former, which can draw a dendrogram with results. In text mining and analyses, the groupable data can be the frequency of words in the texts in question. Relative frequencies are calculated based on raw frequencies from the texts; then, the computer locates each value, calculates the distance between each value, and groups close-distant data into the same groups.

There are multiple choices for calculating distance (e.g., Euclidean distance, Canberra distance, and Manhattan distance) as well as from where to calculate the distance (e.g., Ward and Complete methods). Further analyses were conducted using the Euclidean distance and Ward method, commonly employed in stylometric research. When Euclidean distance was employed, a mismatch of values occurred. Although the Euclidean distance value can be positive or negative, the relative frequencies of words would never have a negative value. The relative frequencies were z-scored before the analysis to avoid discrepancies in the values. The conditions of distance and method were applied to each analysis in this paper. PCA is a technique that compresses information into a small number of components when the number of variables is large. By compressing information into a small number of components, it is possible to efficiently extract values that are representative of the information in the data (Ishikawa et al., 2010: 193). Therefore, PCA is one of the multivariate analysis methods for data reduction, along with factor analysis and correspondence analysis. PCA facilitates visually grasping the overall picture of data by aggregating complex relationships among many variables and samples into a small number of comprehensive indicators (i.e., principal components) (Tabata, 1998: 195). In PCA, eigenvalues and eigenvectors can be obtained from a correlation matrix or covariance matrix (Tabata, 2004: 102). This study used a correlation matrix by referencing Tabata (2004) throughout all analyses.

2.2. Data: Original poems

This study focuses on two Victorian poets—Alfred Tennyson and Robert Browning. The original text data were compiled as corpora from the Delphi Poet Series and Ricks (1987). To build the corpora, all the texts were converted to plain text data using optical character read/recognition, which was then manually edited and proofread for problems, including unnecessary spaces, tabs, newlines, and garbled characters.

Table 1 shows the bibliographical data of Tennyson and Browning corpora. The Tennyson corpus contains 600 verse texts, with 423 taken from the Delphi Poet Series 'Alfred, Lord Tennyson' (2013) and the remaining 177 compiled from Ricks (1987). Browning's 204 poems are compiled from the Delphi Poets Series 'Robert Browning' (2012). In Table 1, Tennyson's number of works is almost three times that of Browning; however, Browning's total number of tokens is nearly double that of Tennyson. The data suggest that Browning wrote more lengthy/wordy poems than Tennyson did.

Table 1: Bibliographical data of Tennyson and Browning corpora

Authors	Num. of works	Num. of total tokens	Type tokens
Tennyson	600	355,235	21,661
Browning	204	590,315	36,159

The plain text data in Tennyson and Browning corpora are further converted to the data described in the international phonetic alphabets (IPA) following the method in Section 2.3.

2.3. Methodology: IPA conversion

To investigate the phonological elements in Tennyson and Browning, the original poetry data were converted to phonetic symbols—IPA. The online resource EasyPronunciation.com (https://easypronunciation.com/en/) was subscribed to and used for its conversion. The subscribed version of EasyPronunciation.com allowed us to translate 10 thousand alphabetical characters of English and other languages into phonetic transcription per run. For English translation, moreover, there are options for American, British, and Australian English accents. Hence, the English dialect of the target corpora is British English, and British English was chosen for the translation of this study (Figure 1). Some words, such as proper nouns, could not be converted because they are not in the dictionary of the tool, but these words are surrounded by voluntary symbols (e.g., #Camelot#). Unconverted words were excluded from further analysis.

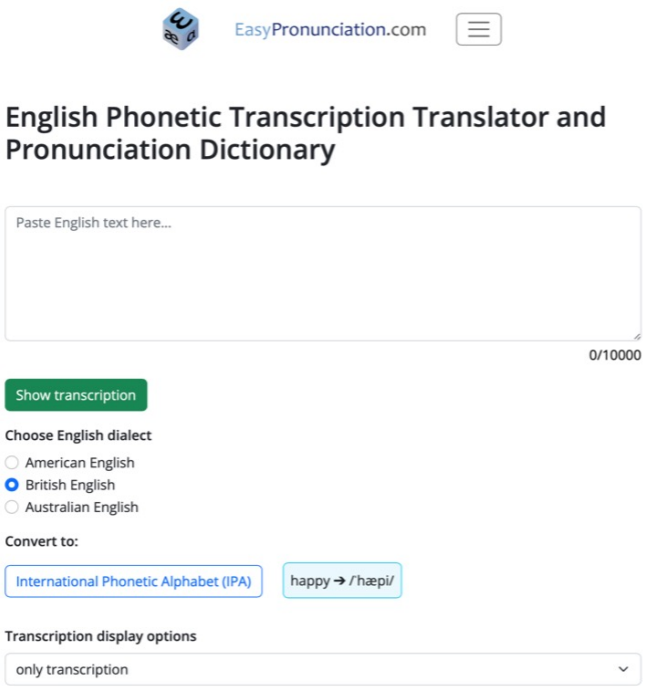

Figure 1: The top page of EasyPronunciation.com

2.4. Data: Converted to IPA

Table 2 lists the information on IPA-converted corpora. There were slight changes in the original text data. The difference in the data in Tables 1 and 2 is due to the deletions of unconverted words along with homophonic and heterophonic spelling.

Table 2: Bibliographical data of the IPA Tennyson and Browning corpora

Authors	Num. of works	Num. of total tokens	Type tokens
Tennyson	600	356,311	20,178
Browning	204	591,750	34,359

The IPA text data were further edited for more detailed analysis. Punctuations and other symbols for accents and quotations were deleted. Vowels and consonants were separated with spaces, but the contiguous vowels and consonants remained together. By doing so, we were able to investigate as small units of sound as possible while the group of vowels/consonants kept holding the sounds themselves, as explained with plural phonetic symbols (e.g., diphthongs /laɪk/ (like), long vowels /lɔːd/ (Lord), and consonant sounds expressed by a few alphabets /tʃɛːtʃ/ (church)). In this study, therefore, the word 'consonant(s)' does not necessarily mean that the legitimate phoneme is the smallest sound unit. However, it widely mentions the consecutive consonant group(s) found in a word. Here are the original poem lines (1) and the same lines of phoneme-divided IPA text (2):

(1) Strong Son of God, immortal Love,
Whom we, that have not seen thy face,

(2) str ɒ ŋ s ʌ n ə v g ɒ d ɪ m ɔː t ᵊ l l ʌ v
h uː m w iː ð ə̲ t h ə̲ v n ɒ t s iː n ð aɪ f e ɪ s

(Tennyson, *In Memoriam A.H.H.*, 1850: ll.1–2, underline added)

As can be seen in the IPA text of *In Memoriam A.H.H.* in (2), where underlines are added, the half (weak) vowel schwa /ə/ appears as superscripted in the first line, but two /ə/ in the second line are not superscripts. The conversion tool EasyPronunciation.com translates the schwa in nonstressed syllables into superscripts. Additionally, the syllabic (vocalic) consonant /l̩/ and nonsyllabic consonant /l/ are translated as separate sounds. Although there may be some disagreement about treating superscript /ᵊ/ and /ə/, syllabic /l̩/, and nonsyllabic /l/ as separate sounds, this paper treats them as different sounds each. Table 3 shows the bibliographical data of IPA text data, which has been divided into vowels (groups) and consonants (groups).

Table 3: Number and type of vowels and consonants of Tennyson and Browning corpora

Authors	Num. of vowels	Type tokens of vowels	Num. of consonants	Type tokens of consonants
Tennyson	487,139	20	597,193	293
Browning	828,936	19	1,025,789	266

Interestingly, the type tokens of Tennyson's consonants are 27 larger than Browning's, though Tennyson's total tokens and the number of consonants reach only about 60% of Browning's. The consonants found in Tennyson but not in Browning are listed in (3) with their raw frequencies in []:

(3) /ðh/ [1], /dθs/ [2], /fd/ [28], /fh/ [1], /ksd/ [34], /kslz/ [1], /ksp/ [1], /kstr/ [1], /kʃ/ [1], /lpd/ [4], /sh/ [1], /lʃ/ [2], /lθs/ [1], /mb/ [1], /mbr/ [1], /mfd/ [1], /mpd/ [5], /mpst/ [1], /mt/ [6], /mθs/ [1], /ndj/ [1], /ndlz/ [1], /nk/ [1], /nldz/ [2], /ntj/ [1], /ntld/ [1], /ntln/ [1], /ntʃd/ [22], /nw/ [1], /nzl/ [1], /ŋgd/ [1], /ŋgld/ [3], /ŋkd/ [17], /ŋr/ [1], /pb/ [1], /pd/ [28], /rl/ [28], /rlz/ [2], /skd/ [38], /spd/ [21], /sθ/ [1], /ʃd/ [207], /ʃm/ [1], /tb/ [1], /tʃd/ [133], /tt/ [1], /vl/ [1], /vm/ [1], /vnz/ [1], /θd/ [6]

On the contrary, the consonants found in Browning but not in Tennyson are shown in (4) with their raw frequencies in square brackets:

(4) /bld/ [1], /dnd/ [4], /dnz/ [1], /dʃ/ [1], /fθs/ [1], /ksθs/ [1], /ktr/ [1], /lb/ [2], /ldʒd/ [11], /lfθ/ [4], /lfθs/ [1], /ltʃ/ [8], /ltʃt/ [9], /mft/ [8], /mpts/ [34], /msn/ [6], /msnz/ [1], /ŋg/ [1], /ŋkst/ [1], /snt/ [11], /ss/ [1], /stlz/ [1], /vw/ [2]

The difference in vowel types causes the occurrence of the open/low-back unrounded vowel /ɑ/ in a Tennysonian poem, 'Stanzas' (1827, l. 14):

(5) In heav'n a more exalted throne,
ɪ n #heav'n# ə m ɔː ɪ g z ɑ l t ɪ d θr ə ʊ n

While the long open-back unrounded vowel /ɑː/ appears much more commonly in both Tennyson (raw frequency is 11,650) and Browning (raw frequency is 15,777), the short /ɑ/ appears only once in a work of Tennyson. Therefore, it seems reasonable to say that the use of extremely low frequency is more of a coincidence, or at least a feature of the work, than Tennyson's characteristics.

2.5. Trial PCA results

For further analysis, the IPA data were separated into the vowel and consonant folders of Tennyson and Browning. Before setting the number of target poems for the analyses, a trial PCA was conducted on the whole vowel and consonant data. Figures 2 and 3 show the results of PCA on vowels and consonants, respectively.

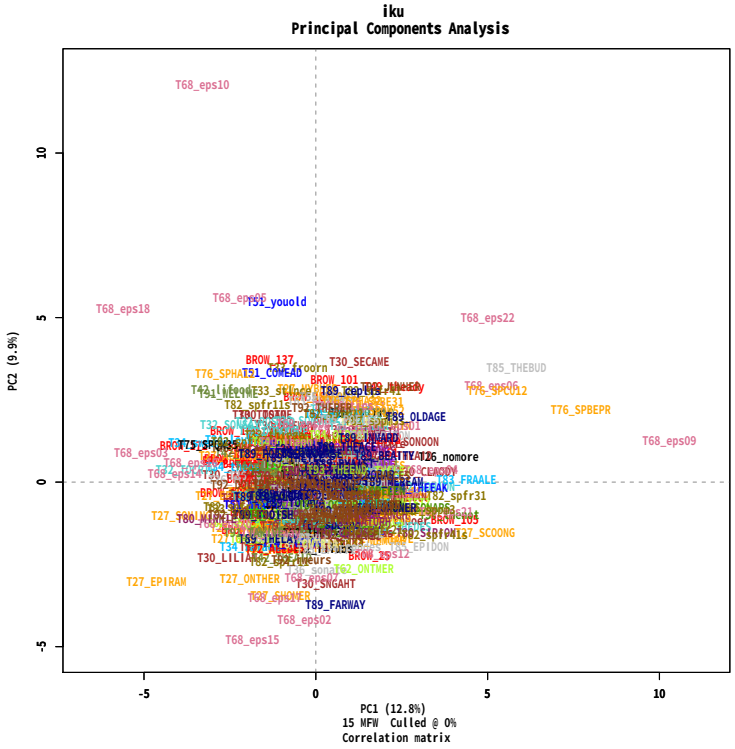

Figure 2: PCA results of the whole vowel data (804 works of Tennyson and Browning)

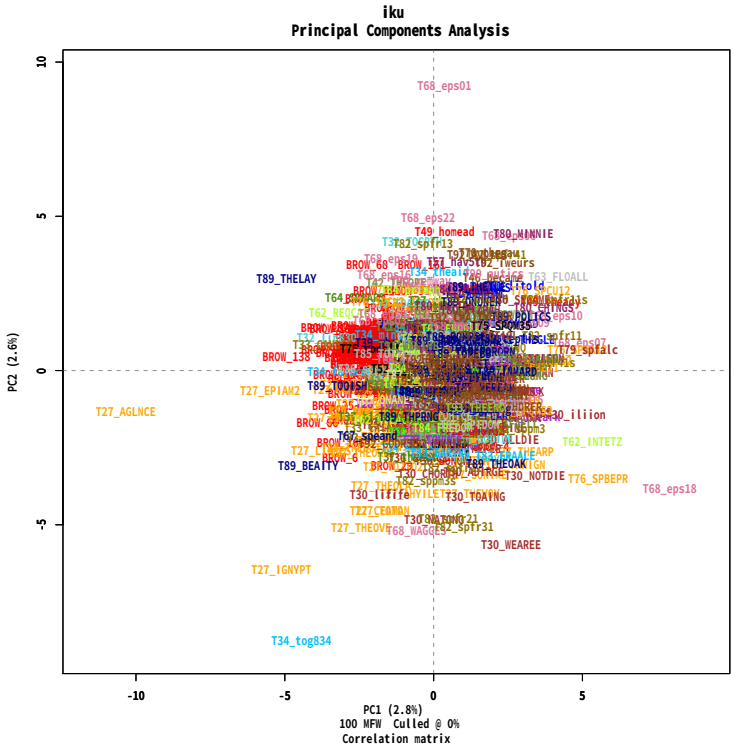

Figure 3: PCA results of the whole consonant data (804 works of Tennyson and Browning)

In both Figures 2 and 3, most of the variables (the title of works) congregate around the origins of the plots. The variables located apart from the origins are the short poems of Tennyson, whose total tokens are less than 100 words. Moreover, the contribution[1] ratios of each component were quite low (PC1 12.8% and PC2 9.9% in Figure 2; PC1 2.8% and PC2 2.6% in Figure 3); thus, it was estimated that calculating the entire 804 works at once was inapplicable. Considering the results that Tennysonian short poems scattered around while most of the other variables gathered around the origins, it is deemed that the number of poems for analysis should be reduced first, and second, the number of tokens of poems should not be too small and should not diverge too much. Therefore, the analyses in this study use 50 works from each poet, whose total tokens of vowels and consonants are the largest to 50th largest among the data.

3. Results and Discussion

This section illustrates the results of CA and PCA on the vowels (Section 3.1) and consonants (Section 3.2) of Tennyson's and Browning's top 50 poems.

3.1. Analysis of vowels in the top 50 works

This section presents the extracted results of CA and PCA for the vowels. Figure 4 shows the heatmap representation of the CA results. The shorter the length of each branch at the top and left

[1] Contribution (ratio) is a measure of how much of the variation in the data can be explained by this component alone. The higher the ratio is, the more a component describes the variation of the data.

sides of the plot, the shorter the distance between variables (i.e., the higher the similarity between variables). Conversely, when the length of each branch is longer, the distances between variables are farther; thus, the variables located farther are not alike. The heatmap is drawn by running CA on tabular data of vowel frequencies per work using CasualConc 3.0.4 (Imao, 2022). The poetry works are located horizontally, and the 20 vowels are located vertically. As the legend on the left top of the figure suggests, the color of a cell explains how significant the vowel in the poem where the cell interjects is. The legend has eight colors; the dark gray represents the most substantial, the first significant is brown, and the green explains the least significant. The dendrogram at the top of Figure 4 was broadly divided into two clusters with the two longest branches at the highest. A group on the left side of the figure is mostly clustered by Tennyson, while a group on the right side is mostly clustered by Browning (see Appendices 1 and 2 for the corresponding table of abbreviated file names and the original poem titles along with their publication years). However, some exceptions are located in the other poet's clusters, e.g., BROW_14 in the left, Tennyson's cluster: T30_SUPELF, T33_THEENT, T33_THETER, T42_THESIN, T42_WILOCK, T55_THEOOK, T62_ENODEN, T64_AYLELD, T68_LUCIUS, T80_SIRHAM, and T80_THEERS in the right, and Browning's cluster, which indicates that CA detected differences between the two authors' vowel usages. Figure 5 shows the results of PCA using the `stylo` package of `R`. The contribution ratio of principal component (PC) 1 is 22.0% and PC 2 is 12.6; therefore, it indicates that 34.6% of the data can be explained by the plot. The variables of Browning tend to scatter in the first and fourth quadrants; meanwhile, Tennysonian variables are inclined to be located in the second and third quadrants. Similar to the CA result, this PCA result shows the classification of the two authors' poems. Furthermore, it is possible to observe which vowels are the characteristics of which poets from the heatmap representation in Figure 4. The vertical black line in Figure 4 separates the plot from Tennyson's and Browning's clusters. The horizontal dotted line divides the vowels according to the dendrogram depicted on the left, which classifies the vowels in question. The heatmap indicates that the eight vowels, /aʊ/, /aɪ/, /ʊ/, /ə/, /ɑː/, /ɛː/, /iː/, and /ɔː/, are more significant in Tennyson than Browning. However, the other 11 vowels, /æ/, /ɒ/, /ɑ/, /ə/, /i/, /ɪ/, /u/, /ɔɪ/, /ʌ/, /uː/, and /e/, rather notably appear in Browning's poems. The result suggests that Tennyson tends to use diphthongs and long vowels, which are pronounced with a retracted and mid-retracted shape of the mouth, whereas Browning prefers short vowels, especially mid-retracted and rounded ones.

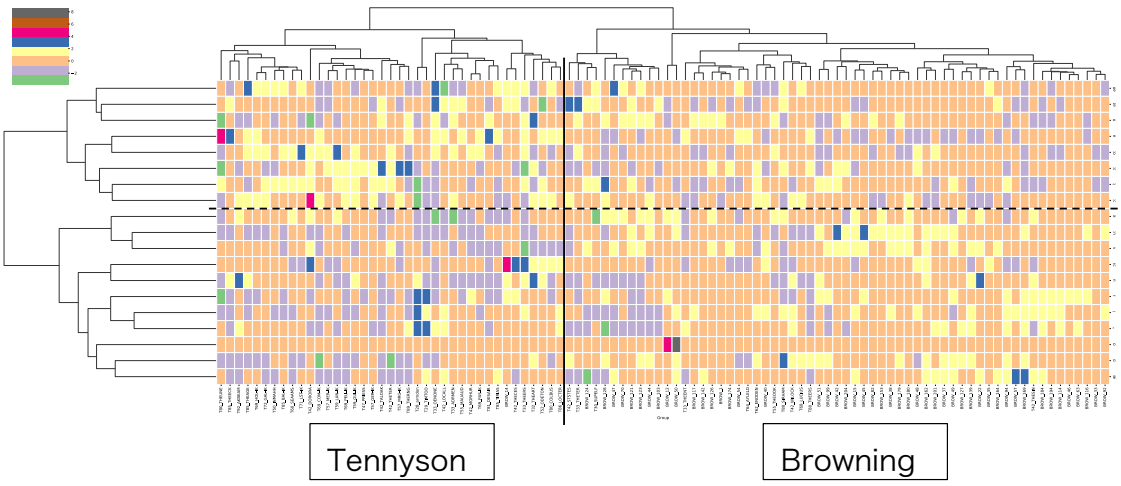

Figure 4: Heatmap representation of the results of CA on vowels

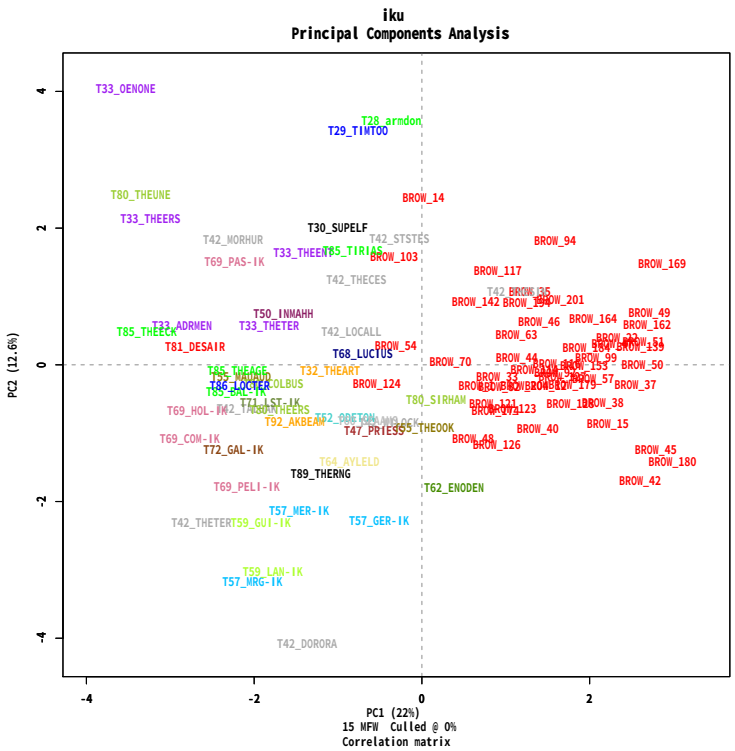

Figure 5: Results of PCA on vowels

Plamondon (2005: 163–164, italics added) assumed that 'musical' Tennysonian poem 'should have a greater proportion of pleasant-sounding *vowels*,' whereas 'harsh' Browning's poem 'should have a greater proportion of harsh-sounding *consonants*.' However, considering the results of CA, the musicalness and harshness of Tennyson and Browning may not be due only to the difference in vowels and consonants; rather, the musicalness and harshness of both vowels and consonants

have been caught by previous critics' and researchers' ears. That is, the long monophthongs and diphthongs introduce the musicalness of Tennyson, while the short vowels carry the harshness of Browning. Regarding musicalness and harshness, Section 3.2 discusses the consonant usages of the two poets. The `stylo` is also available to conduct CA. When CA is run on `stylo`, it draws a figure to facilitate reading the results and outputs the numerical results. Although the CA dendrogram is not used in this paper, the numerical result enables us to see how these variables (poems) relate and how similar they are by depicting a network graph (Figure 6).

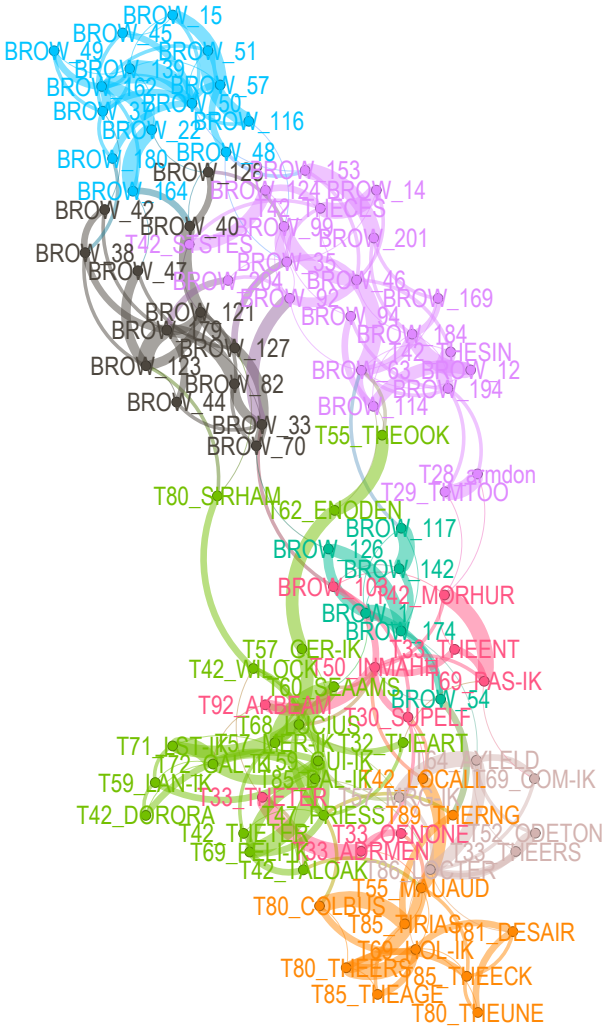

Figure 6: Network representation of vowels from the CA results

Figure 6 is drawn by `Gephi` using a numerical output of CA. The color differences depend on modularity, which indicates how unified the data constituents are. Thus, the elements colored the same are the close-knit poem groups based on the CA output. The degree of similarity is indicated by the thickness of the edges connected to each constituent. The thicker the lines, the more similar the elements are. Figure 6 illustrates that Tennyson's and Browning's poems are catego-

rized distinctly. Some works of Tennyson and Browning, as shown in the CA and PCA results, are confounded with the other poet's works; however, most of Browning's works are colored light blue, purple, brownish gray, and emerald, while Tennyson's are light green, reddish pink, beige, and orange. Tennysonian T28_armdon and T29_TIMTOO, colored purple, are connected with thick edges, one after another, but other lines from T28_armdon and T29_TIMTOO to Browning elements are narrow, suggesting that even though T28_armdon and T29_TIMTOO are categorized within the same group as Browning's, the two poems share more similarity than the other Browning poems. Tennyson's T42_MORHUR ('Morte d'Arthur, 1842) and T69_PAS-IK ('The Passing of the Arthur,' 1869) are edged thickly in reddish pink as both are the stories of the death of King Arthur, and thus, much of their words are distributed in common. Apart from T42_MORHUR and T69_PAS-IK, eight of the thirteen poems from the *Idylls of the King* series are classified in light green. The rest of the *Idylls of the King* series poems, T57_MRG-IK and T69_COM-IK, are in the same beige color group; T69_HOL-IK belongs to the category of orange alone. Although further detailed investigation is needed, the results suggest that the tabular data of vowels can distinguish not only author divergence but also the characteristics of each poem. A more in-depth investigation is encouraged so that the features found in a specific work or in the poetry of a specific poet can help reveal phonetic symbols or the effects that the authors rendered on specific sounds.

3.2. Analysis of consonants in the top 50 works

CA and PCA revealed intriguing results for consonants in Tennyson and Browning. Figures 7 and 8 show the CA and PCA results, respectively. In the heatmap plot (Figure 7) from CA, the top of Figure 7's dendrogram is roughly divided into two groups with the black line in the figure, with the two longest branches being the tallest. Most of Tennyson's works are gathered in a group on the right side of the picture, whereas Browning's works are clustered on the left. On the CA of the consonants, none of the poet's works merged into the other one's cluster; that is, only Tennyson's poems were classified into the groups at the right, and only Browning's works were grouped into the cluster at the left. In the PCA result in Figure 8, the contribution ratios of PCs 1 and 2 are 18.8% and 7.0%, respectively, indicating that the plot can account for 25.8% of the data. While Browning's variables are likelier to be found in the second and third quadrants, Tennyson's variables tend to be dispersed in the first and fourth quadrants. The consonants in question are roughly separated into three groups with the red, blue, and orange rectangles numbered one, two, and three, respectively, in Figure 7. According to the heatmap representation in Figure 7, the consonants in the red rectangle are the significant elements of Browning, and the consonants in the blue rectangle are the remarkable ones of Tennyson. The orange rectangle elements commonly appear in both poets' poems or only occur spottily in a few poems.

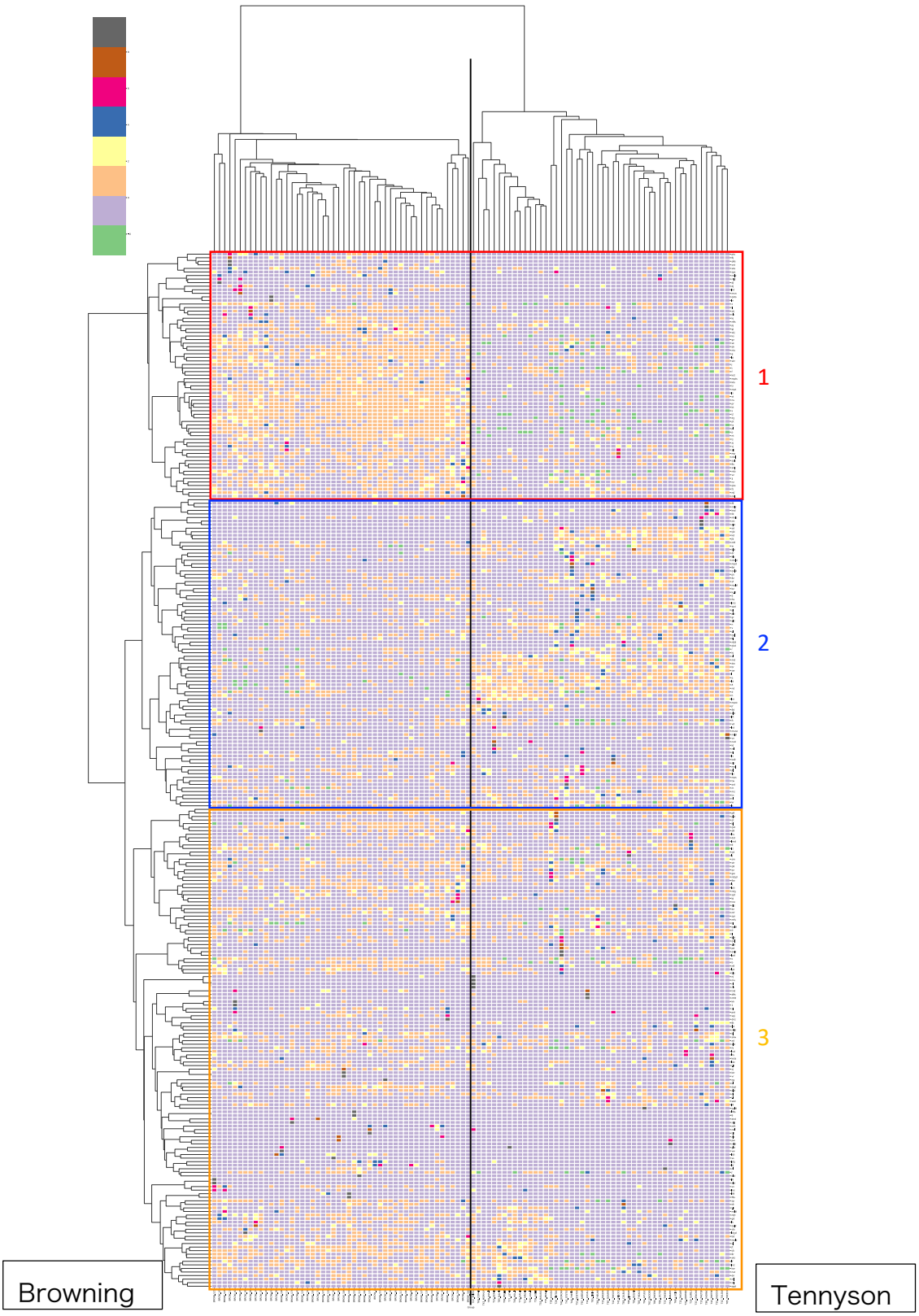

Figure 7: Heatmap representation of the CA result on consonants

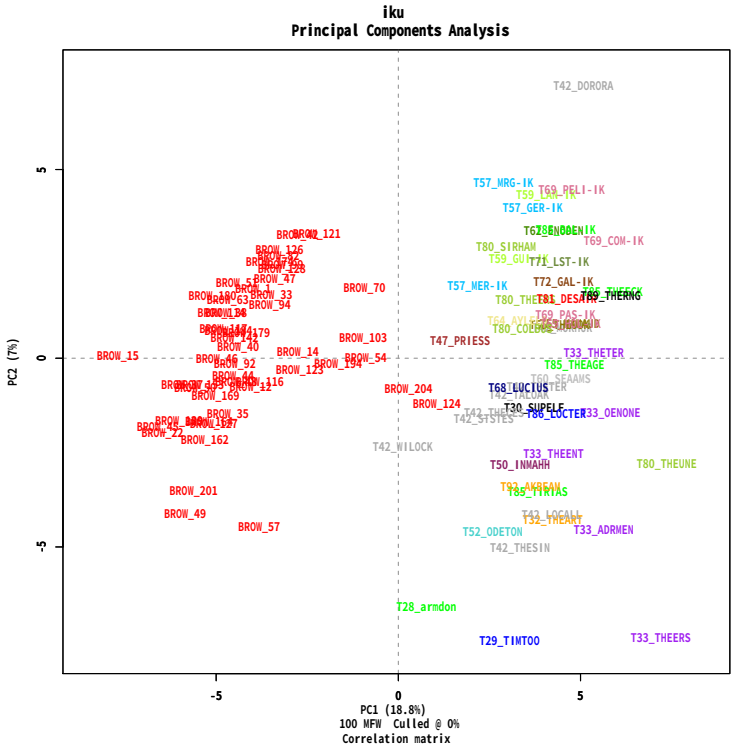

Figure 8: Results of PCA on consonants

Although Figure 7 indicates that CA detected differences in the consonants of the two poets, the dispersion of relative frequency values and the number of elements prevented the plot's readability. Moreover, the relative frequency can unnecessarily accentuate the value or mislead the differences hidden behind the calculated relative frequency value when the raw frequency of words is very low. For example, the word X occurs once in both Tennyson and Browning. When the total tokens of Tennyson's poem are 500, the relative frequency (per thousand words) is two. However, if the total tokens of Browning's poem were 1,000, the relative frequency of the word X would be one. Thus, even though the word X appears once in the two poets' works, the calculated relative frequency does not fully reflect the fact shown by the raw frequency. Notably, low-frequency words, such as those that appear only once or twice in the entire text, might not rise proportionately with the length of the target text, meaning that normalized values might not adequately reflect their usage. The crucial question for words that appear only once or a few times in an author's entire text data is identifying where, how, and what words they collide with. Hence, even though the frequency indicates the same value of 1, such words will be weighted differently depending on why they appear infrequently and inconsistently. To avoid incongruency, this study further applied CA to the top 24 raw frequency consonants, i.e., /t/ [108,233], /n/ [108,209], and /d/ [81,420] (values in brackets are the sum of the raw frequencies). As the raw frequency of 25th consonant, /dʒ/, is 8,890, and the digits get down from five to four, the elements whose raw frequency is lower than

10,000 were cut off. Figure 9 shows the heatmap representation of the top 24 consonants in the top 50 poems of Tennyson and Browning.

Figure 9: Heatmap representation of the CA result on the top 24 consonants

The top of Figure 9's dendrogram is separated into two groups by the black line in the figure. On the left side of the plot, Browning is primarily grouped beside Tennyson's poem, T42_WILOCK, whereas Tennyson is primarily grouped on the left, except for one of Browning's poems (BROW_54). The 24 consonants are generally divided into two groups by the dendrogram on the left side of the figure, where the horizontal black line is depicted. The consonants above the horizontal line appear more frequently in Browning's works because the color of cells at the left hand of the vertical line indicates higher values than Tennyson's works at the right hand of the vertical line. However, the consonants below the horizontal line appear more frequently in Tennyson because the color of the cells at the right hand of the vertical line indicates higher values than Browning at the left hand of the vertical line. The consonants that characterize each poet are listed in Table 4.

Table 4: Consonants classified by CA for each poet

Authors	Consonants
Browning	/t/, /s/, /f/, /k/, /b/, /p/, /st/, /g/, /j/, /tʃ/
Tennyson	/r/, /ʃ/, /z/, /θ/, /m/, /d/, /h/, /w/, /ð/, /nd/, /l/, /ŋ/, /n/, /v/

The tendency of the characteristic consonants of each poet listed in Figure 10 is more noticeable when the consonants are explained in tabular form in Figure 10. In Figure 10, the remarkable consonant symbols of Browning are in blue, while the prominent consonant symbols of Tennyson are in red. Figure 10 is depicted by the author of this paper by referencing Table 5 in Kubozono (1998: 43). The author of this work also added sonority to the right side of the figure by referencing Kubozono (1998: 51). Sonority is an indicator that measures how much the sound can be heard at a given distance; thus, for example, a weak sonority sound, the labial stop /p/, cannot be heard

while the strong sonority sound, the velar approximant /w/, can be heard at the same distance. This is not indicated in Figure 10, but the sounds with the strongest sonority are vowels.

Manner of articulation	Voiceless/ Voiced	fortis/lenis	Point of articulation							Sonority
			Labial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal	
Stop (Plosive)	Voiceless	fortis	p		t			k		Weak
	Voiced	lenis	b		d			g		
Affricate	Voiceless	fortis				tʃ				
	Voiced	lenis				dʒ				
Fricative	Voiceless	fortis	f	θ	s	ʃ			h	
	Voiced	lenis	v	ð	z	ʒ				
Nasal	Voiced	lenis	m		n			ŋ		
Lateral	Voiced	lenis			l					
Approximant	Voiced	lenis				r	j	w		Strong

Figure 10: Consonant map

The characteristics of Tennyson tend to spread below voiced fricatives, where the sonority of sounds is stronger than the stops, affricates, and voiceless fricatives in Figure 10. In contrast to Tennyson, Browning's features tend to be located on weak sonority consonants (i.e., stops, affricates, and voiceless fricatives). The harshness and musicalness in Browning and Tennyson, mentioned on vowels in Section 3.1, can be further explained with sonority here about consonants. Furthermore, Plamondon (2005) compared Browning and Tennyson and hypothesized that Browning should have more 'harsh' consonants, while Tennyson should have 'musical' vowels. Plamondon (2005: 165) further thought that Browning's significant 'harsh' consonants should have been plosives (stops). Plamondon also highlighted that the proportion of plosive consonants in Browning was only 0.6% higher than in Tennyson. Hence, the difference was 'probably too small to justify citing them as responsible for the relative musicality of the poems' (166). However, the CA results indicated that five out of six plosive consonants in English are significant in Browning. The shortage of proportion differences in the Plamondon could be caused by the small number of poems he used in his analyses. The new aspect suggested by CA in Figure 9 and Figure 10 supports what Plamondon has tried to reveal: Tennyson prefers to use 'less-harsh' consonants compared to Browning. Moreover, besides plosives, the postalveolar affricate consonant /tʃ/, labial fricative /f/, and alveolar fricative /s/ are notable consonants of Browning. The three consonants are voiceless, while the voiced sounds of the two—labial and alveolar fricative—among the three are remarkable in Tennyson. Although Plamondon has not acquired sufficient differences to support his hypotheses, the results of this study on both vowels and consonants support his hypotheses.

Before concluding this paper, possible compelling findings that should be intensified in future research are highlighted. `Gephi` was used to create Figure 11, utilizing the numerical output of the CA generated by stylo. Similar to Figure 6, the color differences rely on modularity, which represents how well the data's parts are united. Close-knit poetry groupings based on CA output are represented by components of the same color. The thickness of the edges connecting each item revealed the degree of similarity. More similarities between the components are thought to

exist when the lines are thicker. Figure 11 shows how the poetry of Tennyson and Browning is categorized in different ways. Future attention should focus on the nodes and edges, colored in orange and light blue in Figure 10. The large amount of work in orange and light blue corresponds to the variables in the first quadrant of Figure 8, including T59_LAN-IK, T57_MRG-IK, T57_GER-IK, T55_MAUAUD, and T47_PRIESS. On the one hand, the association of the works can be translated as the *Idylls of the King* series and some narrative poems by Tennyson chiefly converged in the classification when only the light blue network is considered. On the other hand, to dig further, the poems in the two orange and light blue groups will likely be characterized by their female characters. Although female characters are not necessarily the main, leading characters of the poems, they play an important role in the stories. In addition, some other works, such as T55_MAUAUD ('Maud,' 1855), T47_PRIESS (*The Princess*, 1847), and T42_DORORA ('Dora,' 1842), put female characters in key positions as protagonists. Thus, the CA and PCA results suggest that there is perhaps a feature of consonant usage differences in these poems with female characters. Although thorough research is necessary, this study findings suggest that sound features found in certain literary works or poets exhibit a degree of consistency with the poetic characters portrayed in the poems.

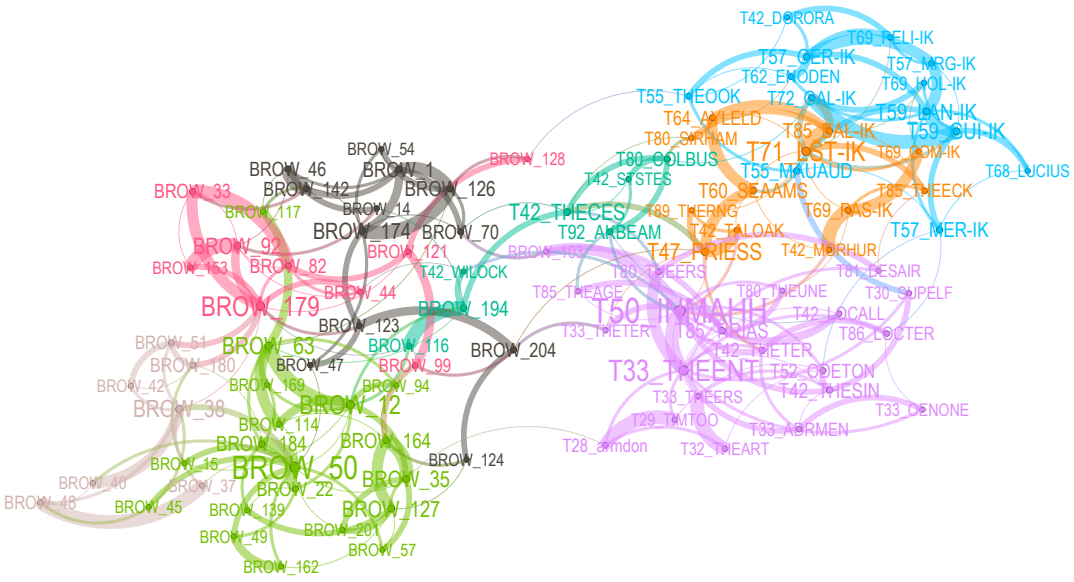

Figure 11: Network representation of consonants from the CA results

4. Conclusion

This study primarily aimed to clarify whether tabulated phoneme frequencies could show elements of Tennyson's and Browning's poetry and to highlight which phonemes are the characteristics of each poet in case the tabulated data detect differences between the two poets. Considering the limitations of prior studies, the top 50 phoneme frequency poems from Tennyson and Browning

were analyzed to thoroughly investigate their respective poetry collections. As determined by CA and PCA, the tabulated data of the IPA were efficient in finding the differences between the two authors and capturing the characteristics in their poetry. Although further in-depth research is necessary, this study findings corroborated the previous study's hypotheses while revealing a new prospect that the sex of characters in poems could coincide with the sound use of poets.

Acknowledgements

This work was supported by the Japan Science and Technology Agency Support for Pioneering Research Initiated by Next Generation [grant number JPMJSP2138].

Bibliography

[1] Bastian, M., M. Heymann and M. Jacomy. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International ICWSM Conference*, San Jose, California, USA, 3/1: 361–362.

[2] Burrows, J. (1987) *Computation into Criticism: A Study of Jane Austen's Novels*. Oxford: Clarendon Press.

[3] Delphi Poets Series (2013) *Alfred, Lord Tennyson*. East Sussex: Delphi.

[4] Delphi Poets Series (2012) *Robert Browning*. East Sussex: Delphi.

[5] Tawfiq, H. H. (2020) A Study of the Phonological Poetic Devices of Selected Poems of Robert Browning and Alfred Tennyson. *English Language and Literature Studies*. 10/4: 16–33.

[6] Eder, M., J. Rybick and M. Kestemount. (2016) Stylometry withR: A Package for Computational Text Analysis. *The R Journal*. 8/1: 107–121.

[7] Hoover, L. D. (2012) The Tutor's Story: A Case Study of Mixed Authorship. *English Studies*. 93/3: 324–339.

[8] Imao, Y. CasualConc (ver. 3.0.4). Accessed in December 30, 2022. https://sites.google.com/site/casualconc/

[9] Kubozono, H. (1998) *Onsei-gaku, On-in-ron. (Phonetics and Phonology)*. Tokyo: Kuroshio.

[10] Plamondon, R. M. (2005) Computer-Assisted Phonetic Analysis of English Poetry: A Preliminary Case Study of Browning and Tennyson. *TEXT Technology*. 14: 153–175.

[11] Ricks, C. (Ed.) (1987) *The Poems of Tennyson*. vol. I–III. (Second edition) London: Longman.

[12] Tabata, T. (1998) Corpus-ni Motoduku Buntairon Kenkyu (Corpus-based Stylistic Studies). In *English Corpus Linguistics*. Tokyo: Kenkyu-sha.

[13] Tabata, T. (2004) *–ly* Fukushi-no Seikihinndo Kaiseki-niyoru Bunntai Shikibetu: Koresupondensu Bunnseki-to Syuseibunn Bunnseki-niyoru Hikaku Kennkyuu (Style Identification of*–ly* Adverbs by the Occurrence Frequencies: Comparing the Correspondence Analysis and Principal Component Analysis. *Denshika Gengo Shiryou Bunseki Kenkyu 2003*. 97–114.

[14] Tabata, T. (2012) Textometry of Co-authorship: Dickens, Collins, and their collaborations. *Denshika Gengo Shiryou Bunseki Kenkyu 2011–2012*. 3–17.

Appendices

Appendix 1: Corresponding table of abbreviated file names and poem titles of Browning

Abbreviated file names	Poetry titles
BROW_1	'Tertium Quid'
BROW_103	'Saul'
BROW_114	'Juris Doctor Johannes-Baptista Bottinius Fisci et Rev Cam Apostol Advocatus'
BROW_116	'Easter-Day'
BROW_117	'The Book and the Ring'
BROW_12	'SORDELLO BOOK THE FIRST'
BROW_121	'Pompilia'
BROW_123	'Ivàn Ivànovitch'
BROW_124	'PAULINE'
BROW_126	'The Other Half-Rome'
BROW_127	'La Saisiaz'
BROW_128	'Caliban upon Setebos'
BROW_139	'WITH FRANCIS FURINI'
BROW_14	'Cleon'
BROW_142	'The Ring and the Book'
BROW_15	'WITH GEORGE BUBB DODINGTON'
BROW_153	'FIFINE AT THE FAIR'
BROW_169	'Pacchiarotto, and How He Worked in Distemper Pacchiarotto'
BROW_174	'Half-Rome'
BROW_179	'Guido'
BROW_180	'Mr Sludge, 'The Medium''
BROW_184	'RED COTTON NIGHT-CAP COUNTRY, OR, TURF AND TOWERS'
BROW_194	'Christmas-Eve'
BROW_201	'WITH CHARLES AVISON'
BROW_204	'James Lee's Wife I'
BROW_22	'FUST AND HIS FRIENDS'
BROW_33	'Count Guido Franceschini'
BROW_35	'Jochanan Hakkadosh'
BROW_37	'Pietro of Abano'
BROW_38	'THE INN ABLUM'
BROW_40	'Clive'
BROW_42	'In a Balcony FIRST PART'
BROW_44	'Fra Lippo Lippi'
BROW_45	'WITH DANIEL BARTOLI'
BROW_46	'The Pope'
BROW_47	'Ned Bratts'
BROW_50	'Aristophanes' Apology'
BROW_51	'Bishop Blougram's Apology'
BROW_54	'The Flight of the Duchess I'
BROW_57	'WITH GERARD DE LAIRESSE'
BROW_63	'Dominus Hyacinthus de Archangelis Pauperum Procurator'
BROW_70	'BALAUSTION'S ADVENTURE'
BROW_82	'Giuseppe Caponsacchi'
BROW_92	'PRINCE HOHENSTIEL-SCHWANGAU, SAVIOUR OF SOCIETY'
BROW_94	'An Epistle'
BROW_99	'A Death in the Desert'

Appendix 2: Corresponding table of abbreviated file names and poem titles of Tennyson

Abbreviated file names	Poetry titles
T28_armdon	'Armageddon'
T29_TIMTOO	'Timbuctoo'
T30_SUPELF	'Supposed Confessions Of A Second-rate Sensitive Mind Not In Unity With Itself'
T32_THEART	'The Palace Of Art'
T33_ADRMEN	'A Dream Of Fair Women'
T33_OENONE	'Oenone'
T33_THEENT	'The Lovers Tale A Fragment'
T33_THEERS	'The Lotos Eaters'
T33_THETER	'The Millers Daughter'
T42_MORHUR	'Morte Darthur'
T42_STSTES	'St Simon Stylites'
T42_TALOAK	'The Talking Oak'
T42_THECES	'The Two Voices'
T42_THESIN	'The Vision Of Sin'
T42_THETER	'The Gardeners Daughter'
T42_WILOCK	'Will Waterproofs Lyrical Monologue Made At The Cock'
T47_PRIESS	*The Princess*
T50_INMAHH	*In Memoriam A.H.H*
T52_ODETON	'Ode On The Death Of The Duke Of Wellington'
T55_MAUAUD	'Maud'
T55_THEOOK	'The Brook'
T57_GER-IK	'Geraint And Enid'
T57_MER-IK	'Merlin And Vivien'
T57_MRG-IK	'The Marriage Of Geraint'
T59_GUI-IK	'Guinevere'
T59_LAN-IK	'Lancelot And Elaine'
T60_SEAAMS	'Sea Dreams'
T62_ENODEN	*Enoch Arden*
T69_COM-IK	'The Coming Of Arthur'
T69_HOL-IK	'The Holy Grail'
T69_PAS-IK	'The Passing Of Arthur'
T69_PELI-IK	'Pelleas And Ettarre'
T71_LST-IK	'The Last Tournament'
T72_GAL-IK	'The Round Table Gareth And Lynette'
T80_COLBUS	'Columbus'
T80_SIRHAM	'Sir John Oldcastle Lord Cobham'
T80_THEERS	'The Sisters'
T80_THEUNE	'The Voyage Of Maeldune'
T81_DESAIR	'Despair'
T85_BAL-IK	'Balin And Balan'
T85_THEAGE	'The Ancient Sage'
T85_THEECK	'The Wreck'
T85_TIRIAS	'Tiresias'
T86_LOCTER	'Locksley Hall Sixty Years After'
T89_THERNG	'The Ring'
T92_AKBEAM	'Akbars Dream'