مجلة المختار للعلوم الإنسانية
Al-Mukhtar Journal of Social Sciences
41 (1): 113-143, 2023
pISSN: 2791-1608; eISSN: 2791-1616
الصفحة الرئيسية للمجلة: https://omu.edu.ly/journals/index.php/mjssc/index

تصنيف الجودة الشعرية في شعر شعراء الفترة الرومانسية الإنكليزية "تحليل أسلوبي رقمي"

رفعت الجميلي[1]*
باحث مستقل

DOI: https://doi.org/10.54172/mjssc.v41i1.1305

المستخلص: تستهدف الدراسة الحالية تصنيف الجودة الشعرية لستة شعراء أوربيين (إنكليز) في فترة الازدهار الشعري الرومانسي أواخر القرن الثامن عشر، ومطلع التاسع عشر، وهم: وليام بلايك، وليم وردزوورث، صامويل تايلر كولوريدج، شيلي، جون كيتس، و بايرون. وترى هذه الدراسة أن تحليل الجودة للنصوص الشعرية وتقييمها لا تتحقق فقط عبر أدوات الناقد الأدبي، بل على العكس من ذلك، يمكن التجديد والانتقال إلى حوسبة التحليل اللغوي للنصوص الشعرية، ورقمنة الجودة الشعرية، وتحويل معاييرها الكمية والنوعية إلى معطيات إحصائية رقمية مرئية، باستخدام التحليل الإحصائي والرقمي؛ بهدف تعزيز النقد الأدبي، وتوسيع حدود أبحاثه التقليدية، وخلق أشكال جديدة من المعرفة؛ تماشيًا مع التغيرات في التعامل مع النصوص الأدبية التي باتت في مجملها رقمية. وتكونت عينة الدراسة من قصائد بلغت (108)، اختيروا بعناية فيما يخص عدد السطور الشعرية، بغض النظر عن أجناسها أو مواضيعها. وقد استخدم الباحث تحليل الانحدار اللوجستي الخطي الثنائي، عن طريق نمذجة الجودة الشعرية بوصفها متغيرًا ثنائيَّ الحد، بدلالة عشرة متغيرات متوقعة، وهي مجموعة من المهيمنات اللغوية والصوتية التي قام الباحث بتعدينها واستخلاصها من القصائد المختارة للشعراء المشار إليهم أعلاه؛ بهدف بناء نموذج إحصائي؛ لتصنيف وتبوء احتمالية الجودة الشعرية لكل واحد منهم. وقد أظهرت نتائج الدراسة تباينًا بسيطًا بين الشعراء فيما يتعلق بالجودة الشعرية، إلا أن الشاعر وليام بلايك حصل على جودة شعرية عالية إجمالا في عناصر القافية، وكثافة الأصوات الانفجارية، وكثافة الأصوات الاحتكاكية مقارنة بالشعراء الآخرين.

الكلمات المفتاحية: الجودة الشعرية، الشعر الرومانسي، مهيمنات لغوية، الانحدار اللوجستي الخطي الثنائي، تصنيف، منحني الانحدار الخطي.

Predicting and Classifying the Poetic Quality of the Early and the Late Generation of Romantics: A Computational Stylistic Study

Refat Aljumily
Independent Researcher

Abstract: Poetic quality is an important value for a poet's literary craft, and it is considered as a cultural heritage of a society. Poetry is made with language and poets use language features in a particular way to create effects and to convey meaning. Measuring poetic quality would appear to be an intrinsically qualitative endeavor. However, it is possible to objectify some of the language features to some degree using a quantitative criterion, which is here proposed. The study represents a preliminary literary computational approach in relation to Romantic poetry which is dominated by a few names: William Blake, William Wordsworth, Samuel Taylor Coleridge, Percy Bysshe Shelley, John Keats, and Lord Byron. There has been little or no research with literary computing techniques focused on measuring and classifying the quality levels of those poets' works. The study builds on a set of ten measurable stylistic features extracted from one hundred and eight poems regardless of genre or subject. The study carries out an analysis using Binary Logistic Regression Modeling (BLRM) as one of the most commonly used predictive modeling techniques. I used the stylistic features as predictive variables and the quality as the outcome variable indicative of the stylistic characteristics of each of the poems selected. In this way, this study is able to measure the stylistic features in the selected poems and, independently, using Binary Logistic Regression to predict the general poetic quality levels for each of the poets examined. The study is also able to make connections between some of the poets which enabled a more detailed view of the subcomponents' usage and occurrence in the poems.

Keywords: poetic quality, romantic poetry, diction, stylistic features, binary logistic regression, classification, logistic curves

*Corresponding author: E-mail addresses: Refat.A.70@outlook.com

1. Introduction:

Poetry is a superior form of creation that uses language features to create effects and convey ideas and feelings. There are plenty of memorable poems from different eras and from different poets that eloquently appeal to many generations and emotionally capture many audience. There are many others that make a little deal of sense eloquently, emotionally, and pleasantly, whether or not the poets intended them to. So, what constitutes poor, medium, or high quality poetry? Answers on this question vary widely. Answers can be subject to personal taste or how the reader responds to a poem. In reader reaction to poetic texts approach, one can read a poem and dislike the ideas or themes but acknowledge that the poetic language was good or effective. Another one can read the same poem and love the ideas or themes but feel that the poetic language was weak or ineffective. Answers can also be subject to an objective interpretation of the data by using quantitative methods. The latter answer is the most popular in the literary computing approach and thus the scope of interest here. The field of literary computing has developed considerably in the past few decades and has attracted growing interest among digital humanists and computational linguists. Digital humanists and computational linguists are increasingly seeking to develop more reliable and systematic approaches of examining aspects or elements of many different literary texts to gain useful insights or help answer specific literary inquiries like classification of literary texts, authorship attribution, measuring poetic quality of different poets, etc.

Measuring poetic quality is still in its beginnings, but it can now be systematically considered as part of overall literary computing approach emerged with many different applications, including literary stylistic studies, evaluation of frequencies of different phenomena, word or sentence lengths, word associations, polysemy values, measurement of differences, classifications, measurement of grammatical structures, rankings, diversifications, and denotative structures (Popescu et al., 2015; Altmann and Fengxiang, 2008). At the same time, demand for measuring the poetic quality is growing fast or mounting among literary computational critics and linguistic researchers. The increasing interest in quantitative measurement of poetic quality has been accompanied and supported by the growing numbers of electronic literary texts, driven, amongst others, by significant developments in information technology and associated advances in natural language processing tools. Perhaps most of us recognize that without quantitative measurement it is difficult to estimate poetic quality, as it is impossible to identify high or low quality poets without reliable information about the amount and the type of the stylistic features of the same poet

occurring in different poems or even in the different contexts of the same poem. Measuring quality of poetry is important for a range of different disciplines within literary studies, and it builds the basis for numerous poetic quality methodologies to confirm results or yield new results that were not originally estimated. In particular, drawing some conclusions about the overall quality of poetic texts rely heavily on the availability of reliable information about the features that define the poems. Common to all research methodologies is that without a vigorous measurement of poetic quality, it is impossible to determine to a fair degree of accuracy how a poet's use of certain stylistic features help him/her make their language memorable or more evocative and explain how his/her popularity has evolved over time. The remainder of this paper is organized as follows: the theoretical foundations that explain the problem that drives the current research is introduced in Section 2, the stylistic features used in poetry is presented in Section 3, the characteristics of the Romantic period poetry is briefly introduced in Section 4, data used in the study are presented in Section 5, empirical results of the study are presented and interpreted in Section 6 and, finally, I draw conclusions in Section 7.

2. Theoretical foundations and motivation:

Poetic quality is a broad area that covers several stylistic features that poets use to add flavor and texture to poetry. These stylistic features may be phonological (e.g. rhythm, patterns of sounds, rhyme, meter, alliteration, etc.), syntactic (e.g. types of sentence structure, couplets, stanzas, parallelism, etc.), lexical (e.g. nouns, verbs, adjectives, etc.), or rhetorical (e.g. personification, climax, irony, simile, metaphor, allusion, and so on). In the traditional literary analysis domain, these among many others, are the key features when literary critics criticize or evaluate a work of poetry. Depending on the ultimate effect that they hope to create, nearly all poets employ the same or similar stylistic features to magnify the beauty or make a poem more meaningful, but how these features are used can differ greatly from one poet to another. Linguistically, this approach initiates the need to search for the amount of elements of poetry or a description of quality characteristics indicative of poetic quality in the selected poems even when they don't appeal to personal taste. This study focuses on the work of six romantic poets—William Blake, William Wordsworth, Samuel T. Coleridge, Percy B. Shelley, John Keats, and Lord Byron. My choice of those poets for an examination of poetic quality is not surprising, given that the Romantic period in poetry saw associations of the variety of styles and genres in which the Romantic poems were produced, which can be an ideal target for analysis on trends and relations among poems and poets. However, there

is lack of quantitative studies looking at or examining the degree of poetic quality of those poets in that literary period. To date studies have been largely qualitative, with little attention given to quantitative criteria, and are largely inductive proposals (for example Al erjan, 2022; Zia , 2017; Zahida et al, 2016; Mehreen, 2015; Frida, 2012). In the present study, I suggest a quantitative criterion that facilitates an objective interpretation of the data. The method of processing the data I used for this study is the tokenization, which breaks the 108 raw text poems into a sequence of sounds, a sequence of words, and a sequence of sentences called tokens. The method of quantifying the tokens (sound tokens, word tokens, and so on) I used is the text analysis tool, and in particular, one of its functions, provides statistics on word frequency, sound frequency, and character count. It also provides information about the most frequent words and frequencies of sounds present in the selected texts of the six poets. All the information I extracted is modelled into vector space and trained to Binary Logistic Regression method to select those features that are sufficiently significant in order to build a model that predicts patterns of poetic quality for each poet from the selected poems. The aim is to determine how frequent/infrequent these stylistic features in the selected poems are and–perhaps more to the point–how good the poetic quality of those poets who produced these poems are. Underlying my engagement with this research is the assumption that the stylistic features those poets use to showcase creative thought in a poem may constitute what I can identify as good poetry through literary computational analysis. In fact, I took my assumption idea from S. T. Coleridge's maxim: "Poetry is the best words in the best order" and gave it a prominent role in the current study. That makes it possible for me to connect the stylistic features to poetic quality. The literature shows that remarkable attention has been given to the impact of stylistic features in evaluative processes of poetic texts and suggests that the stylistic features that are used by a particular author may be adapted to show a poem's style that describes a quality characteristics of that poem (e.g. Crosbie, 2016; Kao and Jurafsky 2012; Kao, 2011; Peer, 2008; Kaplan, 2006).

Having considered the impact of stylistic features on the quality characteristics of poetry, a research statement may now need to be considered: poetic quality can be predicted by a set of stylistic features. So, in line with this statement, good poetry can be defined as poetry that contains varied poetic diction and sound devices with a flexible attitude to elements of grammar that demonstrate the poet's mastery of language. These are some elements of poetry writing which can be measured to assess a work of poetry for that matter. Thus, according to this definition, poetic

quality can be assessed through measuring how often those six romantic poets use specific stylistic features; that is, what words, phrases, grammatical structures, and the sounds of the letters they consciously and perhaps sub-consciously choose and combine which can help to predict the degree of poetical quality in the selected poems. These features are not guaranteed to create a good quality poem but they are measurable and their values will certainly give an objective indication that aligns with the research's purpose. To that end, I created a Romantic poetry corpus that includes a selection of one hundred and eight poems from the six Romantic poets and I measured and compared the stylistic feature values related to the main categories that recurred in these poems, then I identified the significant stylistic features that define these poems to predict their quality using Binary Logistic Regression analysis.

3. Feature indicators for poetic quality:

In this study, measuring poetic quality refers to the process of analyzing poetic text corpus and investigating the use of measurable style features in it in order to classify and predict creative writing quality. The language of poetry has different stylistic features such as rhyme, sound and rhythm, meter, syntax, figurative language, imagery, diction, etc., and poets use these features to create images and effects and to convey meaning. Through this study, the task of measuring poetic quality is based on the assumption that stylistic features have impact on the poetic quality and that good quality poetry must have different stylistic features that demonstrate the poet's creative use of language. So, a poet's use of certain features (e.g. certain words, sound patterns, etc.) might thus play an important role in the prediction of that poet's poetic quality. Over the years, many different stylistic features have been proposed (e.g. Bekmirzaev and Kim, 2017; Crosbie, 2016; Peer, 2008; Miles 1946 & 1957 & 1967; Biber, 1988; Kurland, 2000; Flesch, 1948 & 1949). The majority of these features are based on word forms and their frequency of occurrence and patterns of repetition in sound. And these are the ones that this study will focus on because these features indicate the importance of words or sound patterns in the selected poems by measuring how often certain features appear which are therefore objective (measurable) indicators for poetic quality. At lexical and grammatical levels, previous studies (e.g. Kao and Jurafsky, 2012; Kao, 2011; Breland, 1996; Graves et al., 1987) demonstrate that word frequency is highly correlated with word difficulty and, therefore, it has been used to estimate word difficulty (e.g. Tomayo, 1987; Marks & Carolyn et al., 1974). Specifically, low-frequency words are found to have clear impact on superior-level written language skills. This is true for tasks such as automated essay scoring system

used to assign grades to (well-performing) essays and other written works in educational setting (Ben–Simon & Bennett, 2007; Burstein et al., 2004). Because low frequency words are related to writing skills and assessing essays and other written works, it is not surprising that good quality poems may have a lower average word frequency. Previous studies also demonstrate that word frequency is associated with word specificity when only a large corpus is available (Jones, 1972; Joho and Sanderson, 2002; Caraballo and Charniak, 1999). This points to the connotations of a word and its associations (sensations, attitudes, and emotions) that it evokes, which indicates that low-frequency words tend to have particular connotations of specific words to convey specific, particular or detailed idea or image that a poem talks about (Kao and Jurafsky, 2012 and Kao, 2011). I would, therefore, expect a set of lower-frequency words to be found only in specific contexts in high quality poems. Word frequency has also been used to estimate word variation as an additional feature of vocabulary difficulty. One basic measure for doing so is to define the type/token ratio. The type/token ratio is widely used in literary studies (e.g. Holmes, 1992) and in studies that monitor developments and changes in the use of lexical words in children and adults (e.g. Williamson, 2014; Biber et al., 2002). A high type/token ratio indicates a large amount of lexical variation and a low type/token ratio indicates a relatively little lexical variation. This type/token information may have implications for poetic quality analysis and, therefore, I would expect that good quality poems might have larger and more varied word types when a poet tries to avoid using the same word several times throughout a poem. Further, word frequency has been used to estimate the lexical density of written texts by making a distinction between different lexical words (parts of speech) and grammatical words (function words). A high lexical density indicates a large amount of information-carrying words and low lexical density indicates relatively few information- carrying words. This feature is used to monitor improvements in the use of lexical items (information carrying-words) in children with under-developed vocabulary and/or word finding difficulties. It can also be used to measure how much information is contained within a particular text (e.g. Williamson, 2014; Biber et al., 2002; Hewings et al. 2005). Across this area of research, I would think more broadly about the possibility that good quality poems may have a much denser pattern of words than low quality poems. However, word frequency is not the only feature present in the previous literature. In some cases, sentence length may also play a role. This type of feature is one that may have an impact on the relationship between the poetic line (including its length and positioning and how it fits into other lines) and the content of a poem. In addition to

this, information about sentence length may give insight into how a writer's poem writing considered the effects of unusually long or short sentences. The other feature one can look at have some interesting information about the length of words in each poem. This feature has long been considered as a discriminator of different genres and registers (Aljumily, 2015). It seems more under an author's control in which using lengthy or complex words and using too many of them may add difficulty to the overall grammatical quality of a given text. Previous research on predicting text quality shows that text quality is also related to readability (Louis, 2013). Readability refers to how easy it is to read and understand a given text based on its stylistic features. It is often used in assessing the suitability of a text for an audience. As a result, readability might be an interesting feature that plays a role in how easy someone will find a given poem to read and understand, which may then indicate how well poets can write and communicate meaning and message. There are different readability metrics used for estimating readability, such as the Flesch-Kincaid and Gunning Fog index (Flesch 1948 & 1949). Many work by counting words, sentences and syllables while others use lists of already scored words. A high readability score means a particular piece of text is easy to read, increasing the chances that someone will read it from the beginning to the end and understand it. A low readability score means a particular piece of text is harder to read and is likely considered more complex to understand. At the phonological level, my literature search shows that the use of word sounds, such as alliterations, assonance, rhymes, and other sound devices, to measure the aesthetic value of poetic texts has been widespread since the early days in the field. Admittedly, Birkhoff (1933) produced a method, he called "aesthetic measure" that based on phonemic features, seeking out the density of the elements of order in the aesthetic object that could define and assess a poem's beauty by calculating the ratio of an object order to its complexity. However, this does not amount to an agreed methodology in the domain of language, since the method doesn't capture the subtlety of word choice or richness of meaning in poetry. In light of these observations, it has been deemed appropriate to consider poetic diction along with the aftermentioned features. However, since the suggestion of the phonemic features in 1933, there have been several methodological attempts in the context of developing programs to quantity phonemic features such as rhyme, meter, and word sounds (e.g Genzel et al., 2010; Green et al., 2013; Kaplan, 2006; Hayward, 1996; Logan, 1988). On the basis of this development, researchers have conducted many studies to examine the writing style of poems and to assess their poetic beauty. Many of these studies have focused on alliteration and

consonance, perfect and slant end rhyme, and assonance (e.g. Kao and Jurafsky, 2012; Kao, 2011; Lea et al., 2008; Burroway, 2007; McGlone and Tofighbakhsh, 2000; Rubin, 1995). Other studies have examined rhythmic patterns in poetry (e.g. Agirrezabal et al., 2016) and rhyme categories and patterns in poetic verse (e.g Kavanagh, 2007). In summary, studies have not conclusively determined or identified which types of stylistic features will be most important for making a poem beautiful. It is likely that a large number of stylistic features and aspects are involved in a given poem, each of which makes only a small contribution to a writer's poem. In linguistic terms, the use of language and its various stylistic features with which the poet has communicated the meaning and created the effects is the criterion by which poetic quality is judged.

4. Romantic poetry and famous romantic poets:

Romantic poetry refers to the verses composed during the Romantic era in English (1789-1850). It opposed the objectivity of neoclassical poetry which was connected to intellect and reason. The Romantic poets wanted to escape what they saw as the world's horrible realities while at the same time appealing to nature and naturalism. Thus Romantic poetry tries to stress the importance of both strong human emotions and the individual consciousness: enhancing the instincts of self-gratification and search for pleasure and sensual delights. The writing style of the Romantic poets was varied but the stress was put on simplicity by using a spontaneous and natural poetic diction. In Romantic poetry, there is a tendency to use supernatural elements to give the atmosphere or a sense of wonder and mystery. There is also a growing nostalgia for the past or simple rural life, for a town life in which people lived and worked together with nature, and a new interest in the attitudes and experiences of the lives of the common people. The appreciation for natural man uncorrupted by commercialism and the evils of civilization was accompanied by an interest in innocence and childhood, an interest that is most evident in the themes of many poems. Imagination is a dominant feature in Romantic poetry. It acts as a source of creativity and to see images of the objective realities of life (Ferber, 2010).

The best known Romantic poets in the history of English poetry can be divided into two widely recognized generations. The first generation Romantic poets include William Blake, William Wordsworth, and Samuel Taylor Coleridge. The second generation Romantic poets include John Keats, Percy Bysshe Shelley and Lord Byron. This section gives an initial overview of those two generations of the Romantics:

4.1 William Blake (1757-1827):

Blake is one of the earliest Romantic period poets. His collections of poetry entitled Songs of Innocence (1789) (19 poems) and Songs of Experience (1794) (26 poems), are two of his most famous poems that clearly utilize religion, poverty, racism, child abuse and labour, and the harsh corruption nature of adulthood as a theme. Unlike his fellow poets, Wordsworth and Coleridge, Blake's poetry is quite different from them, both in terms of aesthetic style and poetic modes and in his preoccupation. Blake is not a nature poet in the same way that his fellow Romantics are: he seldom demonstrates a great appreciation or reverence for the world of natural feeling and familiar situations and countryside as his fundamental theme, but he celebrates its physical beauty that links all living things. For Blake nature is closely connected to the material world and it is useful to represent real existence and being of man. This means that Blake uses nature to frame his poetry to create symbolic objects or characters, for example, the tiger, the lamb, or the rose and the worm to convey his main ideas, emotions, and themes about different things such as sin, religion, shame, cruelty, evil, etc. In poetic form and language, Blake's poetry tends to be simple. He organizes his thoughts and ideas into short lines or four-line stanza (usually tetrameter, i.e. containing four feet). But his poetic imagery (similes, metaphor, personification) is often diverse and filled with complicated images which requires deeper analysis to understand their hidden or solve their multiple meanings. Whatever the theme or genre of Blake's poetry one can easily acknowledge the rich language and lyrical quality of his poetic vision (Marsh, 2012).

4.2 William Wordsworth (1770-1850):

Wordsworth is widely regarded as the central figure in the Romantic poetry. During his lifetime, Wordsworth wrote an estimated 387 poems, including private and unpublished poems. He is best known for *Lyrical Ballads* (1798 and 1800), a collection of poetry, he co-authored with Coleridge. He is also widely known for his autobiographical poem *The Prelude* (1798) that describes Wordsworth childhood's memory. Wordsworth's poetry draws upon a sense of love and appreciation for the permanent forms of nature and through this he emphasizes the relationship between human beings and the natural world. Wordsworth's writing style is marked by the use of vocabulary and speech patterns of common men because they, as he argued, are full of emotions and feelings. For Wordsworth, poetry which should be written in "the real language of men in any situation", is nevertheless "the spontaneous overflow of feelings: it takes its origin from emotion recollected in tranquility" (*Wordsworth theory of poetry*, no date). Wordsworth almost always uses

blank verse and he also shows ability in some verse forms such as sonnets, odes, ballads, and lyrics, with themes include nature, humanity, mortality, religion, memory, and morality. In addition to this, he creates some of the finest poetry during his time, on a variety of themes and in several styles and poetic visions. This has led many critics and poets to pay a clear attention to the distinction between poetic and non-poetic diction and to the poet's role in society to communicate knowledge to common people and to understand their feelings and get better their morality (Mahoney, 2001).

4.3 Samuel Taylor Coleridge (1772-1834):

Coleridge—a poet, translator, literary critic and philosopher—is a leading figure of the Romantic Era. He is the poet of the four most famous poems in the English language *The Rime of the Ancient Mariner* (1797), *Kubla Khan* (1797), *Christabel* (1797-1800), and *Frost at Midnight* (1798), among many other poems and fragments, as well as *Lyrical Ballads*, a collaboration undertaken with Wordsworth. An important feature in Coleridge's poetry is his obvious simplicity, imagination, suggestiveness, and symbolism. Coleridge is well noted for his poetry of the unusual romantic themes and exotic images, and he often relates nature to the mind. Throughout Coleridge's poetry, all the themes of Romanticism can easily be found. Besides this, he also focuses on the themes of mystery and the supernatural. Coleridge challenges the theory that the language of poetry should be abstract and general. He makes a clear distinction between prose and poetry. For Coleridge, prose is "words in their best order" and poetry is "the best words in their best order" (Coleridge, et al., 1884). The most significant feature of Coleridge's word choice in the poems is his use of archaic words and his ability of creating new words. He tends to use language that is more symbolic to attain self-knowledge and moral values. He is also fond of using old fashioned spellings and old verbal endings. Aspects connected to lexis are found in the deliberate use of the same words or phrases multiple times. Other aspects can be found at other levels, such as grammar and sound. In addition to poetry Coleridge writes four plays, *The Fall of Robespierre* (1794), written with Southey, *Osorio* (1797), *Zapolya* (1815), and *Remorse* (1797), and translated the plays *Piccolomini* and *The Death of Wallenstein* (1800) from the German of Schiller into English. Coleridge also writes *Biographia Literaria* (1817), an eclectic prose work combining intellectual autobiography, philosophy, and literary theory, which is widely regarded as a source of literary criticism and analysis (Mays, 2013).

4.4 Percy Bysshe Shelley (1792-1822):

Shelley, who died at the age of 29, is regarded as one of the major English Romantic Period Poets. In his short life, Shelley wrote a quantity of poems, including several long poems and short prose works, which relate to his radical political, religious, and ethical views. Among his best-known poems, short and long, are *Ozymandias* (1818), *Ode to the West Wind* (1819), *To a Skylark* (1820), *The Masque of Anarchy* (1819), *Queen Mab* (1813), and *Alastor*, or *The Spirit of Solitude* (1815), *Julian and Maddalo* (1818-1819), and *Adonais* (1821), *Hymn to Intellectual Beauty* (1816), *Mont Blanc* (1816), and other poems. Like many of the Romantic poets, Shelley employed elements of imagination, supernaturalism, melancholy, subjectivity, and idealism in his poetry, but he most often demonstrates a great reverence and sensibility for the beauty of nature, and he feels closely connected to nature's power. More specifically, man and nature cannot be separated from Shelley's continuous reference to the immediate contact between the power of the human mind and the heroic or the visionary role of the poet. For Shelley the power of the human mind equals to the power of nature. Poetry is "something divine, that centres and the circumference of knowledge" and the poet is "unacknowledged legislator of the world" (Mcmaster, 1971: 7,162). Shelley's poetry style could be described as tactile and lush. He selects his words carefully and he never used ornamental words. Every word is placed in a suitable place, and it carries its significance. He uses sometimes extraordinary poetic diction and most often he employs visual imagery and symbolism. Besides poetry, Shelley's other major literary projects include the visionary poetry drama *The Cenci* (1819), *Prometheus Unbound* (1820), and *Hellas* (1822), and the two Romantic Gothic novels *Zastrozzi* (1810) and *St. Irvyne*; or, *The Rosicrucian* (1810). Shelley also wrote several essays on social, political, and philosophical issues (Morton, 2006).

4.5 Lord Byron (1788-1824):

George Gordon Byron, known as Lord Byron is a poet, social, political and religious satirist. One of the leading figures in the Romantic Movement. Byron is best known for his lengthy narrative poems *Child's Harold's Pilgrimage* (1812 and 1818) and *Don Juan* (1819), and also for his poetry collection *Hebrew Melodies* (1815) which includes many of his shorter poems. Byron used a lot of autobiographical elements or put a lot of himself (i.e. his personal expressions of life and adventures) in many of his poetry. Like other romantic poets, Byron looked curiously backward to a distinctly British cultural past of traditions and customs, but for him the poetic diction of the eighteenth century was a natural language to look speculatively forward. This allowed him to

employ a mixed or unique writing style that incorporated satire, neoclassicism, and romanticism. Byron's writing style could be characterized by using different metres; iambs and anapests, different forms; blank verse, hudibrastics and heroic couplets, terzains, quatrains, sixains, rime royal, Spenserians, ottava rima, and Sapphics, and neo-classic personifications and clichés. Elements of fiction and reality are also present in many of the poems of Byron to portray different points of view. Although he covers many romantic subjects and themes, the major themes in most of Byron poems narrow down to nature, the folly of love, realism, freedom, and the power of art. Byron refused to admit the opinion of the poetic imagination as a creative and transforming power, and as the primary faculty of the poet. He declared that "imagination" and "invention" are the two commonest of qualities" which basically means that a poet describes what he/she sees not what he/she imagines. Besides poetry, Byron wrote a number of essays that criticize the social, political, and religious injustice and hypocrisy of his day (Jerome, 2002).

4.6 John Keats (1795 – 1821):

Keats, who died aged just 25, is a pure poet. He is often regarded as one of the five most important poets who made a significant impact on the Romantic Movement in English literature. During a short but distinctive period of poetic career, Keats published 54 poems in three volumes of poetry: *Poems* (1817), *Endymion*: *A poetic Romance* (1818), and finally *Lamia*, *Isabella*, *The Eve of St Agnes, and other poems* (1820). Keats's famous poems include *Ode on a Grecian Urn* (1820), *Ode on Melancholy* (1819), *Ode to A Nightingale* (1819), *To Autumn* (1820), *Bright Star* (1819), and the fragment *Hyperion* (1820). In his poems, Keats used a wide range of poetic forms such as sonnets, Spenserians, and Miltonic epics to heighten the themes he was trying to portray. All of Keats poems were on themes of love and beauty mixed with his feelings of loss, joy and sorrow. They are characterized by sensual imagination and contain many poetic devices such as alliteration, personification, assonance, metaphors and consonance. In using poetic diction, Keats is known for his connotative use of words and bold and daring writing style. Between 1815 and 1821 (the year he died), Keats also wrote letters to siblings and friends (available at www.gutenberg.org) that concern with moral and artistic problems and define his poetic practices and provide insight into writing in general (Grogan, 2021).

In summary, Romantic poetry has many characteristics and romantic poets were bound to their own literary norms in what they wrote about. They tend to place an emphasis on some of these characteristics over others to create a vision in a certain way, but they have enough similarities in

their writing styles that a poem of one poet might have a certain shared poetic form of another (Aljumily, 2015).

5. Data:

5.1 Corpus:

The literary output of first-generation and second-generation romantic poets varies considerably in number and length, so it was therefore important to prepare it prior to creating a corpus from each poet's poems. As a first step towards building such a corpus, I first limited ten or more poems per poet if oeuvre is large, and to balance length variation. I selected samples from each poet's respective known poems made available online in digital forms which are often considered, by literary critics and modern experts, to be among the most powerful or influential in the Romantic period poetry. I then compared them to the printed versions to determine, if the content had any corrupted words or lines or any transmission errors occurred due to scanning them. Therefore, I stripped any textual inclusions not original to each poet such as editorial comments and footnotes, line numbers, and so on. Next, I converted and saved all the poems in an ASCII (txt.doc) format. Where the name of a given poem is long, I referred only to the first word of that work where necessary. Given the majority of Blake's poems are short indeed in size, I compiled them into a single file entitled "Miscellaneous". Table 1 gives an overview of the poems that are selected for each poet.

Table 1: Selected poems per romantic poet

Blake	Wordsworth	Coleridge	Shelley	Byron	Keats
A Dream	A farewell	A Soliloquy of Full ...	Adonais	Beppo	Bright Star
A Little Blak	A Slumber	Ancient Mariner	Alastor	Childe Harlod	La Belle
A poison tree	Expostulation	Christabel	Epopsychidionx	Darkness	Lamia
A Song	Hart Leap	Dejection	Euganean hillsx	Don Juan	Ode on a Grecian
Holy Thursday	I Wandered	France	From Adonais	she walks in beauty	Ode on Melancholy
Jeurasalem	London	Frost at Midnight	Ginevra	so we will go	Ode to a Nightingale
London	Michael	Kubla Khan	julian	Stanzas for music	Ode to Psyche
Miscellenous	My Heart	The Pain of Sleep	mont blanc	The destruction	On First looking
Never Seek	O Nightingale	This Lime Tree	ode to libertyx	The Isles of Greece	OthoTheGreat
The Clod	Ode	Youth and Age	ozymandias	when we two parted	The Eve
The Garden	Surprised by joy		peterx		To Autumn
The Lamb	The Daffodils		prince		
The Marriage	The Lucy		tale of society		
The Sick	The Prelude		the dirgex		
Tyger	The Solitary Reaper		to a skyx		
	The World		west wind		
	Tintern Abbey				

In order to get a poetical representation for each of the 6 poets, I combined each poet's poems into one text corpus for analysis. To equalize the poem lengths among the different corpora, I sampled the poem texts from 14.000 up to a maximum of 14.999 words to make them comparable to each other or about the same size accordingly. Here I built six corpora of poems that are truly representative of each poet, and the size of each corpus is in harmony with each other prior to analyzing it. The corpora used for poetical quality analysis consists of 108 poems by the six poets distributed as follows: 44 poems for Blake's corpus, 17 poems for Wordsworth's corpus, 10 poems for Coleridge's corpus, 16 poems for Shelley's corpus, 10 poems for Byron's corpus, and 11 poems for Keats's corpus. Table 2 gives an overview of the structure of the six corpora.

Table 2: Size of test corpora

Corpora	Poems	Words	Lines
Blake corpus	44	14063	1950
Wordsworth corpus	17	14885	2093
Coleridge corpus	10	14262	2048
Shelley corpus	16	14948	1974
Byron corpus	10	14377	1814
Keats corpus	11	14398	1863
Total	**108**	**86.933**	**11.742**

After the corpora have been prepared, the next step was to select data from them.

5.2 Features:

The current analysis aims to build a stylistic-based model for the six romantic poets that predicts the poetic quality of their selected poems. The data set contains ten stylistic features, of which five are sound-based. Another feature set is syntax-based. I also have three lexical feature sets and another feature is based on readability information. The choice for the stylistic features is based on this study's aim to get an objective indication of some elements of poetry such as words or sound patterns in each corpus and to measure how often these and other certain features appear in it. The data set contains also one outcome variable: "poetic quality" which represents one category of the stylistic feature. In more detail:

- Syntactic features: is calculated as the number of occurrences of each parts of speech (N., Adj., V., Adv., Prep., Pro., Aux) within each corpus.
- Lexical density: is calculated as the number of lexical words divided by the total number of words per corpus.
- Lexical difficulty: type/token ratio, the frequency of difficult words, average word length, medium word length, average sentence length and medium sentence length. The type/token

ratio is calculated as the number of distinct words (types) divided by the total number of words (tokens) per corpus.

- Assonance: the frequency of assonance, and is calculated by dividing the sum of the relative frequencies of all similar vowel sounds in words that are close to each other in each verse-line by the total number of words per corpus.
- Alliteration: the frequency of alliteration, and is calculated by dividing the sum of the relative frequencies of all similar consonant sounds at the beginning of closely connected words in each verse-line by the total number of words per corpus.
- Consonance: the frequency of consonance, and is calculated by dividing the sum of the relative frequencies of all similar consonant sounds in words that are close to each other in each verse-line by the total number of words per corpus.
- Rhymes: the number of rhymes, and is calculated by counting the number of words which have the same last two and three letters at the end of verse-lines per corpus.
- Plosive densities: the percentage of plosive consonants in the poems, and is calculated by adding all plosive consonant values and dividing by the total number of verse-lines, then multiply by 100 per corpus.
- Fricative consonants: the percentage of fricatives density in the poems, and is calculated by adding all fricative consonant values and dividing by the total number of verse-lines then multiply by 100 per corpus.
- Readability score: the average percentage of the six readability metrics (Flesch-Kincaid Reading Ease, Gunning Fog Index, Kincaid Grade Level, SMOG formula, and Dale-Chall Score and Fry Reading Graph metrics), and is calculated by adding all of the six percentage values and then dividing by the total number of percentages per corpus.
- Poetic quality: it is an estimated measurement which represents one category of the stylistic feature and is coded with 1 if the impact falls below 15 in that category and with 0 if not.

Using the six different poet corpora, and the ten stylistic features, the most frequent features were detected and selected using Poetry Analyzer (Ver3.0) developed by softpedia.com, phoneme counter (Ver5.1) [computer software] developed by Nakanishi (2019), Readability Analyzer developed by datayze.com, and Text Analyzer Software developed by Online-Utility.org, which allowed to find the most frequent sounds, phrases and frequencies of words. I took a list of the most frequent tokens in the six corpora and produced a vector containing frequency of occurrence

of each token in all corpora. The data is sorted in comma-separated value (CSV) format. Each row is a poet corpus, and each column is a feature. Note that I used each single poet corpus for detecting and selecting features, a methodology that gave me the best results and can generally be accepted as better strategy. The basic descriptive statistics of the six corpora developed are shown in Table 3.

Table 3: Descriptive statistics of the data set per corpus

Poet	N	Sum	Mean	Variance	Std.
Blake	10	173.21	17.32	214.21	14.63
Coleridge	10	179.71	17.97	217.07	14.73
Wordsworth	10	187.73	18.77	232.31	15.24
Shelley	10	188.81	18.88	224.98	14.99
Keats	10	190.28	19.02	254.97	15.96
Byron	10	190.18	19.01	215.25	14.67

Before conducting the actual analysis, the predicative variables were tested for the assumptions of linearity and co-linearity to ensure that the final output is valid. To detect linearity, the logistic regression analysis was run and the estimated log values were saved as new predicator variables, then the scatterplots of each predicator variable was run to judge whether the relationship appears to be linear or not. The test of linearity is shown in Figure 1.

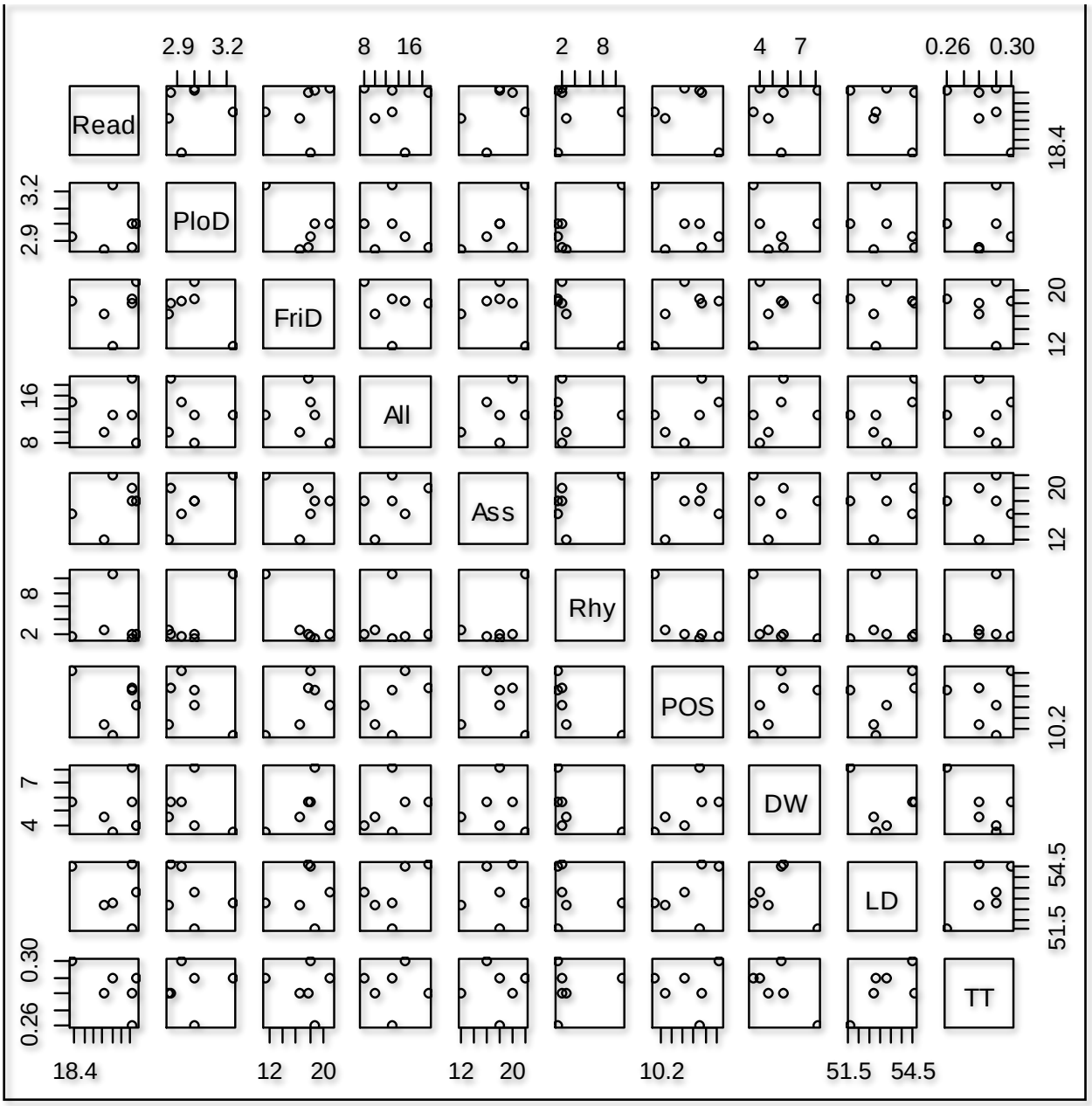
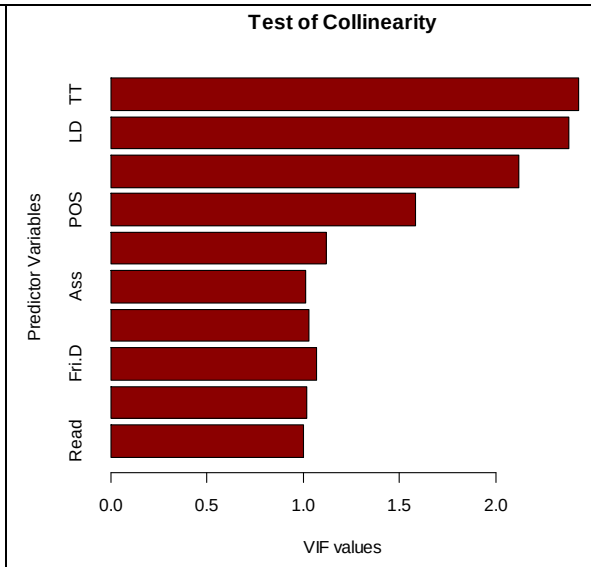


Figure 1: Test of linearity assumption of logistic regression

Figure 1 shows that the data were tested for the assumption of linearity to the logit for the explanatory variables. The logs of the variables are not significant, showing that the linearity assumption is not validated. As for detecting collinearity, a metric known as the Variance Inflation Factor (VIF) was used to measure the correlation and strength of correlation between the explanatory variables in a regression model and the VIF values were barplotted. The test of collinearity along with VIF scores is shown in Figure 2.

Variable	VIF
Read.	1.00
Plo.D	1.02
Fri.D	1.07
All.	1.03
Ass.	1.01
Rhy.	1.12
POS	1.58
DW	2.12
LD	2.38
TT	2.43



VIF values | VIF values bar plot

Figure 2: Test of collinearity assumptions of logistic regression

Figure 2 shows that data were tested for the assumption of collinearity to the logit for the explanatory variables, showing that there is no severe correlation between a given explanatory variable and any other explanatory variables in the model.

5.3 Analysis method:

Given that the aim is to predict and classify poetical quality as an outcome variable in the presence of ten stylistic features as predictor variables, Binomial Logistic Regression is applied on the generated data. The reason for using Binomial Logistic Regression for this data is to find a relationship between the predicator variable and probability of poetical quality. In this study, the outcome variable was poetical quality and is represented by two rating values "1" and "0". I say that observations with a stylistic feature greater than or equal to 15 will be classified as 1 and all other observations will be classified "0". This is simply to modify which rate the outcome variable occupies in each predicator variable χ. The predicator variable was the ten stylistic features and are organized as follows: Read (χ1), Plo.D (χ 2), Fri.D (χ 3), All (χ 4), Ass (χ 5), Rhy.(χ 6), DW (χ7), LD (χ 8), POS (χ9), and TT (χ10).

With this binomial classification, let 'χ' be some stylistic feature and 'Y' be the poetic quality which can be either 1 or 0. The probability that the outcome variable (quality) is 1 given its predictor variable (stylistic feature) can be represented as:

$$P\ (У = 1 \mid \chi)$$

To predict the probability logit function, also called logarithm-odds function is applied. We can express it as:

$$\log\left(\frac{P(\chi)}{1-P(\chi)}\right) = \beta_0 + \beta_1\chi$$

where, log is the logarithm-odds function and $P(\chi)/(1-P(\chi))$ is called odds.

The odds denotes the ratio between probability of good quality poets to the probability of low quality poets. According to this ratio, the linear regression combination of predictor variables are modelled to log odds ratio given the outcome variable is limited or equal to 1. So applying the log-odds function and then its inverse gives us:

$$P(\chi) = \frac{e^{(\beta 0+\beta 1\chi)}}{1+e^{(\beta 0+\beta 1\chi)}}$$

This is called the logistic function or sigmoid function and it always returns a set value of probabilities between 0 to 1 shaped like the letter "S" curve. The parameters of a logistic regression are most commonly estimated by Maximum Likelihood Estimation Method. There can be infinite set of estimates for each of the observations that search for the maximum likelihood. Being a binary solution, the probability for each observation must be Π if it was a good quality and 1- Π if it was a bad quality. Thus, we can calculate the likelihood function as follow:

$$\mathcal{L}(\beta;У) = \prod_{i=1}^{N}\left(\frac{\Pi i}{1-\Pi i)}\right)^{yi}(1-\Pi i)$$

In order to calculate the value of coefficients, the logarithm likelihood is used as it does not alter the features of the function. The logarithm likelihood function is characterized and by using iterative Fisher scoring, values of coefficients that maximize the logarithm likelihood are calculated. And in order to assess the performance of a logistic regression model, Null Deviance (ND) and Residual Deviance (RD) tests are used:

- ND denotes the dependent variable predicated by a model with nothing but an intercept. A lower ND value indicates a better model.
- RD denotes the dependent variable predicated by a model on adding independent variables. A lower RD value indicates a better model.

If RD is significantly lower than the ND, then one can say that the set of coefficients we included in the model improved the fit. We can also assess the performance of a logistic regression model by comparing the goodness of fit of different resulting regression models, a metric known as Akaike information criterion (AIC) can be used and is expressed as:

$$AIC = 2K - 2In(L)$$

where, K is the number of model coefficients and *In*(L) is the logarithm -likelihood of the model. After fitting data with different logistic regression models, we can compare the AIC value of each model. The model with the minimum AIC value provides good fit for the data. (Sperandei, 2014; Baguley, 2012).

Finally, to measure the degree of certainty in the generated results, the odd ratio for each predicator variable is calculated using the formula e^{β}. The 95% confidence interval (CI) for the odds ratio of each predicator variable is calculated using the formula $e^{(\beta+/-1.96*\text{std error})}$. The *P*- values are also used in conjunction with CI metric to measure statistical significance (Starkweather and Moske, 2011 and Hilbe, 2009).

6. Statistical analysis:

In this section, I used the ten stylistic features to build a logistic regression model that predicts the probability of poetic quality for each poet. Given that only ten variables were included in the analysis, I conducted the analysis with the intercept term by testing all the predictive variables each step to see which variable is significant and which is not in a way to select the significant ones to include. The initial step taken in this analysis is model selection. The criteria for comparing the resulting regression models is an Akaike information criterion (AIC). The model with the smallest AIC was selected. The result of our model can be seen below at Table 5.

6.1 Logistic regression results:

In this section I present the logistic regression results, followed by an analysis of goodness of fit of the resulting models. A detailed interpretation of the output models is presented. Beginning with the deviance residuals, important differences are identified between what we observed and what the resulting models predicted. Table 4 shows the deviance residuals for the significant predicator variables and response variable in data set. The residuals are nicely distributed. 1Q/3Q values and Min/Max values are about the same in absolute value, and the Median is close to 0. Also, Min/Max values are less than 3 in absolute value (Agresti, 2002).

Table 4: Summary of predicator variables

	Deviance Residuals				
Poet	**Min**	**1Q**	**Median**	**3Q**	**Max**
Blake	-1.41596	-0.27859	-0.11238	0.08859	2.04172
Coleridge	-1.184	-0.3659	-0.1044	0.3202	2.3764
Wordsworth	-1.32331	-0.41408	0.01679	0.12562	2.23567
Shelley	-1.0771	-0.3625	-0.1082	0.1891	2.3806
Keats	-1.0727	-0.387	-0.2013	0.1526	2.2931
Byron	-1.12635	-0.40602	-0.09767	0.13931	2.42747

Table 5 shows the logistic regression results in terms of the coefficients, *P*-values, and confidence intervals (CI). First let's look at the coefficient of significant variables. It is important to note that these coefficients are statistically significant since they are associated with a p-value < 0.05; Blake 0.01, Coleridge 0.03, Wordsworth 0.02, Shelley 0.02, Keats 0.02, and Byron 0.02, meaning there is a relationship between the predicator variables and the poetical quality which gives evidence for the hypothesis that poetical quality is predicted by a set of ten stylistic features. Notice that the coefficient of Blake in this logistic regression is -6.5288. Blake's probability in this case is very important since its absolute z-value is large. The six resulting models also show the extent to which predictor and outcome variables are related to predict quality. This relationship is shown by averaging variable effect size statistic and identified the minimum and maximum confidence interval magnitudes among the model as an estimation of the largest margin of error. The biggest impact is given to Blake and Wordsworth. The odds for predictor and outcome variables meeting this standard is 17 (9.918-28.102) and 18 (9.589-28.171) respectively. Another biggest impact is given to Shelley and Byron shown by the 95% CI. The odds for the variables fulfilling this standard is 18 (9.324- 28.216) and 19 (8.840-27.100). The last biggest impact is given to Keats and Coleridge shown by the 95% CI. The odds for predictor and outcome variables meeting this standard is 19 (9.128-28.912) and 17 (8.252-26.388) respectively.

The results presented here are based on the 10 stylistic features included in the test that are most likely to be predicative to poetical quality of these six romantic poets. Overall, Blake and Wordsworth consistently demonstrate the best odd ratios among the six poets ranking at or near the top in both *p*-value and 95% CI. Shelley and Byron and Keats also demonstrate well in both *p*-value and 95% CI. Coleridge has the lowest odd ratios, scoring much lower than other poets in the examined data set.

Table 5: The coefficients for the logistic regression models

		Coefficients						
					Intercept			
Poet	**95% CI**	**Null deviance**	**Residual deviance**	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	**AIC**
Blake	(9.918-28.102)	27.726	10.283	-6.5288	1.8105	-2.363	0.0181	14.283
Coleridge	(8.252-26.388)	25.8979	8.8028	-4.2779	3.0097	-2.169	0.0301	12.803
Wordsworth	(9.589-28.171)	27.726	10.702	-4.1513	1.7897	-2.320	0.0204	14.702
Shelley	(9.324-28.216)	27.726	10.473	-4.0394	1.7614	-2.293	0.0218	14.473
Keats	(9.128-28.912)	27.5256	9.8616	-3.7884	1.28998	-2.227	0.0260	13.862
Byron	(8.840-27.100)	27.526	10.901	-4.4734	1.9416	-2.304	0.0212	14.901

Now, let us find out the predicator variables that significantly influence the formation of logistic regression models. Chi-Square test is used to look at the differences with the predictive variables and to assess significance as well. Table 6 shows the Chi-Square results and Table 7 shows the predicted probabilities for the predictor variables by the model.

Table 6: Model fit

Poet	**Chi -Square**	***df***	***P*-value**
Blake	7.3276	8	0.042
Coleridge	1.5112	8	0.032
Wordsworth	0.0407	8	0.031
Shelley	1.0926	8	0.031
Keats	0.7498	8	0.032
Byron	0.759	8	0.021

The results illustrated in table 7 show that the significant predicator variables produce binary logistic regression model are: plosive densities, rhymes, and fricative densities. While the predicator variables that do not significantly affect the model are: readability, alliteration, assonance, parts of speech, lexical density, difficult words. and type/token ratio. Based on these, it can be shown that there are three predictor variables that distinguish between the poetic quality of those six poets. Plosive densities, fricative densities, and rhymes tend to be more likely to be predicative to quality in the six poets' poetry than alliteration, assonance and difficult words. For Blake and Wordsworth, plosive densities and rhymes tend to be higher than alliteration, fricative densities, and difficult words; for Shelley and Keats, plosive densities and fricative densities tend to be higher than rhymes, alliteration, assonance and difficult words; for Byron fricative densities tend to be higher than plosive densities, rhymes, alliteration, assonance and difficult words. Finally, plosive densities are likely to be predicative to poetic quality in Coleridge's poems; as this tends to be higher than other stylistic features. Furthermore, the value of the accuracy of the classification of poetical quality produced using binary logistic regression analysis of 95%. Table 7 shows the predicted probabilities for the predictor variables by the model.

Table 7: Probabilities predicted by the model

Table 7A: Blake's probabilities predicted by the model

Poet	Variables	*P*	95% Confidence Interval
Blake	Read	0.999	(0.0104, 1.0000)
	PloD	0.046	(0.0012, 0.8048)
	FriD	0.075	(0.0016, 0.8014)
	All.	0.999	(0.0210, 1.0000)
	Ass	0.338	(0.0448, 0.8484)
	Rhy	0.016	(0.0000, 0.8584)
	DW	0.913	(0.0990, 0.9990)
	POS	0.594	(0.1003, 0.9506)
	LD	0.998	(0.0233, 1.0000)
	TT	0.997	(0.0287, 1.0000)

Table 7B: Wordsworth's probabilities predicted by the model

Wordsworth	Read	0.999	(0.2551, 0.8881)
	PloD	0.045	(0.2869, 0.8611)
	FriD	0.582	(0.2073, 0.9967)
	All.	0.996	(0.0368,0.9687)
	Ass	0.305	(0.2586, 0.8855)
	Rhy	0.047	(0.2851, 0.8630)
	DW	0.063	(0.2458, 0.8949)
	POS	0.365	(0.2213, 0.8672)
	LD	0.992	(0.2949, 0.8516)
	TT	0993	(0.2502, 0.8917)

Table 7C: Coleridge's probabilities predicted by the model

Coleridge	Read	0.969	(0.2933, 0.9996)
	PloD	0.057	(0.0460, 0.7540)
	FriD	0.343	(0.0750, 0.7709)
	All.	0.777	(0.2994,0.9661)
	Ass	0.913	(0.3143, 0.9959)
	Rhy	0.611	(0.2319,0.8914)
	DW	0.951	(0.03048, 0.9988)
	POS	0.253	(0.0372, 0.7493)
	LD	0.147	(0.0105, 0.7384)
	TT	0.755	(0.2932, 0.9583)

Table 7D: Shelley's probabilities predicted by the model

Shelley	Read	1.000	(0.2516, 0.8235)
	PloD	0.018	(0.2517, 0.8238)
	FriD	0.055	(0.2504, 0.8230)
	All.	1.000	(0.2501, 0.8230)
	Ass	0.412	(0.0000, 1.0000)
	Rhy	0.584	(0.2511, 0.8231)
	DW	0.617	(0.2508, 0.8230)
	POS	0.379	(0.2514, 0.8233)
	LD	0.999	(0.2515, 0.8248)
	TT	1.000	(0.2507, 0.8230)

Table 7E: Keat's probabilities predicted by the model

Keats	Read	1.000	(0.1851,0.9888)
	PloD	0.030	(0.1656,0.8288)
	FriD	0.021	(0.1363,0.8347)
	All.	1.000	(0.0829,0.8536)
	Ass	0.454	(0.3047,0.8919)
	Rhy	0.091	(0.2215,0.8262)
	DW	0.859	(0.1209,0.8389)
	POS	0632	(0.2649,0.9566)
	LD	1.000	(0.2958,0.9221)
	TT	1.000	(0.2820,0.9413)

Table 7F: Byron's probabilities predicted by the model

Byron	Read	0.9990	(0.2494,0.8301)
	PloD	0.006	(0.0733,0.8573)
	FriD	0.030	(0.1528,0.8301)
	All.	0.999	(0.2807,0.9427)
	Ass	0.432	(0.1936,0.8253)
	Rhy	0.667	(0.1839, 0.8259)
	DW	0.632	(0.2174, 0.9805)
	POS	0.421	(0.1790,0.8263)
	LD	0.999	(0.2724,0.9507)
	TT	0.999	(0.2221,0.9789)

Now all I have to do is to model this output. The resulting models are compared and the model with the minimum AIC is selected on the basis that it gives the best fit for the data. The model with Figure 3 shows differences in poetical probability of being predicted by the significant predictive variables respectively. As can be seen the *Sigmoid* curve for each poet is relatively

similar since it takes stylistic predicators of nearly equal or close values. Despite the proximity of their stylistic predicators, the poetic quality for each poet is estimated. Based on the results obtained in figure 3, it can be shown that Blake and Wordsworth have higher quality probabilities of being predicted by plosive densities, fricative densities, and rhymes compared to Shelley, Keats, Byron, and Coleridge. The probability for Blake is relatively higher than the probability for Wordsworth. But Wordsworth has higher quality probabilities compared to Shelley, Keats, Byron, and Coleridge. Further, the lower quality probabilities in Blake and Wordsworth of not being affected by readability, alliteration, assonance, parts of speech, lexical density, difficult words, and type/token ratio resemble the probability level (-0.25) seen among those in, Shelley, Keats, Byron, and Coleridge.

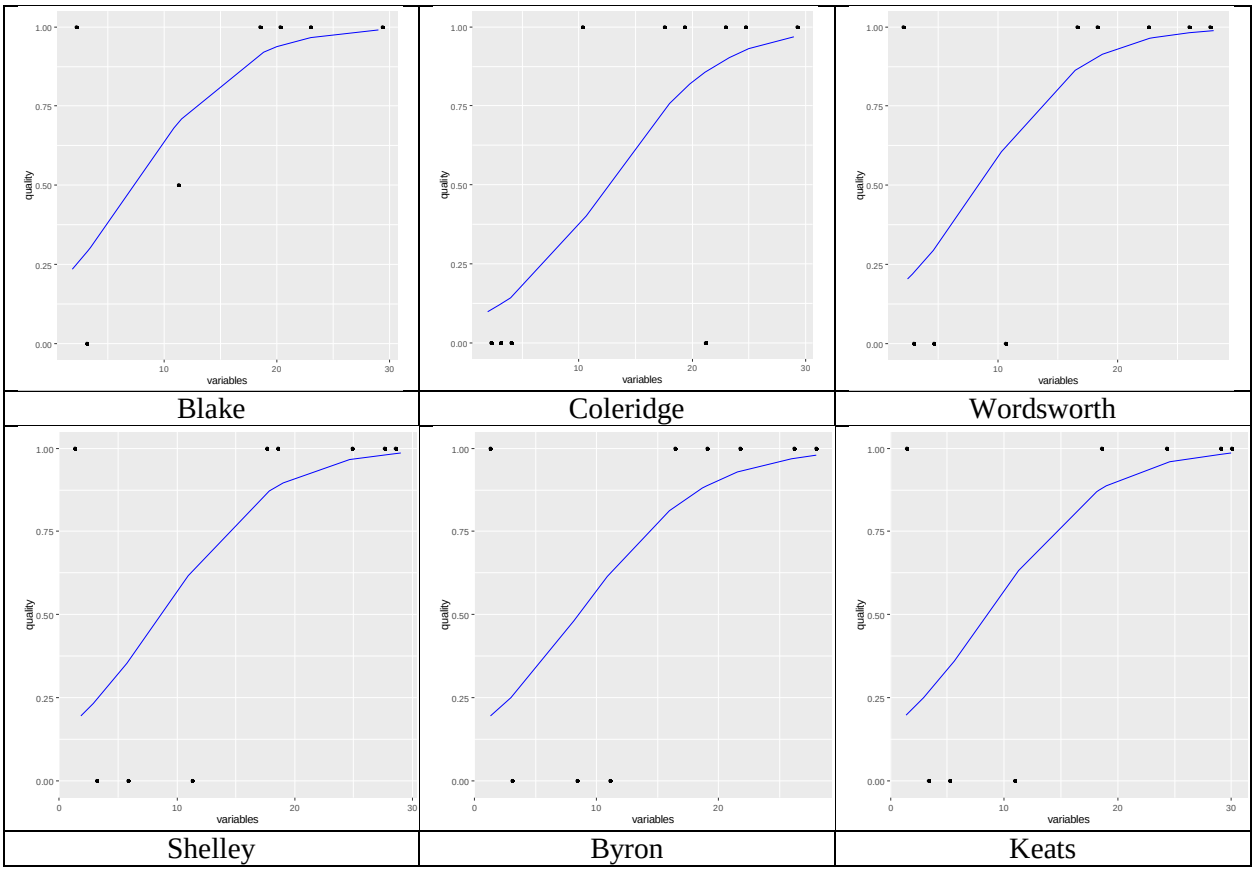


Figure 3: The Probability of poetic quality of Romantic poets modeled with measurement predicator variables.

Finally, I contextualize this outcome in terms of the estimated poetic quality for each poet from the highest to the lowest, and I conclude that those six poets are the quintessential masters of Romantic poetry but the estimated poetic quality for William Blake is relatively higher than that

of the other poets and it is therefore statistically probable to classify him the best English Romantic poet relative to the stylistic criteria used at the present analysis.

Table 8: classification of Romantic poetic quality

Poet	Classification level
Blake	1
Wordsworth	2
Shelley	3
Byron	4
Keats	5
Coleridge	6

7. Conclusions:

In this study, six multiple regression models were built for predicting poetic quality for Blake, Coleridge, Wordsworth, Shelley, Keats , and Byron. A stylistic relationship between a total of 10 stylistic features that were derived from 108 poems written by those poets was established with the specified quality at the specified examining levels. Each model was trained and tested to validate prediction and classification results. The model with the best fit was selected. Literary computing of poetical quality allows us to see alternative approaches to analyzing (classifying or predicting) poetic quality, ones that might be borrowed, or might complement the traditional literary analysis, to build quantitative criteria that yield useful objective insights. The results reinforce the investigation of personal stance in the computational poetic quality since there is not much work in the literature pertaining to this area. As described above, the results suggest that William Blake is the best Romantic poet in the data or the stylistic features included in the analysis. This result is the most likely. Furthermore, because those six English Romantic poets create very unique and different poems using different elements of poetry, the author still believes that more work can be done in the future. This particular study does have limitations: it was conducted only with some elements of poetry using frequency-based features when analyzed the 108 poems, although that is a very diverse criteria which describe many of the characteristics of poetry as a whole. It was also limited in terms of the number of poems that could actually be included for each poet. However, limitations from this study can encourage researchers seeking to expand this research. As the academic community has amply shown, no study has the perfect research design and methodology. Literary computing in humanities is a field in progress; the computation of other

elements of poetry (that create images, simile, metaphor, hyperbole, meter, rhythm, etc) continues to advance. But by learning from what has been involved and what has not in the current research, all researchers have the opportunity to engage and try out new methods with different elements of poetry and literary stylistic devices in order to gain greater insights. Lastly, researchers who may wish to conduct further research or expand this study to other contexts or settings should consider the results I have constructed here.

Duality of interest: The author confirms that he has no duality of interest associated with this manuscript.

Author contribution: The corresponding author listed in this work has solely worked on this manuscript during each development stage.

Funding: This research did not receive any grant from funding agencies in the public, commercial, or non-profit organizations.

References

Agirrezabal, M., Alegria, I. and Hulden, M. (2016). Machine learning analysis of English poetry. *In proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical papers* 772-781.

Agresti, A. (2002). *Categorical data analysis, 2nd edition*. New Jersey: Wiley

Arthur, M. J .(2017). Quantifying the beauty of words: A neurocognitive poetics perspective. Frontiers in Human Neuroscience, 11, 622. https://doi.org/10.3389/fnhum.2017.00622.

Al erjan, R. Sh. (2022). A stylistic analysis of john Keats's poem 'Ode to Psyche'. *Theory and practice in language studies* 12(7): 1378-1385.

Aljumily, R .(2015). *The anonymous 1821 translation of Goethe's Faustus: a cluster analytic approach*. PhD thesis., University of Newcastle Upon Tyne.

Altmann, G., and Fengxiang, F. (2008). *Analyses of script: properties of characters and writing systems*. Berline- New York: Mouton de Gruyter.

Baguley, T. (2012). *Serious Stats: A Guide to Advanced Statistics for the Behavioral Sciences*. UK: Red Globe Press.

Batool, Z., Kiran, S., and Azhar, M. A. (2016). Sstylistic analysis of William Wordsworth's poem 'Daffodils'. *International journal of basic and applied sciences* 5(2): 95-98.

Bekmirzaev, S. and Kim, T. (2017). Pairwise similarity analysis and quality estimation on classical Chinese poetry ancient Korea in 15the century. *International Journal of Applied Engineering Research* 12(23): 13884-13890.

Ben-Simon, A., and Bennet, R. E. (2007). Toward more substantively meaningful automated essay scoring. *The journal psychology* 17(2): 3339-353.

Birkhoff, G. D. (1933). *Aesthetic Measure*. USA: Harvard University Press.

Biber, D .(1988). *Variation across speech and writing.* New York: Cambridge University Press.

Biber, D., Conrad, S. and Leech, G. (2002). *The Longman student grammar of spoken and written English*. Harlow: Longman.

Breland, H. M. (1996).Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science* 7(2): 96-99.

Burroway, J. (2007). *Imaginative Writing: The Elements of Craft* . New York: Pearson/Longman.

Burstein, J. Chodorow, M., and Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine* 25(3).

Caraballo, S. A., and Charniak, E. (1999). Determining the specificity of nouns from text. *In proceedings of the 1999 joint SIGDAT conference on empirical methods in natural language processing and very large corpora:* 63-70.

Coleridge, S. T., Coleridge, H. N., and Morley, H. (1884). *Table Talk of Samuel Taylor Coleridge: And the Rime of the Ancient Mariner*. George Routledge and Sons: USA.

Crosbie, T . (2016). *A computer assisted analysis of literary text: from feature analysis to judgments of literary merits*. PhD thesis., University of Bedfordshire.

Ferber, M. (2010). *Romanticism: A Very Short Introduction*. Great Britain: Oxford University Press**.**

Flesch, R. (1948). A new readability yardstick. *The journal of applied psychology* 32: 221-233.

……....(1949). *The Art of Readable Writing*. New York and Evanston: Harper & Row Publishers.

Frida, M. N. (2012). Stylistic features in Byron's selected poems. Retrieved from https://cherryvlossomlovers.wordpress.com//

Zia, H., and Zia, M. S. (2017). A Stylistic Analysis of 'Ode to A Nightingale' by John Keats. *International journal of scientific and research publications* 7(8).

Genzel, D., Uszkoreit, J., Och, F. J. (2010). Poetic machine translation: rhyme and meter. *EMNLP*: 158-166.

Graves, M. F., Ryder, R. J., Salter, W. H., and Calfee, R. C. (1987). The relationship between word frequency and reading vocabulary using six metrics of frequency. *The journal of educational research* 81(2): 81-90.

Green, J. R., Wang, J., Samal, A., and Yunusova, Y. (2013). Articulatory distinctiveness of vowels and consonants: a data –driven approach. *Journal of speech language and hearing research* 56(5): 1539-1551.

Grogan, S .(2021). *John Keats: Poetry, Life, and Landscapes*. Yorkshire-Philadelphia: Pen and Sword History.

Hayward, M .(1996). Analysis of a Corpus of Poetry by a Connectionist Model of Poetic Meter. In: Poetics. Retrieved from http:// www.english.iup.edu/mhayward/Metrics/Cormetrics.htm//.

Hewings, A. , Painter, C., Polias, J., Dare, B., and Rhys, M .(2005). *Getting started: Describing the Grammar of Speech and writing book1, E303 English grammar in context.* Milton Keynes: Open University press.

Hilbe, J. M .(2009). *Logistic Regression Models*. USA: Taylor and Francis group CRC Press.

Holmes, J .(1992). *An introduction to sociolinguistics*. London: Longman.

Jerome, M .(2002). *Byron and Romanticism*. Cambridge: Cambridge University Press.

Joho, H., and Sanderson, M. (2002). Document frequency and term specificity. Retrieved from http://www. dcs.gla.ac.uk// (accessed 20/05//22).

Jones, S. K .(1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28 (1): 11-21.

Kao, J. T. (2011). A computational analysis of poetic craft in contemporary professional and amateur poetry. Retrieved from http://dl.icdst.org

Kao, J. T., and Jurafsky, D. (2012) A computational analysis of style, affect, and imagery in contemporary poetry. In *proceedings of the NSSCL-HLT*: 8-17.

Kaplan, D. M. (2006). *Computational analysis and visualized comparison of style in American poetry*. PhD thesis, Princeton university.

Kurland, D.J. (2000). What is Critical Reading? Retrieved from http://www.criticalreading.com/critical_reading.htm .

Lea, R., Rapp. D., Elfenbein, A., Mitchel, A., and Romine, R. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychology Science* 19(709).

Logan, G. D. (1988). Automaticity, resources and memory: theoretical controversies and practical implications. *Human factors* 30: 583-598.

Louis, A. (2013). *Predicting text quality: metrics for content, organization, and reader interest*. PhD dissertation, Pennsylvania University.

Mahoney, J. L. (2001). *Wordsworth and the Critics: The Development of a Critical Reputation*. NY: Camden House.

Marks, C. B., Doctorow, M., Maleen, J., and Wittrock, M.C. (1974). Word frequency and reading comprehension. *The journal of educational research* 67(6): 259-262.

Marsh, N. (2012). *William Blake: The Poems (Analysing Texts)*. 2nd edition. China: Palgrave Macmillan.

Mays, J.C. C. (2013). *Coleridge's Experimental Poetics*. USA: Palgrave Macmillan.

McGlone, M. S., and Tofighbakhsh, J. (2000). Birds of a Feather Flock Conjointly (?): Rhyme as reason in aphorisms. *SAGE Journals* 11(5):424-8.

Mcmaster, J. R. (1971). *The poet as hero: The hero Archetype in Romantic poetry*. PhD thesis., Univeristy of Alberta.

Michael, F. (2010). *Romanticism: a very short introduction*. Oxford: OUP.

Miles, J. (1964). Major Adjectives in English Poetry: From Wyatt to Auden. *University of California Publications in English* 12(3): 305-426.

……...(1957). *Eras and Modes in English Poetry*. Berkeley: University of California Press.

……...(1967). *Style and Proportion: The Language or Prose and Poetry*. Boston: Little, Brown and Co.

Morton, T. (2006). *The Cambridge Companion to Shelley*. UK: Cambridge University Press.

Nakanishi, N. (2019). Phoneme Counter (Ver5.1) [Computer Software]. Kobe, Japan: Kobe Gakuin University. Retrieved from https://noriko-nakanishi.com/.

Peer, W. V. (2008). *The Quality of Literature: Linguistic studies in literary evaluation*. Amsterdam-Philadelphia: John Benjamin Publishing Company.

Popescu, L., Lupea, M., Tatar, D., Altmann, G.(2015). *Quantitative Analysis of Poetic Texts*. Deutsch: Walter de Gruyter Mouton.

Rubin, D. C. (1995). *Memory in Oral Traditions*: The cognitive psychology of epic, ballads, and counting out rhymes. New York and Oxford: OUP.

Shakoor, M. (2015). A Stylistic analysis of Coleridge's 'Kubla Khan'. *European academic research* 3(8).

Sperandei, S. (2014). Understanding logistic regression. *Biochem Med (Zagreb)* 24(1):12-18.

Starkweather, J., and Moske, K. A. (2011). Multinominal logistic regression. Retrieved from https://it.unt.edu/sites/default/files/mlr_jds_aug2011.pdf .

Statistics Solutions. (2022). What is Logistic Regression?. Retrieved from https://www.statisticssolution.com/free–resources/directory-of-statistical-analyses/what-is-logistic-regression/.

Tomayo, J. (1987). Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational and psychological measurement* 47: 893-902.

Women's college, Agartala.(n.d.).Wordsworth theory of poetry. Retrieved from WORDSWORTH THEORY OF POETRY.pdf (womenscollege.nic.in)

Williamson, G. (2014). SLTinfo. Type-Token ratio of written language. Retrieved from http://sltinfo.com/