

# International Journal of English and Literature

Volume 6 Number 8 August 2015

ISSN 2141-2626



*Academic  
Journals*

## ABOUT IJEL

**The International Journal of English and Literature** is published monthly (one volume per year) by Academic Journals.

**The International Journal of English and Literature (IJEL)** is an open access journal that provides rapid publication (monthly) of articles in all areas of the subject such as African literatures, literature appreciation, cultural studies, literary styles etc.

The Journal welcomes the submission of manuscripts that meet the general criteria of significance and scientific excellence. Papers will be published shortly after acceptance. All articles published in IJEL are peer-reviewed.

### Contact Us

**Editorial Office:** [ijel@academicjournals.org](mailto:ijel@academicjournals.org)

**Help Desk:** [helpdesk@academicjournals.org](mailto:helpdesk@academicjournals.org)

**Website:** <http://www.academicjournals.org/journal/IJEL>

**Submit manuscript online** <http://ms.academicjournals.me/>.

## Editors

### **Dr. Nemat Azadeh**

*Department of Teaching, Jahrom Branch,  
Islamic Azad University,  
Jahrom,  
Iran.*

### **Prof. C.J. Odhiambo**

*Department of Literature, Theatre and Film Studies.  
Moi University. P.O.Box 3900  
Eldoret 30100,  
Kenya.*

### **Dr. Mrudula Lakkaraju**

*Geethanjali College of Engineering and Technology,  
Hyderabad, Andhra Pradesh, +  
India.*

### **Dr. Ruzbeh Babae**

*English language and literature,  
University of Putra Malaysia.  
Malaysia.*

### **Dr. Miguel Fernández**

*Chicago State University, ED 215  
9501 South King Drive  
Chicago,  
USA.*

### **Dr. Sunil Mishra**

*Dronacharya College Of Engineering,  
Gurgaon(Hr)123506,  
India.*

### **Dr. Farid Parvaneh**

*Department of English Language and Literature,  
Islamic Azad University,  
Qom Branch,  
Iran.*

### **Dr. Arezou Zalipour**

*School of Language Studies and Linguistics,  
Faculty of Social Sciences and Humanities,  
National University of Malaysia (UKM)  
Banda Baru Bangi, 46300 Bangi  
Selangor, Darul Ehsan.  
Malaysia.*

## Editorial Board

**Dr. Kevin Moore**

154A Hicks Street, 3F  
Brooklyn, N.Y. 11201  
(917) 202-4565  
(718) 595-5758.  
USA.

**Dr. Shweta Garg**

School of Humanities and Social Sciences  
IIT Mandi, Mandi  
Himachal Pradesh,  
India.

**Dr. Mohammed I. Mahameed**

Department of English Language  
and Literature,  
Faculty of Arts,  
Tafila Technical University,  
Tafila-Jordan.

**Dr. Yildiz Turgut**

Department of School of Teaching and Learning  
University of Florida,  
USA.

**Dr. Jitendra Kumar Mishra**

English Communications and Soft Skills,  
Faculty of Science and Technology,  
The ICFAI University Tripura.  
India.

**Dr. Patil Anand Balu**

202 Pratiksha, Survey No 96/97,  
Plot No 217, Nr Bhimsen Joshi Park,  
Right Bhusari Colony, Kothrud,  
Pune, State: Maharashtra,  
India 411038.

**Dr. Ream Fathi Fares Odetallah**

Department of English Language and Literature,  
Jordan University/Amman,  
Jordan.

**Dr. Parul Mishra**

Department of English,  
Banasthali Vidyapeeth University,  
India.

**Dr. Dare Owolabi**

Department of English and Literary Studies,  
Ekiti State University, Ado Ekiti,  
Nigeria.

**Dr. Ahmed ElShiekh**

Department of Translation and Linguistics,  
Zarqa Private University,  
Jordan.

**Prof. B. Abirami**

Sri Krishna College of Technology,  
Coimbatore,  
India.

**Dr. Vahid Nowrozi**

Department of publication and Centre Sanjeshetakmili, and  
Translator in Corporate Social Responsibility Development  
Centre  
Tehran,  
Iran.

**Dr. Juan José Varela**

Fernando III El Santo 7-8th,  
Santiago de Compostela 15706,  
Spain.

**Dr. Amir El-Said Ebrahim Al-Azab**

Bossat Karim El-Deen-Sherbin-Mansoura.  
Egypt.

**Dr. Maya Khemlani David**

Faculty of Languages and Linguistics  
University of Malaya.  
50603 Kuala Lumpur,  
Malaysia.

**Dr. Kanwar Dinesh Singh**

Government PG College, affiliated to HP University,  
Post Box # 5, G.P.O. SHIMLA:  
171001 HP India.

**Dr. Ruzbeh Babae**

Department of English language and literature,  
University of Putra,  
Malaysia.

**Dr. Sindkhedkar**

P.S.G.V.P's Mandal's A.S.C College, Shahada 425409,  
India.

**Dr. Lakshmi**

Department of English,  
Veltech Technical University ,  
Avadi, Chennai-54.  
India.

**ARTICLE**

**Research**

<b>Wordsworth's linguistic simplicity over Pope: A text analytic study</b>	<b>140</b>
Md. Jahurul Islam	

*Full Length Research Paper*

## Wordsworth's linguistic simplicity over Pope: A text analytic study

Md. Jahurul Islam

Department of Linguistics, Georgetown University, Washington, DC, USA

Received 29 July, 2014; Accepted 10 August, 2015

Conventionally, the judgment of the degree of linguistic complexity in a text relies much on the subjective interpretation of the readers. Besides being less generalizable across readers, the subjective approach has to remain limited to only a limited number of texts that human eyes can scan through. This study investigated the issue of linguistic simplicity of the poetical works of William Wordsworth over the ones of Alexander Pope by means of some computational tools for text analysis. In trying to investigate if Wordsworth's language was simpler, this study took a corpus consisting of a larger chunk of text than usually taken for subjective analysis; it consisted of the poems from the Lyrical Ballads and The Prelude by Wordsworth, and a collection of poetical works by Pope. The works of Wordsworth and Pope were then compared using a number of computational measures: the extent of overlapping vocabulary, type-token ratio, word frequency, hapax-token ratio, word recycling rates, and so on. Results indicated that the language of Wordsworth was not significantly simpler than that of Pope.

**Key-words:** Romantic poetry, text analysis, computational measures, digital humanities, linguistic complexity.

### INTRODUCTION

This study attempted a text analytic investigation of the language of poetry as employed by William Wordsworth and Alexander Pope in their works. While traditionally we rely on our subjective judgment to judge whose language is more, or less, complex, this study employed quantitative methods involving computational techniques to examine and compare the degree of complexity between Wordsworth and Pope with respect to their use of language in poetry.

The rationale of this study derives from the long standing debate over the claim by Wordsworth himself in his Preface to Lyrical Ballads that the language of his

poetry was completely different from that of his predecessors (including Pope). Some critics, e.g. Austin (1989) supports the view that Wordsworth's language was crucially 'simpler' than Pope's. As Crocco (2008) (p.112) contends, Wordsworth employs the "use of simple language" in Lyrical Ballads. Sarker (2003) (p. 362), similarly, supports the claim that Wordsworth preferred "simple and natural diction" to 'so-called poetic diction' of his predecessors.

However, there are critics (Grin, 1987) who maintained that Wordsworth's claim of linguistic simplicity in his poetry was motivated more by Wordsworth's intention to

\*Corresponding author. E-mail: jahirju30@juniv.edu.

undermine the contemporary popular conventions with a view to paving the way for his own poetry. This could be an indication that Wordsworth's language may not necessarily be linguistically less complex as he claimed to be.

While any reader is capable of passing a subjective judgment regarding the validity of Wordsworth's claims by going through a number of poems, this study attempts a different avenue through the use of computational techniques to find an objective way to contribute to the debate.

The methodology of this study draws profound inspiration from the proceedings of two authors: Ramsay (2008) and Jockers (2014). Ramsay comes up with the concept of 'algorithmic criticism' which puts emphasis on the benefits that literary study can extract by integrating computational techniques; to him, this can allow critics to ask questions that they find inaccessible without computational tools. Jockers (2014), on the other hand, provides a practical introduction to quantitative analysis of literary texts using R (R Core Team, 2014); here he introduces how the notion like 'algorithmic criticism' can be brought into practice using R. Thus, with the assistance of the concepts and techniques introduced by these authors, this study endeavored to ask the old question of linguistic simplicity in Wordsworth.

There may be at least two clear advantages to take from a computational approach to analyze literary texts. First, computational tools permit the analysis of a practically large amount of text, which might often go beyond the capability of bare human eyes and brain. What might take years for a human to scan through, computer can handle quite easily, in seconds. Being able to access such large chunk of text so easily brings about a second advantage; we are now able to ask new questions which were previously inaccessible to us only because of our limitation to handle bigger data (just to echo Ramsay (2008)). The final advantage might be identifying the 'objective basis for our subjective responses' to poetry, or any text in general (Dalvean, 2013); we now have at least another way to verify a claim made subjectively.

Though there has so far not been much interest in the computational measures of linguistic complexity for analyzing poetry, a number of studies can be referred to which take interest in analyzing poetry. For example, Kaplan and Blei (2007) used a quantitative method to analyze style in poetry. The authors took a sample of 81 poems by 18 poets, and 2 quantitatively measured the extent of 84 metrical dimensions of the stylistic features grouped into three categories: orthographic, syntactic and phonemic. Based on their measures, they came up with the concept of placing each poem on a quantitative vector space by means of Principal Component Analysis (PCA).

Another remarkable endeavor in computational

approach to poetry analysis is found in Kao and Jurafsky (2012) who studied how computational measures of linguistic features can be used to differentiate between professional and amateur poetry. The authors constructed logistic regression models with 16 pre-selected feature variables with a view to find which of them have significant effect on bringing about the difference between professional and amateur poetry. And, their study concluded that professional poems have significantly higher type-token ratio and more references to concrete objects, and fewer perfect end rhymes, instances of alliteration, negative emotional words, etc. than the amateur poems.

In line with the study of Kao and Jurafsky (2012), Dalvean (2013) attempted to demonstrate how computational measures may be capable of creating a continuum of features of being a professional or amateur in nature, so that the poems can be ranked on a linear graded scale. While Kao and Jurafsky (2012) focused on only 'successfully distinguishing' the professional poems from the amateur ones by means of computational measures, Dalvean (2013) went further to fine tune with the measures so that each of the poems can be ranked with respect to the others. Dalvean took a much wider range of linguistic and psycholinguistic variables than Kao and Jurafsky, and offered a fully ranked list, based on the Logit score from robust statistical models using the target variables, of a representative sample of contemporary American poems.

Though a number of studies, including the ones cited above, have targeted the computational analysis of the stylistic features in poetry, there have not been much research on analyzing linguistic complexity in poetical works from computational perspective so far; and, this study took interest in addressing this very area.

## Objective

Basically, this study aims to investigate whether Pope used a more complex language in his poetry than Wordsworth, by means of the computational measures. In other words,

Does computational measure of linguistic complexity indicate that Wordsworth is linguistically simpler/more complex than Pope?

If yes, what aspects of language contribute to the simplicity/complexity?

## METHOD

### Text corpus

The corpus for this study consisted of some selected digitized texts by the poets concerned. The texts were collected primarily from Project Gutenberg archive ([www.gutenberg.org](http://www.gutenberg.org)). The corpus for

Wordsworth consisted of his poems in the 1800 edition of *Lyrical Ballads* (Wordsworth, 1800a, 1800b), and the fourteen books of *The Prelude* collected from Bartleby online archive (Wordsworth, 1850). Pope's texts in the corpus included all the poems from one single volume of collected poems as available on the Project Gutenberg archive (Gutenberg, 2014). The total number of poems for Wordsworth was 90, and for Pope, it was 38; however, though Wordsworth had greater number of poems in the whole corpus, most of Pope's poems were greater in length than the individual poems by Wordsworth.

## Tools

R (R Core Team, 2014) was used as the primary software for processing the texts as per the requirement of analytical framework, and for running statistical tests where necessary. The R packages 'ggplot2' (Wickham, 2009) and 'gridExtra' (Auguie, 2012) were used for plotting purposes. Alongside, the Stanford POS Tagger (Toutanova, Klein, Manning, & Singer, 2003) was used to tag the parts of speech of the words in the corpus.

## Pre-analysis text processing

Since the texts from Project Gutenberg archive were just a collection in 'Plain Text UTF-8' format without any patterned encodings to differentiate the texts of the individual poems, the corpus was manually inspected to mark up (with some consistent codes) the beginnings and endings of individual poems. Then, any metadata available were excluded, and the texts of the poems were extracted from the corpus using regular expressions in R.

Since stop words (mostly the function words including word categories like prepositions, articles, determiners, etc.) do not contribute much to the linguistic complexity, the study excluded the stop words from the primary analysis. There-fore, a custom list of stop words was generated, so that the words in the corpus that matched with the words in the custom stop word list could be excluded from the analysis when necessary.

After that, two types of raw-frequency tables of the words in the corpus for each poet were generated: one with the overall frequency of words in the whole corpus; and the other one contained the word frequencies for individual poems (this one was generated to investigate how the findings from the overall corpus for each poet were consistent with their individual poems, throughout the whole corpus). Also, the raw frequency values for all words were converted to relative frequency values to make them comparable between the poets.

Furthermore, the words in the corpus were tagged with their Parts of Speech, using the Stanford POS Tagger (Toutanova et al., 2003). The primary interest in POS-tagging was to see if Wordsworth differed from Pope with regard to the extent of the use of nouns.

After processing the texts, the words from both the corpus were analyzed and compared with each other from a number of perspectives, putting through statistical tests where relevant. The analysis of the texts primarily involved the following: differences/similarities of vocabulary distribution of vocabularies across different parts of the corpus lexical richness cumulative frequency of the top words use of nouns

## RESULTS AND DISCUSSION

### Differences/similarities of vocabulary

This section reports the differences or similarities

**Table 1.** Tokens and types in Pope and Wordsworth.

Author	Tokens	Working Vocabulary(WV)	Vocabulary match
Pope	33076	7995	4116
Wordsworth	41430	9190	

between the vocabularies employed by the two poets, as found in the study.

'Tokens' vs. 'working vocabulary' for individual poets: Table 1 reports the raw counts of words used by the two poets; here, 'tokens' refers to the raw count of all the words in the corpus; 'working vocabulary (WV)' refers to the number of individual word-types (ignoring their multiple occurrences of words) in the corpus; and 'vocabulary match' refers to the number of words that overlap between the Wordsworth and Pope.

Here, we can see that Wordsworth had a larger size of working vocabulary than Pope. This is an indication that Wordsworth was more diversified than Pope in choosing his words. However, the difference in the size of WV is about 13% (in relation Wordsworth's WV) which does not seem to be a huge one considering the difference of the corpus size. The effect of the size of the text on the size of the WV for both the poets has been investigated in further detail in 'Correlation between corpus size and the WV' section.

The most frequent words. Table 2 lists the top 20 words for both Pope and Wordsworth, along with their relative frequency values. Interestingly, the word 'man' tops the lists for both the poets; also a number of words like 'love', 'nature', 'life', 'day' are common to both. The most noticeable thing about them seems to be that the relative frequency values for most of the words were higher for Wordsworth than Pope. This, in other words, indicates that Wordsworth's recycling rate of the top words was consistently higher than that of Pope.

This very notion can be more conveniently visualized in graphical form in Figure 1 which plots the relative frequency values for the top ten words for both the poets. It is evident that the top words in Wordsworth are much more frequent than those in Pope. Therefore, though Wordsworth had a larger size of WV, he tends to have used some of them (the most frequent ones) at a much higher rate than Pope.

**Extent of vocabulary match.** Table 1 reports that Pope and Wordsworth shared 4116 words; this number represents 51% of Pope's and 45% of Wordsworth's vocabulary. Again, relative frequency values for these 'matched' words were consistently higher for Wordsworth than for Pope (Table 3).

To verify the consistency of this claim that Wordsworth's use of the 'matched' vocabularies was consistently higher

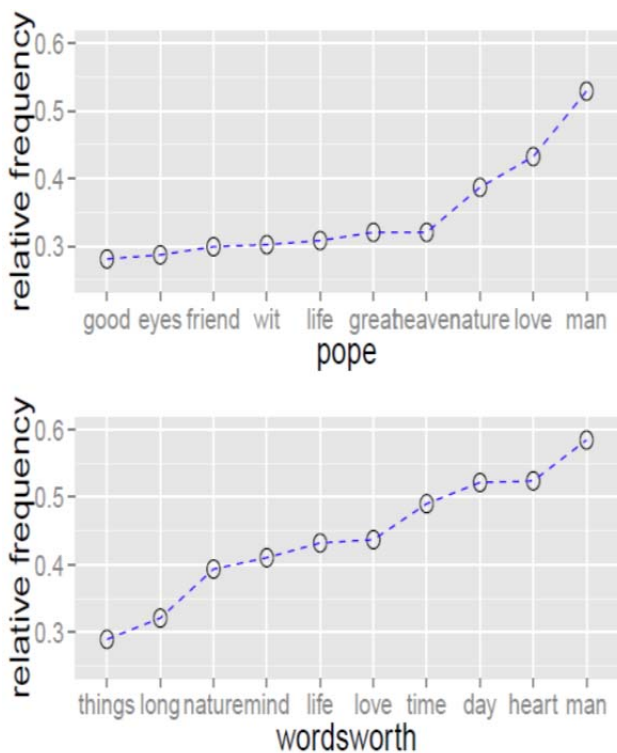


**Table 2.** Most frequent words

	Word	Relative Frequency (Pope)	Word	Relative Frequency (Wordsworth)
1	Man	0.53	Man	0.58
2	Love	0.43	Heart	0.52
3	Nature	0.39	Day	0.52
4	Great	0.32	Time	0.49
5	Heaven	0.32	Love	0.44
6	Life	0.31	Life	0.43
7	Wit	0.30	Mind	0.41
8	Friend	0.30	Nature	0.39
9	Eyes	0.29	Long	0.32
10	Good	0.28	Things	0.29
11	God	0.28	Side	0.27
12	Virtue	0.26	Power	0.26
13	Day	0.26	Made	0.25
14	Fair	0.25	Thought	0.25
15	Long	0.25	Thoughts	0.25
16	Pride	0.23	Years	0.24
17	World	0.23	Earth	0.24
18	Soul	0.22	Left	0.23
19	Rise	0.21	Men	0.23
20	Fame	0.21	Soul	0.23

**Table 3.** Top overlapping words in both Wordsworth and Pope

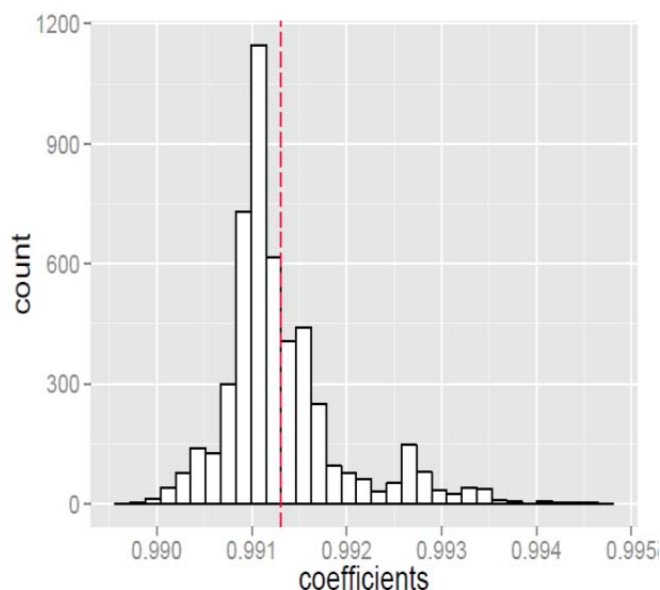
	Word	Relative Frequency (Pope)	Relative Frequency (Wordsworth)
1	Man	0.53	0.58
2	Love	0.43	0.52
3	Nature	0.39	0.52
4	Great	0.32	0.49
5	Heaven	0.32	0.44
6	Life	0.31	0.43
7	Wit	0.30	0.41
8	Friend	0.30	0.39
9	Eyes	0.29	0.32
10	Good	0.28	0.29
11	God	0.28	0.27
12	Virtue	0.26	0.26
13	Day	0.26	0.25
14	Fair	0.25	0.25
15	Long	0.25	0.25
16	Pride	0.23	0.24
17	World	0.23	0.24
18	Soul	0.22	0.23
19	Rise	0.21	0.23
20	Fame	0.21	0.23



**Figure 1.** Most frequent words in Wordsworth and Pope.

than Pope’s, a ‘test of correlation’ was performed, which returned a coefficient of 0.99, indicating almost perfect positive correlation between them. That means, the relative frequency of the overlapping words by Wordsworth was consistently higher than that of Pope. However, just to ensure that the coefficient value was reliable and valid, and not due to chance, the same test of correlation was performed 5000 times with random subset of 90% of the data each time (Figure 2). The results indicate that the distribution of the coefficients was consistently clustered around 0.99; this confirms that the ‘matched words’ were used at a consistently higher rate by Wordsworth.

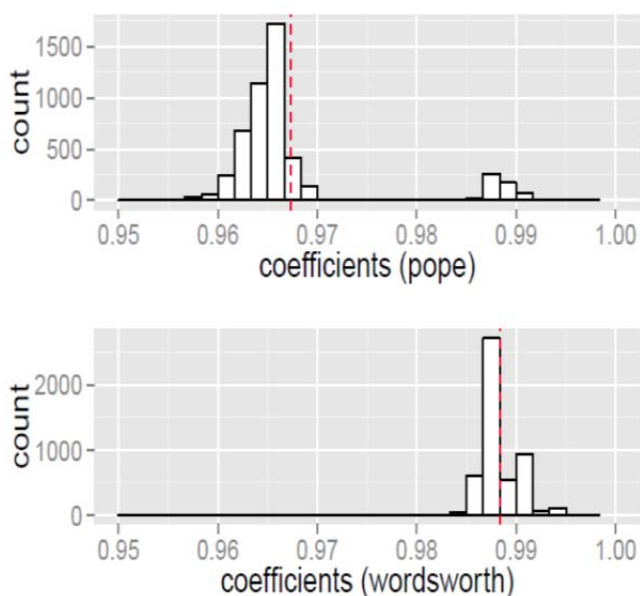
**Correlation between corpus size and the WV.** This section reports the extent of correlation between the corpus size and the size of the working vocabularies (WVs) for the poets. Basically, it tries to compare the correlation between the two poets. Results (Table 4) revealed that Wordsworth had a higher degree of correlation (again, the correlation results were confirmed with 5000 iteration with random sub-sets of the 90% of the whole data (Figure 3); this indicates that Wordsworth was more prone to introduce new words as the volume of the texts increased.



**Figure 2.** 5000 iterations of correlation between relative frequencies of overlapping words in Wordsworth and Pope.

**Table 4.** Correlation between corpus size and working vocabulary.

Author	Correlation coefficient
Pope	0.97
Wordsworth	0.99



**Figure 3.** 5000 iterations of the correlation between corpus size and WV.

### Distribution of vocabulary across the corpus

Looking into the dispersion of the working vocabularies for an author can reveal if different parts of a text (or corpus) are different in the degree of vocabulary richness. It can also be considered as an indication of linguistic complexity throughout the text/corpus for an author.

This section investigates the distribution of the WV across individual corpus for the two poets. For this purpose, the corpus for each poet was divided into ten equal parts, and calculations were made to find what percentage of words from the WV occur in each of the ten parts.

It was revealed (Figure 4) that Wordsworth had greater degree of variance in drawing words from his WV in different parts of his corpus; Pope had a much less degree of variance. Nearly one half of the corpus for Wordsworth drew a very low percentage of words from his WV, but the later parts incorporated very large number of new words. This variance might be a contribution of different pieces of work by a poet. In fact, Wordsworth’s corpus for this study included both Lyrical Ballads and The Prelude, and the text of The Prelude was in the later part of the corpus. Therefore, this might be an indication that the language of The Prelude was much richer in terms of vocabulary than Lyrical Ballads.

### Lexical richness

A useful measure for linguistic complexity can be the degree of lexical richness which is further represented by a number of measures including type-token ratio, word frequency, hapax-token ratio, etc. This section reports results of these measures.

**Type-token ratio.** Type-token ratio (TTR) is a measure of linguistic complexity calculated from the size of the WV and the number of tokens in a corpus. The size of the WV is divided by the number of tokens, and then multiplied by 100 to obtain a value in the form of a percentage. A lower value of TTR is indicative of lower degree of linguistic complexity.

It was found (Table 5) that Pope had a little higher mean TTR than Wordsworth, which is indicative of Pope’s being a little more linguistically complex than Wordsworth, on the whole. However, since these were mean TTR values which represent the corpus as a whole, the TTR values for individual poems for both poets were plotted to examine the differences of their distributions (in relation to the length of the poems) between the poets. A visual inspection of the distributions (Figure 5) was not indicative of any obvious difference between the corpora of the two poets. For both, the TTR value decreases as the length of the poem increases; also, the slopes for

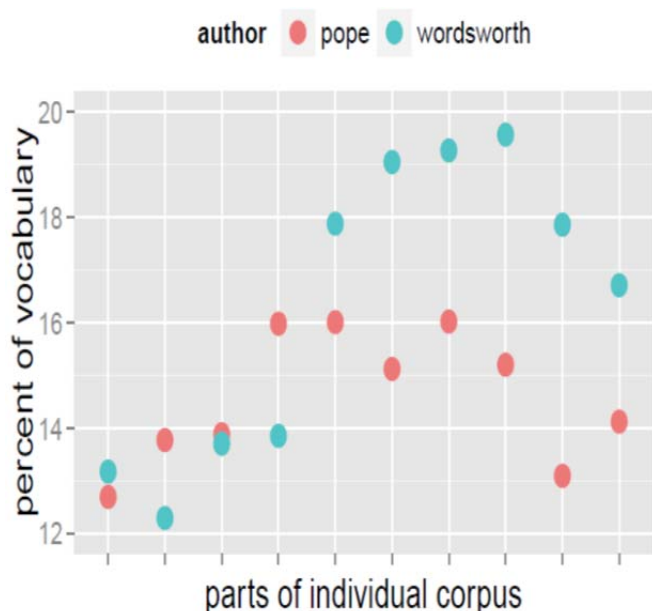


Figure 4. WV across the corpus.

Table 5. Overall type-token ratio

Author	Mean TTR (%)
Pope	53.16
Wordsworth	51.02

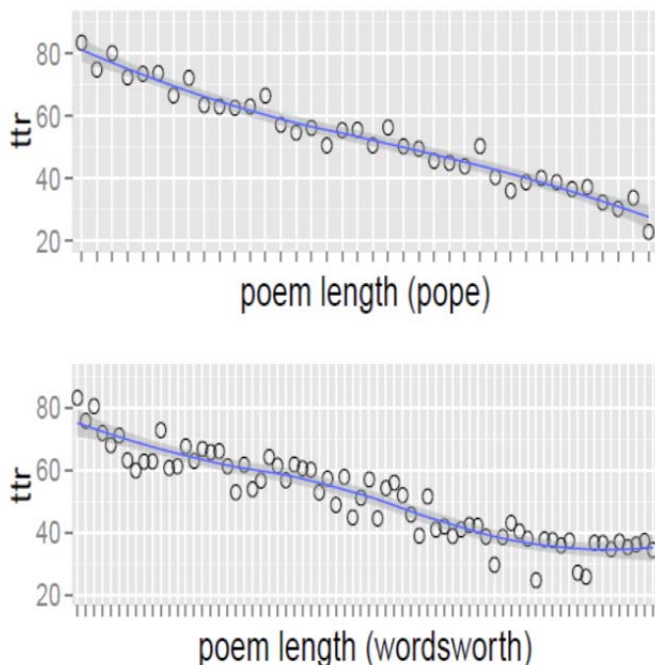


Figure 5. Type-token ratio in individual poems.

Table 6. T-test of TTR values between the two poets

WELCH TWO SAMPLE T-TEST
t = 0.7196, df = 70.463, p-value = 0.4742
95 percent confidence interval
-3.789289 8.067444

Table 7. Overall mean word frequency

Author	Overall mean word frequency
Pope	2.06
Wordsworth	2.12

both poets seemed to be very close to each other in visual inspection.

Since the difference was not obvious from initial inspection, a Welch two-sample t-test was performed to determine if the distribution of the TTR values for the two poets were significantly (statistically) different or not; and the results (Table 6) revealed no significant p-value (not less than 0.05, the conventionally accepted reference value). The 95% interval included 0 (zero), which confirmed that the difference was not statistically significant. Therefore, based on these results, the distribution of the TTR values indicate no significant difference in terms of linguistics complexity.

### Word frequency

An alternative way to express the notion of linguistic complexity is ‘mean word frequency’ (or just ‘word frequency’), which is basically the mean word-recycling-rate for an author. This is quite similar to TTR mathematically; it is calculated by dividing the number of tokens by the size of WV. Findings (Table 7) indicate that Wordsworth had a higher overall mean word frequency value than that of Pope; the distribution of the mean frequency values across different poems showed that there was no salient difference between the poets (Figure 6) with regard to mean word frequency.

Again, t-test of the distribution of the mean word frequency values for the two poets confirmed the lack of any statistically significant difference (p-value=0.47).

### Hepax legomena

Hapax legomena (henceforth ‘hapax’) are the ‘one-time words’, i.e. the words in the corpus that were used only once. To Jockers (Jockers, 2014), the ratio of hapax is an

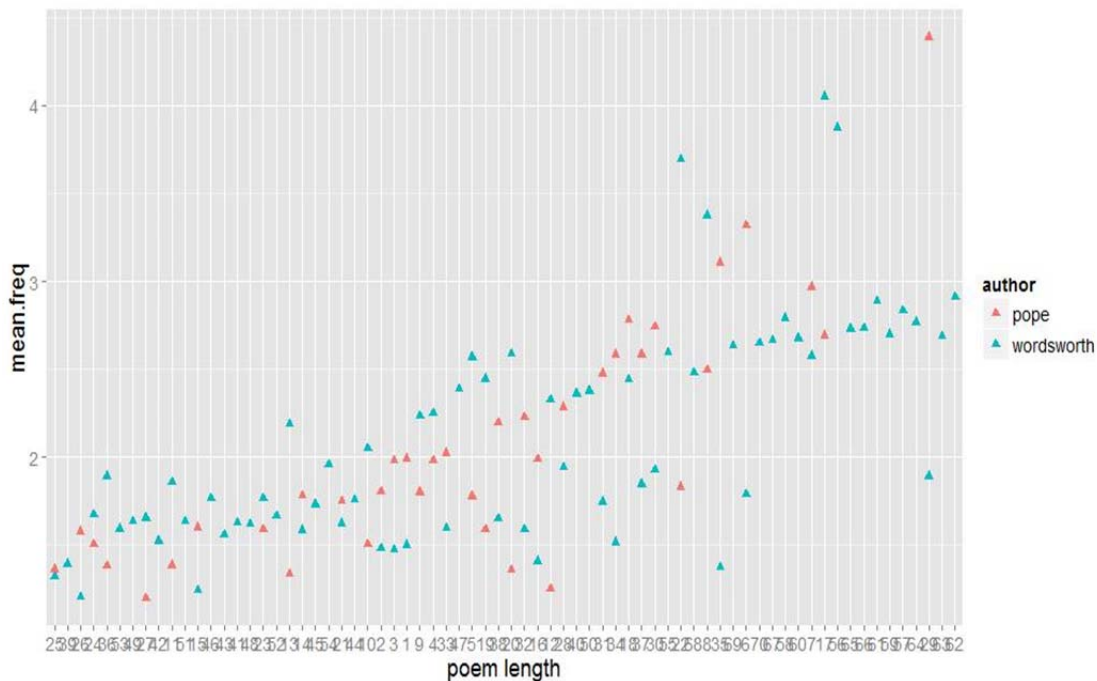


Figure 6. Mean word frequency and poem length.

Table 8. Overall hapax-token ratio in the two corpus.

Author	Mean hapax-token ratio (%)
Pope	40.51
Wordsworth	37.54

indicator of lexical richness; a lower hapax-token ratio is indicative of lower degree of linguistic complexity.

Here, Wordsworth was found to have a relatively lower value of hapax-token ratio (Table 8); though the ratio across different poems (plotted in figure 7) did not seem to make the different much salient.

Therefore, a t-test of the hapax-token ratio values across different poems for the two poets was performed here again, and the results did not return any statistically significant p-value (0.56).

### Two more measures

Alongside the relatively newer computational measures presented above, the study took interest in two more traditional measures of linguistic complexity (or simply 'read-ability'), which have been there for a long time; following are results from Gunning Fog Index test and Flesch Reading Ease Formula test.

**Gunning Fog Index.** The Gunning Fog Index Readability Formula (aka FOG Index) is a formula introduced by Robert Gunning back in 1952; it has been in use since then as a device to measure the level of complexity of texts for quite a long time. On this scale, a higher index indicates more complexity. The following table reports the Fog index (calculated from <http://gunning-fog-index.com/>) for the given corpora (Table 9).

**Flesch Reading Ease Formula.** Flesch Reading Ease Formula is yet another widely used measure to determine the complexity of a text. The range of score is usually between 0 and 100, and a higher score indicates easier read-ability in this scale. Table 10 reports the Flesch Reading ease scores for the two corpora (calculated from <https://readability-score.com/>).

Now, it is interesting to find that both of the above measures indicate that the texts from Wordsworth were rather more complex and 'less readable' than those of Pope. Thus, these traditional measures of readability also seem to be in line with the findings from the previous sections here.

### Cumulative frequency of the top words

So far the quantitative measures indicate that Wordsworth did not, in fact, have a significantly less complex

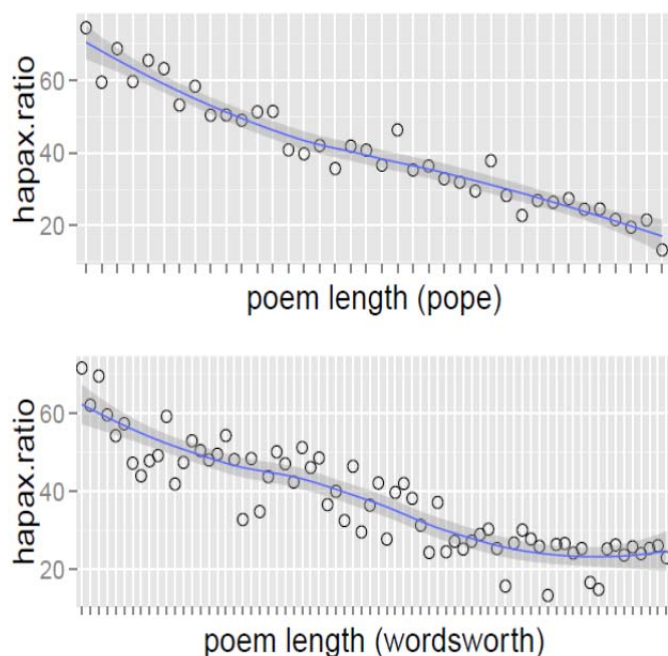


Figure 7. Hapax-token ratio in individual poems.

Table 9. FOG index for the authors

Author	FOG Index (%)
Pope	7.27
Wordsworth	10.57

Table 10. Fleisch-Kincaid score for the authors

Author	Fleisch-Kincaid score (%)
Pope	76.6
Wordsworth	63.5

Table 11. Cumulative word frequency of the top words.

Author	First 50	First 100	First 150	First 200
Pope	37.92%	46.56%	51.38%	54.82%
Wordsworth	41.04%	50.17%	55.09%	58.79%

vocabulary than Pope.

However, Wordsworth's propagation of the idea that his language is simpler may have derived from some other components of language. With the assumption that the cumulative word frequency of the top words and of the stop words might have a contribution to this notion, this

Table 12. Use of stop words

Author	Stop words (%)
Pope	53.34
Wordsworth	63.30

study took interest in analyzing them, too.

As Table 11 reveals, Wordsworth's rate of using (or recycling) the top words were consistently higher than Pope. This indicates to the fact that readers of Wordsworth encountered some of the vocabularies at a significantly higher rate (in relation to the other less occurring words) than they did in Pope's texts. Recycling a certain group of words more and more may have to ability to lead to the perception of linguistic simplicity since the reading is encountering the similar words again and again which might trigger their memory of those words encountered in previous sections. Therefore, if someone is prone to term Wordsworth's writing as 'less complex' than Pope's. This factor could be a potential motivation behind that.

Another factor that might have contributed (to those who talked in favor of Wordsworth) to the perception of simplicity is the predominance of stop words (as opposed to content words) in Wordsworth. While stop words do not contribute to the linguistic complexity in practice, their presence in a greater quantity might contribute to the sense of less complex language. And, investigation into this matter really indicated that Wordsworth had a much higher percentage of stop words (Table 12); almost 63% of Wordsworth's corpus was occupied by the stops words, which was about 53% for Pope. Thus, the predominance of the stop words might be cited as an explanation behind any perception that Wordsworth's language was less complex.

### Use of stop words

### Use of nouns

Frances (1989) claimed that Wordsworth used a lot of nouns which contribute to the 'simplicity' of Wordsworth's language. To investigate the credibility of such a claim, all the words in the corpus was tagged with their parts of speech using the Stanford POS-tagger (Toutanova et al., 2003). All of the tagged nouns in the corpus were then extracted, and the overall ratio plotted. The overall ratio (Figure 8); however, was rather indicative that it was Pope who used more nouns.

To explore how far this tendency was generalizable to individual poems (across the corpus, in other words) the ratio of nouns in individual poems were plotted also (Figure 9). As the graph indicates, most of the values for

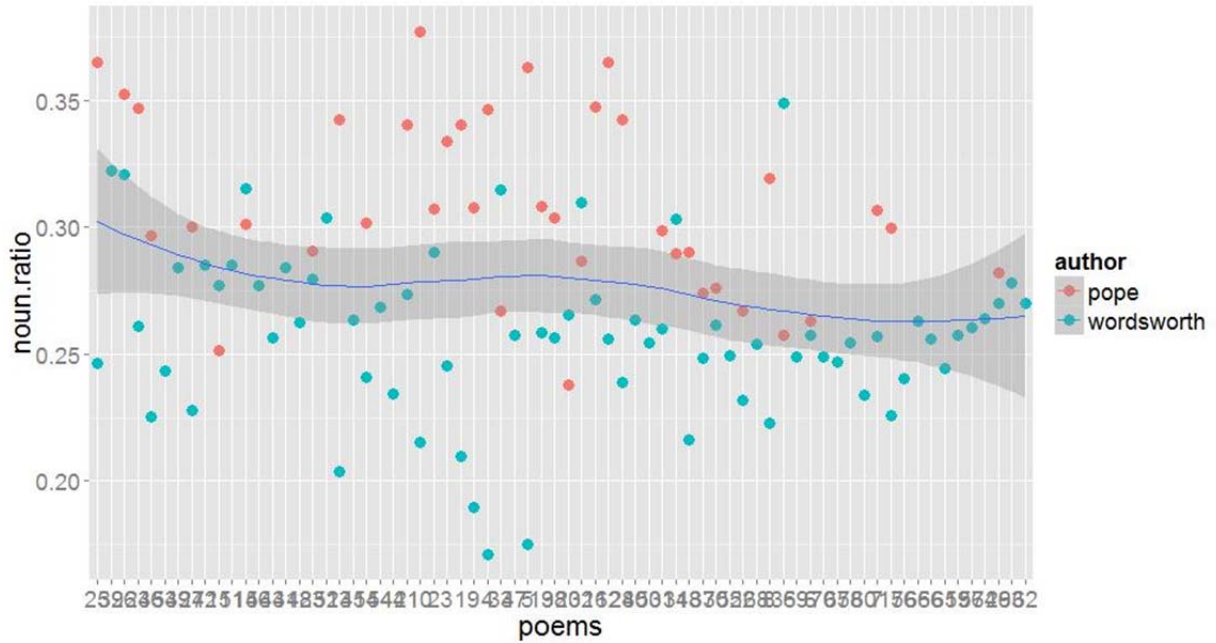


Figure 8. Ratio of noun use.

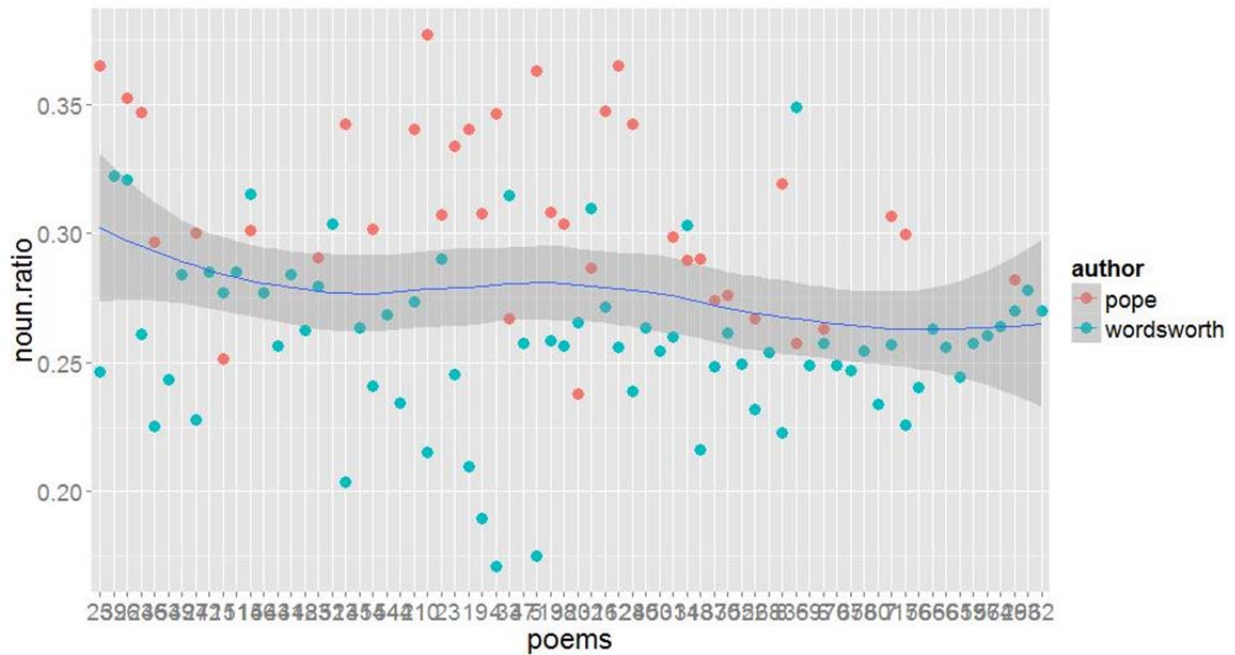


Figure 9. Ratio of noun use in individual poems.

Wordsworth were below corresponding local mean values (the blue line), while most of those for Pope were higher than the mean values. Thus, it appears that Wordsworth

did not have a higher rate of using nouns in comparison to Pope, and the claim of Francis does not seem to be well grounded, at least for the given corpus.

## Conclusion

The findings reported in this study did not provide evidence that Wordsworth's poetic language was simpler than Pope. The objective measures like type-token ratio, mean word frequency and hapax-token ratio revealed no significant differences. Rather, the FOG index and Flesch Reading Ease scores indicated an opposite trend. Also, the claim by Francis that Wordsworth had higher ratio of nouns, which contribute to the simplicity of his language, could not be validated through computational measures.

However, some factors seemed to have contributed to the somewhat popular perception regarding the Wordsworth's being linguistically simpler than Pope. The high rate of re-cycling the top words in the frequency hierarchy (e.g. table 2, 3) was possibly an important factor; this can be also validated from the cumulative percentage of the top words which were consistently higher for Wordsworth (table 12). Also, Wordsworth's language incorporated a considerably higher rate of non-content words, which naturally could have contributed to the perception of 'simplicity' in his language.

One thing to be noted that the quantitative measures of linguistic simplicity, as employed in this study, is capable of measuring 'lexical complexity' only, and it does not make take into concern the grammatical simplicity. Poetry has a natural tendency to deviate from the usual grammatical structures of natural language, a fact that is capable of influencing the perception of the language being more or less complex. However, this study tries to assert that Wordsworth's language of poetry is no simpler than Pope's, at least lexically.

Finally, it might often be somewhat inappropriate to holistically label the language of a poet to be less or more complex since the degree of linguistic complexity might differ even in different parts of the same piece of writing (as indicated in Figure 4). Furthermore, Pope died (1744) way before the *Lyrical Ballads* (published in 1798) and the other significant pieces by Wordsworth came out; so, there a big time gap between the productive years of the two poets. It might well be possible that the language of poetry, alongside the language in the society, was already undergoing changes. The works of Pope were very popular during his age (even when Wordsworth started his writings), which indicates that Pope's diction was still accessible to people. Therefore, Wordsworth's explicit denial of Pope's language to be artificial, and claiming that his own language embodies 'simplicity' seems to have derived from a rather 'political motivation' that Griffin (1987) terms as undermining Pope to pave his own way.

## Limitations and further studies

This study dealt basically with the lexical entries in the

corpus; however, adding syntactic complexity, alongside vocabularies, might be a nice idea to analyze linguistic complexities of the texts concerned. Also, a larger corpus might reveal more generalizable patterns. Changes of linguistic complexity in the writings of an author over time might also be an interesting thing to pursue.

## Conflict of Interests

The authors have not declared any conflict of interests.

## REFERENCES

- Auguie B (2012). gridextra: functions in grid graphics [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=gridExtra> (R package version 0.9.1).
- Austin F (1989). The language of Wordsworth and Coleridge. Macmillan. Retrieved from <http://books.google.com/books?id=5ElrAQAAlAAJ>
- Crocco F (2008). National eyes: Romantic poetry and the rise of British nationalism. City. University of New York. Retrieved from <https://books.google.com/books?id=2dgmcs2DYjUC>
- Dalvean MC (2013). Ranking contemporary American poems. *Literary and Linguistic Computing*.
- Griffin RJ (1987). Wordsworth's Pope: The language of his former heart. *ELH*, 54(3):695-715. Retrieved from <http://www.jstor.org/stable/2873227>
- Gutenberg (2014). The poetical works of Alexander Pope, Vol. 2. <http://www.gutenberg.org/cache/epub/9601/pg9601.html>. (Retrieved on Nov. 10, 2013).
- Jockers M (2014). Text analysis with R for students of literature. Springer International Publishing. Retrieved from [http://books.google.com/books?id=K4\\_IAwAAQBAJ](http://books.google.com/books?id=K4_IAwAAQBAJ)
- Kao J, Jurafsky D (2012). A computational analysis of style, affect, and imagery in contemporary poetry. In NAACL 2012 Workshop on Computational Linguistics for Literature.
- Kaplan DM, Blei DM (2007). A computational approach to style in American poetry. *IEEE*:553-558.
- R Core Team (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Ramsay S (2008). A companion to digital literary studies. In Schreibman S, Siemens R (Eds.), (chap. Algorithmic Criticism). Oxford: Blackwell. <http://www.digitalhumanities.org/companionDLS/>.
- Sarker SK (2003). A companion to William Wordsworth. Atlantic Publishers & Distributors (P) Limited. Retrieved from <https://books.google.com/books?id=CLzRPS-JatgC>
- Toutanova K, Klein D, Manning C, Singer Y (2003). Feature-rich part-of-speech tagging with a cyclic dependency network.
- Wickham H (2009). ggplot2: elegant graphics for data analysis [Computer software manual]. Springer New York. Retrieved from <http://had.co.nz/ggplot2/book>.
- Wordsworth W (1800a). Lyrical Ballads with Other Poems, 1800 Vol.1. <http://www.gutenberg.org/cache/epub/8905/pg8905.html>. (Retrieved on Nov. 03, 2013).
- Wordsworth W (1800b). Lyrical Ballads with Other Poems, 1800 Volume II. <http://www.gutenberg.org/cache/epub/8912/pg8912.html>. (Retrieved on Nov. 03, 2013).
- Wordsworth W (1850). The Complete Poetical Works: William Wordsworth. <http://www.bartleby.com/145/>. (Retrieved on Nov. 17, 2013).

# International Journal of English and Literature

## Related Journals Published by Academic Journals

- African Journal of History and Culture
- Journal of Media and Communication Studies
- Journal of African Studies and Development
- Journal of Fine and Studio Art
- Journal of Languages and Culture
- Journal of Music and Dance

**academicJournals**