

The Discovery of Subliminal Manifestation in Language

By Thomas McGrath

June 15, 2018

Abstract

This is just an initial finding and needs to be verified by a larger, more thorough study. Part 1 of this paper demonstrates that we unconsciously select words based on letter sounds that we like or dislike. Part 2 demonstrates that there may be harmony and dissonance in the pattern of frequency of letter usage, at least in the case of the vowels. To the best of my knowledge this is a new idea/discovery. The following paper contains graphs of the data collected from 15 classic works of literature including 4 works of HG Wells and 4 works of Jane Austen. Also a separate analysis is done using just the vowels from Shakespeare's Sonnets and the Eminem Recovery album.

Part 1 – Letter Frequency Preference

“It is impossible to read the compositions of the most celebrated writers of the present day without being startled with the electric life which burns within their words. They measure the circumference and sound the depths of human nature with a comprehensive and all-penetrating spirit, and they are themselves perhaps the most sincerely astonished at its manifestations; for it is less their spirit than the spirit of the age.”

- Percy Bysshe Shelley

Shelley was right when he said writers (as a generalization) are tuned in to the spirit of their generation, but what sounds pleasing to one person may not give the same pleasure to his or her neighbor, and thus there is a variety of poetry and literature even from within a single generation that often express the same concepts in different ways using different words. It seems common sense that each individual has different literary tastes, but

it is my intent to show that different tastes do not confine themselves to subject matter and word choice, but extend to the very use of different letters themselves. That the elemental sounds of words can influence our word choice or be studied as mass trends is not readily apparent. It would seem that, by sheer probability, two authors may use different amounts of each letter to form a book as unique as the individual that wrote it. However, in the example of Jane Austen's literature as well as that of H.G. Wells', we see a consistent pattern of usage of certain letters more than average and others less than average. This pattern is especially interesting because the two authors are opposites of one another in letter preference for nearly half of the alphabet.

In this small study, 4 popular books were chosen from each author to make a total of 8 books, plus an additional 7 books from other authors to make a grand total of 15 books. These additional books are to help us get a more precise grand total average letter frequency as well as see how other books compare to the 4 from each author that we are studying. First, all of the letters from each of the 15 books were counted and a baseline was set for each letter as the average percent of each letter used in context of the total of all of the letters used. For example, the letter A was used a total of 536,812 times across all 15 books. There were 6,635,567 total letters used, therefor the letter A represents about 12.361 percent of the total letters used in all 15 books. Then, the total letter usage for each individual book was done in similar fashion, the result being the percent of each letter when the total of all of the letters for the individual book are taken into account. The percent of the individual letter as used in relation to the book is subtracted from the average percent of that letter as is used in all 15 books, and this is done for every letter. Some results are negative and some positive. For the negative results, the letter was used more than average, while the positive numbers show a difference that is positive because the letter was used less than average, and this subtracted from the average yields a positive result. The accompanying graphs are thus counter intuitive

as I repeat: positive values are letters used less often and negative values are letters used more often than average.

For the purposes of this paper you can pretty much just eyeball the graphs to see the difference between Jane Austen and H.G. Wells, but this is because they have been set up to be a difference from the average in the fashion described above. How big is this difference? Each 0.1 percent represents a certain number of times the letter has been used. This is a different number of letters for each book, however the average for letter A is 537 times = 0.1 percent. So the number of letters that vary for the letter A can be thought of as somewhere in this ballpark. It would be slightly less for Wells' shorter works and slightly more for Austen's longer ones, and completely different for the letter B.

First level differences in letter choice are unanimous throughout the 4 books of one author and are opposed unanimously to the four books of the other, and this across the boundary of average letter usage. These letters are A, D, G, K, and Q, with O, T, and Y so close that I am counting these as first level as well. Second Level differences are almost there if it weren't for that one pesky book that throws it all off, but you can still see the difference. Letters J, P, R, and V.

There are several differences between the work of Jane Austen and that of H.G. Wells that may account for the difference in letter usage.

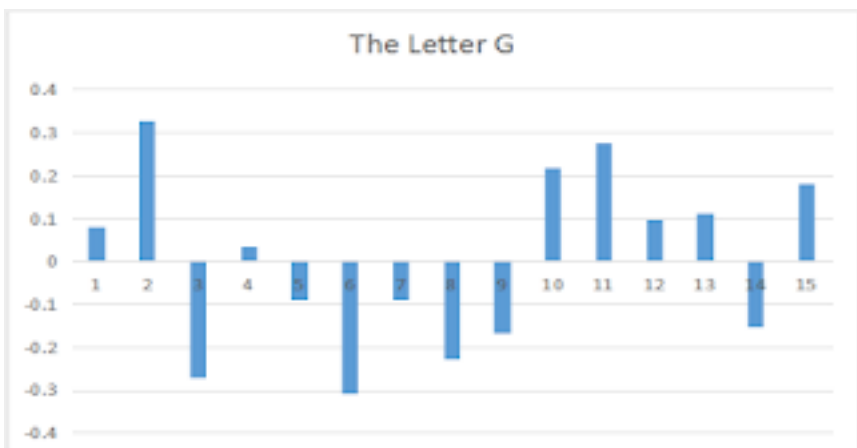
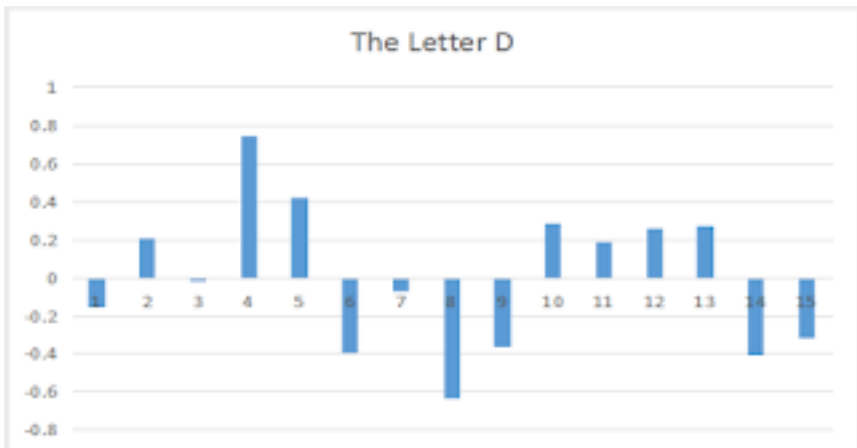
1. The length of their respective works are different with Austen averaging 529,442 letters per book and Wells averaging 200,921 per book.
2. The authors are different individuals with different tastes.
3. Gender
4. Genre
5. Time Period (Austen = turn of the 19th century while Wells = turn of the 20th)
6. Other (Including but not limited to several factors combined.)

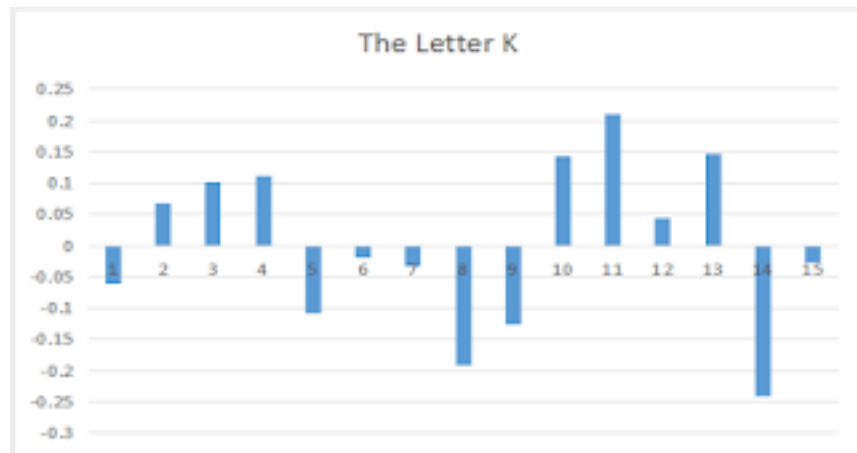
We can attempt to look at some of the other authors listed for insight into these factors. Moby Dick, for example, is a longer book than those of Jane Austen, and yet for first level difference letters A, G, K, and T, and second level difference letters J,R and V, Moby Dick letter usage is in the range of the works of Wells. This is, however, just one example and further study is needed.

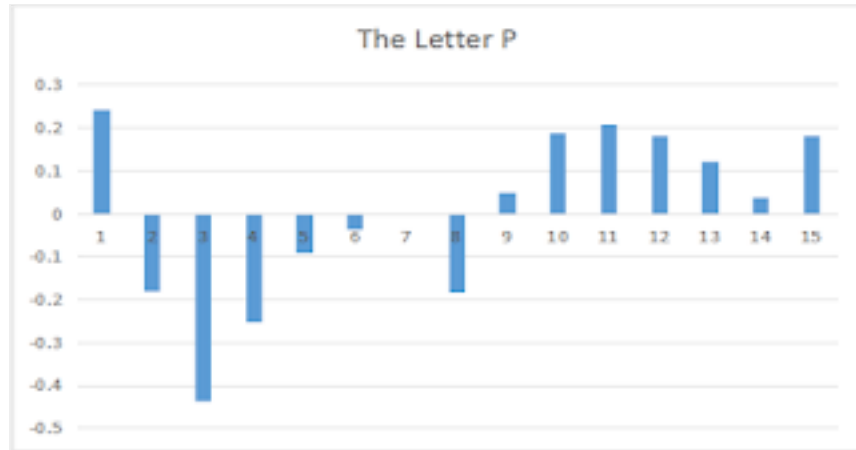
Hopefully this short report on some of the works of H.G. Wells and those of Jane Austen will spark some interesting and more comprehensive research. The rest of this paper shall consist of graphs for each letter. It may also be of worth to note that Austen and Wells are also diametrically opposed to one another in consonant to vowel ratio also across each of their 4 books and across the baseline average of the totals of 15 books.

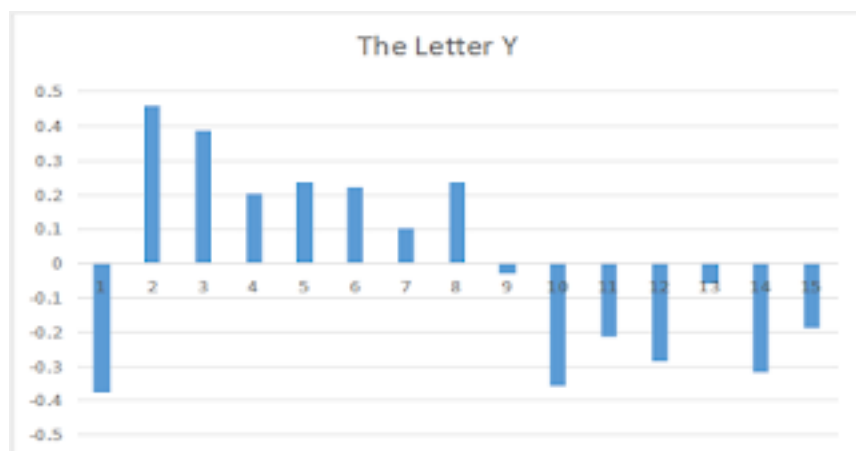
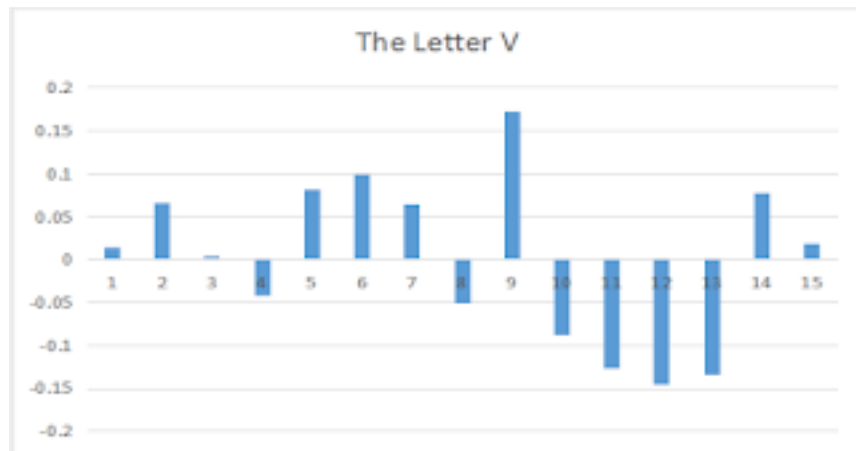
The order of books throughout the following graphs is presented below. For purposes of graph analysis, books 6-9 are written by Wells and books 10-13 by Austen. The difference letter graphs of note are presented below.

- 1.The Picture of Dorian Gray by Oscar Wilde
- 2.20,000 Leagues Under the Sea by Jules Verne
- 3.Around the World in 80 Days by Jules Verne
- 4.Billy Budd by Herman Melville
- 5.Moby Dick by Herman Melville
- 6.War of the Worlds by H.G. Wells**
- 7.The Time Machine by H.G. Wells**
- 8.The Invisible Man by H.G. Wells**
- 9.The Island of Dr. Moreau by H. G. Wells**
- 10.Pride and Prejudice by Jane Austen*
- 11.Sense and Sensibility by Jane Austen*
- 12.Emma by Jane Austen*
- 13.Persuasion by Jane Austen*
- 14.Little Women by Louisa May Alcott
- 15.Jane Eyre by Charlotte Bronte









Part 2 - Letter Frequency Composition

This essay shows the subliminal harmonies and dissonances between letter frequency distribution across Shakespeare's Sonnets and compares them with those of the verses on Eminem's Recovery album. The concept of harmony and dissonance as it relates to letter frequency will be made clearer through use of graphs in which the average use of each letter per verse is presented. The verses from each author's respective works are compared to the other verses in the order in which they were published, which is important as the harmonies and dissonances phenomenon can span multiple verses in a row and were most likely not intentionally organized this way consciously. In the future, people may want to have a professional linguist analyze everything from political speeches to movies, music albums, and books, to maximize the overall aesthetic effect of language in the subliminal dimension of letter frequency as it is presented in this essay.

If you look at figure 1, you will see a chart of the vowels as they are used in each of Shakespeare's Sonnets as a percent of total letters used in each sonnet. Of note is the stratification between the first level, letter E, the second level, letters A, I, and O, and the third level, letters U and Y with letter Y being considered always a vowel in computer analysis of the letters. This stratification is sometimes harmonic in that you can see it distinctly for as much as 10-15 sonnets in a row, and is sometimes dissonant in that the second and third levels jumble together also for as much as 10-15 sonnets in a row. This can also be referred to in the mathematical sense as a phase space, because we are not yet sure as to whether they are actually harmonic and dissonant sounding to the ears. All we know so far is that there is a clear distinction for multiple sonnets in a row.

This study of Shakespeare's Sonnets shows us what we are looking for in figure 2, the same data for Eminem's Recovery album. This time the stratification is different, with the harmonies less easily seen until closer to the end of the album, though they are also noticeable earlier on if you know what you are looking for (a 3 level stratification as described in the last paragraph). Interestingly, the dissonances are jumbled differently in Eminem's verses with the 2nd level merging up with the 1st level instead of down with the 3rd level as in the sonnets. We are looking here at two completely different styles of composition that remain mostly consistent in how they differ throughout the body of work, but the harmonies in the two respective bodies of work are of the same stratification.

In composition, it's not that a dissonance is undesirable, in fact the areas of dissonance can make the areas of harmony more powerful and vice versa. We see in Eminem's album composition slight harmonies towards the beginning in verses (8-13) and (16-18) that build to more powerful harmonies towards the end (21-22, 25-26, 28, 31-32, 38-39, 42, and 44). Most of his singles are in the latter half of the album. In fact, of his four singles, the two that became number 1 hits on the billboard hot 100 chart follow the same pattern of dissonance in two verses followed by a harmonic verse. These are Not Afraid (19-21) and Love the Way You Lie (40-42) on figure 2. His other singles are Space Bound (27-29) which has a dissonance harmony dissonance structure and No Love (25-26) which arguably has a harmony-harmony structure. The album ends on a harmonic tone.

This is just an initial investigation looking at results that could possibly be noticed subconsciously as they appear subconsciously organized in entire bodies of work. For this reason the general trends of stratification are what we are primarily interested in, though it may be worth further study to see how much and to what level of detail these trends can be picked up on a subconscious level. This investigation of how vowels are stratified by frequency shows us that the organization of letter frequency by harmony

and dissonance is a real phenomenon with practical application and deserves further attention.



Fig 1

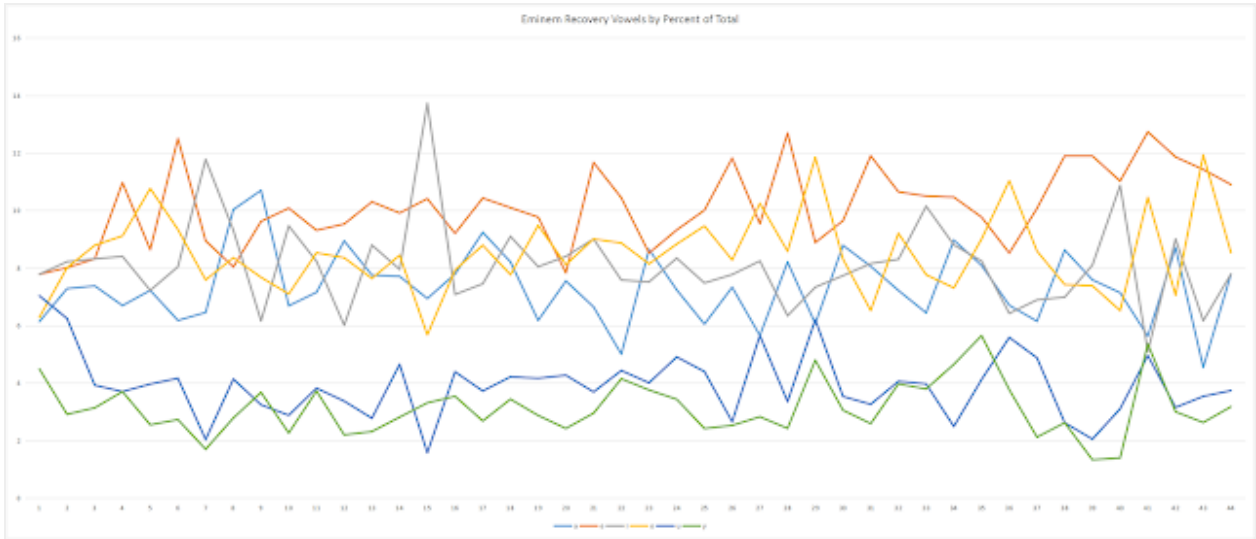


Fig 2