

Mendenhall selected many random samples of words from Shakespeare's prose and poetry from *The rape of Lucrece and Venus and Adonis* to study shifts in the peaks for frequency distributions of the number of letters per word. He discovered that although the two distributions for this characteristic were fairly different, they had the peak at four-letter words. The prose comprised relatively more shorter words (of three or less letters), but there were fewer words of five or more letters. Incidentally, this analysis raised a question whether this was to be thought as a general feature of that time, but further studies revealed no evidence supporting it. Samples of about 400,000 words were taken nearly from all his most famous plays, and the frequency curve was compared with the one based on the first 5,000 words. Two curves differed very little with respect to the characteristic under consideration.

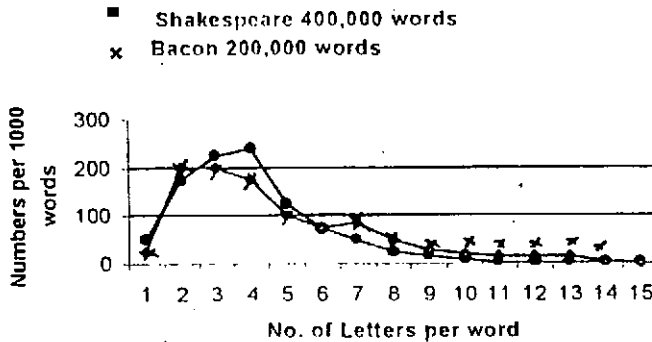


Figure. 1

The curve for Shakespeare in Fig. 1 represents an account of 400,000 words. The curve rises up to four-letter word and then registers a persistent drop. The other curve represents the distribution of the number of letters per word based on 200,000 words from a number of Francis Bacon's works including *Henry VII* and *Advancement of Learning*. The two distributions give an idea of the use of words of various sizes by these authors.

Based on such curves, it is not difficult to compare two more authors, and answer questions as to who uses more shorter words, larger

words, words of medium size, etc. studies of this nature have also been made for other famous authors.

G.B. William, a British statistician, estimates the following distribution of Shakespeare's word length

No. of letters per word	No. of words per 1,000	No. of letters per word	No. of words per 1,000
1	47.6	8	31.6
2	175.8	9	18.4
3	225.0	10	9.0
4	237.6	11	3.4
5	124.4	12	2.0
6	71.2	13	1.0
7	52.6	14	0.4

The style of one author differs from that of another living in the same milieu, writing in the same language and specializing in the same genre of literature. For example, a novelist may prefer to use longer words for one character and shorter words for another character. With similar vertical and horizontal axes, Fig. 2 shows the difference between two works of the same age and the same genre – Charles Dickens' *Oliver Twist* and Thackeray's *Vainety Fair*.

The calculations here are based on a much larger sample than that depicting the curve in Figure 1. There is a close agreement between the above information and the one provided in Figure 1.

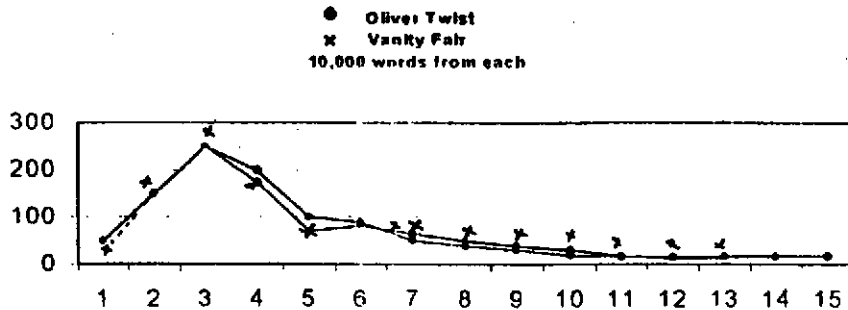


Figure 2