



Научно-учебная группа

Исследование частотных
характеристик языка

Санкт-Петербург
2023

Нарах legomenon

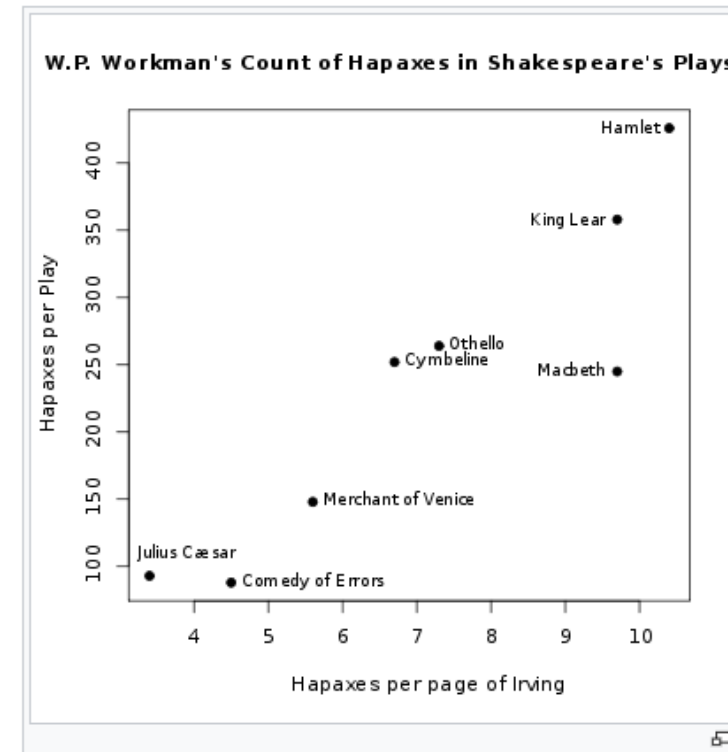
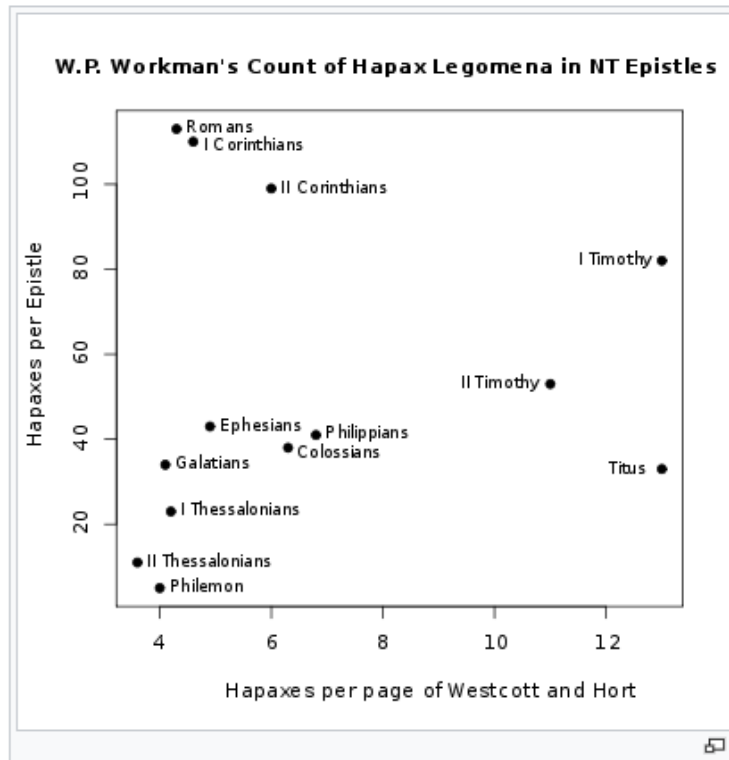
Спикер: Крайторов Михаил Владимирович



Что влияет на количество гапаксов в тексте?

- Длина текста;
- Тема текста;
- Текстовая аудитория;
- Время написания текста.

Что влияет на количество гапаксов в тексте?





Математическое выведение нарак legomena



Journal of Quantitative Linguistics



ISSN: 0929-6174 (Print) 1744-5035 (Online) Journal homepage: <https://www.tandfonline.com/loi/njql20>

Study of Optimal Text Size Phenomenon in Zipf–Mandelbrot’s Distribution on the Bases of Full and Distorted Texts. Author’s Frequency Characteristics and derivation of Hapax Legomena

Olga G. Gorina, Natalya S. Tsarakova & Sergey K. Tsarakov

To cite this article: Olga G. Gorina, Natalya S. Tsarakova & Sergey K. Tsarakov (2019): Study of Optimal Text Size Phenomenon in Zipf–Mandelbrot’s Distribution on the Bases of Full and Distorted Texts. Author’s Frequency Characteristics and derivation of Hapax Legomena, Journal of Quantitative Linguistics, DOI: [10.1080/09296174.2018.1559460](https://doi.org/10.1080/09296174.2018.1559460)

To link to this article: <https://doi.org/10.1080/09296174.2018.1559460>



Математическое выведение narax legomena

Несколько важных обозначений:

V - размер частотного словаря или списка слов

N - сумма частот всех слов - сумма всех слов текста (объем текста)

$F(i)$ - абсолютная частотность i слова

$p(i)$ - относительная частотность i слова = F_i/N

K - константа

$$F_i = \frac{k}{i}$$

- Закон ципфа, где k - константа



$$p_i = \frac{K}{i}$$

- Разделим левую и правую сторону на N

Математическое выведение нарах legomena

Представим себе частотный словарь. Последнее слово в частотном словаре будет иметь частоту вхождения равную 1. Предыдущее слово, тоже будет иметь частоту вхождения равную 1. И так будет продолжаться до какого то значения J при котором слово на позиции V-J будет иметь частоту вхождения 2. Наша задача найти это значение J, которое и будет равняться количеству слов в словаре с частотой вхождения равной 1

$$V_1 = \frac{(V + B)}{2}$$



$$F_V = \frac{K \cdot N}{(V + B)} = 1 \quad \longrightarrow \quad F_{V-j_1} = \frac{K \cdot N}{(V - j_1 + B)} = 2 \quad \longrightarrow \quad \frac{K \cdot N}{(V - j_1 + B)} = \frac{K \cdot N}{(V + B)} + 1 \text{ or } K \cdot N \cdot (V + B) = (K \cdot N + V + B) \cdot (V - j_1 + B).$$

- где B - константа