

Міністерство освіти й науки України
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
з предмета: Комп'ютерна лінгвістика

Звіт
про виконання лабораторної роботи № 4
«Закон Гіпса для слів в окремих текстах»

Виконав:
Студент групи
Фес-32с
Бойко Кирило

Львів 2024

Завдання

Використовуючи програму +proj6stats&plots, на лінгвістичному рівневі слів дослідити виконання закону Гіпса для обраних Вами текстів у лабораторній роботі №1. Провести дослідження для однограм ($n = 1$) за різних опцій (умов).

Хід виконання роботи

Підготував тексти з першої лабораторної роботи

Clemencia Novela de costumbres by Fernán Caballero

DONALD J. TRUMP January 20, 2017

Запустив програму +proj6stats&plots і завантажив текстові файли для проведення дослідження закону Гіпса

Результати

NGramm v3.2

Відкрити текст **Порахувати**

Параметри

☒ Ігнорувати регістр

Довжина n-грам 1

Буквені Символьні Словесні

Додати стоп-символи:

..?!;

☐ Враховувати розділові знаки

☐ Common Rank

Спільна статистика

Статистика для n-грам

від 1 до 1

Побудувати

Показати n-грами

Зберегти статистику

Статистика

для n-грам порядку 1

Показати n-грами

Зберегти статистику

Графіки

Нарахунок

для n-грам порядку 1

Гіпс

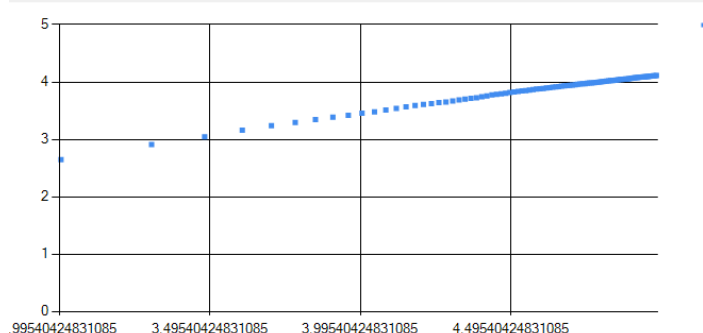
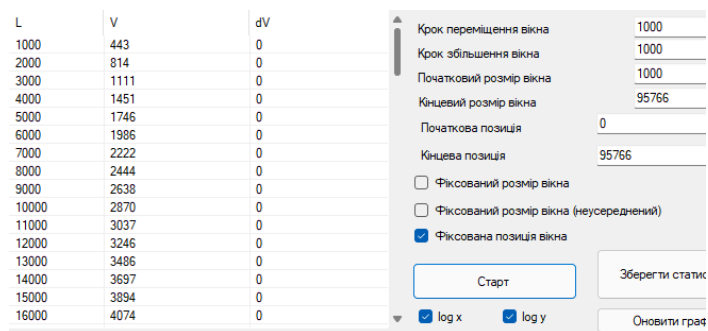
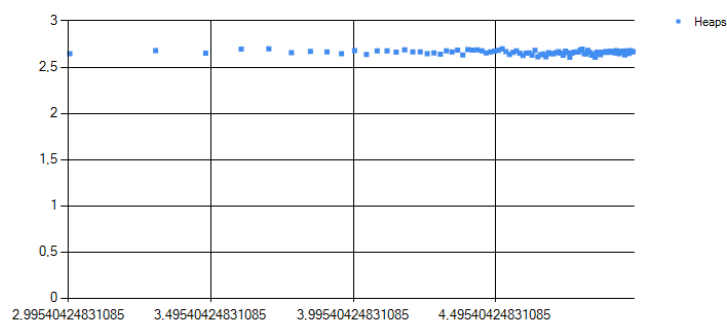
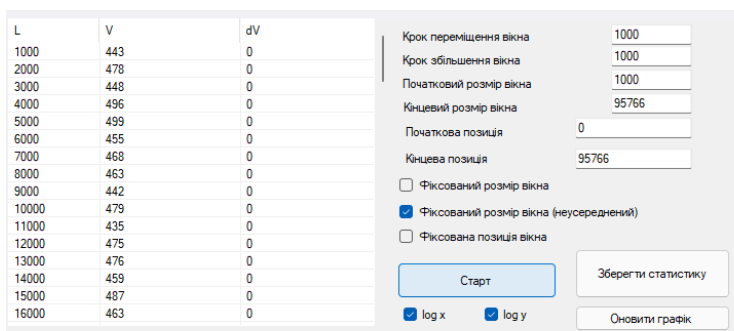
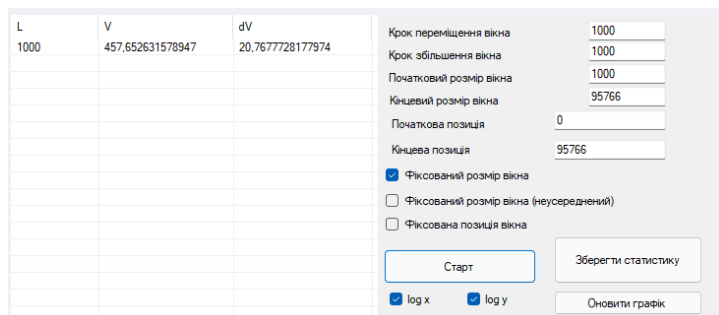
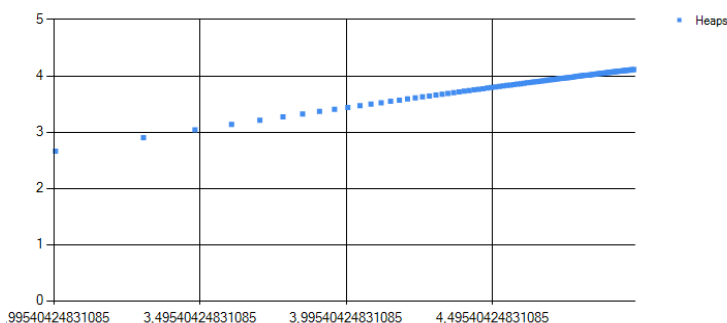
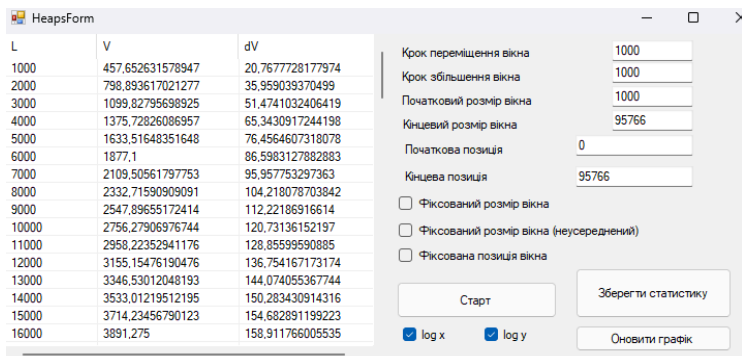
Швидкодія

Потоки 11

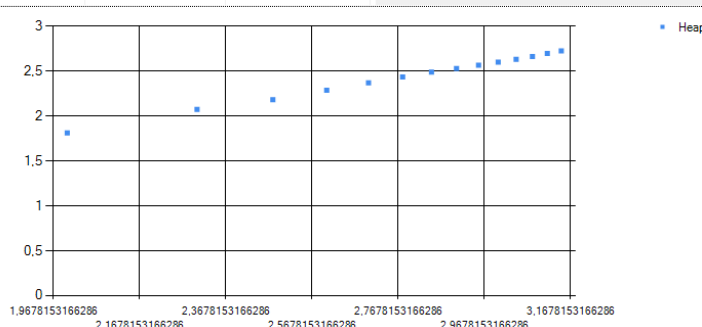
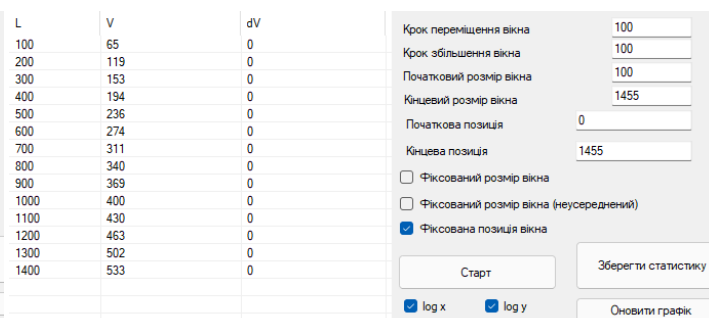
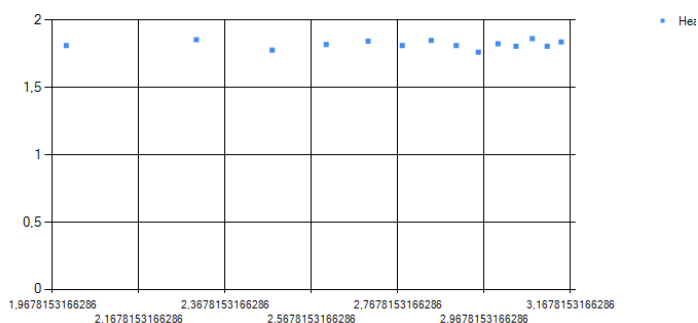
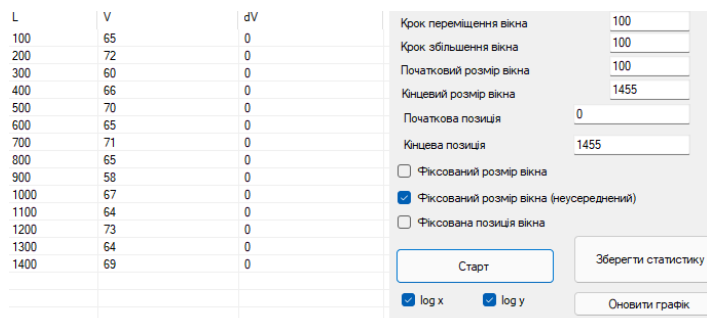
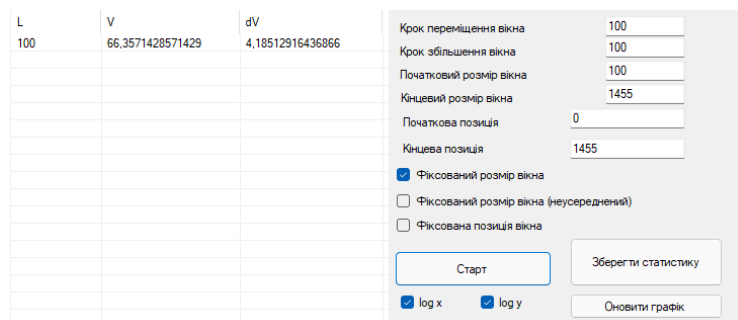
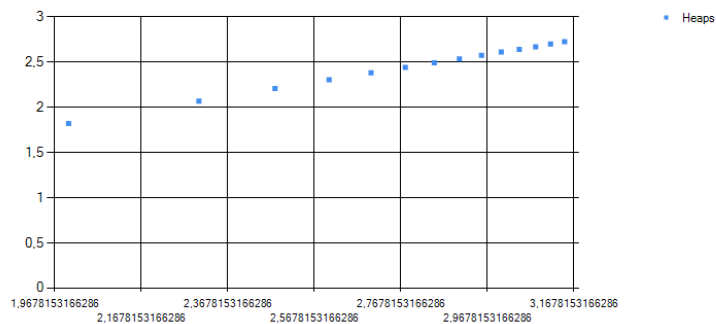
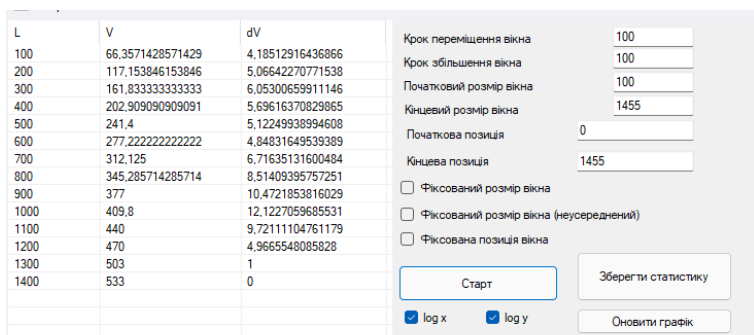
Поріг статистики 1

00:01:24.73

Розмір тексту: 95766



Результати виконання закону Гіпса для файлу ***Clemencia Novela de costumbres by Fernán Caballero***



Результати виконання закону Гіпса для *DONALD J. TRUMP January 20, 2017*

Кожен з текстів має характерну експоненційну форму кривої, що підтверджує відповідність розподілу частот закону Гіпса. Це означає, що найбільш вживані слова мають значно вищу частоту, ніж менш вживані, що свідчить про ієрархічну структуру лексики.

Малі відхилення від ідеального експоненційного розподілу можуть бути зумовлені стилістичними особливостями текстів та авторським слововживанням. Однак загальна тенденція розподілу частот у обох текстах підтверджує відповідність закону Гіпса, що свідчить про те, що лексика кожного з текстів підпорядковується цьому закону.

Висновок: У цій лабораторній роботі я дослідив виконання закону Гіпса для текстів на рівні однограм. Результати показали, що частоти слів у текстах загалом слідує експоненційному розподілу, що підтверджує виконання закону Гіпса. Хоча на графіках можна помітити деякі відхилення, вони, ймовірно, зумовлені стилістичними особливостями авторів та структурою текстів. Отже, проведене дослідження підтверджує, що частоти слів у літературних текстах відповідають закономірностям, описаним законом Гіпса