

Assignment 1:

First person frequency

Assigned 6/25

Due 7/2/19

Paul

1 Introduction: first person singular pronouns

Much of what we want to determine about an individual's underlying personality given an elicited sample of text can be determined merely by the frequency of certain important function words (an, the, because, he). Of these function words, among the most revealing is first person singular pronouns (1PSP)—‘I’, ‘me’ and ‘my’—which can indicate different things about the author of the text depending on the context, including self esteem, honesty and more.

2 The task

2.1 Part 1

Read Chapters 1-3 in NLTK book.

2.2 Part 2

Write a short program that counts the joint frequency of 1PSPs and measures the distance between occurrences of those pronouns—how many characters apart, how many words apart.

Apply your program to 100 cover letters that you find for different occupations. Set up a matrix of the different categories of occupations and the frequency of 1PS pronouns and compare the frequency distributions across categories.