Eng 2020: In-Class Demo on Evidence

Due: in class

Joshua Bowles

Your Name:

Jan. 27, 2010

In many cases when making an argument or a claim we are compelled to give supporting evidence. However, in many cases the supporting evidence we give is NOT based on empirical data, but is instead based on impressionistic data. Nowhere is this more apparent than when people are talking about language.

As an example, we will look at some of the statistical distributions of words in certain texts. Imagine that the questions you are answering are in fact propositions from a thesis statement. That is, for problem 1 we will argue that

> U.S presidents have a tendancy to use a large percentage of words that end with *-ly*. I argue that this is becuase more educated people tend to use a larger proportion of adjectives and adverbs formed by *-ly*. I also argue that given the fact that U.S. presidents use the word "America" more than any other word, there is significant pre-occupation with the identity of the U.S. in presidential addresses.

Imagine that problem 2 also has a thesis statement that proposes exactly the questions we are asking.

Problem 1

Directions: These questions relate to the collection (55 texts) of all U.S. Presidential Inagural addresses (1789 -- 2009).

(1) Guess the percentage of words that end in -ly (i.e.: "concurrent-**ly**", obvious-**ly**", etc.)?

(2) Which word occurs more: "citizen", "democracy", "America", "we"?

Problem 2

Directions: These questions refer to the book of Genesis in the Bible

(3) How diverse is the vocabulary used?[1]

(4) Which word occurs more: "God", "beginning", "said", "serpent"?

Problem 3

What is the most frequent word in English?

---

[1]One convention in religious writing is repetition of phrases, clauses, and words; if this text falls within convention, then its lexical diversity should be low. Lexical diversity = the number of unique words divided by the total number of all words. Here, we are looking for the number of words ending in *-ly* divided by the number of all unique words (no duplicates are counted).