

Austin James

CSCE 578

Assignment 1 Analysis

02/04/2019

In this project I found that there is a large difference in the reporting verbs that appear in the freshman English essays and in the formal academic essays. This is to be expected as the education level and expertise level vary significantly between the two groups of authors. The verbs used in the student essays are more like those that you hear in everyday conversation. Words like feel, seem, and show appear a lot. In the academic essays the reporting verbs used have much stronger connotations, and are typically verbs we use in casual conversation. These verbs, like demonstrated and reported, are much more pointed and are usually related to hard data.

I think that the most striking difference between the sets of the most commonly used reporting verbs in each essay is that the verbs used in the COCA data set have a much more confident connotation. These words make the information they are reporting sound much more reliable and accurate. This is something I expected, as someone writing an academic paper most likely has much more confidence in what they are writing than a freshman does in his or her English essay. They convey a certainty missing from those in the freshman English essays.

My results are not clean and should be viewed with the following information in mind. I did not remove stop words or punctuation from my data set. Therefore the frequencies are not completely accurate. A more refined data set would have removed punctuation, stop words, and numbers in order to calculate a more accurate number of valid tokens. I did clean up the verb set to only include instances of reporting verbs that were taken to be verbs, so we can say that

every recorded instance of a reporting verb was as valid as the raw data set can suggest.

However, no parser use to determine parts of speech is completely reliable.