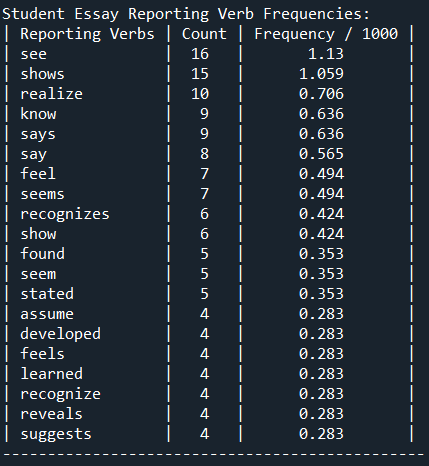
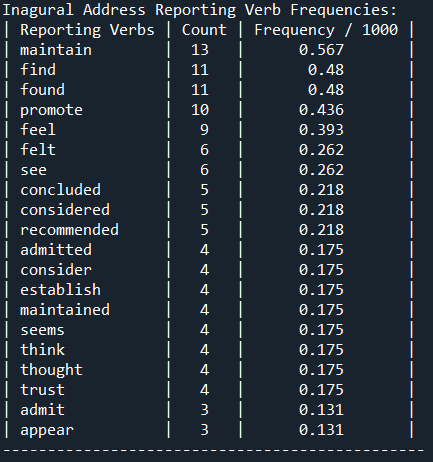
In this report we analyze the frequency of reporting verbs across a corpus of both first-year student essays and presidential inaugural addresses. We hypothesize that these bodies of text will have different types of reporting verbs used most frequently. To test this hypothesis, we built a python program which outputs the 20 most common reporting verbs in each corpus, their frequency per 1000 words, and the average length of these 20 reporting verbs. From this we can start to examine the validity of our hypothesis, though we will note areas of further research that would be necessary to understand the relationship more fully between the reporting verbs used in each corpus.

To start, our program generates the following tables of the top 20 reporting verbs in inaugural addresses and student essays, respectively:



To start, we can observe that only four reporting verbs are shared between our most common lists for each corpus: 'found,' 'feel,' 'see,' and 'seems.' In addition, we can observe that students seem to re-use reporting verbs more often. The frequency per 1000 of the top student reporting verb ‘see’ is approximately double the frequency of the inaugural address’ top reporting verb ‘maintain.’ In fact, the top 5 student reporting verbs are all more relatively common than the singular top inaugural address reporting verbs. Finally, the top 20 reporting verbs in the inaugural addresses are longer on average, with an average length of 6.40 characters compared to 5.26 in the student final essays.

In further research and refinement of this analysis, we might take several factors into account. To start, the size of each corpus was significantly different, with the 10 inaugural addresses containing 22910 tokens, and the 10 final student essays containing only 14158 tokens.

Speaking of tokens, we might consider a more complex filtering of tokens. This analysis relies on the assumption that both corpuses will utilize punctuation and stop words at similar rates, and thus does not filter them out of the total token count. A more refined approach might test the difference in results with these extraneous tokens removed.

Overall, our analysis offers some valuable insights into the usage of reporting verbs in inaugural addresses and student final draft essays. We saw that there was little overlap in the top 20 reporting verbs between both corpuses, there were longer reporting verbs in inaugural addresses, and that students tend to repeat their common reporting verbs at a greater rate than in inaugural addresses. However, it is still an incomplete analysis, and as discussed there is room for improvement.