Zachary Shorts

LING 360

Assignment 4

I chose to analyze the number of contractions, modals, and pronouns in each register. Here are some hypotheses that I came up with: 1. Contractions were going to be more prevalent in LY than other registers. 2. Pronouns were going to be more prevalent in the HI register. 3., Modals were going to be more prevalent in registers that had more opinions, such as the ID and OP registers.

My hypotheses were primarily correct as we can see from the measurements. The average number of contractions between all registers was 2.5 per 1000 words and in the lyrics register it was over 11. The average number of modals per 1000 words in all registers was 9.97 and in the ID register is was 13.55 and in the OP it was 11.08. The average number of pronouns per 1000 words was 121.27 and in the HI it was 130.52.

I ran a test on 1-2 of the files in the corpus and my precision and recall was measured at 100% for each. I matched all of the occurrences of each linguistic feature and every match that I made was a valid match,