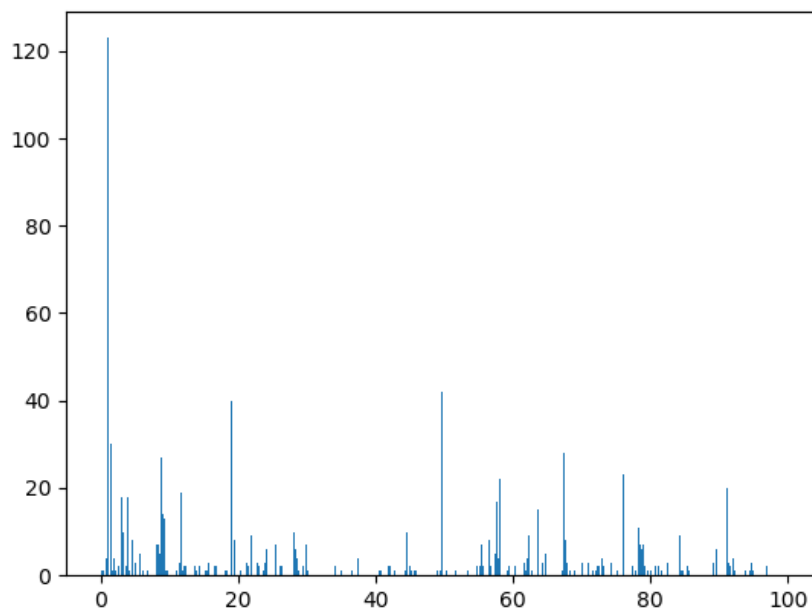


I believe that the language of the “mystery.txt” file is Spanish. I came to this conclusion by comparing the ngram-frequency histogram of mystery.txt to histograms of the other files, and this allowed me to narrow my choice down to either Spanish or Portuguese. I proceeded to compare the output of the Top 10 most frequent ngrams. The Spanish and Portuguese outputs contained 4 and 3 ngrams, respectively, from the Top 10 most occurring ngrams in myster.txt. Because Spanish had more of the frequently occurring ngrams, and because I believe the peaks of the mystery histogram line up better with Spanish than Portuguese, I concluded that the mystery file contains Spanish text.

Below, you may see the histogram and Top 10 ngram outputs of the mystery file.



Mystery.txt:

[(' de', 123), ('de ', 102), ('el ', 66), ('os ', 56), (' co', 56), ('as ', 53), (' la', 49), ('do ', 49), (' el', 45), ('ia ', 44)]