

In this project, our goal was to analyze the use of the word break in respect to people, before and after world war one. We pulled three books written before and three books written after world war one. We looked into the use of the word “break” and other synonyms in relation to humans, or pronouns, in our case. Our goal was to analyze the change of frequency of the word “break” before and after the catastrophic world event.

Our code works by breaking apart the text into sentences. In our code, we wrote two lists: one of made up of pronouns, the other made up of synonyms of the word “break”.

The library that we used was Project Gutenberg. We changed the names of the books in our file names so anyone that wishes to run the code must either change their file names or change the code. We used lists and dictionaries. We created a list of all the sentences in the book and then looked if the instances of two of our lists occurred at any point within a sentence.

We found that our initial hypothesis was incorrect, at least in the way that we evaluated. There is some room for error as the principle books that we evaluated were significantly longer pre-1913. Also, we may have added too many synonyms and lost sight of our initial inquiry. Our program output the total number of sentences containing our target phrase. {'Once a Greech': 0, 'War and Peace': 246, 'Metamorphosis': 16, 'Heart of Darkness': 25, 'Siddhartha': 16, 'Les Miserables': 574}. This list (actually dictionary) is deceptive as Les Miserables is nearly 1500 pages where as Metamorphosis is only 100. All the same, Siddhartha has far fewer examples than I was expecting at a full 200+ pages whereas Heart of Darkness which is shorter has 25.

We had a great success at taking our split sentences and searching within them for the key areas of interest. We could have improved on our allotted time (we were both extremely busy) and on the depth to which we went. We ran out of time and didn't get to our visual representation of the data although we might have if we had more time. For our skill level, this was a good introductory project as both of us are beginners. A glorified “control f” was really all that we were hoping for. We did unit test at the beginning inserting sample sentences into the sentence parser before we changed the code to enter in entire books. It would have probably been better practice to separate those two functions.