Chart, histogram

Description automatically generatedChapter 2

4. They all stay relatively the same but the gap between ‘men’ and ‘women’ gets tighter. People also has a peak.

5. Text

Description automatically generated

Nouns usually have more part meronyms than anything else.

7.

A picture containing text, plaque

Description automatically generated

However is almost always followed by a comma and usually at the start of a sentence or after a pause.

9. The word ‘monstrous’ in Sense and Sensibility and in Moby Dick have two very different meanings: in Sense and Sensibility, the word is more used with ‘very’ and ‘heartily’ where in Moby Dick, the word is more used with “Christian’ and ‘abundant’. Text

Description automatically generated

12. 123455 distinct words, 7.4853% of the words have more than one pronunciation

17. Graphical user interface, text, application

Description automatically generated

Moby dick: ['whale', 'one', 'like', 'upon', 'man', 'ship', 'Ahab', 'ye', 'old', 'sea', 'would', 'head', 'though', 'boat', 'time', 'long', 'said', 'yet', 'still', 'great', 'two', 'seemed', 'must', 'Whale', 'last', 'way', 'Stubb', 'see', 'Queequeg', 'little', 'round', 'whales', 'say', 'three', 'men', 'thou', 'may', 'us', 'every', 'much', 'could', 'Captain', 'first', 'side', 'hand', 'ever', 'Starbuck', 'never', 'good', 'white']

18. Sense and Sensibility:

[('Colonel', 'Brandon'), ('Sir', 'John'), ('Lady', 'Middleton'), ('every', 'thing'), ('Miss', 'Dashwood'), ('said', 'Elinor'), ('John', 'Dashwood'), ('dare', 'say'), ('every', 'body'), ('thousand', 'pounds'), ('every', 'day'), ('said', 'Marianne'), ('said', 'Mrs'), ('Miss', 'Steeles'), ('Miss', 'Steele'), ('replied', 'Elinor'), ('Elinor', 'could'), ('young', 'man'), ('cried', 'Marianne'), ('great', 'deal'), ('Miss', 'Dashwoods'), ('Miss', 'Marianne'), ('next', 'morning'), ('next', 'day'), ('young', 'ladies'), ('Berkeley', 'Street'), ('Harley', 'Street'), ('soon', 'afterwards'), ('Miss', 'Morton'), ('almost', 'every'), ('two', 'thousand'), ('would', 'give'), ('Edward', 'Ferrars'), ('would', 'never'), ('took', 'place'), ('could', 'hardly'), ('Barton', 'Park'), ('ever', 'since'), ('low', 'voice'), ('cried', 'Mrs'), ('Combe', 'Magna'), ('said', 'Lucy'), ('short', 'time'), ('one', 'day'), ('could', 'never'), ('could', 'make'), ('young', 'men'), ('went', 'away'), ('Lucy', 'Steele')]

23. Graphical user interface, chart

Description automatically generated with medium confidenceGraphical user interface

Description automatically generated

Yes, we can confirm Zipf’s Law for the written texts as the graph clearly shows the exponential ness of the function. At the left end, the points are far apart and at the end points are closer together.

For the random text, this isn’t shown as clearly but the law sort of still occurs,.

27. 1.766

Chapter 3

20. Bloomberg - Are you a robot? Bloomberg Need help?

from Bloomberg

22. “\b\w\*\b” for matching words or “\w\*\.” For matching sentences

Chapter 6

4. contains(outstanding) = True

contains(seagal) = True

contains(mulan) = True

contains(wonderfully) = True

contains(damon) = True

contains(awful) = True

contains(poorly) = True

contains(flynt) = True

contains(lame) = True

contains(wasted) = True

contains(ridiculous) = True

contains(waste) = True

contains(era) = True

contains(worst) = True

contains(bland) = True

contains(allows) = True

contains(portrayed) = True

contains(dull) = True

contains(unfunny) = True

contains(mess) = True

contains(fantastic) = True

contains(stupid) = True

contains(portrayal) = True

contains(memorable) = True

contains(pointless) = True

contains(badly) = True

contains(jedi) = True

contains(laughable) = True

contains(terrific) = True

contains(boring) = True

A lot of these words represent something very specific that would used in a more specific context (barely any “stopwords” or very common words). Some of these include jedi, flynt, damon, mulan. I didn’t really find anything that surprising because a lot of the words are unique to the case that they are used in.