1. [10 points] From the SnapshotBROWN.pos.all.txt file extract all

word types and their frequencies. Sort the list of word types in

decreasing order based on their frequency. Draw a chart showing the

relationship between the rank in the ordered list and the frequency

(Zipf's Law). Do not stem but do ignore punctuation.

Done. The chart could be seen after ran the program.

-----------------------------------------------------------------

2. [20 points] Generate a Bigram Grammar from the above file. Perform

add-one smoothing. Show the grammar before and after smoothing for

the sentence "A similar resolution passed in the Senate".

Done. Bigram algorithm was based on what we learned in class.

Briefly speaking:  

The probability was calculated Bigram: P(w1n) ≈ Π P(wk|wk-1)

before add-one, the probabilty is 9.193905443730285E-13

after add-one, the probabilty is 4.155756647706474E-21