Bigram Probabilities:

An automatic speech recognition system has provided two written sentences as possible interpretations to a speech input.

S1: The chief executive said that the company's profit was going down last year.

S2: The president said the revenue was good last year.

Using the bigram language model trained on Corpus (provided as Addendum to this homework), find out which of the two sentences is more probable. Compute the probability of each of the two sentences under the three following scenarios:

i. Use the bigram model without smoothing.
ii. Use the bigram model with add-one smoothing
iii. Use the bigram model with Good-Turing discounting.

Write a computer program to:

A. Compute the bigram counts for any given input. Apply your program to compute the bigram counts on Corpus.

B. For each of the three scenarios, construct the tables with the bigram counts for the two sentences above.

C. For each of the three scenarios, construct the table with the bigram probabilities for the sentences.

D. For each of the three scenarios, compute the total probabilities for each sentence S1 and S2.

Due Feb 8th at 2:30pm.

Please submit the source code and results via elearing.