Homework #2

Due February 23, 2015 before the class

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

S1: The company chairman said he will increase the profit next year .
S2: The president said he believes the last year profit were good .

Using the bigram language model trained on Corpus A provided below, find out which of the two sentences is more probable. Compute the probability of each of the two sentences under the three following scenarios:

a) Use the bigram model without smoothing.
b) Use the bigram model with add-one smoothing
c) Use the bigram model with Good-Turing discounting.

Assignment:

1. Write a program to compute the bigrams for any given input.
Apply your program to compute the bigrams you need for this homework.
2. Construct the tables with the bigram counts for the two sentences above.

For each of the three scenarios:

3. Construct the table with the bigram probabilities for the sentences.
4. Compute the total probabilities for each sentence S1 and S2.

[Here is the corpus for this assignment.](#)