Homework 1

PART I:

Exercise 3.4 in Chapter 3:

We are given the following corpus, modified from the one in the chapter:

<s> I am Sam </s>

<s> Sam I am </s>

<s> I am Sam </s>

<s> I do not like green eggs and Sam </s>

Using a bigram language model with add-one smoothing, what is P(Sam |am)? Include <s> and </s> in your counts just like any other token.

Ans: V = {<s>, I, am, Sam, </s>, do, not, like, green, eggs, and}

|V| = 11

P(Sam | am) =