2. ***Q: Estimate (just by eyeballing) the proportion of the word types that occurred only once in this corpus. Do you think the proportion of words that occur only once would be higher or lower if we used a larger corpus (e.g., all 57000 sentences in Brown)? Why or why not?***
It is 80 percent of word types that occurred only once in this corpus just by eyeballing. I think the proportion would be lower if we used a larger corpus, because the number of word types was going to increase while the number of words was going to increase by a lot, and the ratio of the increase of words would be much bigger than the ratio of the increase of word types. Therefore, the proportion would be lower.

4. ***Q: Why did all four probabilities go down in the smoothed model?*** **Now note that the probabilities did not all decrease by the same amount. In particular, the two probabilities conditioned on 'the' dropped only slightly, while the other two probabilities (conditioned on 'all' and 'anonymous') dropped rather dramatically.*****Q: Why did add-δδ smoothing cause probabilities conditioned on 'the' to fall much less than these others? And why is this behavior (causing probabilities conditioned on 'the' to fall less than the others) a good thing?***
Because smoothing as a technique made the distribution "smoother" by decreasing the frequencies of "spikes" in the distribution and manually increasing frequency to those rarely occurred in the context and had very low frequencies. Because the total population of words behind "all" and "anonymous" was relatively small and by applying this technique of smoothing, the frequency of "spikes" became much lower than before. This is good because not given so large data sets, we cannot put much faith into the distributions generated accordingly under the model.

5. ***Q: Which model performed worst and why might you have expected that model to have performed worst?*** **Note the large difference in perplexities between the MLE and smoothed bigram models.** ***Q: Did smoothing help or hurt the model's 'performance' when evaluated on this corpus? Why might that be?***
Unigram model performed worst because it only cared about the occurrence rate of each word in the corpus and never cared about the relationships among words and thus it was a naive model. Smoothing helped model's performance. If we only compared the perplexity of the sentences before and after "smoothing", we might conclude "”smoothing" hurted the the model's performance as it increased the perplexity; however given so small test data we had, we might have run into a situation called "overfitting". The test data set had phrases that used in the training data set and our model yielded a better performance than it should have had when tested in a random test set.