4. Comparing perplexities. The naıve case:
Is Othello a good model for Quijote? If you train on Othello and test on Don Quijote, what is the value of perplexity there? What is the perplexity if you train on Othello and test on the Interviews from the 1970s? Lastly, If you train on Othello, what is the perplexity of The Two Gentlemen of Verona? Why do you think you get the perplexities you get?

The model for Othello is not a good model for Quijote. Unfortunately due to the volatile nature of the texts given, it was hard to compare any of the perplexities since most probabilities ended up reverting to 0. Therefore I made any word not included in the corpus to equal to a very small number. However the perplexity is relatively low for both text, just as they should be given the information on the slide. It is safe to say that using a different models for different texts we would get a very low perplexity. The reason we get these perplexities is because given a certain model, the perplexity will be very very close to 0 because the language is completely different and basing the branching factor of one text given a completely different text would significantly lower and probabilities of the next word.

5. Comparing perplexities with Smoothing (10 pts.)
What is the value of the perplexity when trained on Othello and tested on the same files as above, but without smoothing. Why do you think this happens? What could improve this?

For my model we still get a perplexity lower than 0 or very small for every occurance. This is because without smoothing, the probabilities become less significant and represent a less accurate model of probability. This could be improved by taking any probability that may end up being 0 and replacing it with something very small which I ended up doing above. This gives potential numbers of actual significance, however it is very hard to generate probabilities and perplexities with an incomplete model.