

1. Take the data you tokenized (your language data) in Assignment-1 and build 4 language model as the following:
 - a. Unigram Model
 - b. Bigram Model
 - c. Trigram Model
 - d. Quadrigram Model

Run the code on the full data, not on a sample of sentences.

2. Create the 3 following smoothing techniques:
 - a. Add one Smoothing
 - b. Add K Smoothing
 - c. Add Token Type Smoothing (May not be a probability distribution)
3. Take 1000 sentences from randomly chosen news articles from your language and apply the smoothed language models to compute the probability of those sentences.