1. Take the dataset brown_nouns.txt in Assignment-2. Store all the words in a trie. Implement both prefix and suffix tries. Identify the stem and the suffix for each word stored in the trie. Analyze and find out which type of trie does better in stemming. Your output should look like the following:
   goes=go+es
   kites=kite+s
   [Hint: The node where maximum branching happens can be taken as a suffix and the rest part can be taken as a stem. Try to include a frequency or probability measure in this.]
2. Take the dataset that you tokenized in Assignment-1. Create a frequency distribution on the tokenized dataset where each key represents the word and the corresponding frequency is the value. Do not use any predefined library for this. Plot the most frequent 100 words showing the frequency distribution. From the frequency distribution, identify stop words and remove them. Plot a similar graph for three different thresholds after stop words removal. [Hint: Use a frequency threshold to find the stop words.]