# E1 246 Assignment 1

**Shankararama Sharma R**
**Reg no. 14044**

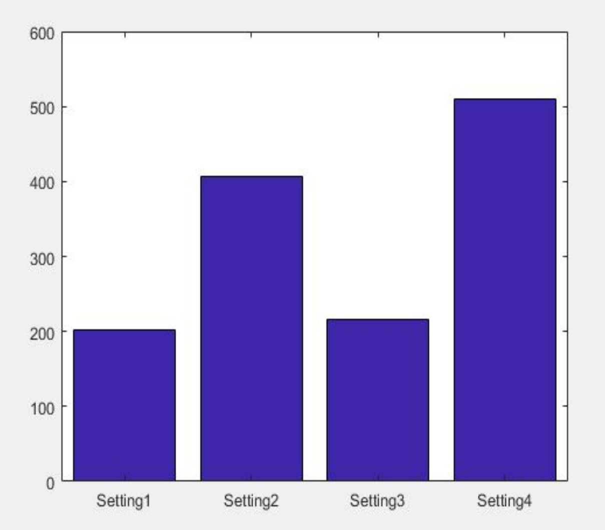
## 1 Task 1

For this task the Language Model that I've used is the Trigram model with stupid backoff. That is, if there is no Trigram available, I look for a bigram. If there is no Bigram, then I go with Unigram. The dat is divided as 80% for training and 20% for testing for both corpora. All probabilities used are log base 10.

**SETTING1:**
Trained and tested on Brown Corpus.
Avg Perplexity = 201.96
Sample Sentence: <s> That guiding principle of divine justice in the liquid helium . </s>
**SETTING2**
Trained and tested on Gutenberg Corpus.
Avg Perplexity = 406.96
<s> I long to Zin , and what her godmother , who would take more expeditiously the dimensions of a little knew how to make it a pair of me . </s>
**SETTING3**
Trained on Brown and Gutenberg and tested on Brown Corpus.
Avg Perplexity = 215.44
<s> " But I say , Who in hell are we unto you , and he made an end in effect of its piping system in the subject of Mrs . - ] </s>
**SETTING4**
Trained on Brown and Gutenberg and tested on Gutenberg Corpus.
Avg Perplexity = 509.74
<s> When Franklin brought the ass .</s>

Setting1 has the best perplexity score. This might be be because the length of sentences in the Brown corpus are relatively much shorter than Gutenberg. The Gutenberg conatins a lot of old English and hence the sentences are quite long. We see that even when we train on both the sets, and test on Brown (i.e. Setting 3), the perplexity is quite good. This is again due to the same reason. The sentences in Brown are smaller and hence have higher probability.

Figure 1: Perplexities of Different Settings



## 2 Task 2

The sentences generated by the model trained in Setting1 are the best and closer to natural language. Some examples are:

- But Mercer hung on , adapt decisions and their subsequent reactions to that of Rousseau .
- It would be greatest .
- They graduated together from all over and learn than to the flat-bottomed boats and rowed up the legal principle of state automobile practices

- 8 ) and continue strong long after education and social , was the real estate salesman , Kern said this Sunday's sessions – including several self-portraits – were stringy , contorted and strangely pathetic .