Assignment 01 Language Model

Chaikesh Chouragade
SR -14358 https://github.com/chaikesh/Ngram_model

1 Perplexity Measure:

Training dataset split is as follows

Tr-G = train on gutenberg dataset (80%)
Br-Train= brown dataset (80%)
Gu+Br-Train= gutenberg and brown dataset (80%)

Test dataset split is as follows:-

1.1 Add-K smoothing

parameter k=0.01

– – –	uni-gram	bi-gram
Br-Train & **Br-Test**	274	141
Gu-Train & **Gu-Test**	339	261
Gu+Br-Train & **Br-Test**	299	142
Gu+Br-Train & **Gu-Test**	335	254

1.2 kneser-Nay smoothing

discunting factor d=0.75

—	bi-gram
Tr-B and **Ts-B**	92
Tr-G and **Ts-G**	183
Tr-GB and **Ts-B**	159
Tr-GB and **Ts-G**	101

2 Sentence Generation :

2.1 add-K smoothing

Example- brown dataset he was a good deal of thought and habits of the united states , and the other hand , the first time in the first two years ago , the " public " buys at the same time , and he had

Example- Gutenberg dataset
he was a great deal of the lord , and the lord . not a word , and he said , " i am sure i should have been a great many more people , and i will not be afraid

Example- Gutenberg dataset + brown dataset he was a very good - natured , and the lord , and he said , " i am sure i should have been a great deal of the lord . not a word , and i will not be afraid

[here we can see UNK tag appear due to sparsity in brown dataset]

2.2 kneser-Nay smoothing

Example- brown dataset
he was a few years ago , and the first , the same time , but the " the most of the united states , a new york , he had been a little more than the new england , in the other hand , as a man , " i

Example- Gutenberg dataset

he was a man , and the lord , that he had been a little joe otter , the king of the children of his own , i will be a great deal of israel , which is the son of a very much as the house of her , but

Example- Gutenberg dataset + brown dataset
i have been a UNK , and the lord , and he said , " i am sure i should be the lord . not been in the land of egypt , and i will not be a great deal of the lord god of israel , and to the lord