

**Output:**

Total Word Types in Test Data: 1248

Total Word Types NOT in Training: 49

Percentage of Word Types Not in Training: 3.926282051282051 %

Total Tokens in Test: 2769

Total Tokens NOT in Training: 50

Percentage of Tokens Not in Training: 1.8057060310581439 %

There are 39502 unique words in the training corpus.

There are 2221290 total tokens in the training corpus

Total Bigram Types in Test Data: 2359

Total Bigram Types NOT in Training: 616

Percentage of Bigram Types Not in Training: 26.112759643916917 %

Total Bigram Tokens in Test: 2769

Total Bigram Tokens NOT in Training: 618

Percentage of Bigram Tokens Not in Training: 22.318526543878654 %

Unigram MLE Log Probability of Given Sentence:

i : -8.400191146154114

look : -11.982303986215943

forward : -12.375607193140258

to : -5.540690135458277

hearing : -13.505537497185392

your : -10.953683308276174

reply : -17.623534706583843

. : -4.811868102497583

</s> : -4.62532894422938

Total: -89.81874501974097

Bigram MLE Log Probability of Given Sentence:

<s> i : -5.631088893700847

i look : -8.875420257009424

look forward : -4.345774836841731

forward to : -2.2644156362321546

to hearing : -13.2203480948755

hearing your : -inf

your reply : -inf

reply . : -inf

. </s> : -0.0848872391724313

Total: -inf

Bigram Add-One Smoothed Log Probability of Given Sentence:

<s> i : -6.1552832892341485

i look : -11.584899882389982

look forward : -10.482231743023195

forward to : -8.825428620790015

to hearing : -13.82744195061677

hearing your : -15.276633330886703

your reply : -15.31000916584291

reply . : -15.270076277195477

. </s> : -0.6693560071854955

Total: -97.4013602671647

Unigram MLE Perplexity of Given Sentence:

1009.8046830192555

Bigram MLE Perplexity of Given Sentence:

inf

Bigram Add-One Smoothed Perplexity of Given Sentence:

855.2106605432591

Unigram MLE Perplexity of Test Data:

1082.014772901339

Bigram MLE Perplexity of Test Data:

inf

Bigram Add-One Smoothed Perplexity of Test Data:

1807.233786592602