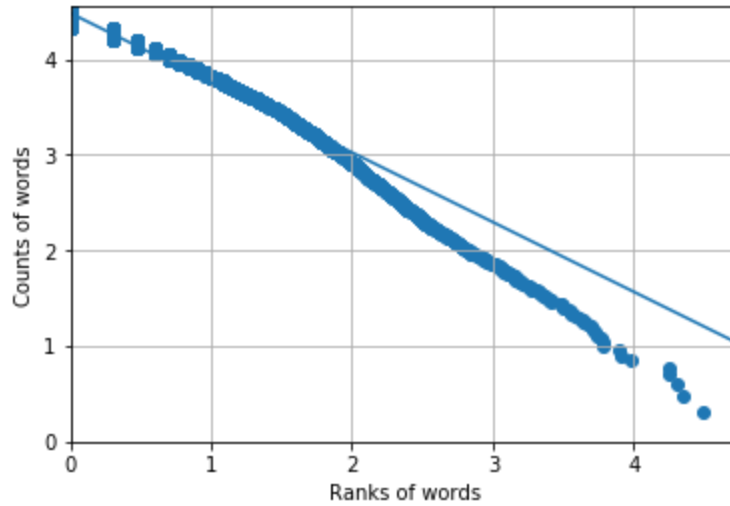


# Report for Assignment 1

## Zipf's law verification for unigrams



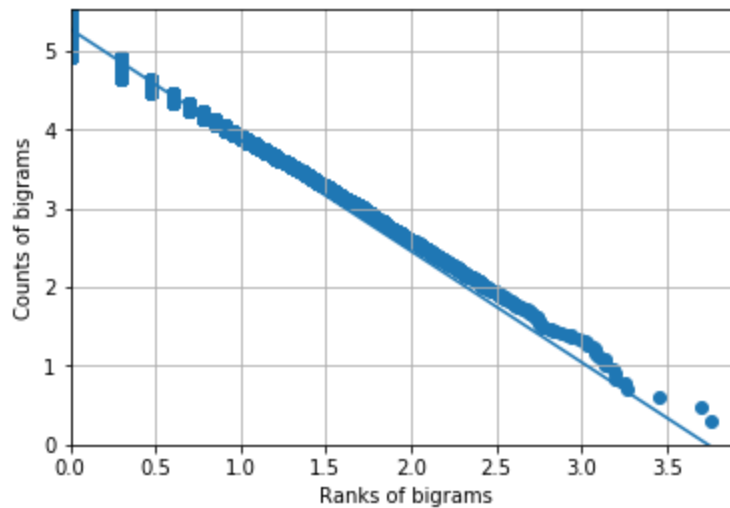
Slope and Intercept for log-log plot for zipfs law in unigrams

-0.7294188985606873 4.485539390443362

Top 10 Unigrams with count

[('the', 56448), ('of', 31276), ('and', 22091), ('to', 20335), ('in', 17696), ('a', 17641), ('is', 9474), ('that', 8240), ('for', 7787), ('it', 6051)]

## Zipf's law verification for bigrams



Slope and Intercept for log-log plot for zipfs law in bigrams  
-1.4104530417477712 5.2802626901581

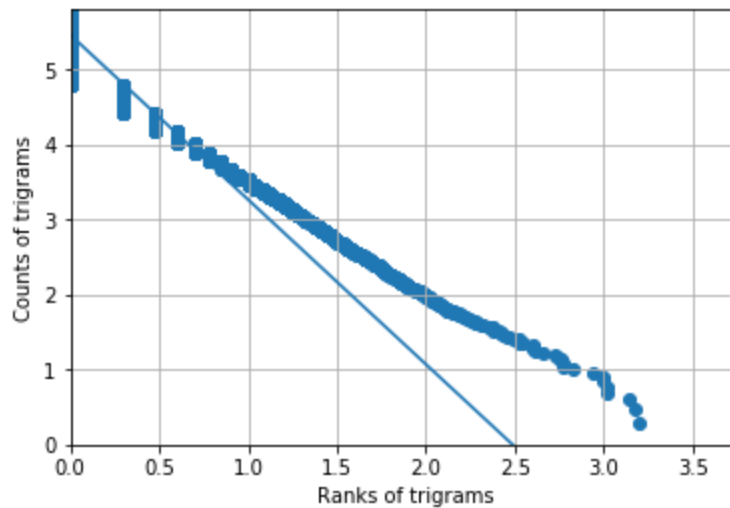
Top 10 Bigrams with count when None allowed

[('of the', 8516), ('None the', 5820), ('in the', 4992), ('to the', 2825),  
('and the', 1864), ('on the', 1824), ('for the', 1592), ('None in', 1589),  
('None it', 1519), ('it is', 1390)]

Top 10 Bigrams with count when None not allowed

[('of the', 8516), ('in the', 4992), ('to the', 2825), ('and the', 1864),  
('on the', 1824), ('for the', 1592), ('it is', 1390), ('to be', 1373),  
('with the', 1262), ('that the', 1244)]

### Zipf's law verification for Trigrams



Slope and Intercept for log-log plot for zipfs law in trigrams  
-2.1975922562310486 5.467006145296135

Top 10 Trigrams with count when None allowed

[('None None the', 5820), ('None None in', 1589),  
('None None it', 1519), ('None None he', 1377),  
('None None this', 1054), ('None None but', 1038),  
('None None None', 998), ('None None a', 980),  
('None None and', 867), ('None None i', 675)]

Top 10 Trigrams with count when None not allowed

[('one of the', 337), ('the united states', 333), ('as well as', 225),  
('some of the', 156), ('the fact that', 154), ('part of the', 131),  
('of the united', 128), ('it is not', 126), ('a number of', 119), ('there is a', 118)]

# N-Gram Model Performance without Smoothing

## Unigram Model Results

Sentence No	log-likelihood	perplexity
1	-33	678
2	-24	393
3	-23	332
4	-27	883
5	-25	462

## Bigram Model Results

Sentence No	log-likelihood	perplexity
1	-27	86
2	-22	80
3	-25	142
4	-inf	inf
5	-21	64

## Trigram Model Results

Sentence No	log-likelihood	perplexity
1	-inf	inf
2	-inf	inf
3	-inf	inf
4	-inf	inf
5	-16	14

# Additive Smoothing Results

## Additive Smoothing with $k=1$

Unigram test with Additive Smoothing

The sequence he lived a good life has unigram log-likelihood of -32.80237583  
The sequence the man was happy has unigram log-likelihood of -24.06386777  
The sequence the person was good has unigram log-likelihood of -23.39853251  
The sequence the girl was sad has unigram log-likelihood of -27.24934479  
The sequence he won the war has unigram log-likelihood of -24.70738218

Bigram test with Additive Smoothing

The sequence he lived a good life has bigram log-likelihood of -43.66615774  
The sequence the man was happy has bigram log-likelihood of -34.8631562  
The sequence the person was good has bigram log-likelihood of -36.54954686  
The sequence the girl was sad has bigram log-likelihood of -40.5430167  
The sequence he won the war has bigram log-likelihood of -34.35564902

Trigram test with Additive Smoothing

The sequence he lived a good life has trigram log-likelihood of -60.0475751  
The sequence the man was happy has trigram log-likelihood of -48.42203672  
The sequence the person was good has trigram log-likelihood of -49.33189563  
The sequence the girl was sad has trigram log-likelihood of -52.35097309  
The sequence he won the war has trigram log-likelihood of -47.3550479

Unigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 706.6073710835825  
The sequence the man was happy has perplexity of 409.92201814722625  
The sequence the person was good has perplexity of 347.10701343216573  
The sequence the girl was sad has perplexity of 908.9919125530047  
The sequence he won the war has perplexity of 481.47078001809484

Bigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 1447.6443920650272  
The sequence the man was happy has perplexity of 1067.0266629370701  
The sequence the person was good has perplexity of 1495.0416918792678  
The sequence the girl was sad has perplexity of 3228.92430907225  
The sequence he won the war has perplexity of 964.0371247036445

Trigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 176287178.9389828  
The sequence the man was happy has perplexity of 1123951.6926896814  
The sequence the person was good has perplexity of 6740498.8655875465  
The sequence the girl was sad has perplexity of 3363921.301414764  
The sequence he won the war has perplexity of 2375816.85352238

## Additive Smoothing with $k=0.1$

### Unigram test with Additive Smoothing

The sequence he lived a good life has unigram log-likelihood of -32.61509462  
The sequence the man was happy has unigram log-likelihood of -23.9164008  
The sequence the person was good has unigram log-likelihood of -23.24210663  
The sequence the girl was sad has unigram log-likelihood of -27.14544629  
The sequence he won the war has unigram log-likelihood of -24.55937915

### Bigram test with Additive Smoothing

The sequence he lived a good life has bigram log-likelihood of -35.36645242  
The sequence the man was happy has bigram log-likelihood of -28.43177263  
The sequence the person was good has bigram log-likelihood of -30.52678498  
The sequence the girl was sad has bigram log-likelihood of -36.49939055  
The sequence he won the war has bigram log-likelihood of -27.77381152

### Trigram test with Additive Smoothing

The sequence he lived a good life has trigram log-likelihood of -53.81814805  
The sequence the man was happy has trigram log-likelihood of -42.15146424  
The sequence the person was good has trigram log-likelihood of -43.58637058  
The sequence the girl was sad has trigram log-likelihood of -47.16902795  
The sequence he won the war has trigram log-likelihood of -37.50198373

### Unigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 680.6300575534524  
The sequence the man was happy has perplexity of 395.08471059428706  
The sequence the person was good has perplexity of 333.79487457158746  
The sequence the girl was sad has perplexity of 885.685191506742  
The sequence he won the war has perplexity of 463.98154949367876

### Bigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 363.0021427064168  
The sequence the man was happy has perplexity of 294.8169117247738  
The sequence the person was good has perplexity of 448.25263693523505  
The sequence the girl was sad has perplexity of 1242.4209176173451  
The sequence he won the war has perplexity of 258.4655210587587

### Trigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 104348786.40802842  
The sequence the man was happy has perplexity of 136772.3848259934  
The sequence the person was good has perplexity of 6942960.366513984  
The sequence the girl was sad has perplexity of 619558.4590200821  
The sequence he won the war has perplexity of 184610.00331814899

## Additive Smoothing with $k=0.01$

### Unigram test with Additive Smoothing

The sequence he lived a good life has unigram log-likelihood of -32.59590537  
The sequence the man was happy has unigram log-likelihood of -23.90128863  
The sequence the person was good has unigram log-likelihood of -23.22608705  
The sequence the girl was sad has unigram log-likelihood of -27.13482949  
The sequence he won the war has unigram log-likelihood of -24.54421099

### Bigram test with Additive Smoothing

The sequence he lived a good life has bigram log-likelihood of -29.71040845  
The sequence the man was happy has bigram log-likelihood of -24.39550934  
The sequence the person was good has bigram log-likelihood of -26.60998627  
The sequence the girl was sad has bigram log-likelihood of -33.98835464  
The sequence he won the war has bigram log-likelihood of -23.42543281

### Trigram test with Additive Smoothing

The sequence he lived a good life has trigram log-likelihood of -48.83736245  
The sequence the man was happy has trigram log-likelihood of -37.36084261  
The sequence the person was good has trigram log-likelihood of -38.76175501  
The sequence the girl was sad has trigram log-likelihood of -42.31297364  
The sequence he won the war has trigram log-likelihood of -27.84595823

### Unigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 678.0229075237785  
The sequence the man was happy has perplexity of 393.59487998587684  
The sequence the person was good has perplexity of 332.4607350041083  
The sequence the girl was sad has perplexity of 883.3375230213098  
The sequence he won the war has perplexity of 462.2254445877776

### Bigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 141.42007860737655  
The sequence the man was happy has perplexity of 131.51249518093405  
The sequence the person was good has perplexity of 204.79249649830578  
The sequence the girl was sad has perplexity of 587.2946939676733  
The sequence he won the war has perplexity of 108.31964832774844

### Trigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 73190751.97588311  
The sequence the man was happy has perplexity of 18819.311702928888  
The sequence the person was good has perplexity of 9073895.750341415  
The sequence the girl was sad has perplexity of 76004.31397580953  
The sequence he won the war has perplexity of 6785.374524733131

## Additive Smoothing with $k=0.001$

### Unigram test with Additive Smoothing

The sequence he lived a good life has unigram log-likelihood of -32.59398169  
The sequence the man was happy has unigram log-likelihood of -23.89977364  
The sequence the person was good has unigram log-likelihood of -23.22448121  
The sequence the girl was sad has unigram log-likelihood of -27.13376548  
The sequence he won the war has unigram log-likelihood of -24.54269038

### Bigram test with Additive Smoothing

The sequence he lived a good life has bigram log-likelihood of -27.32027061  
The sequence the man was happy has bigram log-likelihood of -22.48238605  
The sequence the person was good has bigram log-likelihood of -25.08750661  
The sequence the girl was sad has bigram log-likelihood of -33.03525774  
The sequence he won the war has bigram log-likelihood of -21.35969732

### Trigram test with Additive Smoothing

The sequence he lived a good life has trigram log-likelihood of -46.78854666  
The sequence the man was happy has trigram log-likelihood of -34.29235141  
The sequence the person was good has trigram log-likelihood of -35.26552304  
The sequence the girl was sad has trigram log-likelihood of -38.04641712  
The sequence he won the war has trigram log-likelihood of -20.77957714

### Unigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 677.7620975281864  
The sequence the man was happy has perplexity of 393.4458353233881  
The sequence the person was good has perplexity of 332.32729166813056  
The sequence the girl was sad has perplexity of 883.1025843514434  
The sequence he won the war has perplexity of 462.0497616471592

### Bigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 94.95265843295174  
The sequence the man was happy has perplexity of 89.70057769780212  
The sequence the person was good has perplexity of 151.03344793414027  
The sequence the girl was sad has perplexity of 440.076225130505  
The sequence he won the war has perplexity of 71.66048383987265

### Trigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 230385332.86090907  
The sequence the man was happy has perplexity of 8908.654638105856  
The sequence the person was good has perplexity of 38797135.71155561  
The sequence the girl was sad has perplexity of 16099.383918696338  
The sequence he won the war has perplexity of 264.44860305316354

## Additive Smoothing with $k=0.0001$

### Unigram test with Additive Smoothing

The sequence he lived a good life has unigram log-likelihood of -32.59378927  
The sequence the man was happy has unigram log-likelihood of -23.8996221  
The sequence the person was good has unigram log-likelihood of -23.22432058  
The sequence the girl was sad has unigram log-likelihood of -27.13365906  
The sequence he won the war has unigram log-likelihood of -24.54253828

### Bigram test with Additive Smoothing

The sequence he lived a good life has bigram log-likelihood of -26.78768533  
The sequence the man was happy has bigram log-likelihood of -22.00304228  
The sequence the person was good has bigram log-likelihood of -24.82092856  
The sequence the girl was sad has bigram log-likelihood of -34.11102151  
The sequence he won the war has bigram log-likelihood of -20.87998241

### Trigram test with Additive Smoothing

The sequence he lived a good life has trigram log-likelihood of -47.64450431  
The sequence the man was happy has trigram log-likelihood of -34.40700904  
The sequence the person was good has trigram log-likelihood of -34.71683545  
The sequence the girl was sad has trigram log-likelihood of -35.5428909  
The sequence he won the war has trigram log-likelihood of -17.17786012

### Unigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 677.7360155778854  
The sequence the man was happy has perplexity of 393.4309302405119  
The sequence the person was good has perplexity of 332.31394704061734  
The sequence the girl was sad has perplexity of 883.0790887643913  
The sequence he won the war has perplexity of 462.0321926279583

### Bigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 86.88750379417478  
The sequence the man was happy has perplexity of 81.50044307382983  
The sequence the person was good has perplexity of 143.19190322983454  
The sequence the girl was sad has perplexity of 538.2083762927297  
The sequence he won the war has perplexity of 65.10468302260901

### Trigram Perplexity with Additive Smoothing

The sequence he lived a good life has perplexity of 4216239232.6492457  
The sequence the man was happy has perplexity of 18092.30235000095  
The sequence the person was good has perplexity of 750166528.6054481  
The sequence the girl was sad has perplexity of 14137.194198797833  
The sequence he won the war has perplexity of 23.297434311347686



# Good Turing Smoothing Results

## Bigram test with Good Turing Smoothing

The sequence he lived a good life has bigram log-likelihood of -27.71743189  
The sequence the man was happy has bigram log-likelihood of -20.92721544  
The sequence the person was good has bigram log-likelihood of -24.83631352  
The sequence the girl was sad has bigram log-likelihood of -37.02050632  
The sequence he won the war has bigram log-likelihood of -20.06233468

## Bigram Perplexity with Good Turing Smoothing

The sequence he lived a good life has perplexity of 101.45060014277486  
The sequence the man was happy has perplexity of 65.72261538248301  
The sequence the person was good has perplexity of 143.63318233923422  
The sequence the girl was sad has perplexity of 961.5367459252897  
The sequence he won the war has perplexity of 55.283082341739174

## Trigram test with Good Turing Smoothing

The sequence he lived a good life has trigram log-likelihood of -58.07371875  
The sequence the man was happy has trigram log-likelihood of -40.27026893  
The sequence the person was good has trigram log-likelihood of -40.5898134  
The sequence the girl was sad has trigram log-likelihood of -29.0057064  
The sequence he won the war has trigram log-likelihood of -17.09546651

## Trigram Perplexity with Good Turing Smoothing

The sequence he lived a good life has perplexity of 3389328978366.3647  
The sequence the man was happy has perplexity of 92415.79846559363  
The sequence the person was good has perplexity of 412080362700.0125  
The sequence the girl was sad has perplexity of 89298.03158603702  
The sequence he won the war has perplexity of 8.695864839975235

**Good Turing Method can't be applied to Unigrams because of problem in rank value according to the formula.**

# Interpolation Bigram Results

Interpolation of Bigram Model with lambda value 0.2

The sequence he lived a good life has bigram log-likelihood of -32.40873916  
The sequence the man was happy has bigram log-likelihood of -26.24928962  
The sequence the person was good has bigram log-likelihood of -27.22715572  
The sequence the girl was sad has bigram log-likelihood of -30.10650982  
The sequence he won the war has bigram log-likelihood of -25.96759924

Interpolation of Bigram Model with lambda value 0.5

The sequence he lived a good life has bigram log-likelihood of -29.33361575  
The sequence the man was happy has bigram log-likelihood of -24.05734186  
The sequence the person was good has bigram log-likelihood of -25.93593035  
The sequence the girl was sad has bigram log-likelihood of -28.95662442  
The sequence he won the war has bigram log-likelihood of -23.32978181

Interpolation of Bigram Model with lambda value 0.8

The sequence he lived a good life has bigram log-likelihood of -27.57424179  
The sequence the man was happy has bigram log-likelihood of -22.66459485  
The sequence the person was good has bigram log-likelihood of -25.17318529  
The sequence the girl was sad has bigram log-likelihood of -28.83973905  
The sequence he won the war has bigram log-likelihood of -21.66946273

Bigram Perplexity with Interpolation of lambda value 0.2

The sequence he lived a good life has perplexity of 221.72913544511488  
The sequence the man was happy has perplexity of 190.53919540280435  
The sequence the person was good has perplexity of 231.69715296943434  
The sequence the girl was sad has perplexity of 412.1148052284036  
The sequence he won the war has perplexity of 180.10136788352423

Bigram Perplexity with Interpolation of lambda value 0.5

The sequence he lived a good life has perplexity of 132.8121811600024  
The sequence the man was happy has perplexity of 122.91196559965756  
The sequence the person was good has perplexity of 178.96425097058838  
The sequence the girl was sad has perplexity of 327.4465655561572  
The sequence he won the war has perplexity of 106.2671666232951

Bigram Perplexity with Interpolation of lambda value 0.8

The sequence he lived a good life has perplexity of 99.05814150633329  
The sequence the man was happy has perplexity of 93.02971603117257  
The sequence the person was good has perplexity of 153.643818834768  
The sequence the girl was sad has perplexity of 319.88060292420886  
The sequence he won the war has perplexity of 76.2404792740472