Question 4.3 answers:
Amogha Sekhar, A53301791, CSE 250A

a) MILLION 0.002072759168154815
MORE 0.0017088989966186725
MR. 0.0014416083492816956
MOST 0.0007879173033190295
MARKET 0.0007803712804681068
MAY 0.0007298973156289532
M. 0.0007034067394618568
MANY 0.0006967290595970209
MADE 0.0005598610827336895
MUCH 0.0005145971758110562
MAKE 0.0005144626437991272
MONTH 0.00044490959363187093
MONEY 0.00043710673693999306
MONTHS 0.0004057607781605526
MY 0.0004003183467688823
MONDAY 0.00038198530259784006
MAJOR 0.00037089252670515475
MILITARY 0.00035204581485220204
MEMBERS 0.00033606096579846475
MIGHT 0.00027358919153183117
MEETING 0.0002657374141083427
MUST 0.0002665079156312084
ME 0.00026357267173457725
MARCH 0.0002597935452176646
MAN 0.0002528834918776787
MS. 0.0002389900041002911
MINISTER 0.00023977273580605944
MAKING 0.00021170446604452378
MOVE 0.0002099555498894477
MILES 0.00020596851026319035

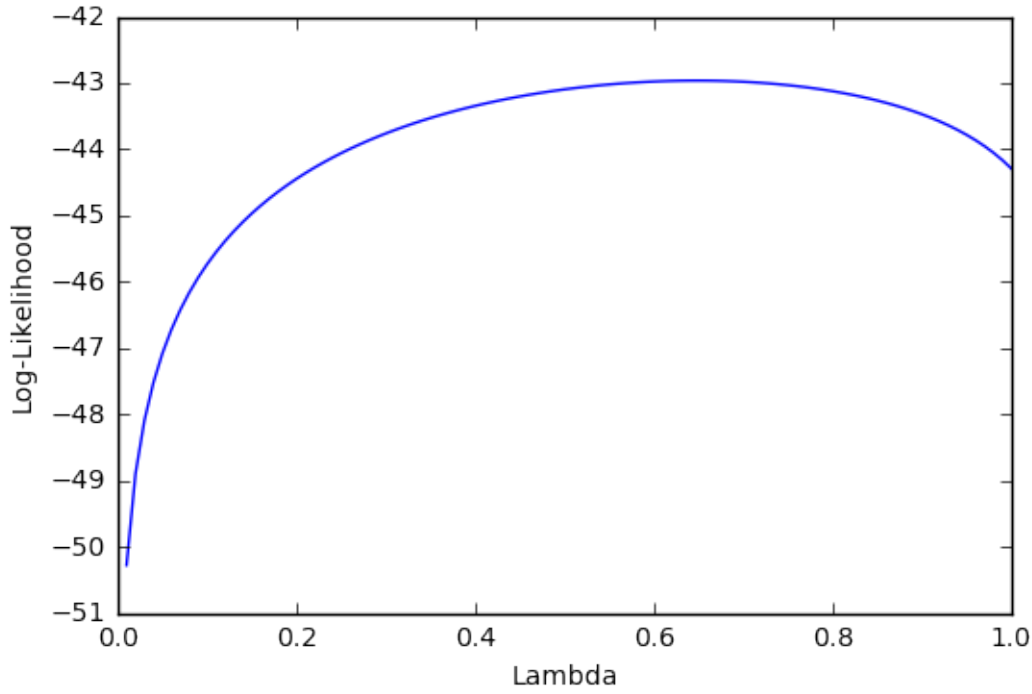
b) The ten most likely words to follow the word “THE”, along with their numerical bigram probabilities is as follows:

Word	**Numerical Bigram Probabilities**
UNK	0.615020
U.	0.013372
FIRST	0.011720
COMPANY	0.011659
NEW	0.009451
UNITED	0.008672
GOVERNMENT	0.006803
NINETEEN	0.006651
SAME	0.006287
TWO	0.006161

c) The unigram model yields a value of -64.50944034364878, whereas the bigram model yields a value of -40.91813213378977. Therefore, the bigram model yields a higher log-likelihood value .

d) The pair {Sixteen, Officials}, {Sold, Fire} is not observed in the training corpus. The result of this is that the probability becomes 0 and since log(0) is not defined, we get an undefined value.

e)



The optimal value of lambda is 0.65.