Question 1:
Word types: 41739

Question 2:
Word tokens: 2568210

Question 3:
percent of word tokens in test corpus that did not occur: 1.603346113628442%
percent of word types in test corpus that did not occur: 3.6028823058446755%

Question 4:
Percent of word tokens and word types in test corpus that did not occur in training: 25.69735006973501%
Percent of word tokens that do no appear in training bigrams: 20.95536959553696%

Question 5:
[i] : -8.450963962476674
[look] : -12.032588480668233
[forward] : -12.403588495460756
[to] : -5.597321004705777
[hearing] : -13.584972612278133
[your] : -11.043218291645285
[reply] : -17.591892026217923
[.] : -4.868854680279238
unigram total: -94.93878209357644

log prob of bigram:[<s> i] : -5.639534583824631
log prob of bigram:[i look] : -8.93447718627382
log prob of bigram:[look forward] : -4.172280422440442
log prob of bigram:[forward to] : -2.2448870591235344
log prob of bigram:[to hearing] : -13.110048238932082
log prob of:[hearing your] : undefined
log prob of:[your reply] : undefined
log prob of:[reply .] : undefined
log prob of bigram:[. </s>] : -0.08460143194821208
sum of all bigram probability: undefined

add one log prob: [<s> i] : -6.142052348726813
add one log prob: [i look] : -11.582788837823436
add one log prob: [look forward] : -10.240859462550432
add one log prob: [forward to] : -8.707188259410588
add one log prob: [to hearing] : -13.725046665121754
add one log prob: [hearing your] : -15.35631440692812
add one log prob: [your reply] : -15.390572037471506
add one log prob: [reply .] : -15.34955768662052
add one log prob: [. </s>] : -0.6451804614204727
sum of all add one log probability: -97.13956016607362

("hearing", "your", "your", "reply" and "reply", ".") are the only pair of words that have a 0 in the bigram training model.

Question 6:
unigram perplexity: 721.0113746656128
bigram perplexity: undefined
add one log perplexity: 839.83145676326

Question 7:
preplexity of test under unigram: 387.84871001426245
preplexity of test under bigram: undefined
preplexity of test under add-one bigram: 1.2842288569947

Every other method except bigram had a defined perplexity. Bigram was undefined because there were some pair of words which did not exist in the training set. For unigram and add-one bigram the preplexity was relatively low which means that the model was able to predict the words with high accuracy. The best performing model for my model was bigram with add-one smoothing.