Name:_____

(1) How many possible tag sequences can the sentence below have? Show your work.
The dog ate his homework.

Possible tags:

the – DT, NN
dog – DT, NN, V
ate – NN, V
his – PRO
homework – NN, V, NNS

Answer: 2*3*2*1*3

(2) **HMM taggers**: For the sentence "Karlsson lives on the roof" and the tag sequence "NN VB IN DT NN", what is the probability assigned to the sentence and the tag sequence by a trigram POS tagger? Note: assume START and STOP symbols.

Answer*: p("Karlsson lives on the roof", "NN VB IN DT NN")=*
p(NN|START)*p(VB|START NN)*p(IN|NN VB)*p(DT| VB IN)*p(NN| IN DT)*p(STOP|DT NN)*
p(Karlsson|NN)*p(lives|VB)*p(on|IN)*p(the|DT)*p(roof|NN)

(3) **Exercise 8**.1 (Jurafsky and Martin, Chapter 8, p. 25).

Answer:
1. Atlanta NNP
2. dinner NN
3. have VBP
4. can MD

(4) **Exercise 3**.2 (Jurafsky and Martin Chapter 3, p. 25). Also calculate the perplexity of the given sentence under each model.
Answer (see pp, 5 and 6 in the chapter for relevant unigram and bigram counts):
Unsmoothed bigram model:
p(i want chinese food)
=p(i|<s>)*p(want|i)*p(chinese|want)*p(food|chinese)*p(</s>|food)
=0.25*0.33*0.0065*0.52*0.68
=.000189

Bigram model with Add-one smoothing
Note: count(<s>,i) can be found from unsmoothed MLE probabilities:
0.25*count(<s>)=0.25*9332=2333
Similarly for count(food,</s>)=count(food)*0.68=1093*0.68=743
|V|=1446

p(i|<s>)=(count(<s>,i)+1)/((count(<s>)+1446)=(2333+1)/(9332+1446)=0.22
p(</s>|food)=(count(food, </s>)+1)/((count(food)+1446)=(743+1)/(1093+1446)=0.29

p(i want chinese food)
=0.22*0.21*0.0029*0.052*0.29
=0.000002

Perplexities:
Unsmoothed bigram model:
l=[math.log(0.25,2)+math.log(0.33,2)+math.log(0.0065,2)+math.log(0.52,2)+math.log(0.68,2)]/5
ppl=math.pow(2,-l)=5.552

Bigram model with Add-one smoothing:
=[math.log(0.22,2)+math.log(0.21,2)+math.log(0.0029,2)+math.log(0.052,2)+math.log(0.29,2)]/5
ppl=math.pow(2,-l)=13.77