Name:_____

Question 1: Write out the joint probability for the following sentence using the chain rule:

p(There, is, only, one, person, who, is, not, ordinary)

Write out the probability above using the second-order Markov assumption.

Question 2: Consider the following training corpus **T** of sentences:

- *START Karlsson is round STOP*
- *START He lives on the roof STOP*
- *START He is happy STOP*
- *START On the roof STOP*
- *START Karlsson lives happily STOP*

(b) Compute the following maximum likelihood parameters:

p(Karlsson|START)=

p(Karlsson|lives,happily)=

p(STOP|happy)=

(c) Compute the probability of the following sentences under the trigram model trained on **T**:

START Karlsson is happy STOP

START Karlsson lives on the roof STOP

Question 3: We have the following training corpus:

the green book STOP
my blue book STOP
his green house STOP
book STOP

Assume we have a language model based on this corpus using linear interpolation with $\lambda_i = 1/3$ for all i. Compute the value of the parameter p(book|the green) under this model. Assume STOP as part of your unigram model.

p(book|the green)