Name:_____

**(1) Question 1:** We have the following training corpus:
**(2)** the green book STOP
**(3)** my blue book STOP
**(4)** his green house STOP
**(5)** 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)

**Answer:**
total_words=14
p(book|the green)
=1/3*count(the,green,book)/count(the,green)+1/3*count(green,book)/count(green)+1/3*count(book)/total_words=1/3*1+1/3*1/2+1/3*3/14=1/3+1/6+1/14

**Question 2: Naïve Bayes**

Consider the task of classifying movie reviews using the Naïve Bayes algorithm. The features used are *bag-of-word* features. Assume the following likelihoods for each word being part of a positive or negative movie review, and equal prior probabilities for each class.

| | **pos** | **neg** |
|---|---|---|
| I | 0.09 | 0.16 |
| always | 0.07 | 0.06 |
| like | 0.29 | 0.06 |
| foreign | 0.04 | 0.15 |
| films | 0.08 | 0.11 |

What class will Naïve Bayes assign to the sentence "I always like foreign films"? Show your work.

**Answer:**

p(pos)*p(S|pos)=0.5*0.09*0.07*0.29*0.04*0.08

p(neg)*p(S|neg)=0.5*0.16*0.06*0.06*0.15*0.11

After simplifying the two products above, we conclude that p(neg)*p(S|neg) is greater than p(pos)*p(S|pos). Thus, the model predicts the class *negative* for this sentence.

**Question 3:**

| | Man | Woman | Child | Total |
|---|---|---|---|---|
| First Class | 10 | 15 | 5 | 30 |
| Second Class | 25 | 30 | 10 | 65 |
| Third Class | 30 | 35 | 15 | 80 |
| Total | 65 | 80 | 30 | 175 |

Given that a passenger selected at random was a man, find the probability that the passenger traveled in second class.

**Answer:**

p(second_class|man)=25/(65)=0.3846