1.1

$P(Sam|am) = \lambda_2 \times P(Sam|am) + \lambda_1 \times P(Sam) = \frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{4}{24} = \frac{5}{12}$

1.2

$$
\begin{aligned}
P(Sam|I) &= \alpha(I) P_{katz}(Sam) \\
\alpha(I) &= \frac{1 - \sum_{w_n : c(w_{n-1} w_n > 0)} P^*(W_n|W_{n-1})}{\sum_{w_n : c(w_{n-1} w_n = 0)} P_{katz}(W_n)} \\
&= \frac{1 - (P^*(Am|I)) - (P^*(do|I))}{P_{katz}(< s >) + P_{katz}(I) + ... + P_{katz}(eggs) + P_{katz}(and)} \\
&= \frac{1 - \frac{3-1}{4} - \frac{0}{4}}{5/6} \\
&= \frac{3}{5}
\end{aligned}
$$

Thus $P(Sam|I) = \frac{3}{5} \times \frac{1}{6} = \frac{1}{10}$

2.3

The performance with a rigorous implementation by using $< Unknown >$ token was rather poor and the model predicts 'r' overwhelmingly. This is due to the fact that movie reviews, because of a way smaller corpus, has a way larger proportion of $< Unknown >$ words. Through multiplication the larger proportion of unknown words will lead to predictions of mainly r. Thus, instead of using $\frac{numberof<Unknown>}{numberoftokens}$ for P_{katz} when a word/phrase is unseen, I use $\frac{1}{numberoftokens}$ to offset the disproportional unknown word ratio between review and plot. With the modified P_{katz}, we found out that the bigram model is performing better than the unigram model. Such better performance is expected as bigrams generally provides a better context about the patterns observable in different styles of writing. One interesting similarity among the models is that the precision around p is surprisingly high. As precision is calculated as $tp/(tp + fp)$, since there are less tokens in the review corpus, P_{katz} when predicted with the review language model will be larger and therefore the model is less likely to predict false positives with regard to p(when real value is r and predicted value is p).

Now we look at the differences in performances between the unigram model and the Bayes model.

Mathematically, according to Bayes' Rule,

$CLASS = argmax_{c \in p,r} P(C) \prod_i P_{ti}|C$

where $\prod_i P_{ti}|C$ is the probability calculated by a given unigram class model. Thus the class is determined by the probability of classes multiplied by the unigram probability, which will further help determine the correct class of the sentence and make the predicted results reflective of the distribution of the training set.

```
For Unigram language models
stats with regard to p
{'Recall': 0.592, 'Precision': 0.9977528089887641, 'F1':
0.7430962343096233}
stats with regard to r
{'Recall': 0.996, 'Precision': 0.4486486486486487, 'F1':
0.6186335403726708}
With Bayes
stats with regard to p:
{'Recall': 0.676, 'Precision': 0.9960707269155207, 'F1':
0.8054011119936458}
stats with regard to r
{'Recall': 0.992, 'Precision': 0.505091649694501, 'F1':
0.6693657219973009}

For bigram language models
stats with regard to p
{'Recall': 0.678, 'Precision': 0.9630681818181818, 'F1':
0.795774647887324}
stats with regard to r
{'Recall': 0.922, 'Precision': 0.4883474576271186, 'F1':
0.6385041551246537}
```