Akash Chaurasia
achaura1@jhu.edu
JHU CS465 Fall 2019
Homework 5

Problem 4:

Output for vtag using entrain and entest:
Model perplexity per tagged test word: 3924.111
Tagging accuracy (Viterbi decoding): 92.20% (known: 96.30% novel: 49.72%)
Tagging accuracy (posterior decoding): 92.37% (known: 96.25% novel: 52.23%)

The improved tagger increased the model perplexity per tagged word, with a
lambda of 5. This is likely because it biases probabilities in favor of new words and thus sacrifices the probabilities of words it knows. However, since there are many novel words, this improves accuracy compared
to the baseline result.

Increasing lambda definitely increased the novel accuracy while decreasing the known
accuracy, which is a worthwhile sacrifice in this situation, where we have
many novel words that need to be tagged accurately for good overall accuracy.

Also, it seems that model perplexity is not necessarily a good inidcator
of accuracy, since here the perplexity increases (which indicates that the model does not fit the data as well), even though we have a better accuracy.

Problem 5:

Output for vtag_em using entrain25k, entest, and enraw:

a) Figure 2 initializes a###(0) and b###(n) because the model defines these as