Probability and N-Grams Quiz

Computational Linguistics

Name: _____

This page is printed on both sides. You do not have to simplify arithmetic expressions.

1. We pick a random word of length n by generating letters according to a unigram model, with each of the 26 letters generated with equal probability.

   (a) How many possible such words are there?

   (b) What is the expected number of times that the letter "a" occurs in the word?

   (c) What is the probability that the letter "a" does not appear in the word?

2. Compute $P(\texttt{hovercraft}|\texttt{my})$ – that is, the probability of the word `hovercraft` given the preceding word is `my` – from these bigram counts.

my hovercraft	20
my eels	10
my house	20
hovercraft is	10
hovercraft eels	5

3. Here are some facts about Canadians.

   - The probability of encountering a Canadian in Hanover is 5%.
   - 80% of Canadians in Hanover use the British spelling, *colour*. The remaining Canadians assimilate to the US spelling, *color*.
   - 95% of non-Canadians in Hanover spell the word as *color* while the remaining non-Canadians spell it as *colour*.

   Your classmate (in Hanover) spells the word as *colour*. Assuming you know nothing else about her, what is the probability that she is Canadian?