

- For all possible n-grams, add the count of one.

$$p = \frac{c + 1}{n + v}$$

- c = count of n-gram in corpus
 - n = count of history
 - v = vocabulary size
- But there are many more unseen n-grams than seen n-grams
- Example: Europarl 2-bigrams:
 - 86,700 distinct words
 - $86,700^2 = 7,516,890,000$ possible bigrams
 - but only about 30,000,000 words (and bigrams) in corpus