Exercise 3.1 – Trigram Probability Estimation

The general form for trigram probability estimation is:

[P(w\_i | w\_{i-2}, w\_{i-1}) = {C(w\_{i-2}, w\_{i-1}, w\_i)}{C(w\_{i-2}, w\_{i-1})}]

Corpus:

* I am Sam
* Sam I am
* I am Sam

Trigram Counts:

* I am: 2
* I am Sam: 2
* am Sam: 2
* Sam I: 1
* Sam I am: 1
* I am: 1

Trigram Probabilities (non-zero only):

* P(am | , I) = 2 / 2 = 1.0
* P(Sam | I, am) = 2 / 3 ≈ 0.67
* P( | am, Sam) = 2 / 2 = 1.0
* P(I | , Sam) = 1 / 1 = 1.0
* P(am | Sam, I) = 1 / 1 = 1.0
* P( | am, I) = 1 / 3 ≈ 0.33