

Populating the Transition Matrix

To populate the transition matrix you have to keep track of the number of times each tag shows up before another tag.

$A =$

	NN	VB	O
π	1	0	2
NN (noun)	0	0	6
VB (verb)	0	0	0
O (other)	6	0	8

<s> in a station of the metro

<s> the apparition of these faces in the crowd :

<s> petals on a wet , black bough .

Ezra Pound – 1913

In the table above, you can see that green corresponds to nouns (NN), purple corresponds to verbs (VB), and blue corresponds to other (O). Orange (π) corresponds to the initial state. The numbers inside the matrix correspond to the number of times a part of speech tag shows up right after another one.

To go from O to NN or in other words to calculate $P(O|NN)$ you have to compute the following:

$A =$

	NN	VB	O	
π	1	0	2	3
NN	0	0	6	6
VB	0	0	0	0
O	6	0	8	14

$$P(\text{NN}|\text{O}) = \frac{C(\text{O}, \text{NN})}{\sum_{j=1}^N C(\text{O}, t_j)} = \frac{6}{14}$$

To generalize:

$$P(t_i | t_{i-1}) = \frac{C(t_{i-1}, t_i)}{\sum_{j=1}^N C(t_{i-1}, t_j)}$$

Unfortunately, sometimes you might not see two POS tags in front each other. This will give you a probability of 0. To solve this issue, you will "smooth" it as follows:

$A =$

	NN	VB	O	
π	1+ ϵ	0+ ϵ	2+ ϵ	3+3* ϵ
NN	0+ ϵ	0+ ϵ	6+ ϵ	6+3* ϵ
VB	0+ ϵ	0+ ϵ	0+ ϵ	0+3* ϵ
O	6+ ϵ	0+ ϵ	8+ ϵ	14+3* ϵ

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i) + \epsilon}{\sum_{j=1}^N C(t_{i-1}, t_j) + N * \epsilon}$$

The ϵ allows you to not have any two sequences showing up with 0 probability. Why is this important?