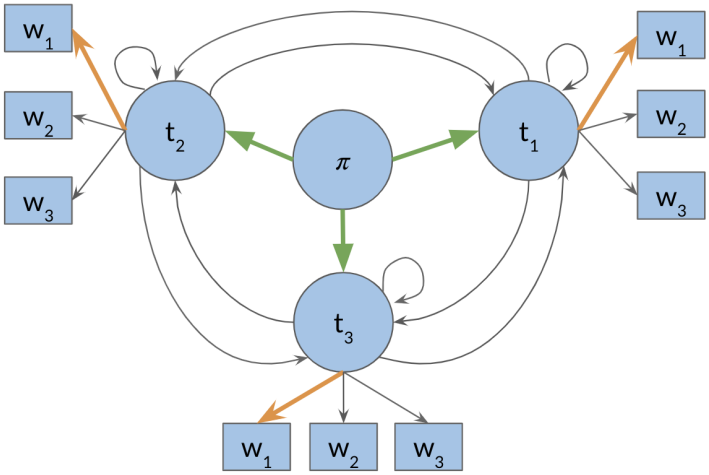
Viterbi Initialization

You will now populate a matrix **C** of dimension (num_tags, num_words). This matrix will have the probabilities that will tell you what part of speech each word belongs to.

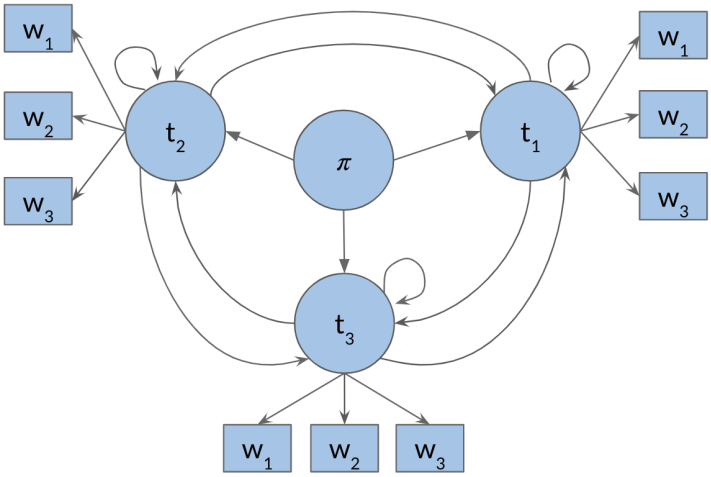


$C =$

	w_1	w_2	...	w_K
t_1	$c_{1,1}$			
...				
t_N	$c_{N,1}$			

$$c_{i,1} = \pi_i * b_{i,cindex(w_1)}$$
$$= a_{1,i} * b_{i,cindex(w_1)}$$

Now to populate the first column, you just multiply the initial π distribution, for each tag, times $b_{i,\text{cindex}(w_1)}$. Where the *i,* corresponds to the tag of the initial distribution and the $cindex(w_1)$, is the index of **word 1** in the emission matrix. And that's it, you are done with populating the first column of your new **C** matrix. You will now need to keep track what part of speech you are coming from. Hence we introduce a matrix **D,** which allows you to store the labels that represent the different states you are going through when finding the most likely sequence of POS tags for the given sequence of words $w_1, \ldots, w_K$. At first you set the first column to 0, because you are not coming from any POS tag.



$D =$

	w_1	w_2	...	w_K
t_1	$d_{1,1}$			
...				
t_N	$d_{N,1}$			

$$d_{i,1} = 0$$

These two matrices will make more sense in the next videos.