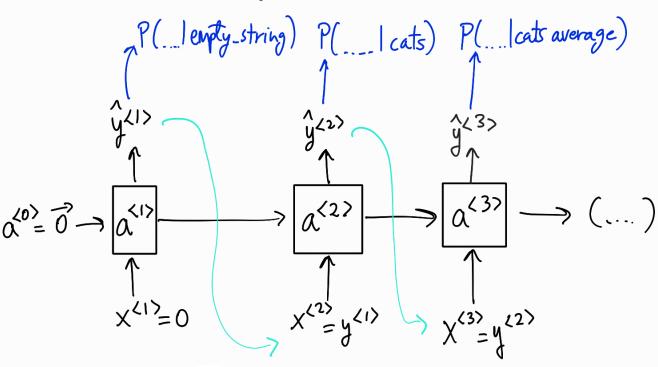We want to generate the sentence:

"Cat average 15 hours of sleep a day."

So, we build the following RNN:

$P(...|empty\_string)$ $P(...|cats)$ $P(...|cats\ average)$

$\hat{y}^{<1>}$ $\hat{y}^{<2>}$ $\hat{y}^{<3>}$

$a^{<0>} = \vec{0} \rightarrow \boxed{a^{<1>}} \longrightarrow \boxed{a^{<2>}} \longrightarrow \boxed{a^{<3>}} \longrightarrow (...)$

$x^{<1>} = 0$ $x^{<2>} = y^{<1>}$ $x^{<3>} = y^{<2>}$

$$\mathcal{L}(\hat{y}^{<t>}, y^{<t>}) = -\sum_i y_i^{<t>} \log \hat{y}_i^{<t>}$$

$$\mathcal{L} = \sum_t \mathcal{L}^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

So, we are calculating the joint probability !!

$$P\left(y^{<1>}, y^{<2>}, ..., y^{<T_y>}\right)$$

Each computed $y^{<i>}$ is usually a softmax distribution, which gives us the chance that the ith word is ... for each word of the vocabulary.

For $y^{<1>}$, we use this distribution and compute np.random() for sampling.