

# The Simplest Language Model: Unigram LM

- Generate text by generating each word INDEPENDENTLY
- Thus,  $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2)\dots p(w_n)$
- Parameters:  $\{p(w_i)\}$   $p(w_1) + \dots + p(w_N) = 1$  (N is voc. size)
- Text = sample drawn according to this **word distribution**



Wednesday  
today  
eigenvalue

$$\begin{aligned} p(\text{"today is Wed"}) \\ &= p(\text{"today"})p(\text{"is"})p(\text{"Wed"}) \\ &= 0.0002 \times 0.001 \times 0.000015 \end{aligned}$$