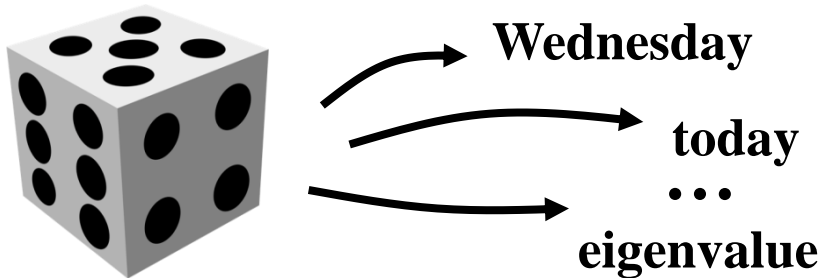# The Simplest Language Model: Unigram LM

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

Wednesday

today

…

eigenvalue

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