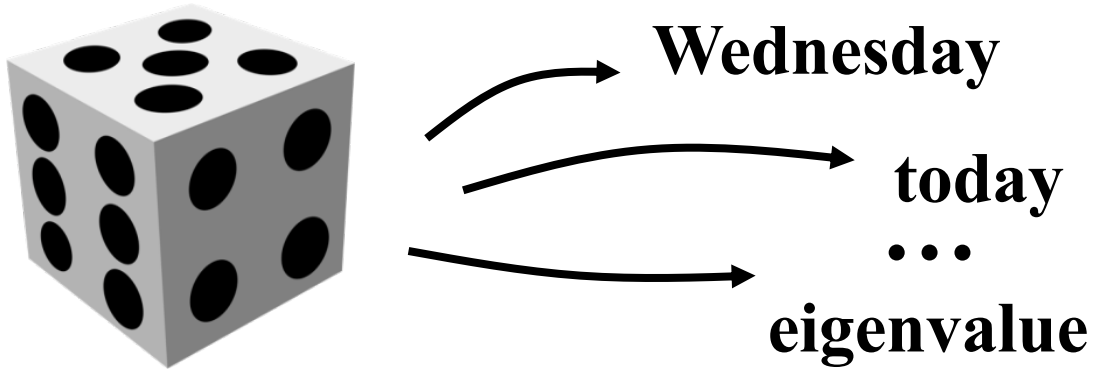
The Simplest Language Model: Unigram LM

- 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: $\{p(t_i)\}$ $p(t_1)+\ldots+p(t_N)=1$ (N is voc. size)
- Text = sample drawn according to this **word distribution**



p("today is Wed")
= p("today")p("is")p("Wed")
= 0.0002 × 0.001 × 0.000015