

# Language Models

A language model is a **probability distribution** over **word sequences**.

- Unigram language model  $p(w | \theta)$ :

$$p(\mathbf{W} | \theta) = \prod_{i=1}^N p(w_i | \theta)$$

- Bigram language model  $p(w_i | w_{i-1}, \theta)$ :

$$p(\mathbf{W} | \theta) = \prod_{i=1}^N p(w_i | w_{i-1}, \theta)$$

- $n$ -gram language model:  $p(w_n | w_1^{n-1}, \theta)$

$$p(\mathbf{W} | \theta) = \prod_{i=1}^N p(w_i | w_{i-n+1}^{i-1}, \theta)$$