Summary of Language Models

- $p(e_1^I)$ should report how "good" sentence e_1^I is.
- We surely want $p(\text{The the the.}) < p(\text{Hello.})$
- How about $p(\text{The cat was black.}) < p(\text{Hello.})$?

  ...We don't really care in MT. We hope to compare synonymic sentences.

LM is usually a 3-gram language model:

$$p(\curvearrowright\ \curvearrowright \text{ The cat was black . } \curvearrowleft\ \curvearrowleft) = \begin{matrix} p(\text{The}|\curvearrowright\ \curvearrowright) & p(\text{cat}|\curvearrowright \text{ The}) & p(\text{was}|\text{The cat}) \\ p(\text{black}|\text{cat was}) & p(.|\text{was black}) & p(\curvearrowleft|\text{black .}) \\ p(\curvearrowleft|.\ \curvearrowleft) & & \end{matrix}$$

Formally, with $n = 3$:

$$p_{\text{LM}}(e_1^I) = \prod_{i=1}^{I} p(e_i|e_{i-n+1}^{i-1}) \tag{4}$$