# Katz backoff

**Problem:**

- Longer n-grams are better, but data is not always enough

**Idea:**

- Try a longer n-gram and back off to shorter if needed

$$\hat{p}(w_i|w_{i-n+1}^{i-1}) = \begin{cases} \tilde{p}(w_i|w_{i-n+1}^{i-1}), & \text{if } c(w_{i-n+1}^{i}) > 0 \\ \alpha(w_{i-n+1}^{i-1})\,\hat{p}(w_i|w_{i-n+2}^{i-1}), & \text{otherwise} \end{cases}$$

where $\tilde{p}$ and $\alpha$ are chosen to ensure normalization.