ML Estimate of N-Gram LM

$$p(w_m | w_{m-n+1},...,w_{m-1}) = \frac{c(w_{m-n+1}...w_{m-1}w_m, D)}{\sum_{u \in V} c(w_{m-n+1}...w_{m-1}u, D)}$$

- Count of long word sequences may be zero!
  - Not accurate
  - Cause problems when computing the conditional probability $p(w_m | w_{m-n+1},...,w_{m-1})$
- Solution: smoothing
  - **Key idea: backoff to shorter N-grams, eventually to unigrams**
  - **Treat shorter N-gram models as prior in Bayesian estimation**