## 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

The Dafa and Information Systems Laboratories at The University of Illinois at Urbana-Champaign

- Cause problems when computing the conditional probability
  p(w<sub>m</sub>|w<sub>m-n+1</sub>,...,w<sub>m-1</sub>)
- Solution: smoothing
  - Key idea: backoff to shorter N-grams, eventually to unigrams
  - Treat shorter N-gram models as prior in Bayesian estimation