- Language models too large to *build*
- What needs to be stored in RAM?
  - maximum likelihood estimation

$$p(w_n|w_1,...,w_{n-1}) = \frac{\text{count}(w_1,...,w_n)}{\text{count}(w_1,...,w_{n-1})}$$

  - can be done separately for each history $w_1,...,w_{n-1}$
- Keep data on disk
  - extract all n-grams into files on-disk
  - sort by history on disk
  - only keep n-grams with shared history in RAM
- Smoothing techniques may require additional statistics