Central Question in LM: $p(w_1w_2...w_m|C)=?$

- Refinement 2: $p(w_1w_2...w_m) = p(X_1=w_1,X_2=w_2,...X_m=w_m)$
  - What assumption have we made here?
- Chaining Rule: $p(w_1w_2...w_m)= p(w_1)p(w_2|w_1)...p(w_m|w_1w_2...w_{m-1})$
  - What about $p(w_1w_2...w_m)= p(w_m)p(w_{m-1}|w_m)...p(w_1|w_2...w_m)$?
- Refinement 3: Assume limited dependence (only depends on the n previous words) ➔ N-gram LM

$$p(X_1=w_1,X_2=w_2,...X_m=w_m)\approx p(X_1=w_1)\ p(X_2=w_2|X_1=w_1)...p(X_n=w_n|X_1=w_1,...,X_{n-1}=w_{n-1})\ ...$$

$$\mathbf{p(w_1w_2...w_m) \approx p(w_1)\ p(w_2|w_1)...p(w_n|w_1,...,w_{n-1})...\ p(w_m|w_{m-n+1},...,w_{m-1})}$$