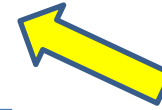# Smoothing of $p(w_m | w_1, \ldots, w_{m-n+1} \ldots, w_{m-1})$

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

What if this is zero?

- How should we define p(w|REF)?

- In general, p(w|REF) can be defined based on any "clues" from the history **h=($w_{m-n+1}, \ldots, w_{m-1}$)**

  - **Most natural: p(w|REF)=p($w_m | w_{m-n+2}, \ldots, w_{m-1}$), ignore w_{m-n+1} ; can be done recursively to rely on shorter and shorter history**

  - **In general, relax the condition to make it less specific so as to increase the counts we can collect (e.g., shorten the history, cluster the history)**