



- Only windows which occur in  $W^O$  are used
- Only tokens which occur after a window  $W^O$  are used
- B represents the weight of each window

$$B = \frac{\text{Occurrences}(W_i^O)}{\text{Total}(W^O)}$$

$$= \sum_{i=0}^n |W_i^U - W_i^O| * B$$

$$W_i^U - W_i^O = \sum_{q=0}^m |P(T_q|W_i^U) - P(T_q|W_i^O)| * P(T_q|W_i^O)$$

- $P(T_q|W_i^O)$  is both the probability of occurrence of token  $T_q$  and the weight of  $T_q$