## Expected Overlap of Words in Context (EOWC)

**Probability that a randomly** Count of word wi in d1 picked word from d1 is wi  $d1 = (x_1, ..., x_N) \quad x_i = c(w_i, d1)/|d1|$  $d2=(y_1, \dots, y_N)$   $y_i = c(w_i, d2)/|d2|$ **Total counts of** words in d1  $Sim(d1,d2)=d1.d2=x_1y_1+...+x_Ny_N=\sum_{i=1}^N x_i y_i$ Probability that two randomly picked words from d1 and d2, respectively, are identical.