

| Count | Adjusted count           |                                      | Test count |
|-------|--------------------------|--------------------------------------|------------|
| $c$   | $(c + 1)\frac{n}{n+v^2}$ | $(c + \alpha)\frac{n}{n+\alpha v^2}$ | $t_c$      |
| 0     | 0.00378                  | 0.00016                              | 0.00016    |
| 1     | 0.00755                  | 0.95725                              | 0.46235    |
| 2     | 0.01133                  | 1.91433                              | 1.39946    |
| 3     | 0.01511                  | 2.87141                              | 2.34307    |
| 4     | 0.01888                  | 3.82850                              | 3.35202    |
| 5     | 0.02266                  | 4.78558                              | 4.35234    |
| 6     | 0.02644                  | 5.74266                              | 5.33762    |
| 8     | 0.03399                  | 7.65683                              | 7.15074    |
| 10    | 0.04155                  | 9.57100                              | 9.11927    |
| 20    | 0.07931                  | 19.14183                             | 18.95948   |

- Add- $\alpha$  smoothing with  $\alpha = 0.00017$
- $t_c$  are average counts of n-grams in test set that occurred  $c$  times in corpus