- Back-off models use only highest order n-gram
  - if sparse, not very reliable.
  - two different n-grams with same history occur once → same probability
  - one may be an outlier, the other under-represented in training

- To remedy this, always consider the lower-order back-off models

- Adapting the α function into interpolated α_I function by adding back-off

$$\alpha_I(w_n|w_1,...,w_{n-1}) = \alpha(w_n|w_1,...,w_{n-1}) + d(w_1,...,w_{n-1})\, p_I(w_n|w_2,...,w_{n-1})$$

- Note that d function needs to be adapted as well