Interpolation smoothing

Idea:

- Let us have a mixture of several n-gram models
- Example for a trigram model:

$$\hat{p}(w_i|w_{i-2}w_{i-1}) = \lambda_1 p(w_i|w_{i-2}w_{i-1}) + \lambda_2 p(w_i|w_{i-1}) + \lambda_3 p(w_i)$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

- The weights are optimized on a test (dev) set
- Optionally they can also depend on the context