

# Log-Linear Model (1)

- $p(e_1^I | f_1^J)$  is modelled as a weighted combination of models, called “feature functions”:  $h_1(\cdot, \cdot) \dots h_M(\cdot, \cdot)$

$$p(e_1^I | f_1^J) = \frac{\exp(\sum_{m=1}^M \lambda_m h_m(e_1^I, f_1^J))}{\sum_{e_1^{I'}} \exp(\sum_{m=1}^M \lambda_m h_m(e_1^{I'}, f_1^J))} \quad (8)$$

- Each feature function  $h_m(e, f)$  relates source  $f$  to target  $e$ .  
E.g. the feature for  $n$ -gram language model:

$$h_{\text{LM}}(f_1^J, e_1^I) = \log \prod_{i=1}^I p(e_i | e_{i-n+1}^{i-1}) \quad (9)$$

- Model weights  $\lambda_1^M$  specify the relative importance of features.