Reminder: Log-Linear Model

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

$$p(e_1^I|f_1^J) = \frac{\exp(\sum_{m=1}^{M} \lambda_m h_m(e_1^I, f_1^J))}{\sum_{e'^{I'}_1} \exp(\sum_{m=1}^{M} \lambda_m h_m(e'^{I'}_1, f_1^J))} \quad (1)$$

- The constant denominator not needed in maximization:

$$\begin{aligned} \hat{e}_1^{\hat{I}} &= \text{argmax}_{I,e_1^I} \frac{\exp(\sum_{m=1}^{M} \lambda_m h_m(e_1^I, f_1^J))}{\sum_{e'^{I'}_1} \exp(\sum_{m=1}^{M} \lambda_m h_m(e'^{I'}_1, f_1^J))} \\ &= \text{argmax}_{I,e_1^I} \exp(\sum_{m=1}^{M} \lambda_m h_m(e_1^I, f_1^J)) \end{aligned} \quad (2)$$