

Relation to Noisy Channel

With equal weights and only two features:

- $h_{\text{TM}}(e_1^I, f_1^J) = \log p(f_1^J | e_1^I)$ for the translation model,
- $h_{\text{LM}}(e_1^I, f_1^J) = \log p(e_1^I)$ for the language model,

log-linear model reduces to Noisy Channel:

$$\begin{aligned}\hat{e}_1^I &= \operatorname{argmax}_{I, e_1^I} \exp(\sum_{m=1}^M \lambda_m h_m(e_1^I, f_1^J)) \\ &= \operatorname{argmax}_{I, e_1^I} \exp(h_{\text{TM}}(e_1^I, f_1^J) + h_{\text{LM}}(e_1^I, f_1^J)) \\ &= \operatorname{argmax}_{I, e_1^I} \exp(\log p(f_1^J | e_1^I) + \log p(e_1^I)) \\ &= \operatorname{argmax}_{I, e_1^I} p(f_1^J | e_1^I) p(e_1^I)\end{aligned}\tag{11}$$