Motivation for Noisy Channel

$$\hat{e}_1^{\hat{I}} = \underset{I, e_1^I}{\text{argmax}}\, p(f_1^J | e_1^I) p(e_1^I) \tag{3}$$

Bayes' law divided the model into components:

$p(f_1^J | e_1^I)$ Translation model ("reversed", $e_1^I \rightarrow f_1^J$)
...is it a likely translation?

$p(e_1^I)$ Language model (LM)
...is the output a likely sentence of the target language?

- The components can be trained on different sources.
  There are far more monolingual data ⇒ language model can be more reliable.