Common Features of PBMT

- Phrase translation probability:
  $h_{\text{Phr}}(f_1^J, e_1^I, s_1^K) = \log \prod_{k=1}^{K} p(\tilde{f}_k|\tilde{e}_k)$ where $p(\tilde{f}_k|\tilde{e}_k) = \frac{\text{count}(\tilde{f}, \tilde{e})}{\text{count}(\tilde{e})}$
  $\Rightarrow$ Are all used units $\tilde{f} \leftrightarrow \tilde{e}$ likely translations?
- Word count/penalty: $h_{\text{wp}}(e_1^I, \cdot, \cdot) = I$
  $\Rightarrow$ Do we prefer longer or shorter output?
- Phrase count/penalty: $h_{\text{pp}}(\cdot, \cdot, s_1^K) = K$
  $\Rightarrow$ Do we prefer translation in more or fewer less-dependent bits?
- Reordering model: different basic strategies (Lopez, 2009)
  $\Rightarrow$ Which source spans can provide continuation at a moment?
- n-gram LM: $h_{\text{LM}}(\cdot, e_1^I, \cdot) = \log \prod_{i=1}^{I} p(e_i|e_{i-n+1}^{i-1})$
  $\Rightarrow$ Is output n-gram-wise coherent?