# Phrase-Based Translation Model

- Captures the basic assumption of phrase-based MT:
  1. Segment source sentence $f_1^J$ into $K$ phrases $\tilde{f}_1 ... \tilde{f}_K$.
  2. Translate each phrase independently: $\tilde{f}_k \rightarrow \tilde{e}_k$.
  3. Concatenate translated phrases (with possible reordering $R$): $\tilde{e}_{R(1)} ... \tilde{e}_{R(K)}$

- In theory, the segmentation $s_1^K$ is a hidden variable in the maximization, we should be summing over all segmentations: (Note the three args in $h_m(\cdot,\cdot,\cdot)$ now.)

$$\hat{e}_1^{\hat{I}} = \text{argmax}_{I,e_1^I} \sum_{s_1^K} \exp(\sum_{m=1}^M \lambda_m h_m(e_1^I, f_1^J, s_1^K)) \quad (3)$$

- In practice, the sum is approximated with a max (the biggest element only):

$$\hat{e}_1^{\hat{I}} = \text{argmax}_{I,e_1^I} \max_{s_1^K} \exp(\sum_{m=1}^M \lambda_m h_m(e_1^I, f_1^J, s_1^K)) \quad (4)$$