

# Encoder-Decoder: Training Objective

For output word  $y_i$  we have:

- estimated conditional distribution  $\hat{p}_i = \frac{\exp t_i}{\sum \exp t_j}$  (softmax function)