Encoder-Decoder: Training Objective

For output word y_i we have:

- estimated conditional distribution $\widehat{p}_i = \frac{\exp t_i}{\sum \exp t_j}$ (softmax function)
- unknown true distribution p_i, we lay $p_i \equiv \mathbf{1}\left[y_i\right]$