

# Greedy Decoding

- In each step, the model computes a distribution over the vocabulary  $V$  (given source  $\mathbf{x}$ , the previous outputs  $h$ , and the model parameters  $\theta$ ).

$$p(y|h) = g(\mathbf{x}, h, \theta)$$