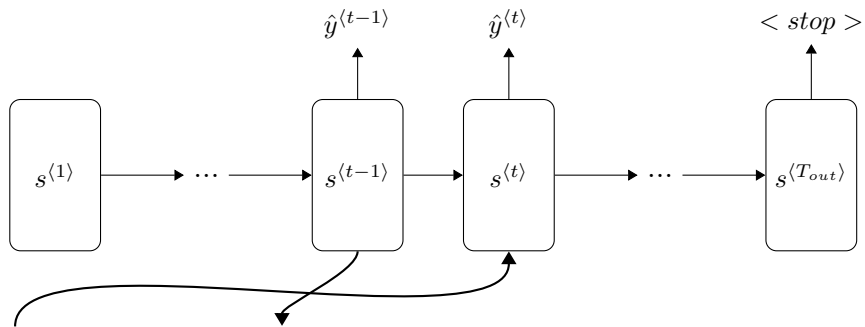


Decoder



$$u^{(t)} = \text{Attention}_t[z_q = s^{(t-1)}, z_k = (v^{(1)}, \dots, v^{(T_{in})}), v = (v^{(1)}, \dots, v^{(T_{in})})]$$

Encoder

