

**Input:** Sequence  $x_1, \dots, x_T$

**Output:** Sequence  $y_1, \dots, y_T$

**Encoder:**  $h_t = f_W(x_t, h_{t-1})$

From final hidden state predict:

**Initial decoder state**  $s_0$

**Context vector**  $c$  (often  $c=h_T$ )

