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

**Output**: Sequence $y_1, \ldots, y_{T'}$

**Decoder**: $s_t = g_U(y_{t-1}, s_{t-1}, c)$

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

From final hidden state predict:
**Initial decoder state** $s_0$
**Context vector** $c$ (often $c = h_T$)