# Put all together

$$p(y_1, \ldots y_J | x_1, \ldots x_I) = \prod_{j=1}^{J} p(y_j | v_j, y_1, \ldots y_{j-1})$$

- Still encoder-decoder architecture with RNNs:

$$h_i = f(h_{i-1}, x_i) \qquad s_j = g(s_{j-1}, [y_{j-1}, v_j])$$

- But the source representations differ for each position j of the decoder.