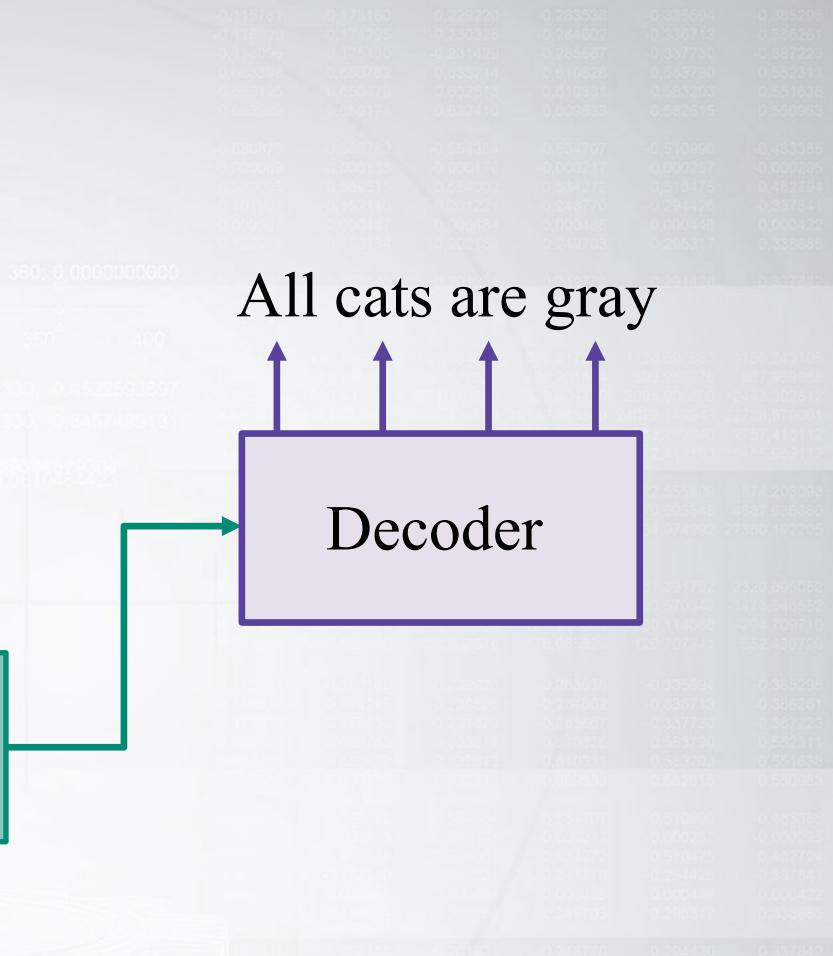
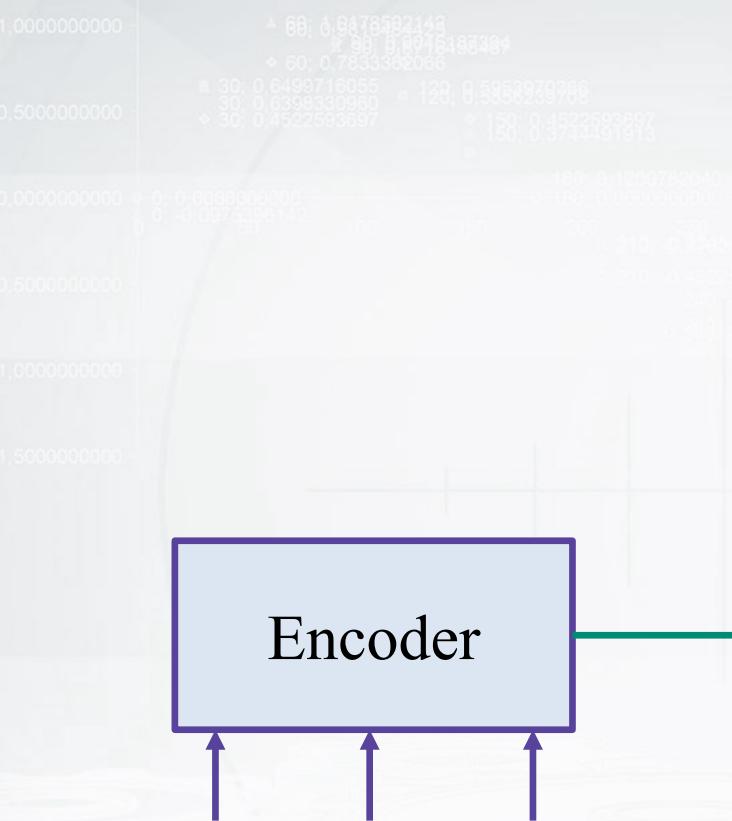


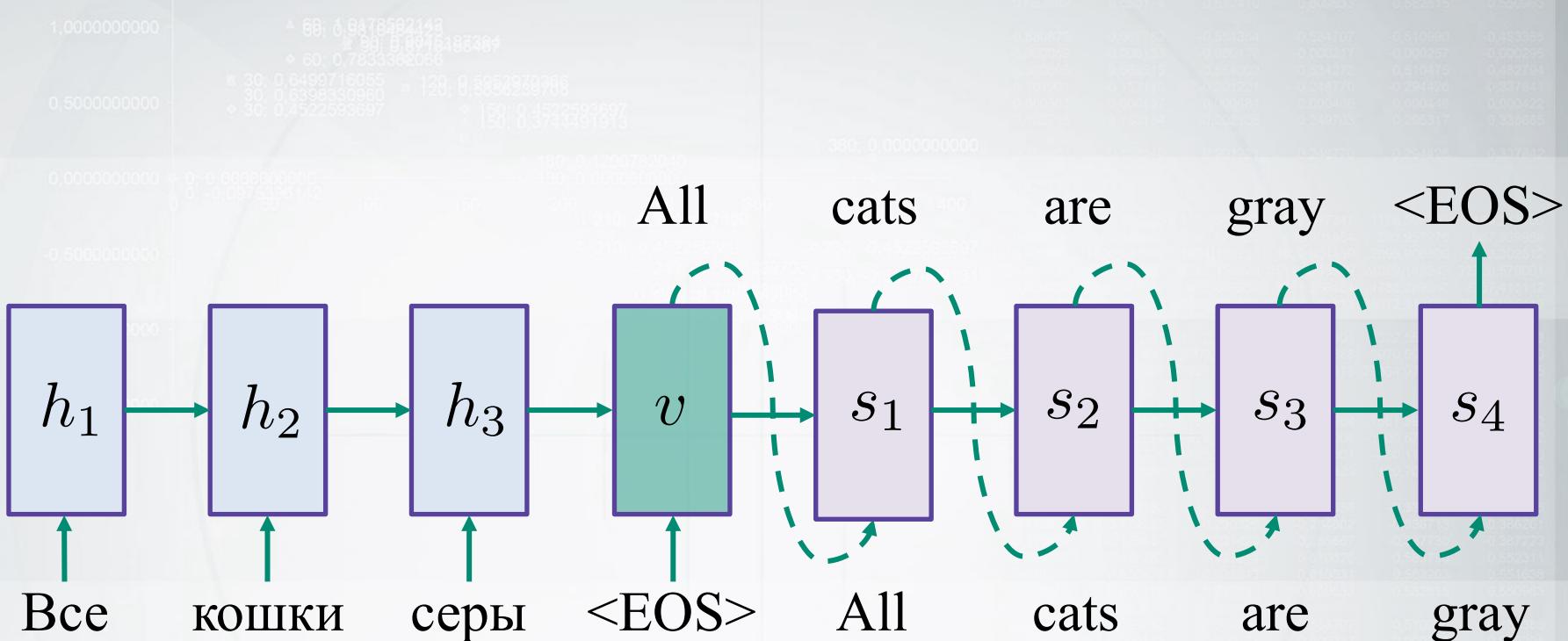
Encoder-decoder architecture



Sequence to sequence

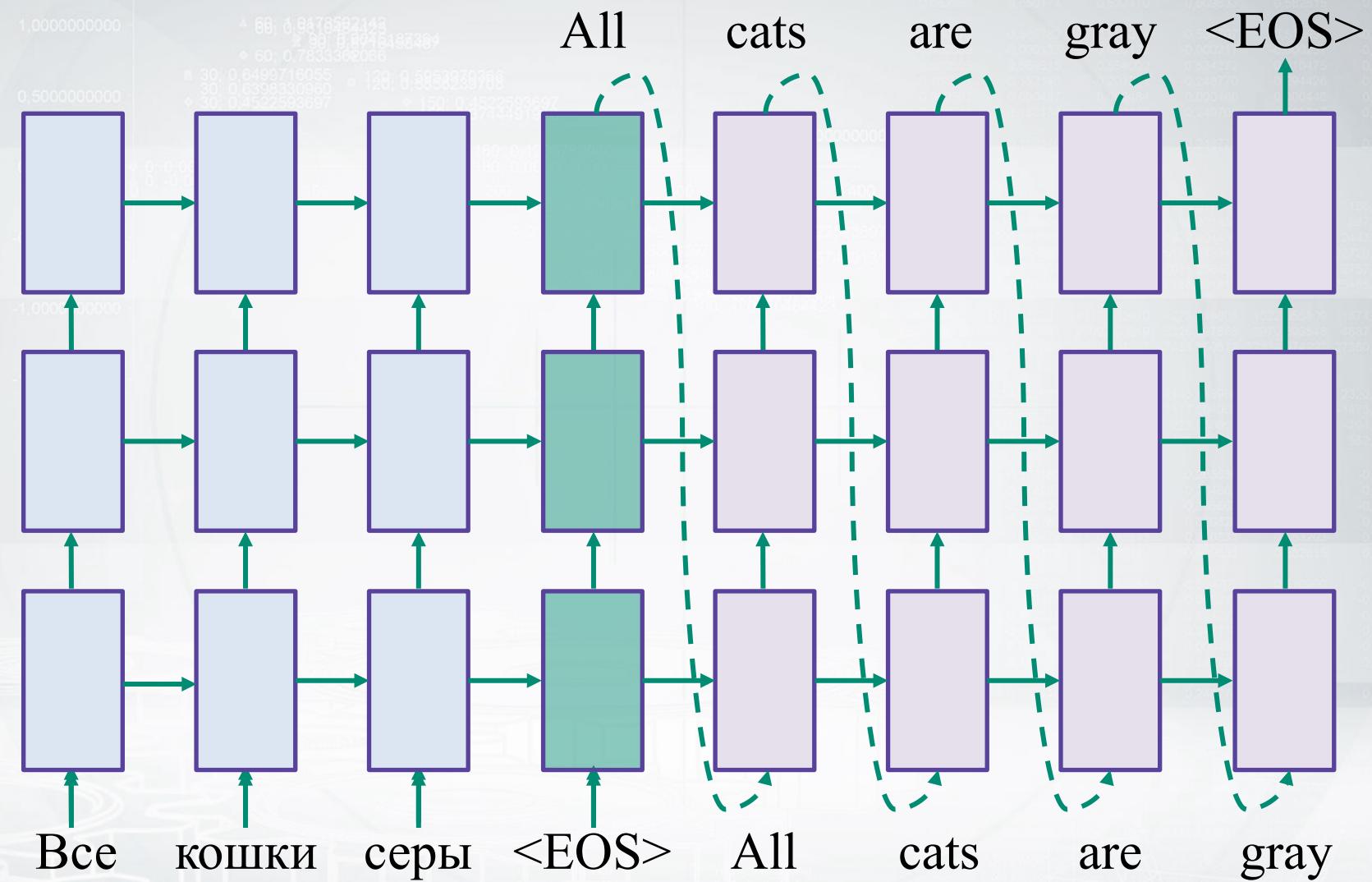


Sequence to sequence

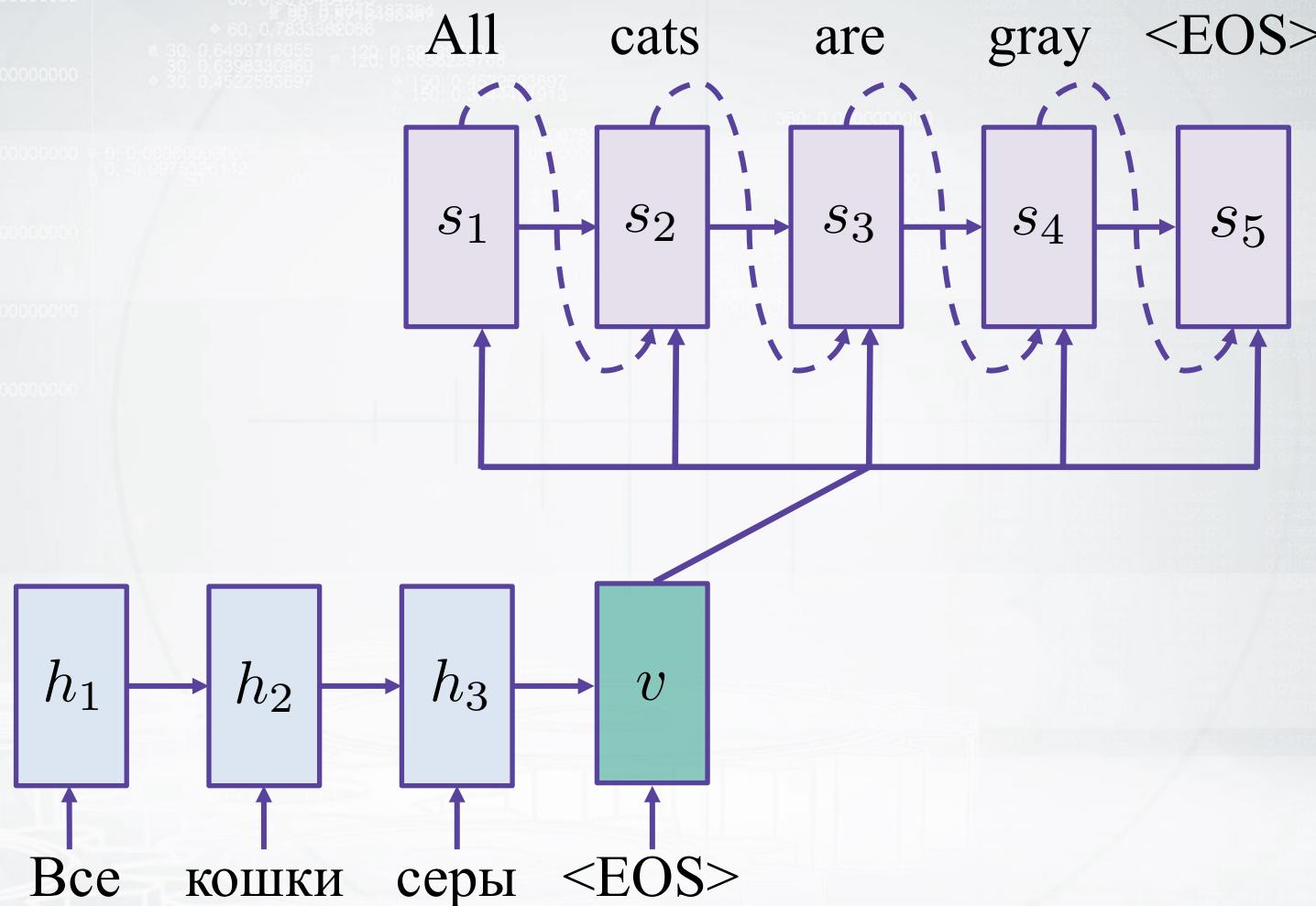


Ilya Sutskever, Oriol Vinyals, Quoc V. Le. Sequence to Sequence Learning with Neural Network, 2014.

Sequence to sequence



Sequence to sequence



Cho et. al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, 2014.

Sequence to sequence

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

- **Encoder:** maps the source sequence to the hidden vector

$$\text{RNN: } h_i = f(h_{i-1}, x_i) \quad \mathbf{v} = h_I$$

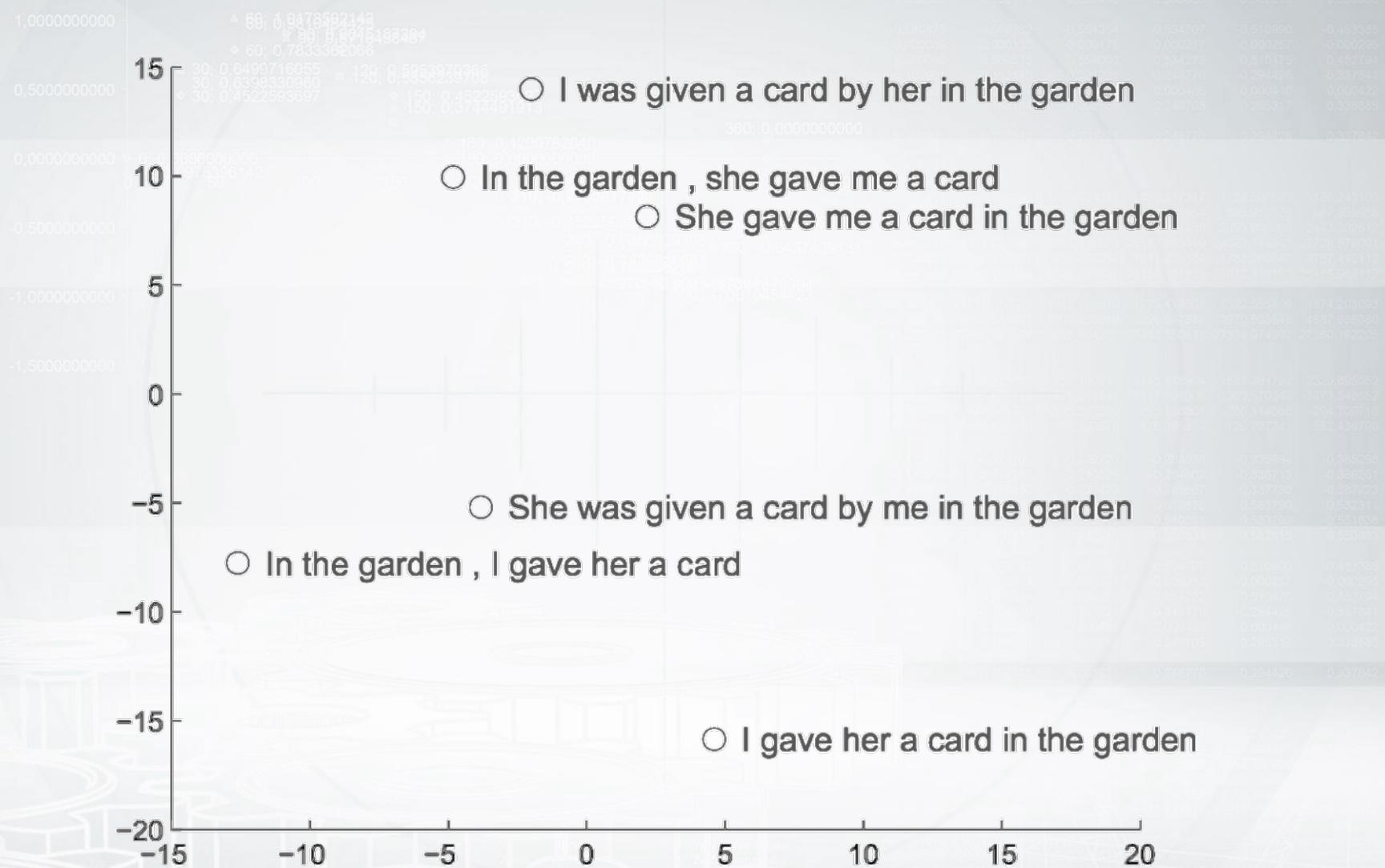
- **Decoder:** performs language modeling given this vector

$$\text{RNN: } s_j = g(s_{j-1}, [y_{j-1}, \mathbf{v}])$$

- **Prediction** (the simplest way):

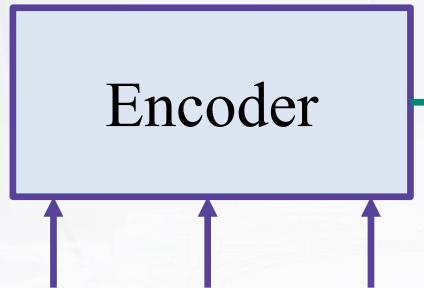
$$p(y_j | v, y_1, \dots, y_{j-1}) = \text{softmax}(Us_j + b)$$

Hidden representations are good...



Ilya Sutskever, Oriol Vinyals, Quoc V. Le. Sequence to Sequence Learning with Neural Network, 2014.

... but still a bottleneck



Все кошки серы

All cats are gray

Decoder

Bottleneck!