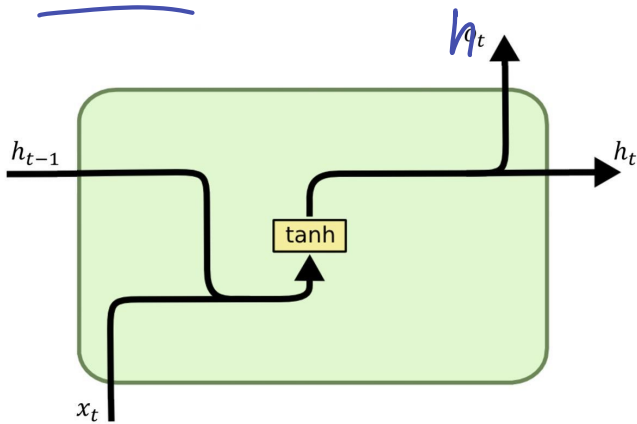


# Рекуррентные нейросети

## RNN



$$h_t = \sigma(W_h[h_{t-1}, x_t] + b_h)$$

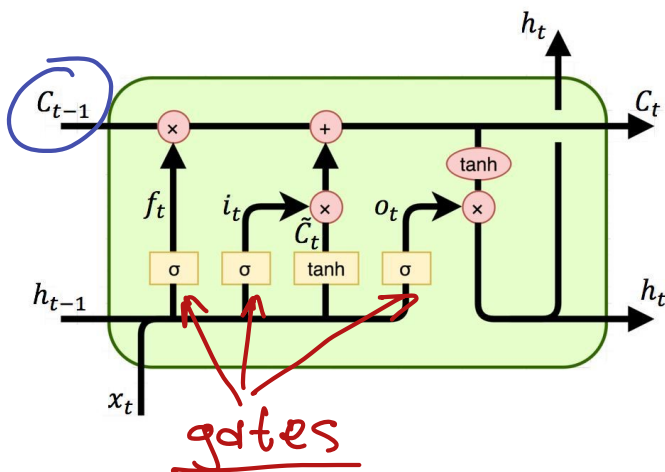
$$O_t = \sigma(W_o h_t + b_o)$$

$$p_t = \text{softmax}(O_t)$$

$$\nabla_{h_{t-n}} L(p_t, y_t) = \frac{\partial L(p_t, y_t)}{\partial h_t} \cdot \prod_{i=t-n}^{t-1} \frac{\partial h_i}{\partial h_{i-1}}$$

↗  
затухающий градиент.

## LSTM (1997)



$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

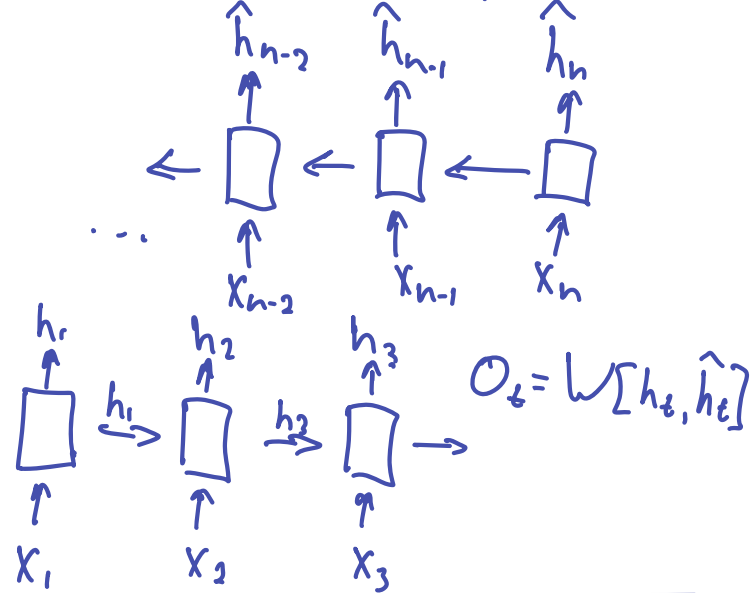
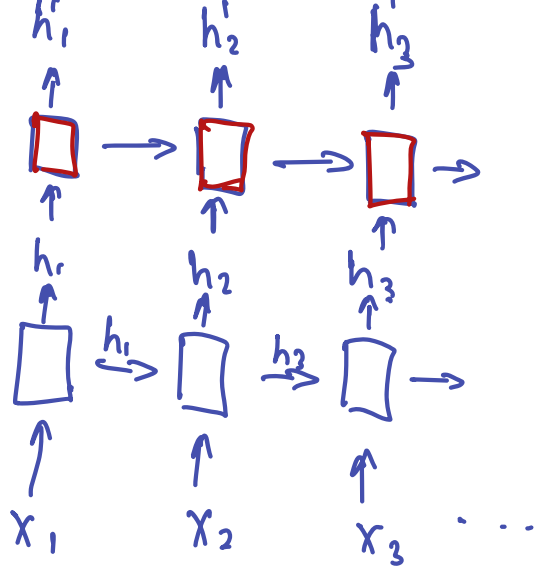
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

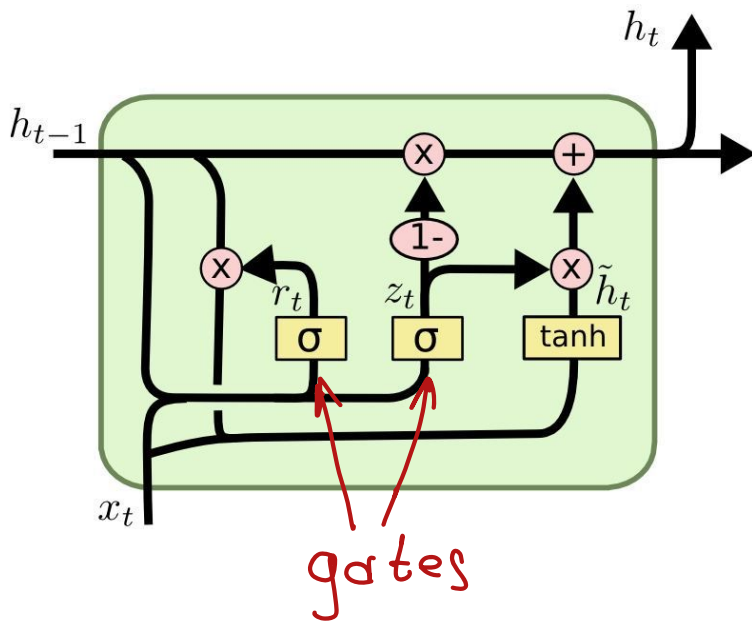
$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$h_t = O_t \odot \tanh(C_t)$$



## GRU (2014)



$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r)$$

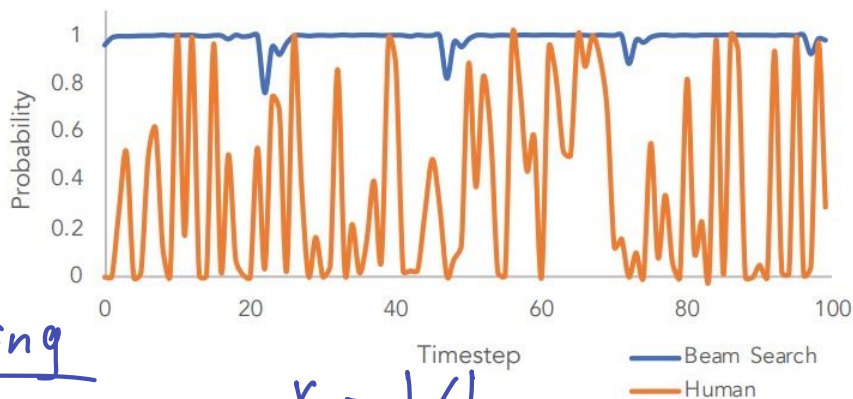
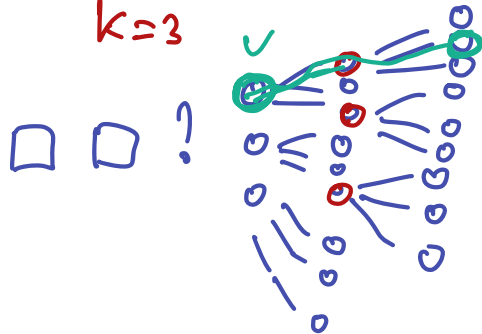
$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$$

$$\tilde{h}_t = \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

## Как семплить?

- 1) Greedy — берем токен с наиб. вероятностью
- 2) Beam-search — находим последовательности токенов с наиб. вероятностью

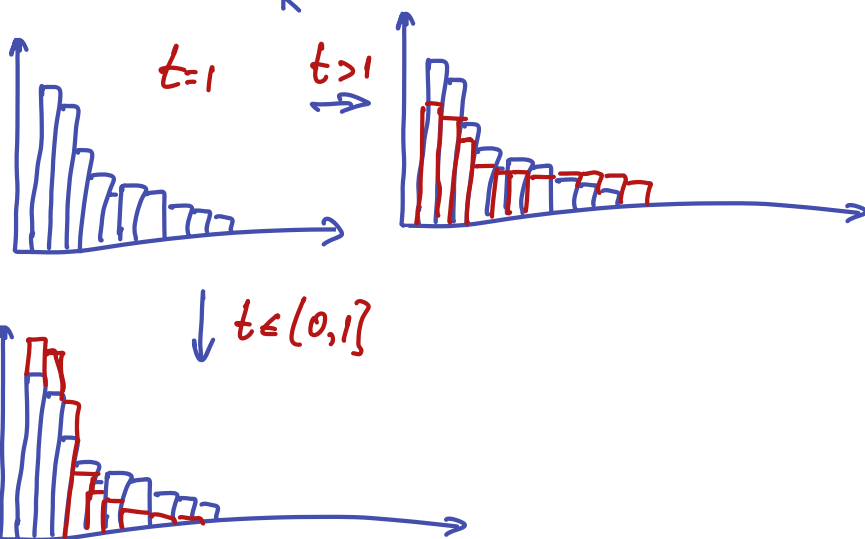


### 3) Temperature Sampling

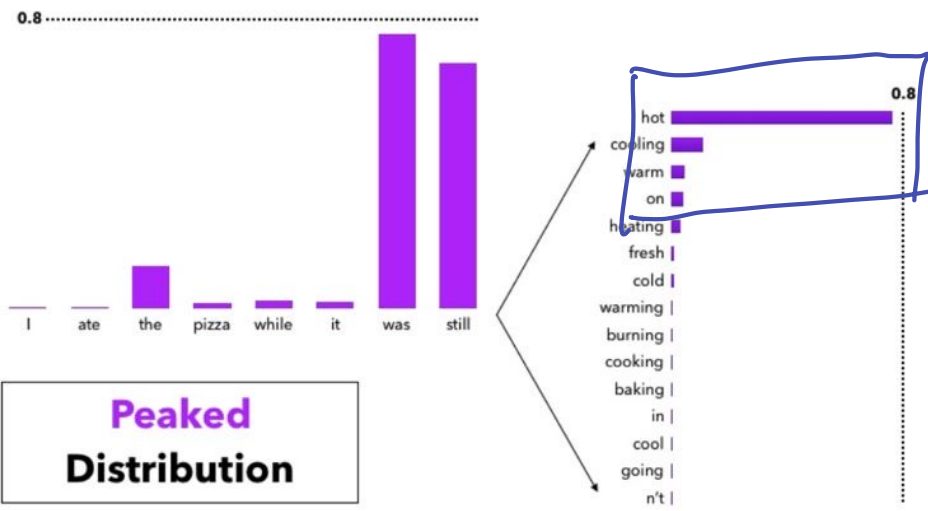
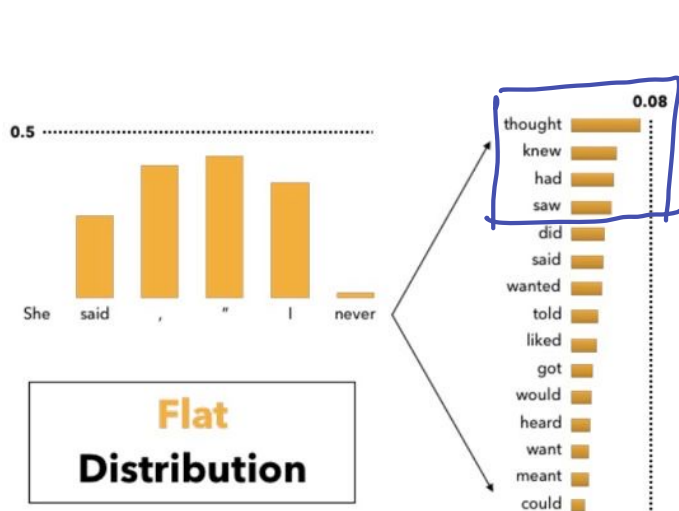
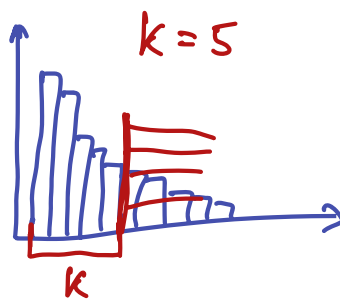
$$p(x_i) = \frac{e^{x_i/\tau}}{\sum_{j=1}^M e^{x_j/\tau}}$$

$$x = w \cdot h$$

$$\uparrow_{R_{LM}}$$



### 4) Top k Sampling



## 5) Nucleus Sampling (2020)

(Top P)

$$\sum_{x \in V} P(x|x_{<t}) \geq p$$

$$p \sim 0.95$$

Берем мин. число токенов,  
сумма вер-тей которых  $\geq p$

