# HOW CAN MONOLINGUAL DATA BE USED TO IMPROVE NMT?

**Andrey Danilov, Polina Proskura, Dmitry Pyrkin, Ekaterina Adamenko, Anna Kuzmina, Elena Zherdeva, Alexandra Volokhova**

October 25, 2018

Yandex School of Data Analysis

# Shallow Fusion

$$\log p = \log p^{TM} + \beta \log p^{LM}$$

# Deep Fusion

$$\log p \propto y_t^T (W_o f_o(s_t^{LM}, s_t^{TM}, y_{t-1}, c_t) + b_o)$$

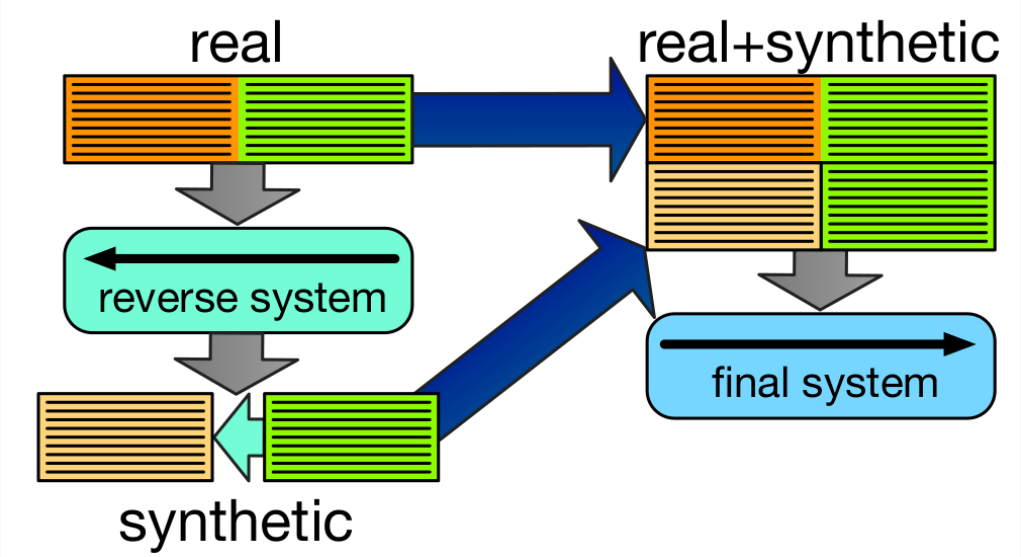
$s_t^{TM}$ - hidden state of translation model

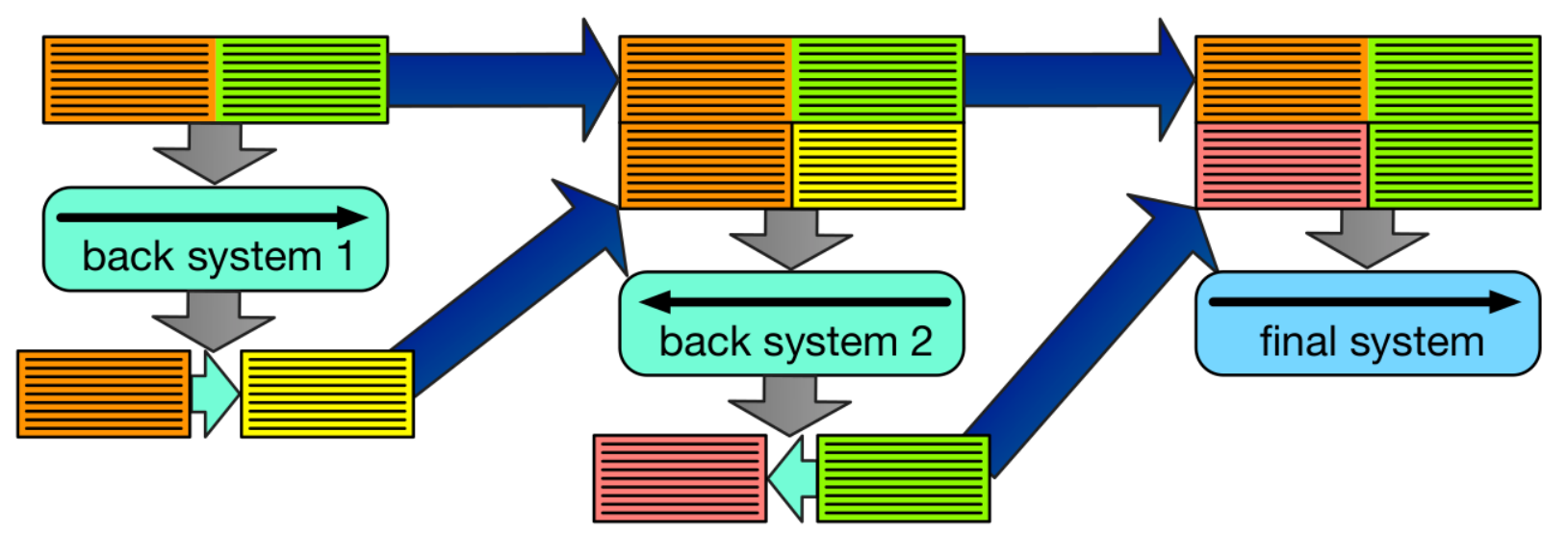
$s_t^{LM}$ - hidden state of language model

$f_o$ - single layer NN

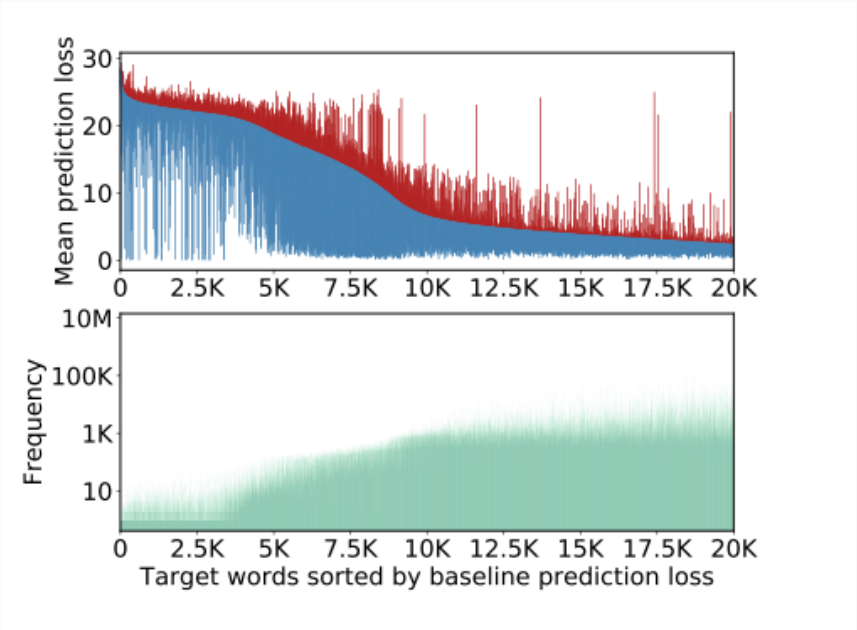
$c_t$ - attention vector

# Iterative Back-Translation for Neural Machine Translation



# Back-Translation Sampling by Targeting Difficult Words




$$\mathfrak{D} = \{\forall y_i \in V_t \colon freq(y_i) < \eta\}$$

$$\mathfrak{D} = \{\forall y_i \in V_t \colon \hat{\ell}(y_i) > \mu \wedge \sigma(\ell(y_i)) > \rho\}$$

$$\mathrm{Sim}(C_m, C_p) = cos(\mathbf{v}(C_m), \mathbf{v}(C_p))$$