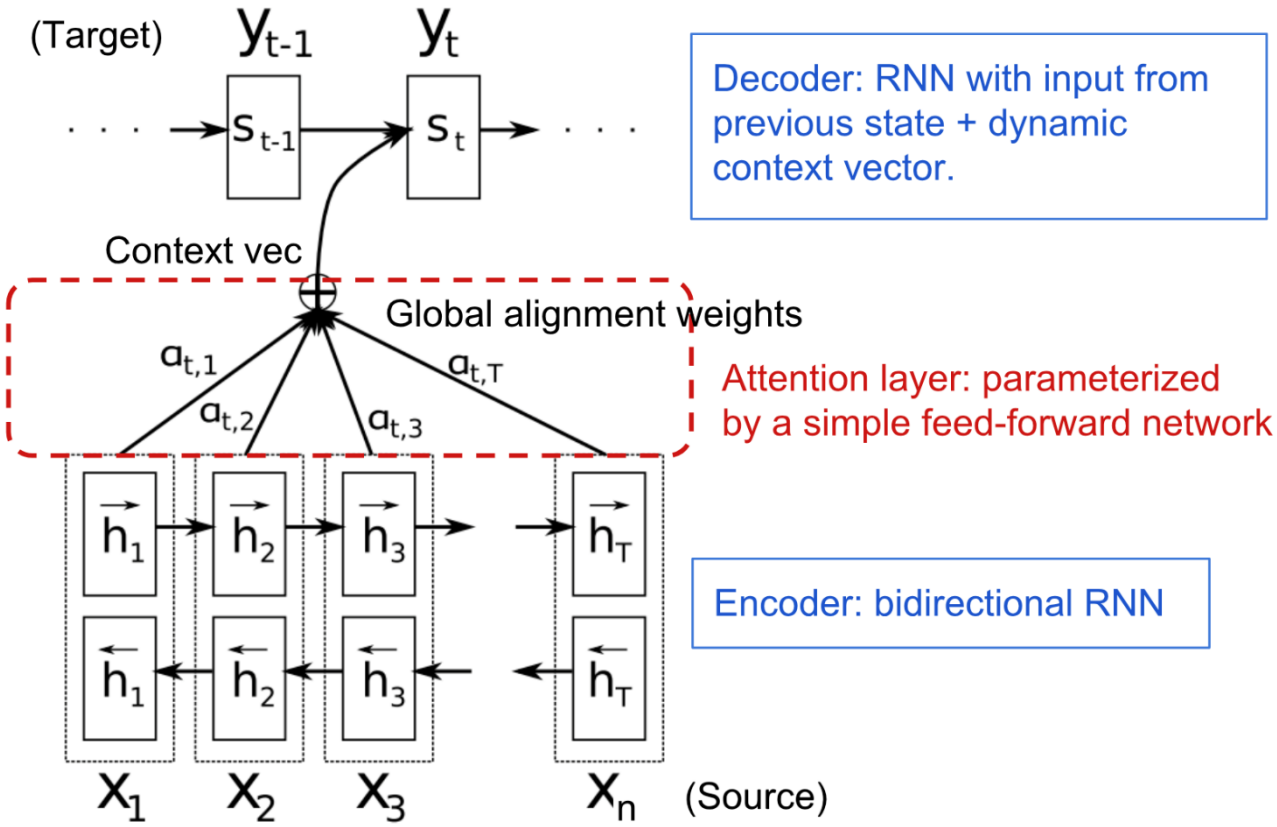
Attention



Decoder: RNN with input from previous state + dynamic context vector.

Attention layer: parameterized by a simple feed-forward network

Encoder: bidirectional RNN

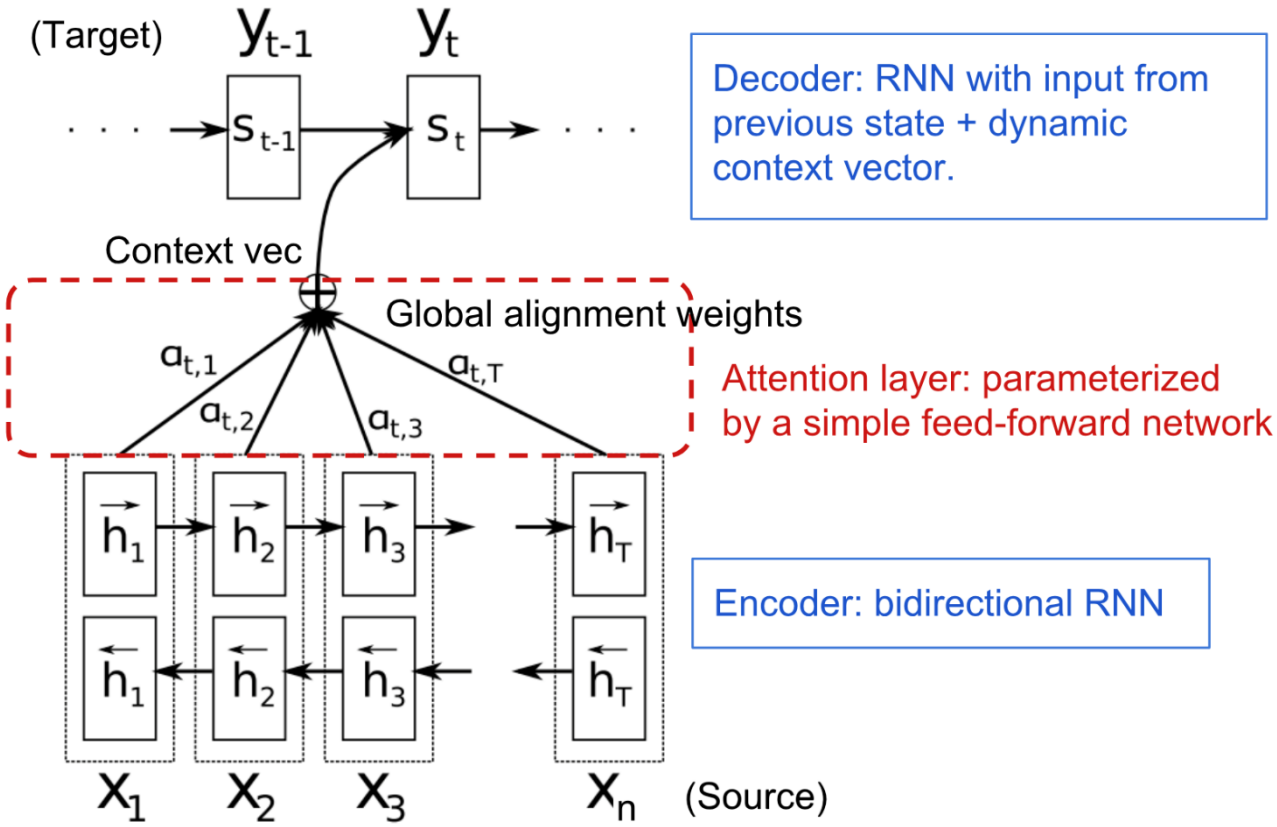
$$e_{ij} = a(s_{i-1}, h_j)$$

$$s \rightarrow \quad h \rightarrow \quad \rightarrow a$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

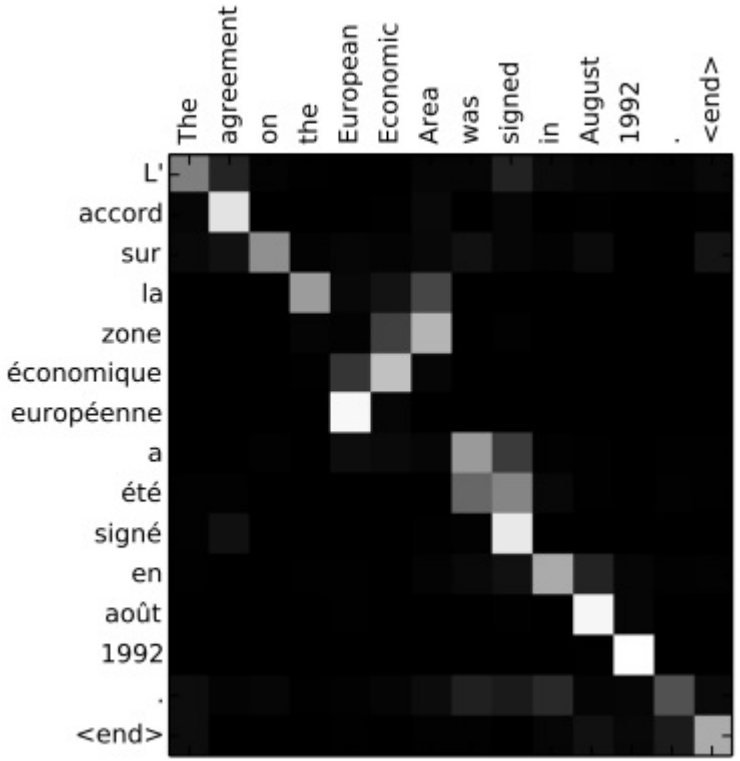
Attention



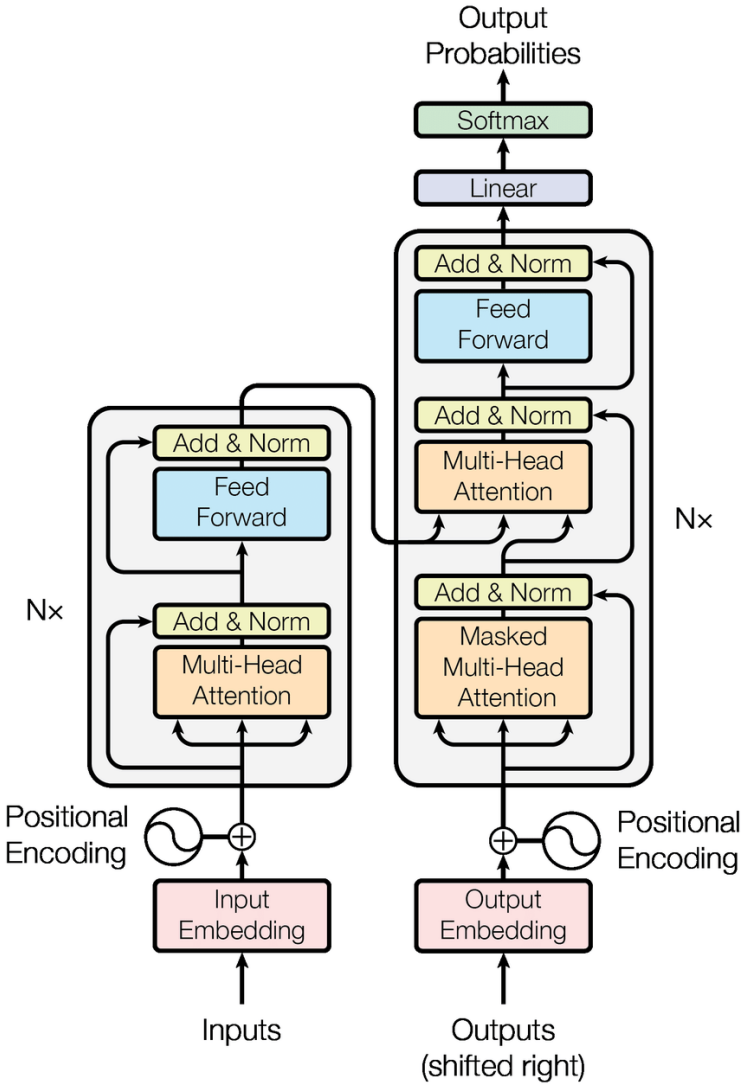
Decoder: RNN with input from previous state + dynamic context vector.

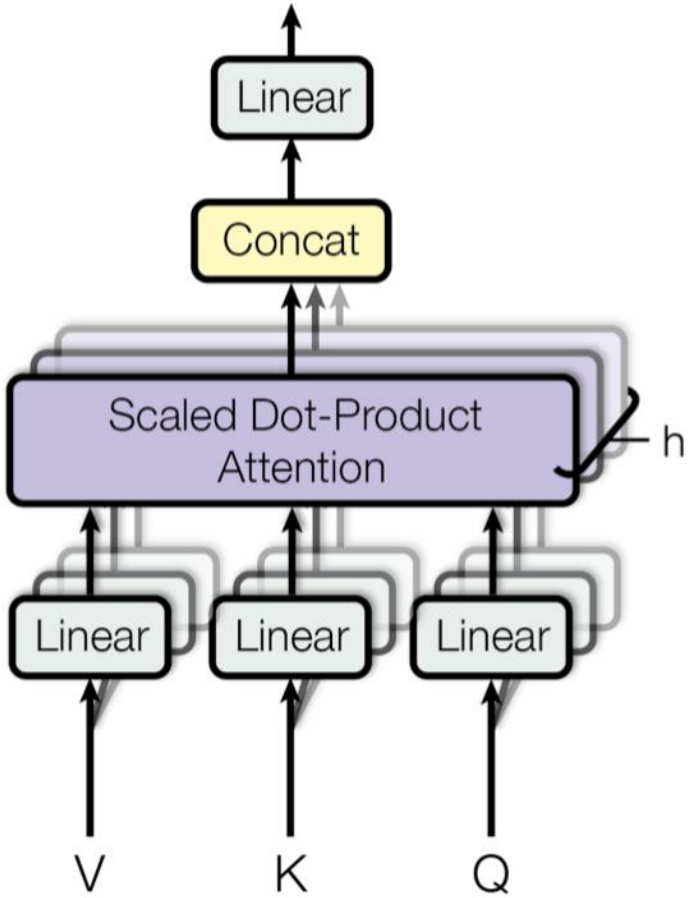
Attention layer: parameterized by a simple feed-forward network

Encoder: bidirectional RNN

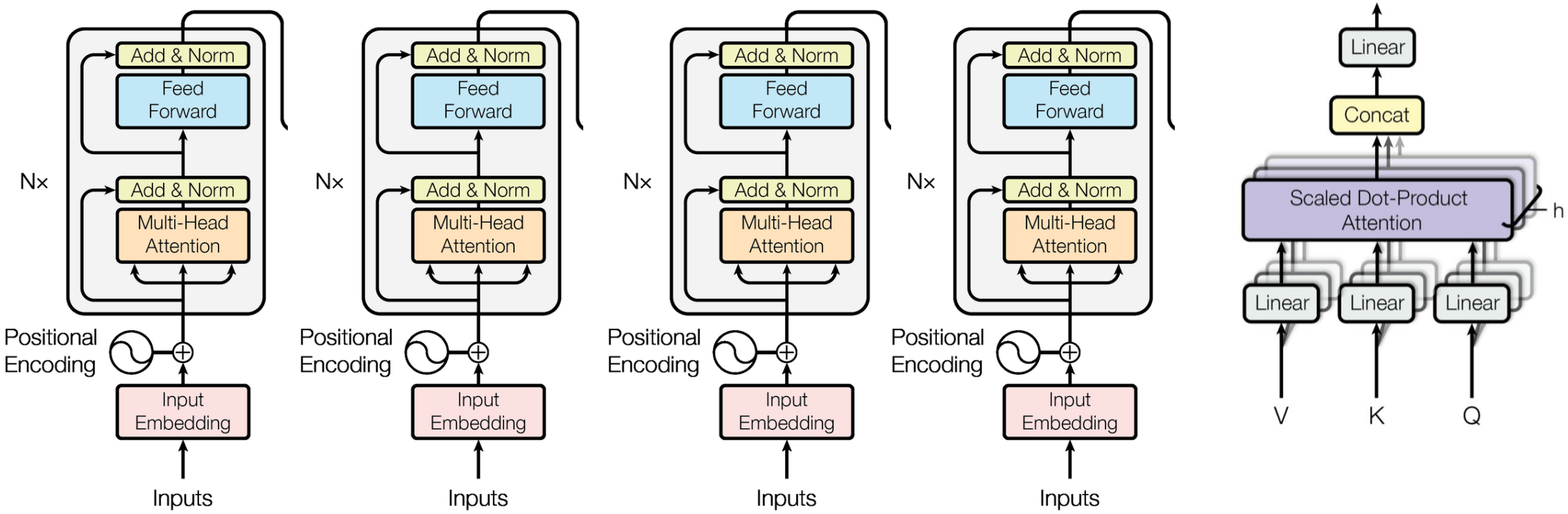


Transformers

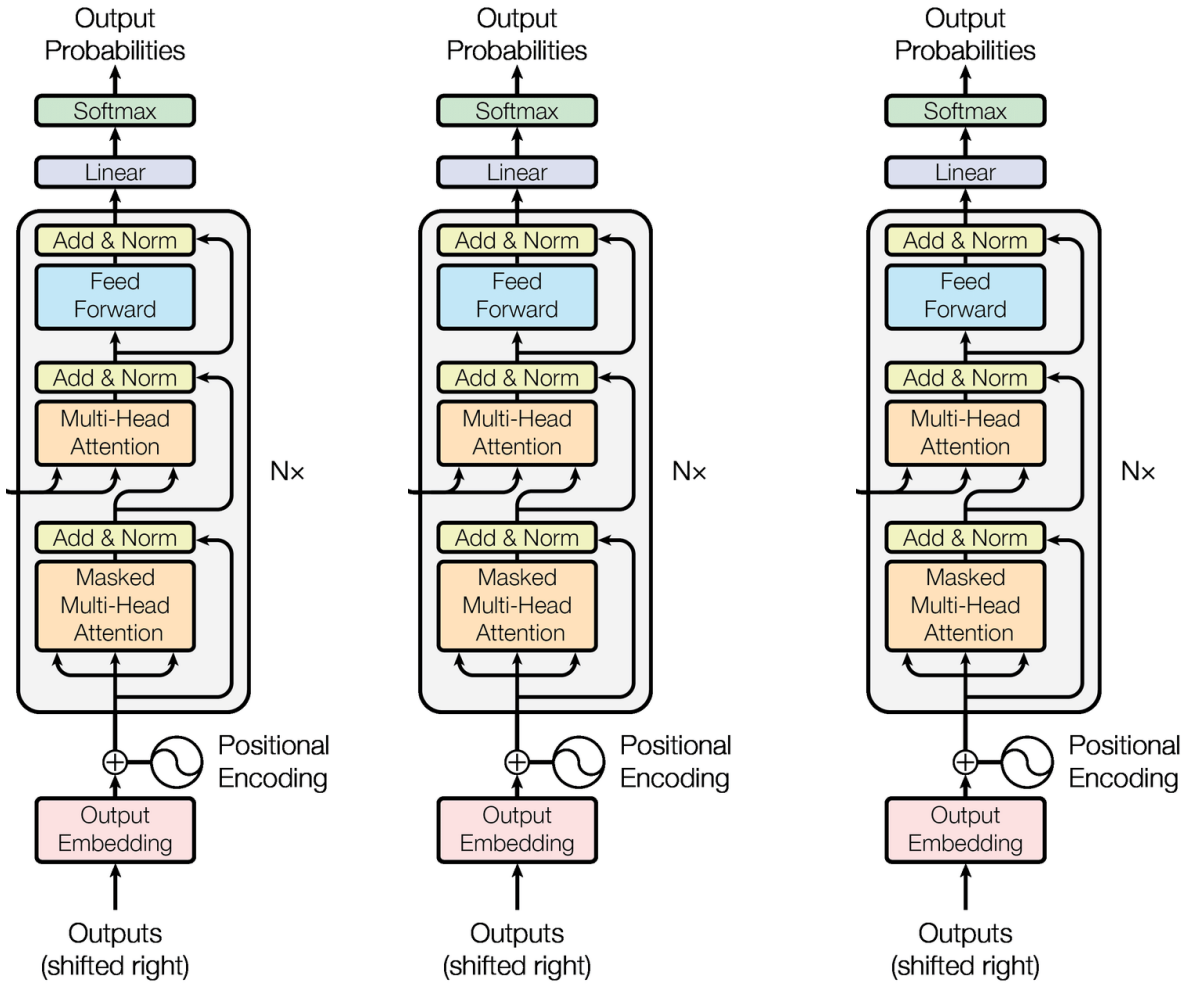




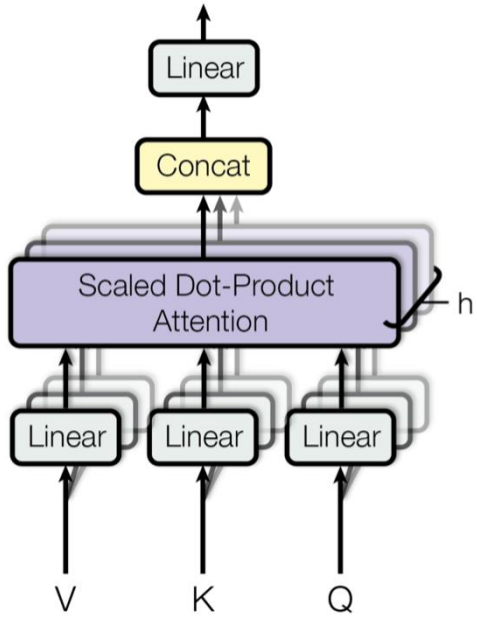
Transformers. Encoder



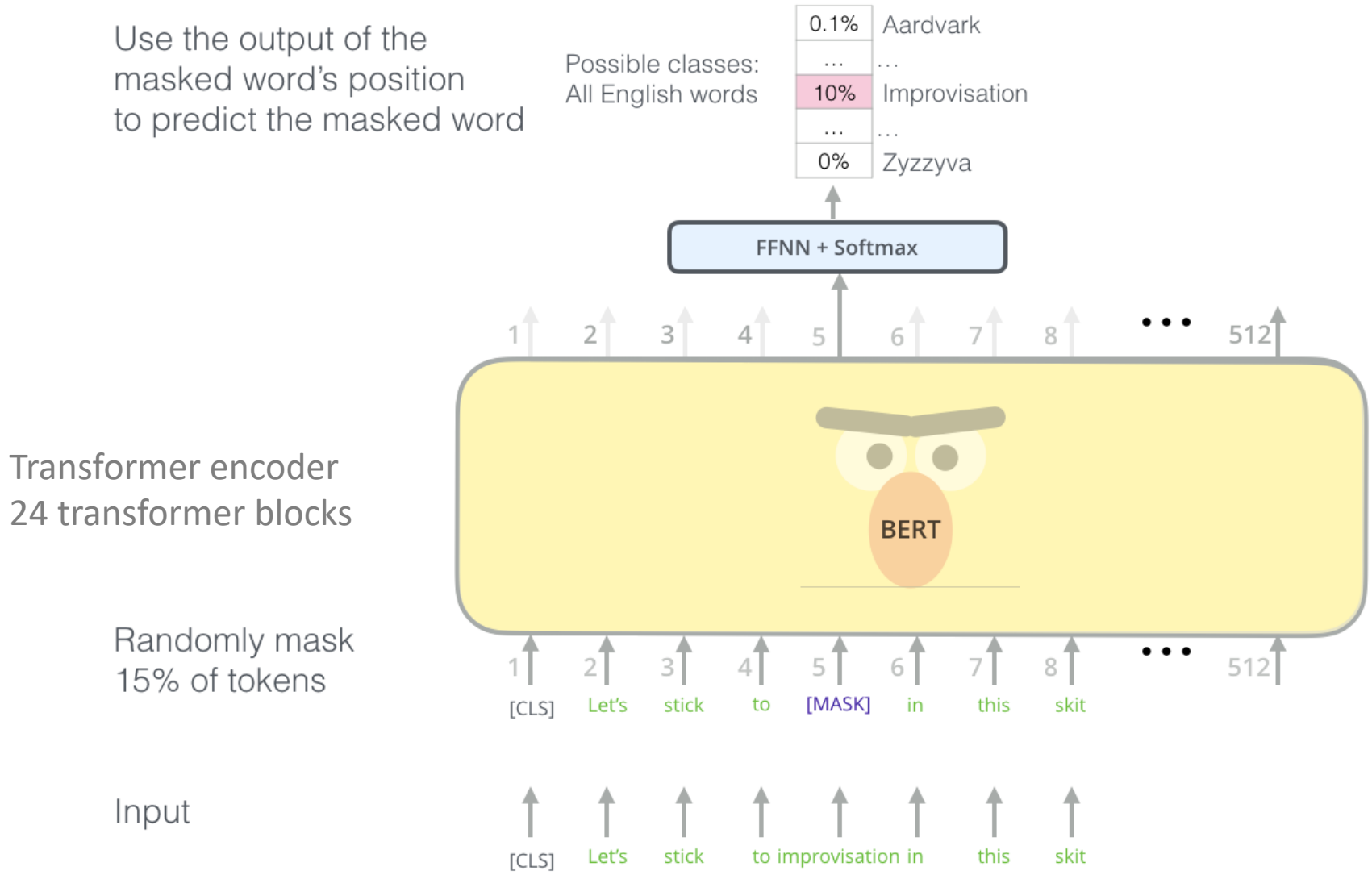
Transformers. Decoder



<BOS> I want …<EOS>



Bidirectional Encoder Representations from Transformers (BERT)



http://jalammar.github.io/illustrated-bert/

Byte Pair Encoding (BPE)

astronomy
astrology
topology
topography
geology
geometry
geography
astronomist
topologist
geologist

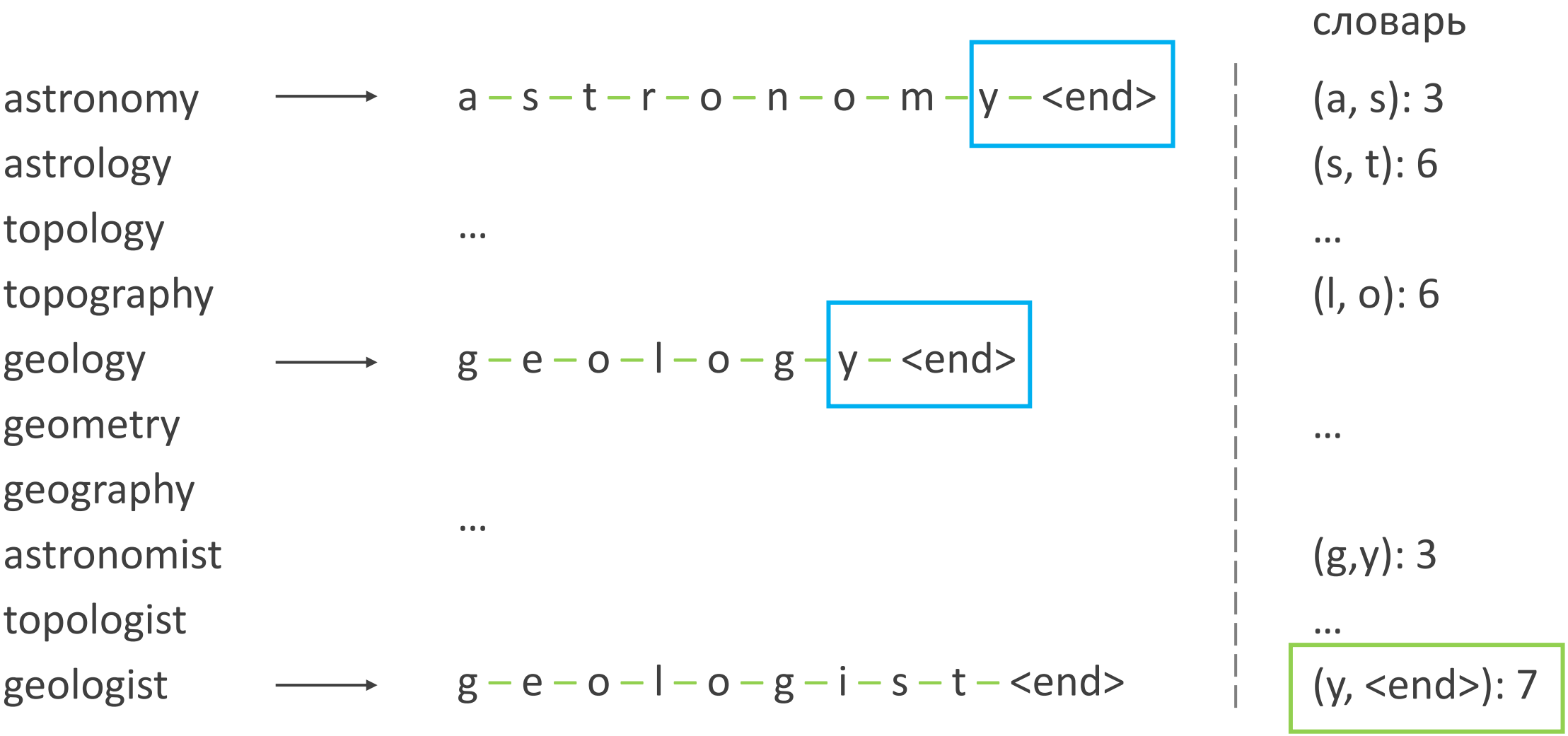
ast
astr
astro
astrol

top
topo
topol
topolo

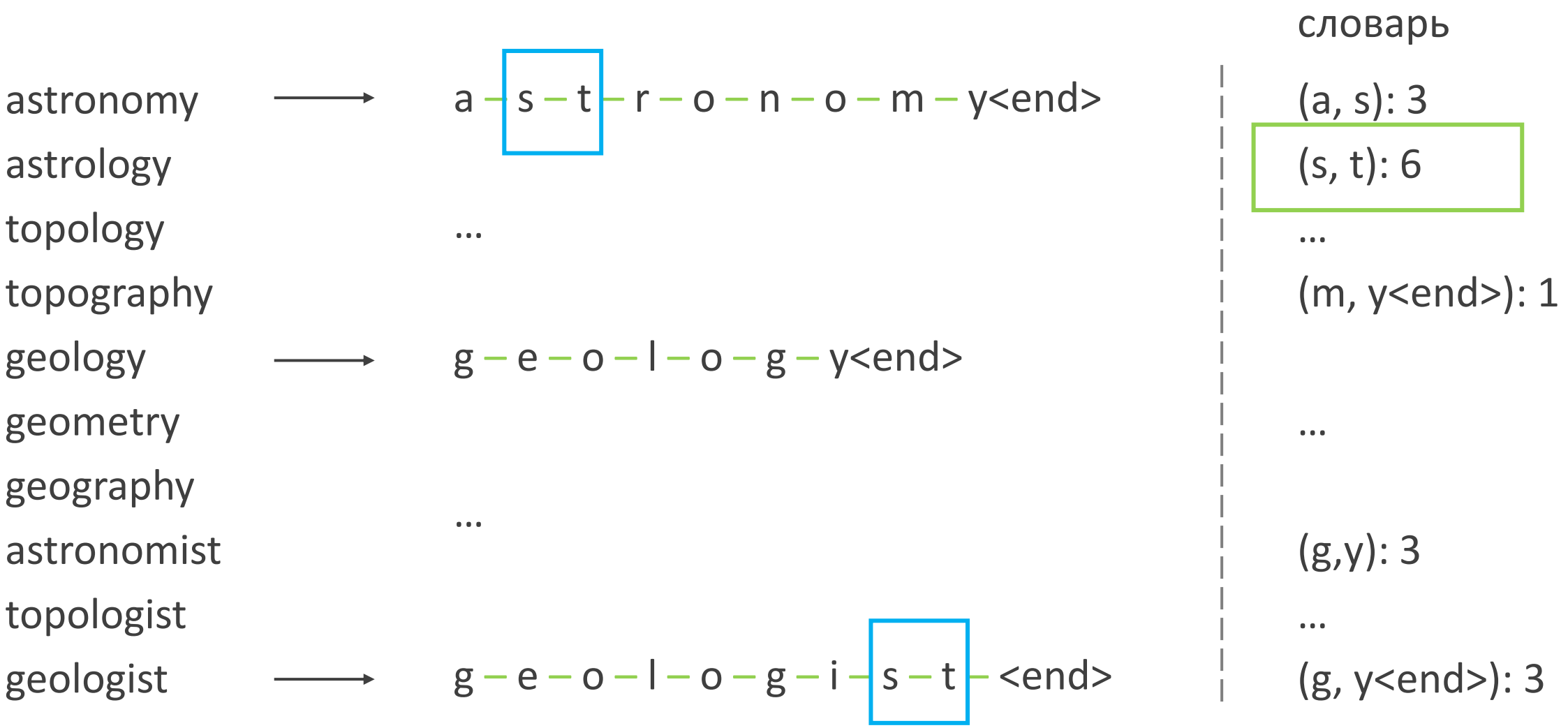
logist
ogist
gist
ist

n-граммы букв могут передавать **смысл** слова

Byte Pair Encoding (BPE)



Byte Pair Encoding (BPE)



Byte Pair Encoding (BPE)

словарь

astronomy → a – st – r – o – n – o – m – y<end>

astrology

topology …

topography

geology → g – e – o – l – o – g – y<end>

geometry

geography

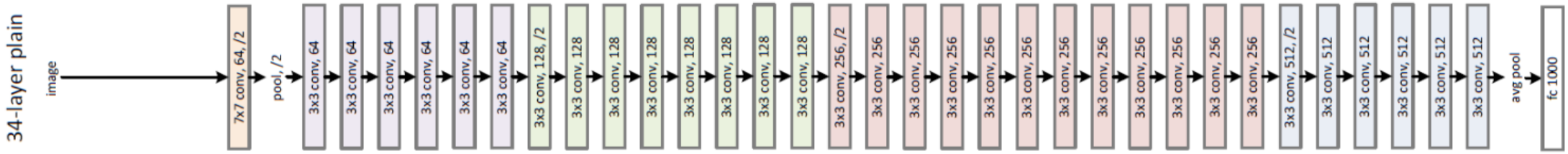
astronomist …

topologist

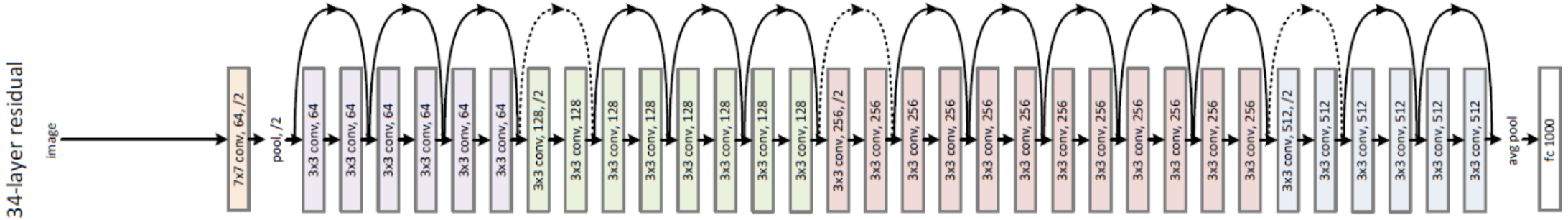
geologist → g – e – o – l – o – g – i – st – <end>

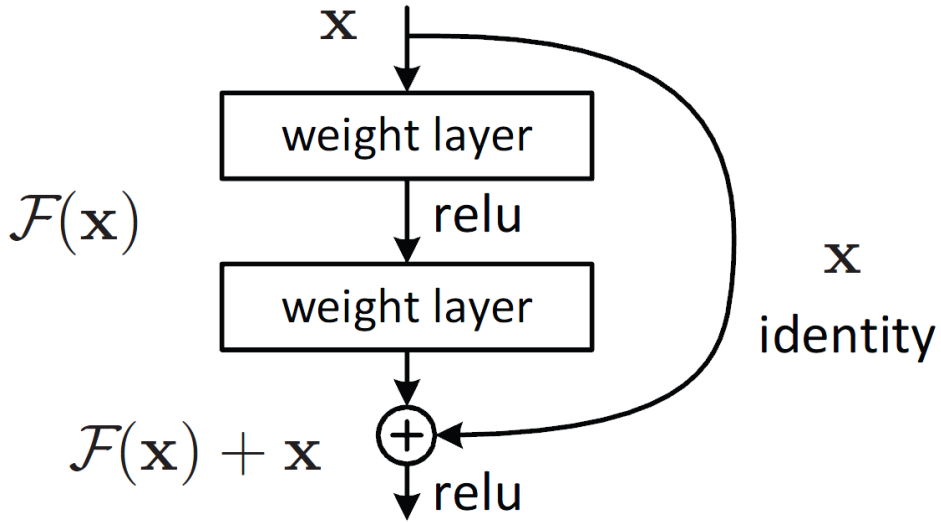
продолжать пока размер словаря допустимый

Residual Networks (ResNet)



Residual Networks (ResNet)





Residual Networks (ResNet)