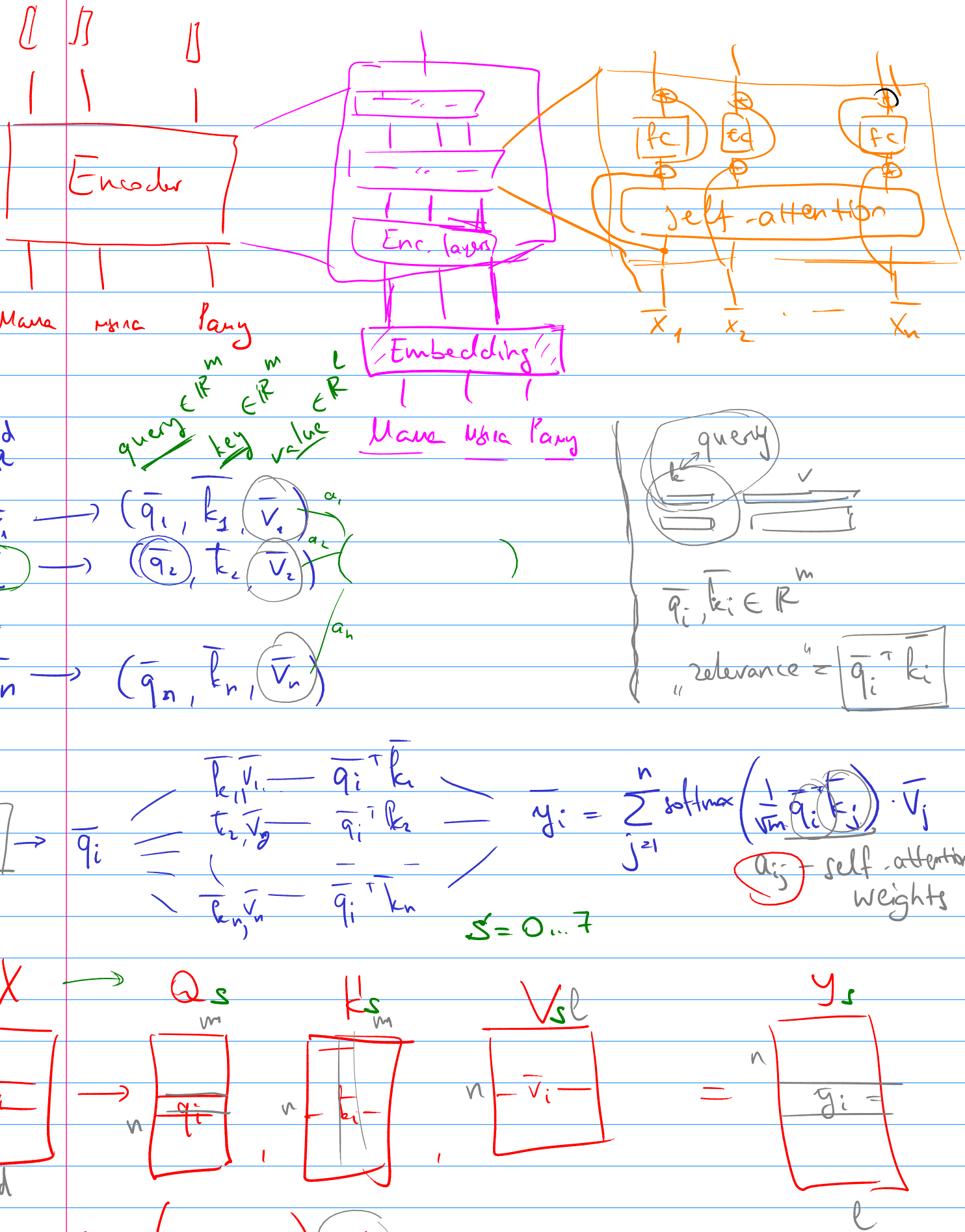
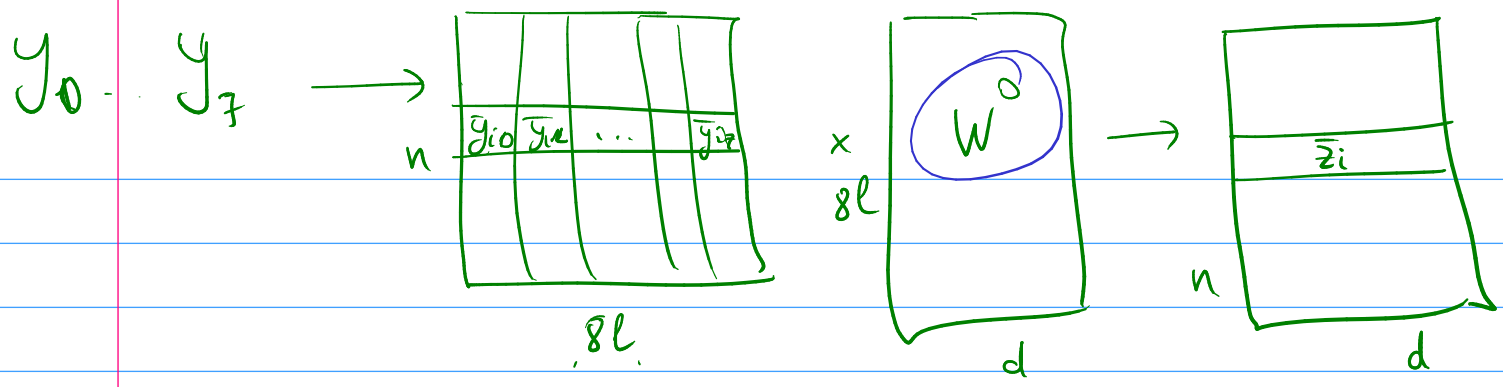

Encoder
Enc. layer
Embedding
fc
fc
fc
self-attention
Мама мыла раму
query
key
value
query
relevance
softmax
self-attention weights
s = 0 ... 7
softmax



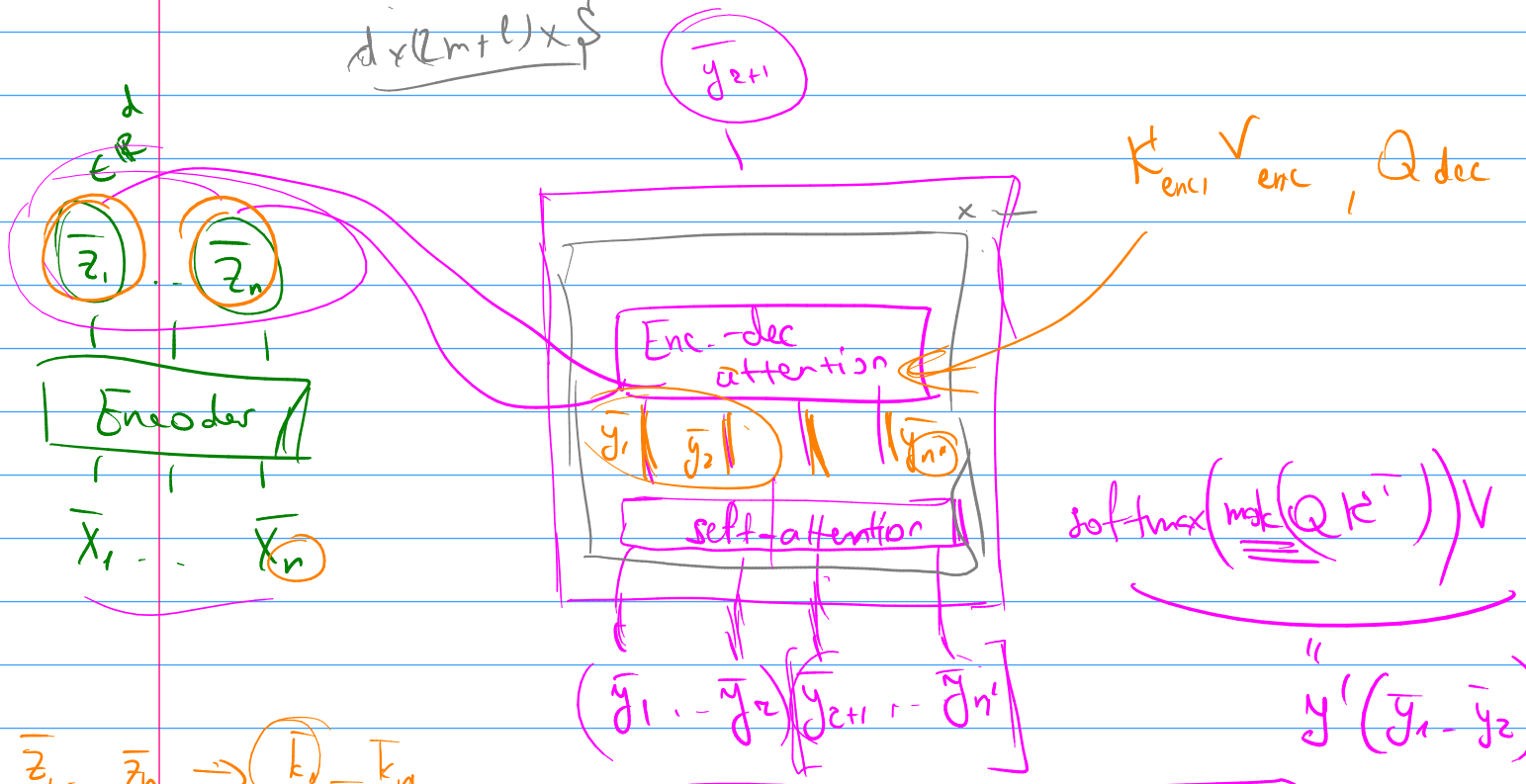
$$XW_0^q = Q_0, \quad XW_0^k = K_0, \quad XW_0^v = V_0$$

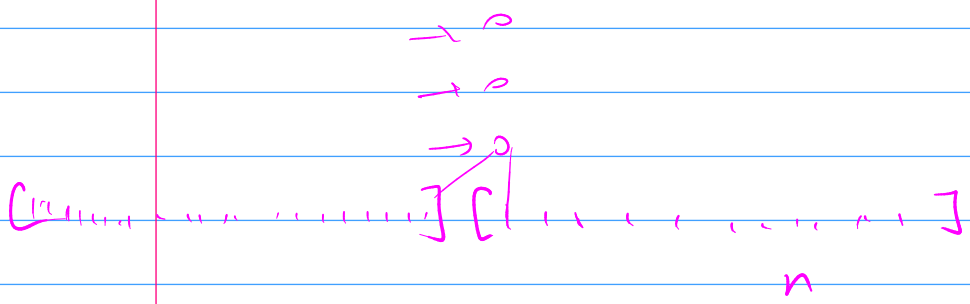
$$X \rightarrow XW_7^q = Q_7 \qquad XW_7^v = V_7$$

Reca: W_s^q, W_s^k, W_s^v, $s = 0 \dots S-1$, W^0

$d \times m$ $d \times m$ $d \times l$ $\quad Sl \times d$

$d \times (2m+l) \times S$




n