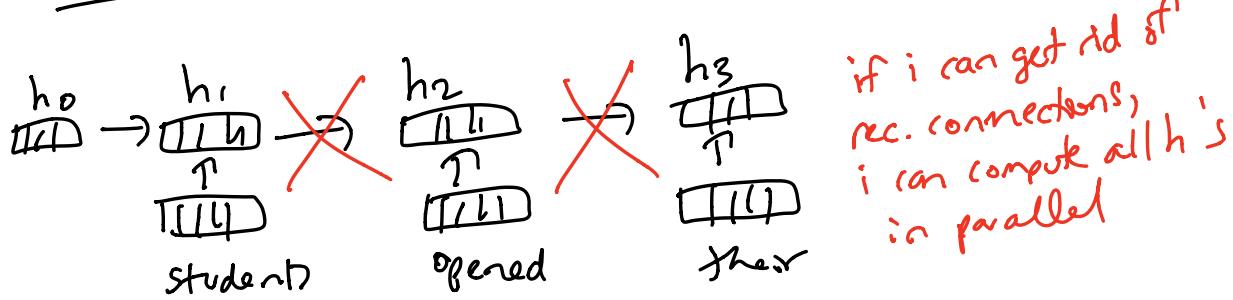
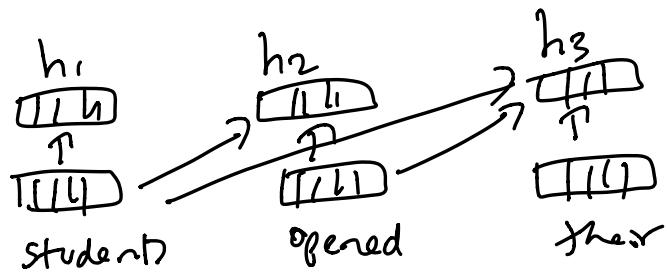


## Self-attention:

motivation: get rid of recurrent computation



goal:



Self-attr  
① 3<sup>rd</sup> pos. of  
say

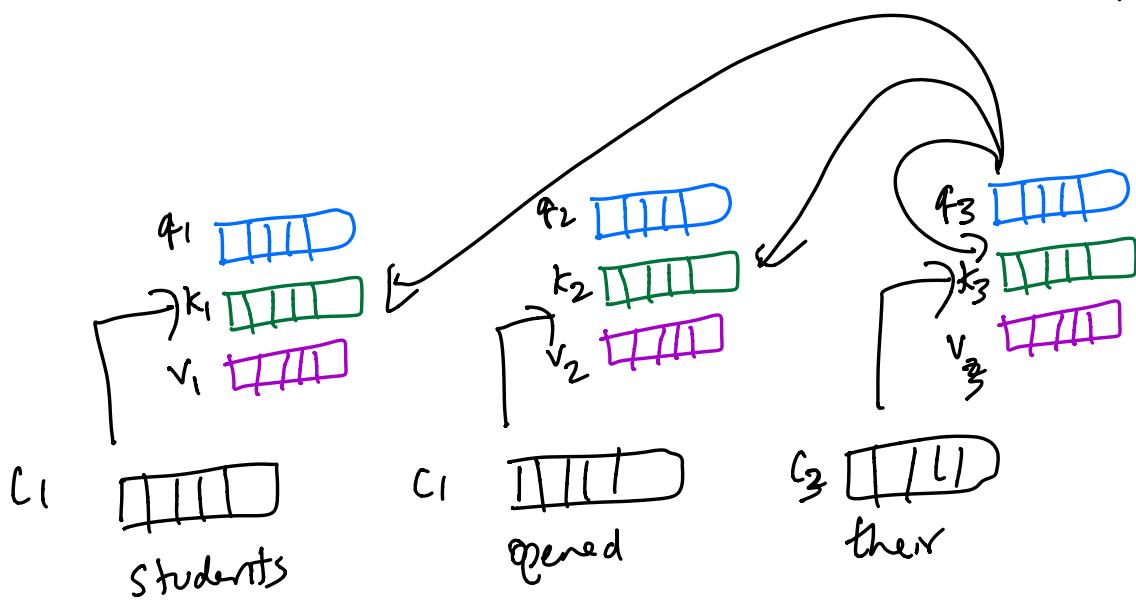
$$h_3 = 0.3 \cdot v_1 + 0.5 \cdot v_2 + 0.2 \cdot v_3$$

$$= \begin{array}{|c|c|c|} \hline & & \\ \hline \end{array}$$

↳ predict books

$q_3 k_1 \quad q_3 k_2 \quad q_3 k_3$

$$\text{softmax}((q_3 k_1, q_3 k_2, q_3 k_3))$$



$$q_1 = f(W_q c_1)$$

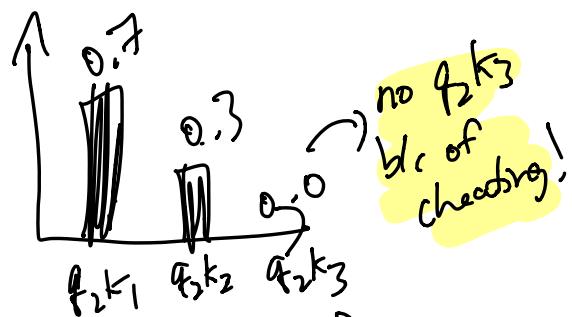
$$k_1 = f(W_k c_1)$$

$$v_1 = f(W_v c_1)$$

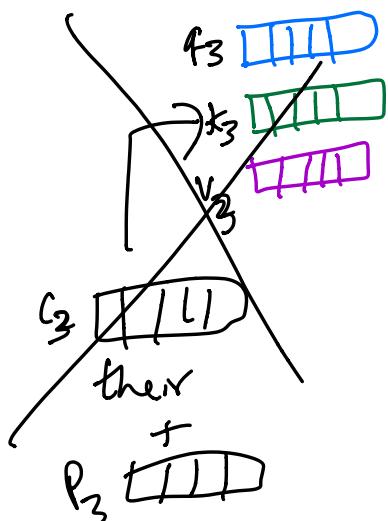
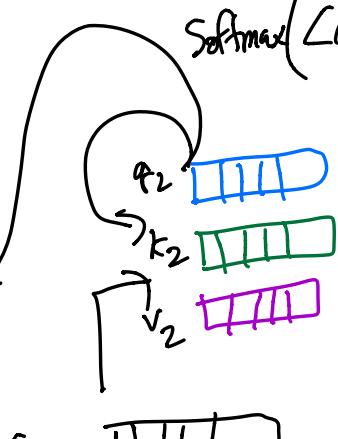
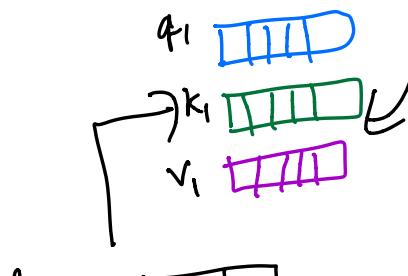
Self-attention  
@ 2<sup>nd</sup> pos of seq

$$h_2 = 0.7 \cdot v_1 + 0.3 \cdot v_2$$

~~= 1111~~ → predict "their"



$$\text{softmax}([q_2 k_1, q_2 k_2])$$



$P_1$  [white hatching]

students

$P_2$  [white hatching]

$P_3$  [white hatching]

Now, there's no dependencies between  $h_1, h_2, h_3$

how can we compute  $h_i$ 's in parallel?

$q_1$    
 $q_2$    
 $q_3$  

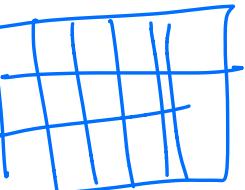
  $k_1$ ,  
  $k_2$ ,  
  $k_3$

attn scores

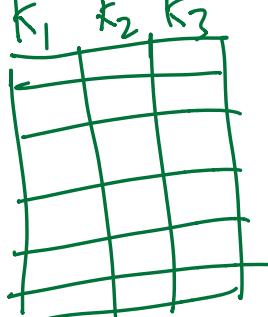
$$a_1 = \langle q_1, k_1 \rangle$$

$$a_2 = \langle q_2, k_1, q_2, k_2 \rangle$$

$$a_3 = \langle q_3, k_1, q_3, k_2, q_3, k_3 \rangle$$

$f_1$    
 $f_2$   
 $q_3$

$\times$

$k_1 \quad k_2 \quad k_3$   


$= f_1 \quad k_1 \quad k_2 \quad k_3$   
 $f_2 \quad \cdot \quad \cdot \quad \cdot$   
 $q_3 \quad \cdot \quad \cdot \quad \cdot$

$k_1 \quad k_2 \quad k_3$   
 $q_1 \quad \cdot \quad \cdot \quad \cdot$   
 $f_2 \quad \cdot \quad \cdot \quad \cdot$   
 $q_3 \quad \cdot \quad \cdot \quad \cdot$

these cells contain  
info about future words,  
need to exclude!

Softmat  $\left( \begin{array}{c|ccc} & k_1 & k_2 & k_3 \\ \hline q_1 & \cdot & \cdot & \cdot \\ f_2 & \cdot & \cdot & \cdot \\ q_3 & \cdot & \cdot & \cdot \end{array} \right) \cdot \left( \begin{array}{c|ccc} & k_1 & k_2 & k_3 \\ \hline 1 & -\infty & -\infty \\ 1 & 1 & -\infty \\ 1 & 1 & 1 \end{array} \right) =$

$$\begin{array}{c} \begin{array}{c} k_1 & k_2 & k_3 \\ \hline f_1 & \cdot & 0 & 0 \\ f_2 & \cdot & \cdot & 0 \\ f_3 & \cdot & \cdot & \cdot \end{array} \times \begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} \end{array} =$$

$$\begin{array}{c} h_1 \\ h_2 \\ h_3 \end{array} \begin{array}{|c|c|c|c|} \hline & | & | & | \\ \hline \end{array}$$