

Attention (P-3) (LLM notes)

- Ex from last page:

- in our Ex. imagine

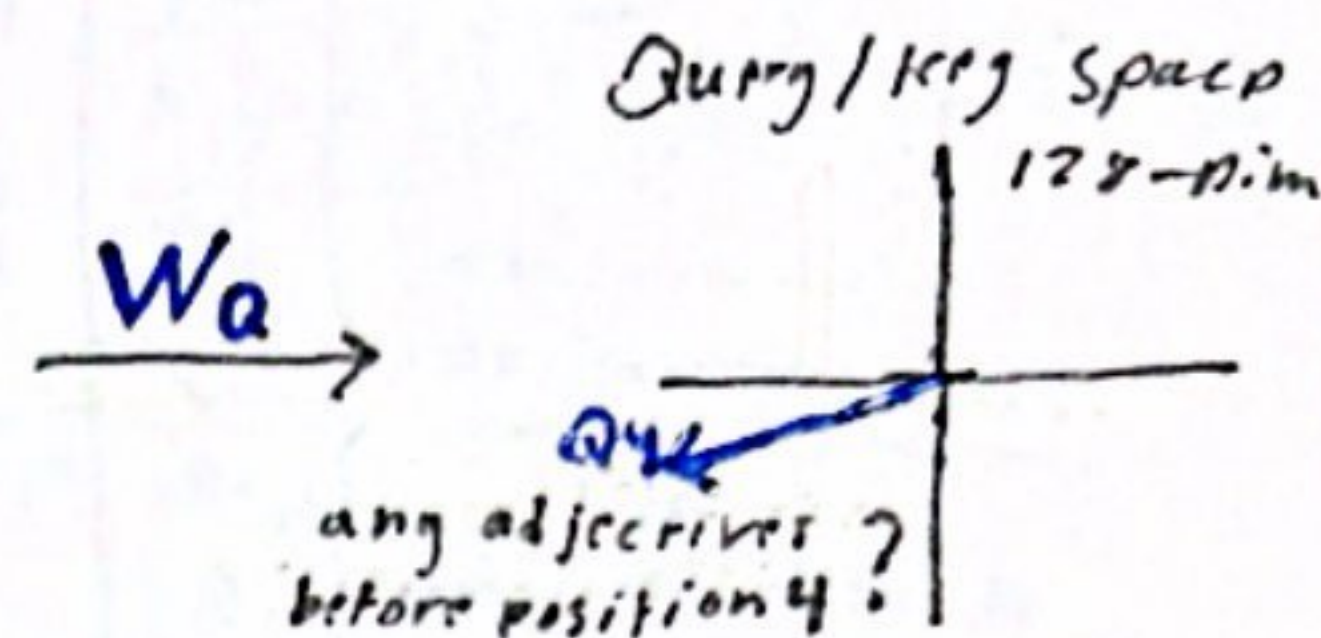
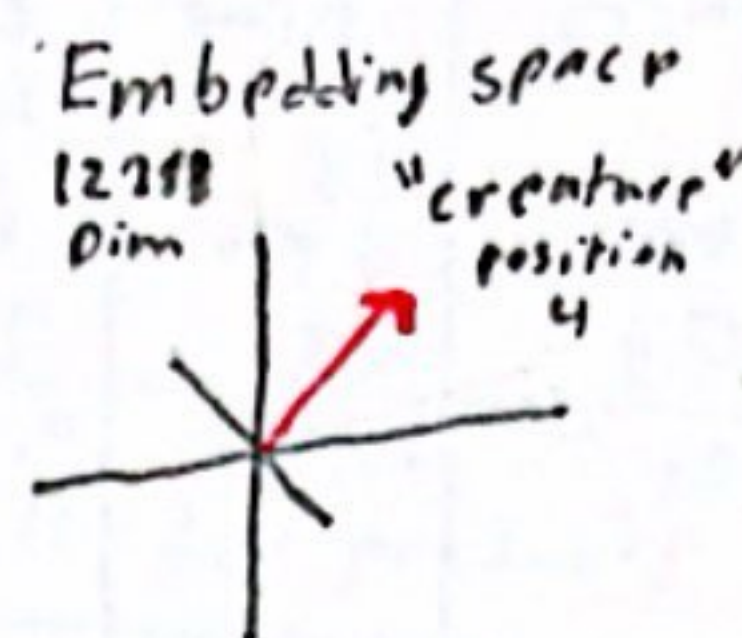
a example where we might hope that the W_Q would map the embeddings of nouns to certain directions in the Query space, that somehow encodes the notion of looking for adjectives in preceding positions. as for non-nouns what W_Q does is unknown.

- At the same time associated with this is a second matrix called the Key Matrix that you also multiply by all the embeddings that produces a second sequence of vectors called the Keys. Conceptually think of the keys as answering the question (query) also W_K is learned in training and behavior is unknown in real life. and just like W_Q it maps embedding to a smaller space.

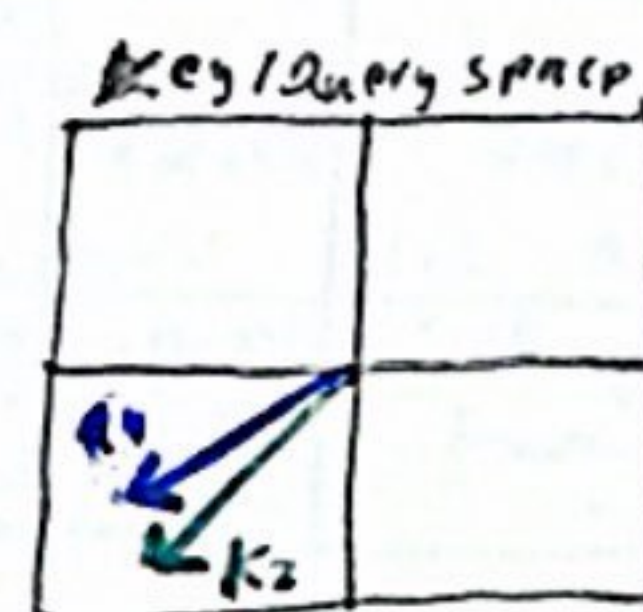
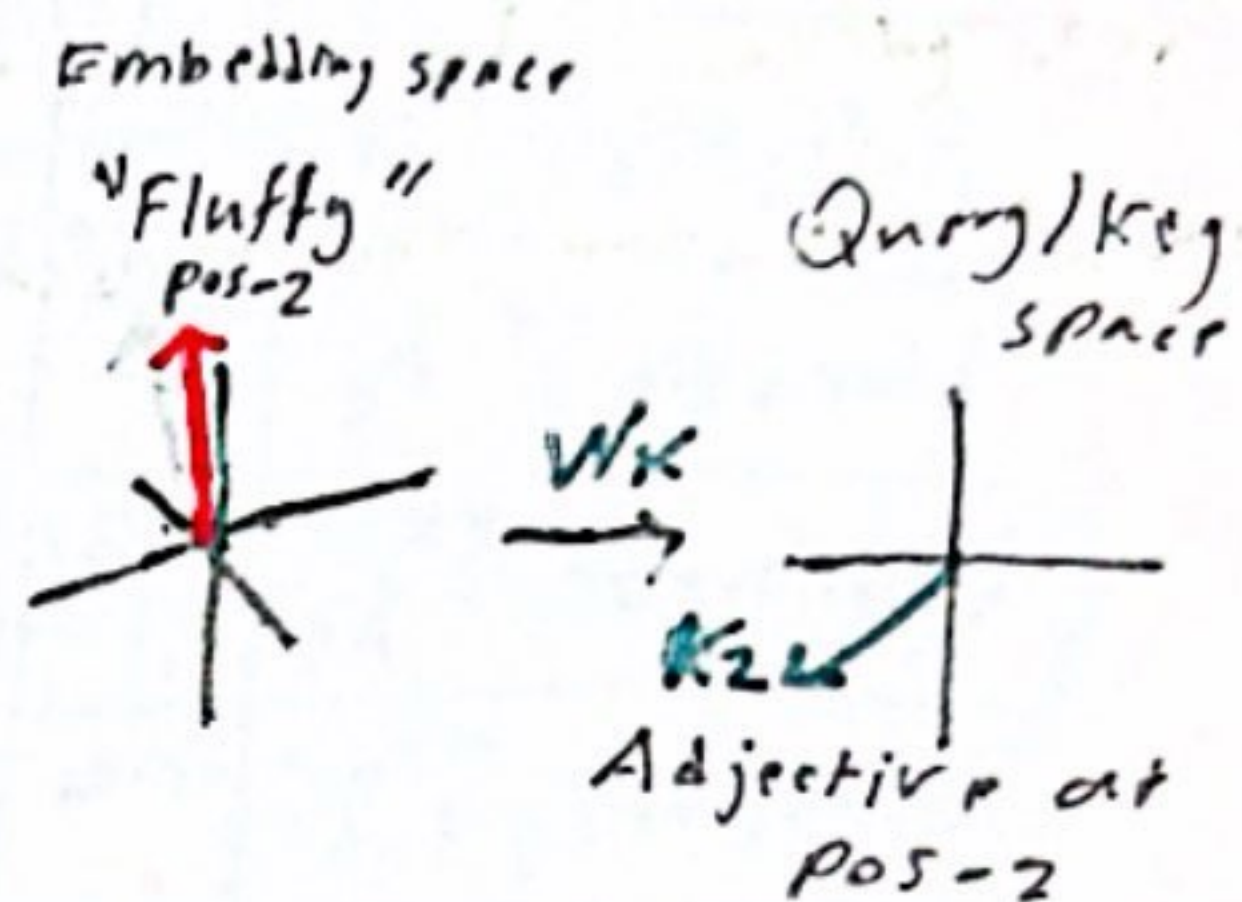
$$\begin{bmatrix} 1 & 2 & 9 & \dots & 10 & 11 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \xrightarrow{W_K \text{ is a matrix}} E_i \xrightarrow{W_K} K_i$$

means $E_1 \cdot W_K = K_1$

a Fluffy blue creature roamed the verdant Forest
 $\downarrow W_Q \quad \downarrow W_Q \quad \downarrow W_Q \quad \downarrow W_Q \quad \downarrow W_Q \quad \downarrow W_Q \quad \downarrow W_Q \quad \downarrow W_Q$
 $Q_1 \quad Q_2 \quad Q_3 \quad Q_4 \quad Q_5 \quad Q_6 \quad Q_7 \quad Q_8$



$a \rightarrow E_1 \xrightarrow{W_K} K_1$
 Fluffy $\rightarrow E_2 \xrightarrow{W_K} K_2$
 Blue $\rightarrow E_3 \xrightarrow{W_K} K_3$
 creature $\rightarrow E_4 \xrightarrow{W_K} K_4$
 ...



\rightarrow Query
 \rightarrow Key

$$K_2 \cdot Q_4$$

- you think of key as matching the Query whenever they closely align with one another in our ex you would imagine that key matrix maps adjectives "fluffy, blue" to vectors that are closely aligned with the Query produced by the nouns "creature"
- to measure how well each key matches each Query you compute a dot product between each Key Query vector pair. The larger the dot product the closer they align is more confidence in answer (key)