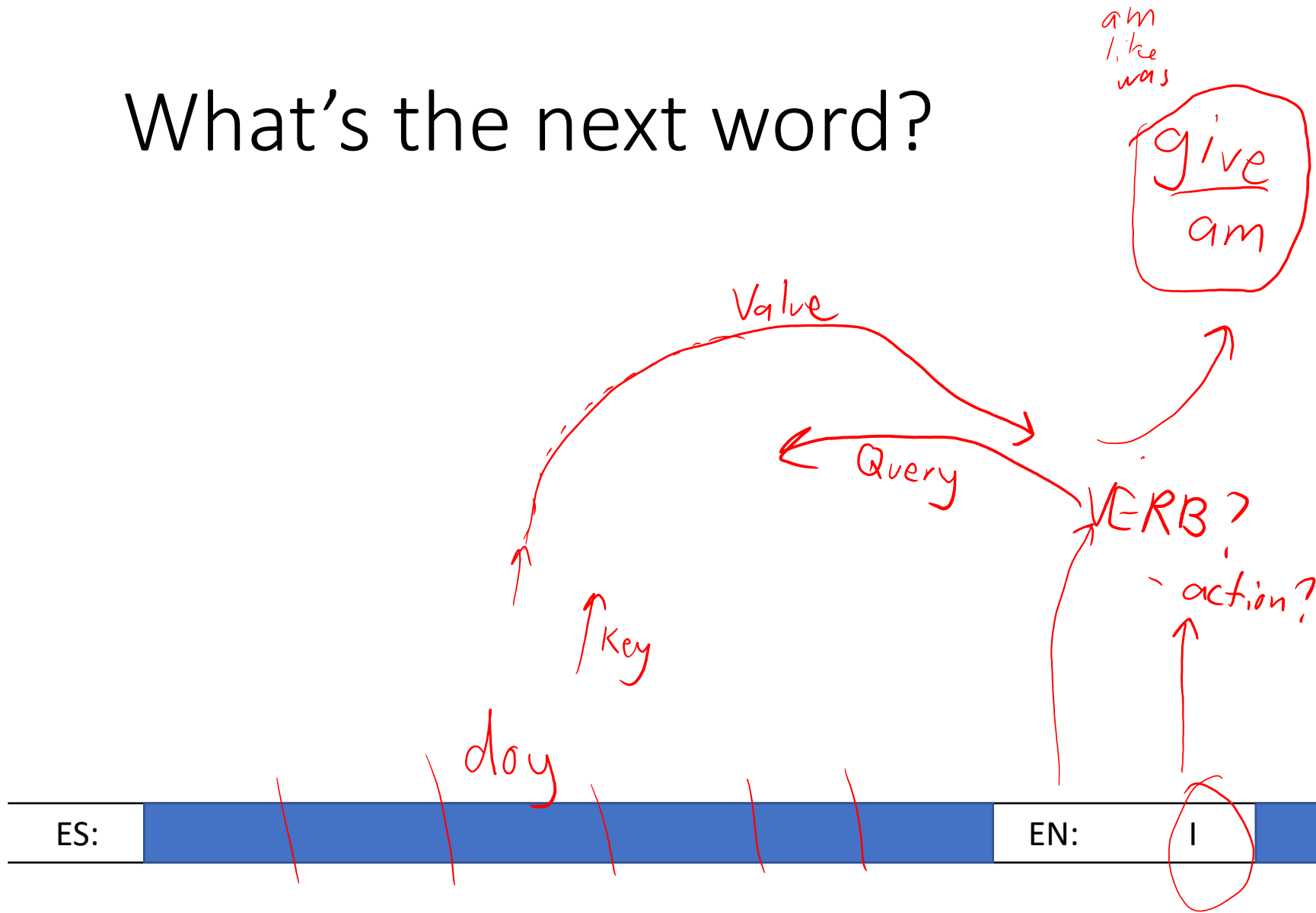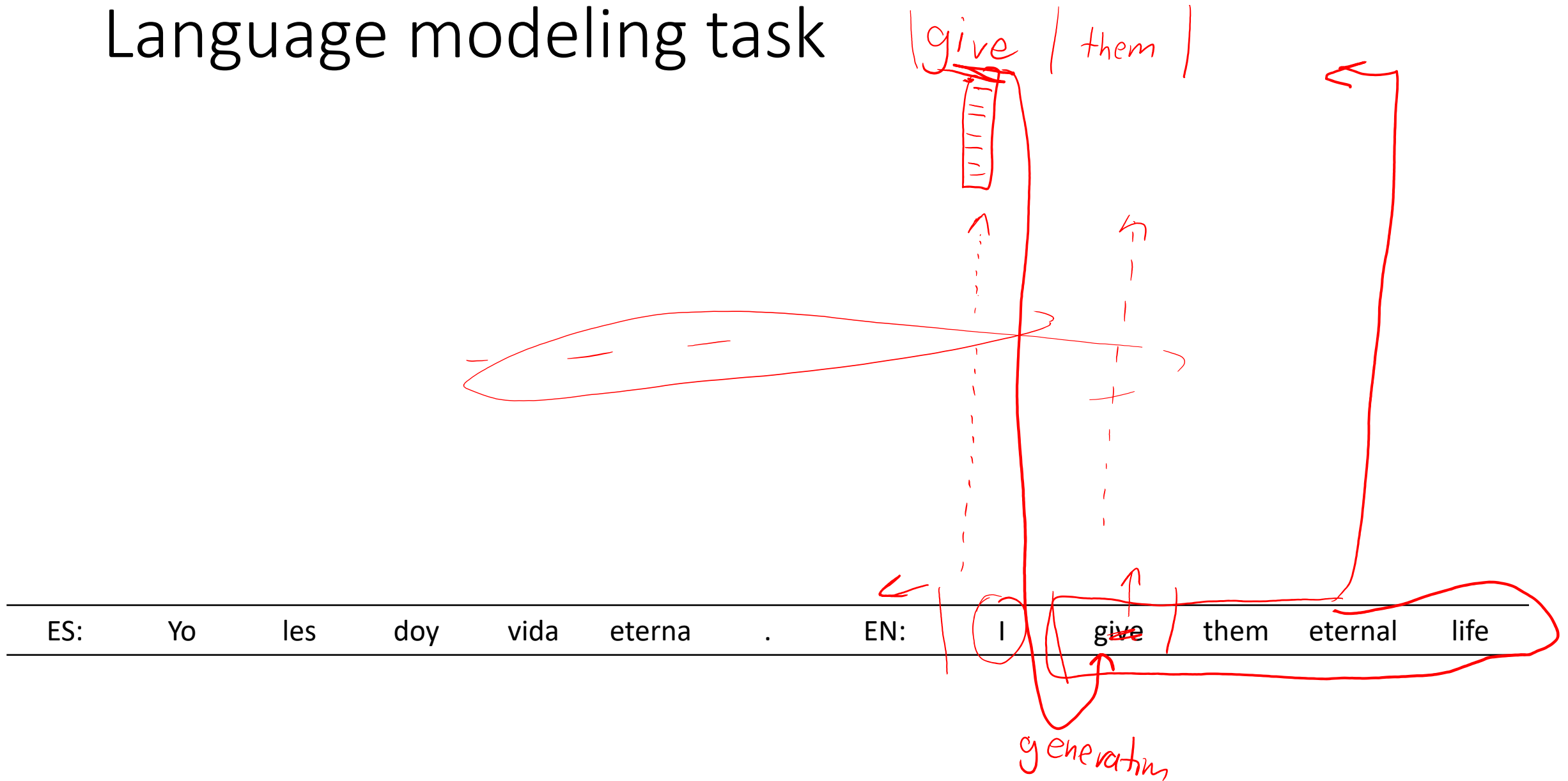# Transformers

2021 - 04 - 21

Objectives

- Trace the flow of data through a Transformer model
- Explain how Transformers use position information
- Explain the purpose of the Query, Key, and Value vectors for attention

What's the next word?

Language modeling task

P(give | them)

ES: Yo les doy vida eterna . EN: I give them eternal life

generation

query
comparison?

$$\sum_i w_i V_i$$

$$w_i = Q K_i$$
verb?   verblyness

give

.99

W_Q W_V W_k 500   verb?   K   V   verb?   Q   500   V   500   500   K   500   I

ES: Yo   les   doy   vida → eterna   .   EN: I   give   them   eternal   life
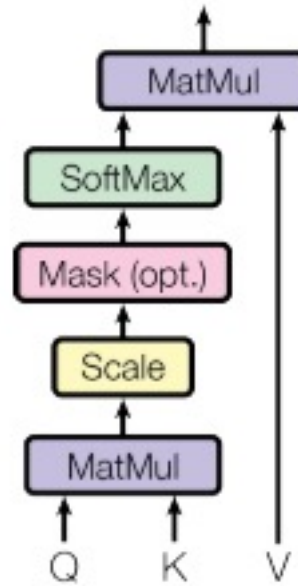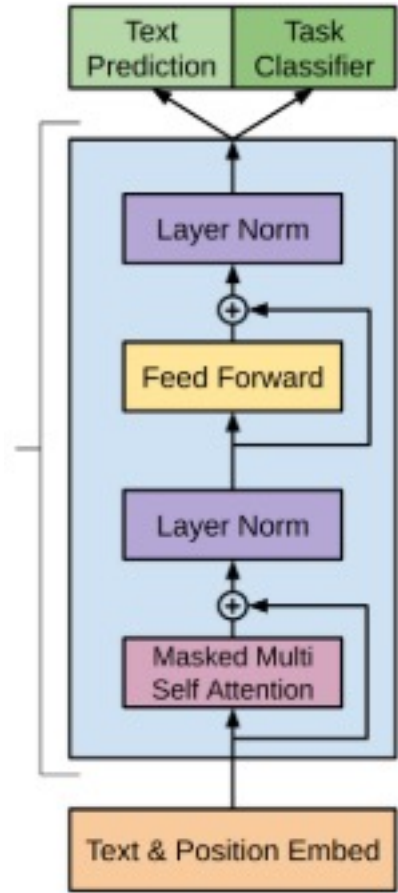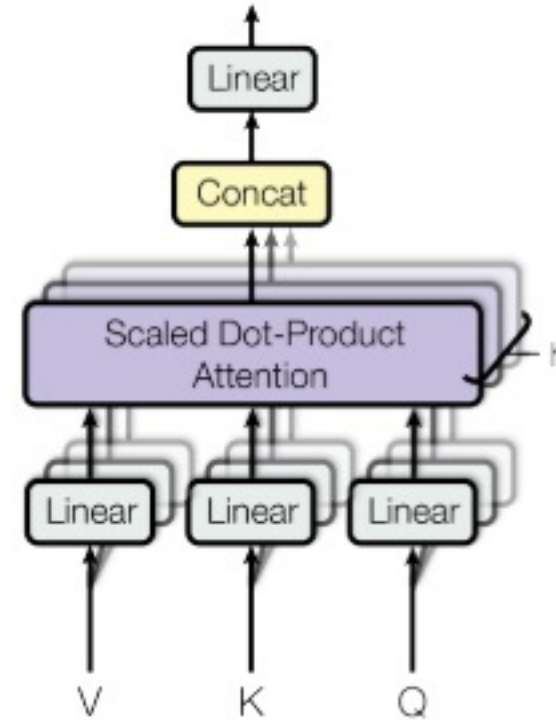
I give them eternal life .

ES: Yo les doy vida eterna . EN: I give them eternal life

GPT-2 adds…

- Feed-forward layers
- Multi-head attention
- Normalization
- Residual connections
- Dropout

- More: https://amaarora.github.io/2020/02/18/annotatedGPT2.html

Extensions

- GPT-3
- Image generation (Taming Transformers, DALL-E)
- VideoGPT

Wrap-up:
https://forms.office.com/r/jr2zcZYYRe