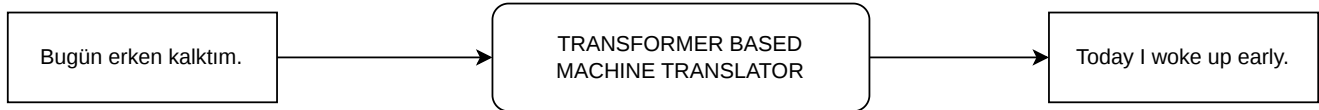
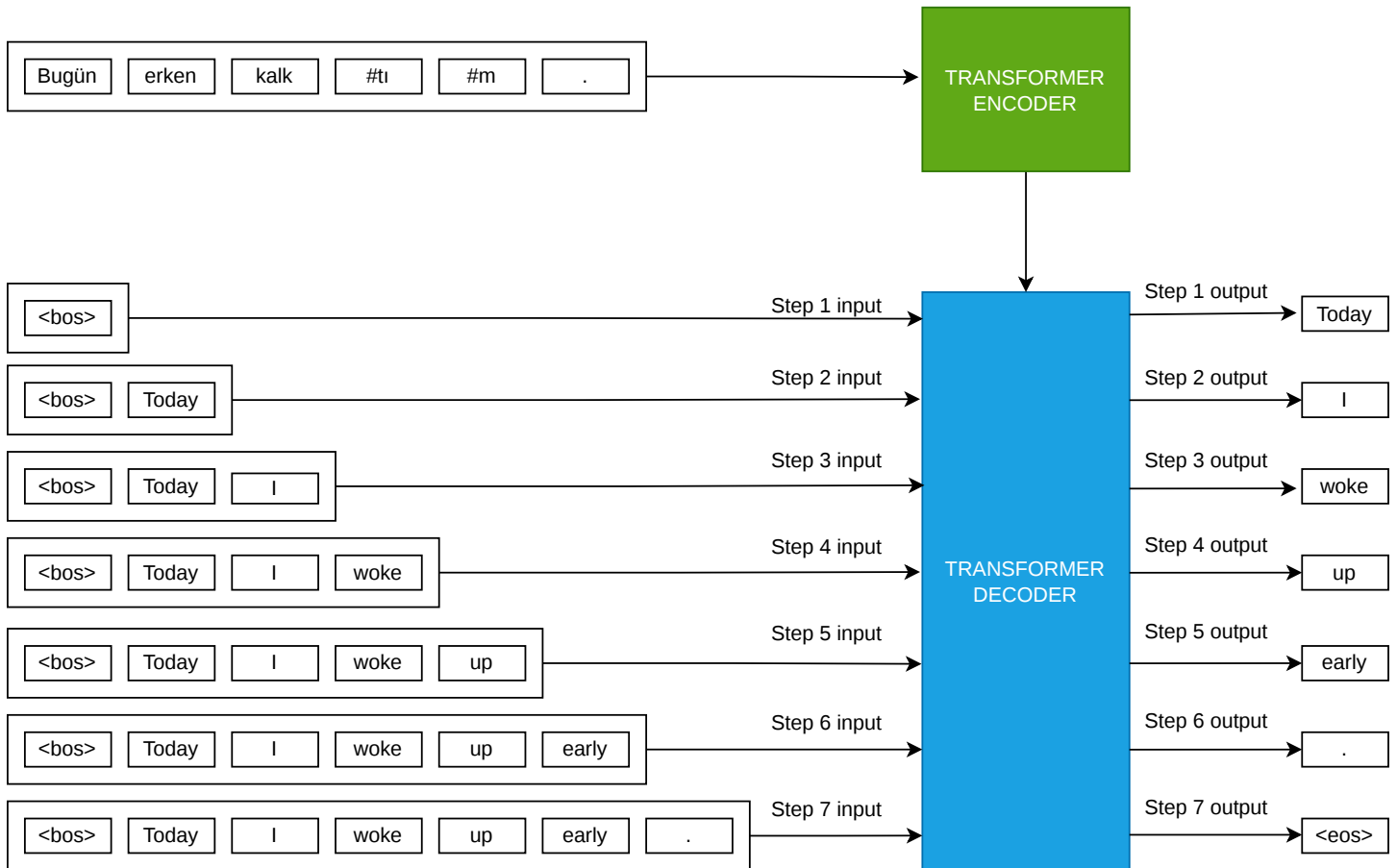


LEVEL 1

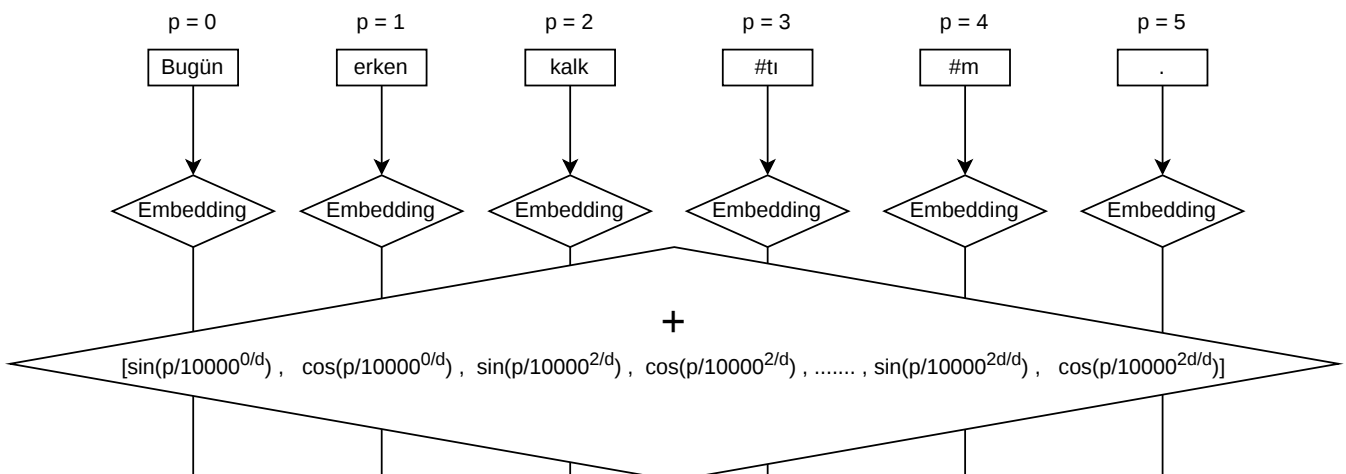


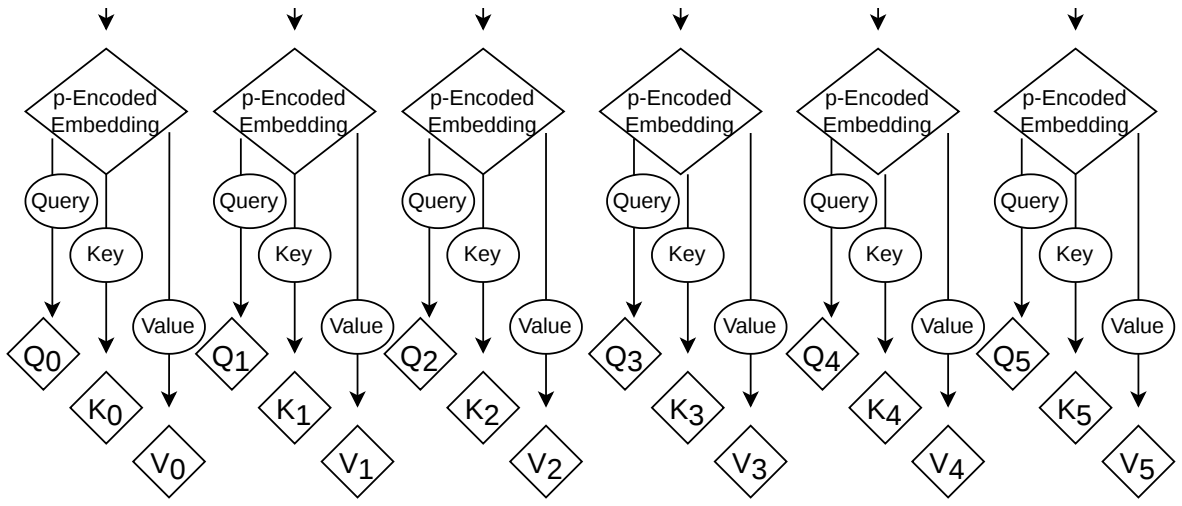
LEVEL 2



LEVEL 3 - PART 1: ENCODER LAYER

Squares are tokens, diamonds are vectors, ellipses are matrices (they multiply the input vector and spit out the output vector)





$$Attention_p = \sum_i \text{softmax}\left(\frac{Q_p \cdot K_i}{\sqrt{d_k}}\right) V_i$$

