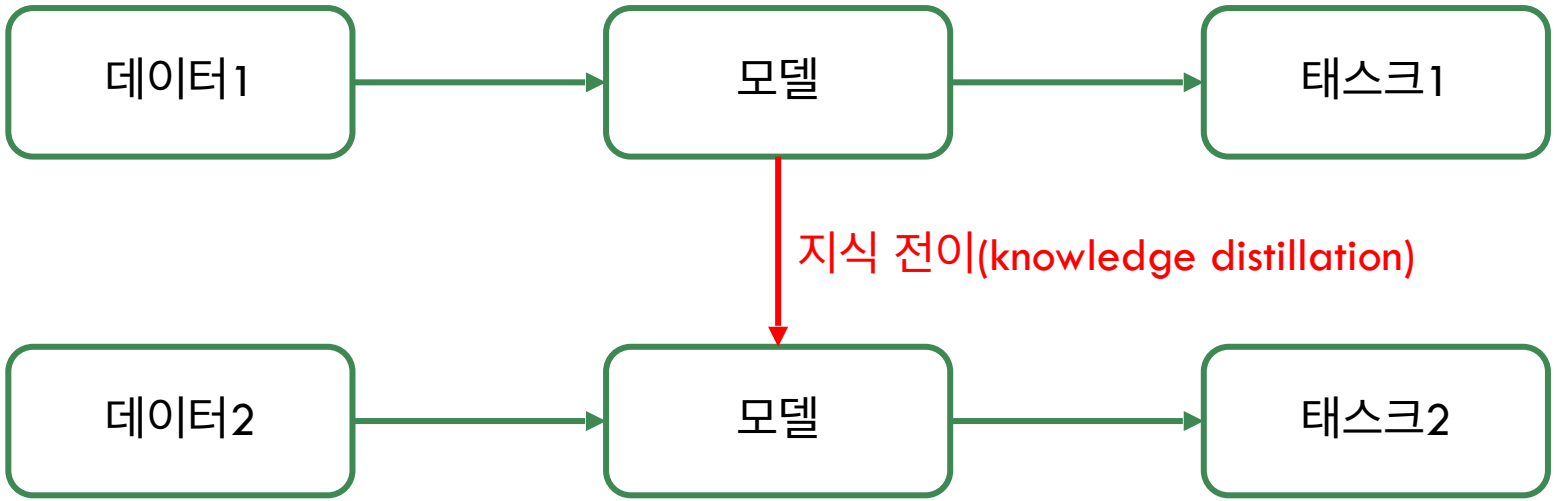

TRANSFORMER

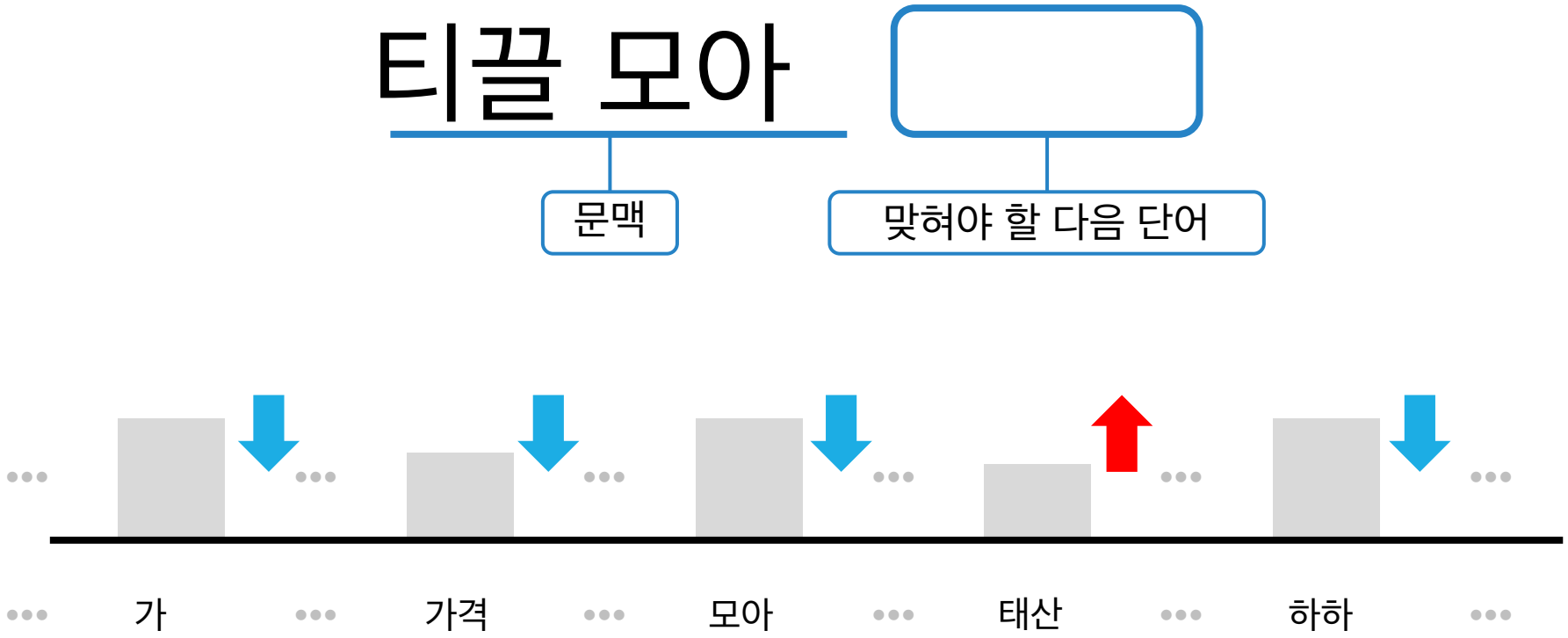
TRANSTER LEARNING



UPSTREAM

- task1과정을 upstream이라고 함
- upstream은 다음 단어 맞히기, 빈칸 채우기 등 대규모 말뭉치의 문맥을 이해하는 과정을 말함

UPSTREAM



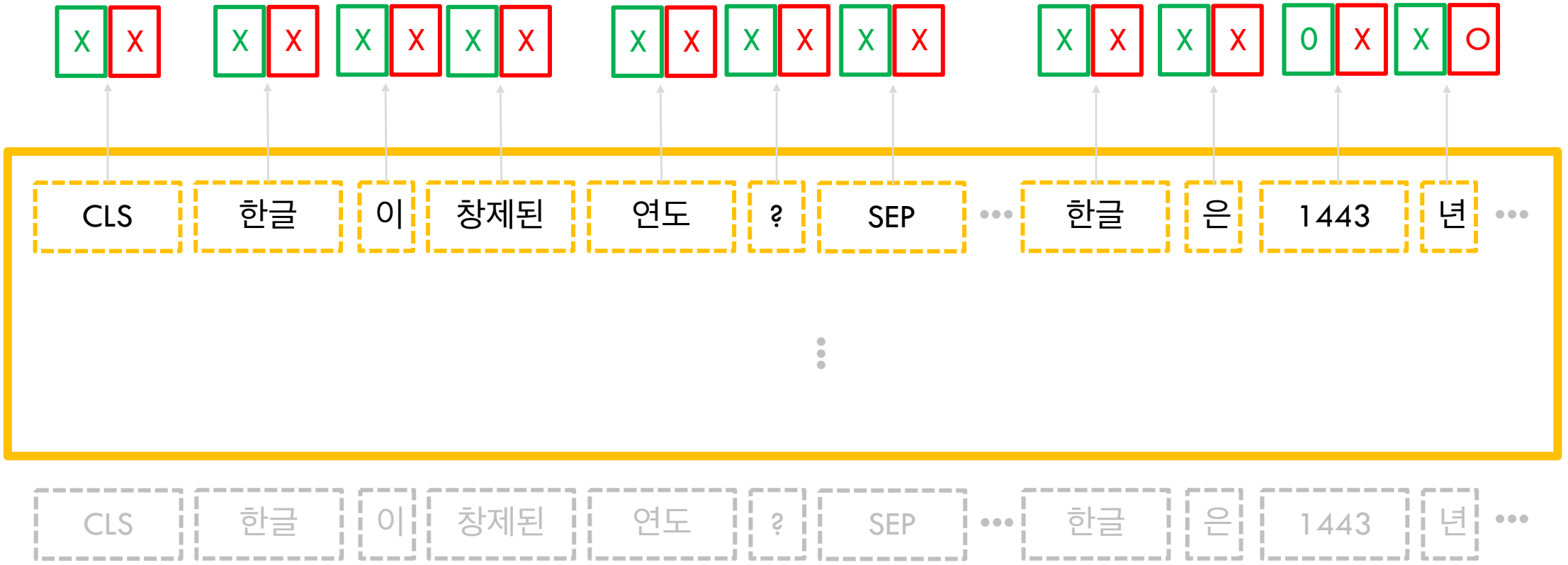
DOWNSTREAM

- pretrain을 마친 모델을 task에 맞게 업데이트는 하는 방식을 말함
- downstream은 문서 분류, 개체명 인식 등 자연어 처리의 구체적인 문제들을 푸는 과정을 말함

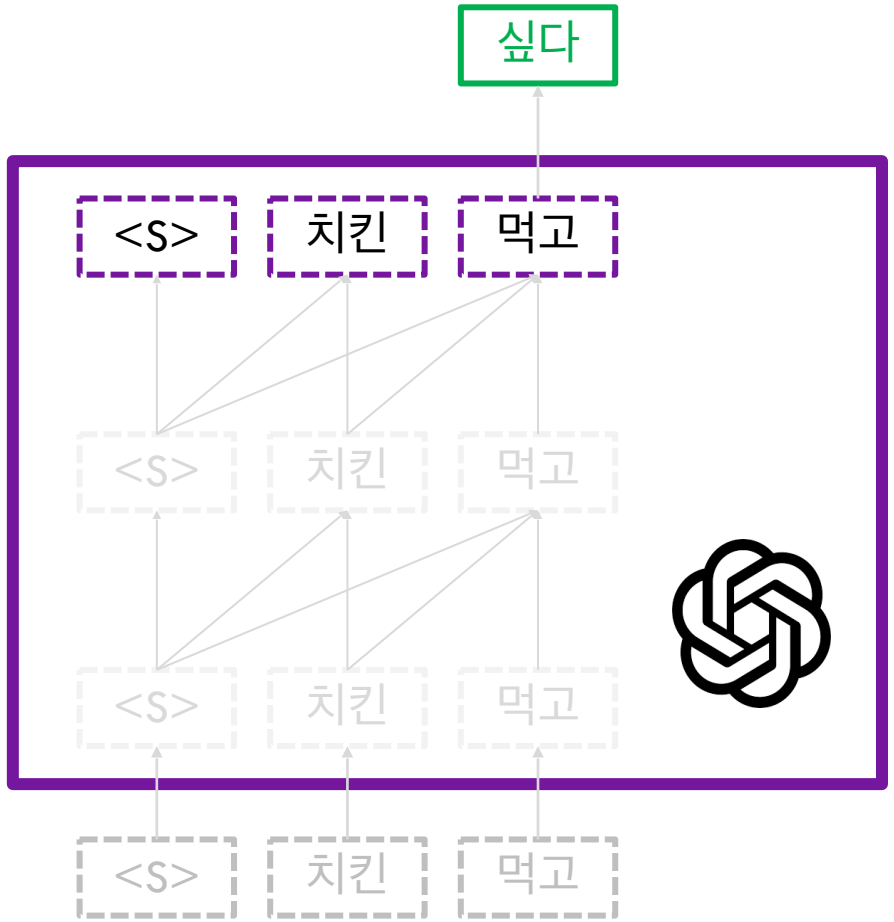
DOWNSTREAM(문서 분류)

DOWNSTREAM(문서 분류)

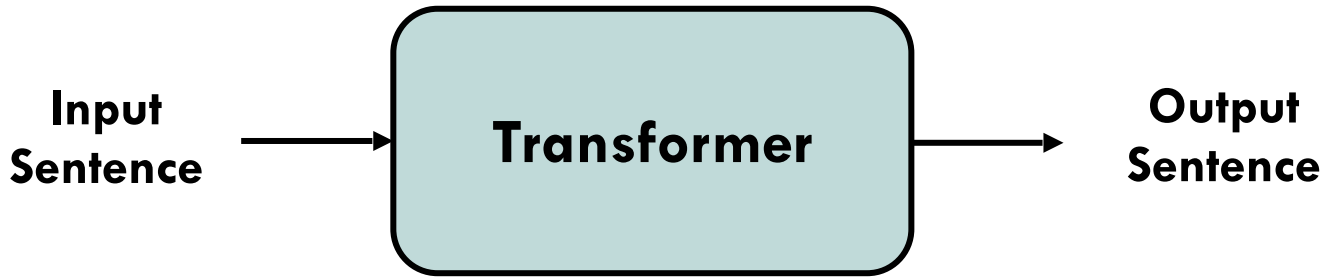
DOWNSTREAM(질의 응답)



DOWNSTREAM(문장 생성)



TRANSFORMER의 개괄적인 구조

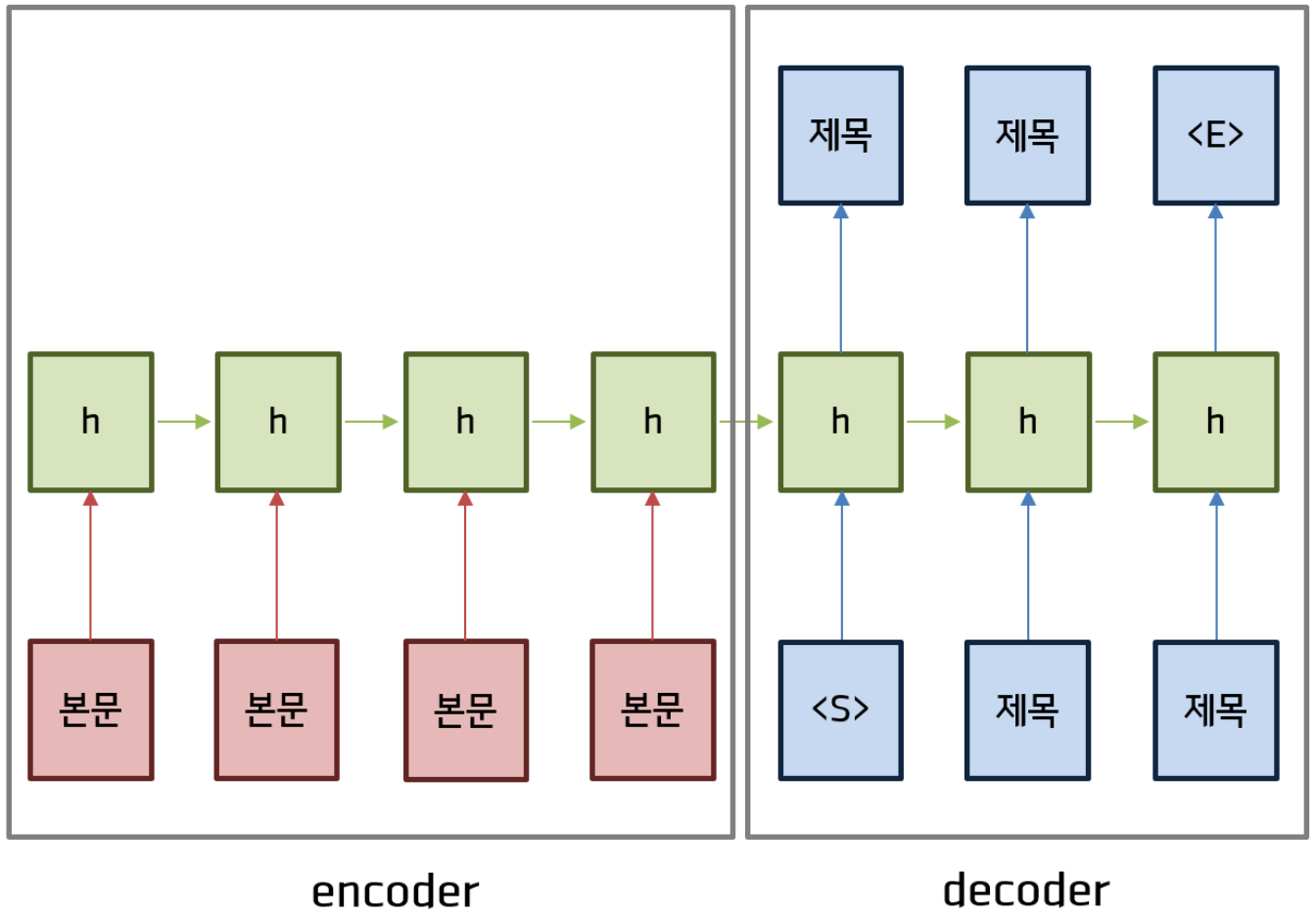


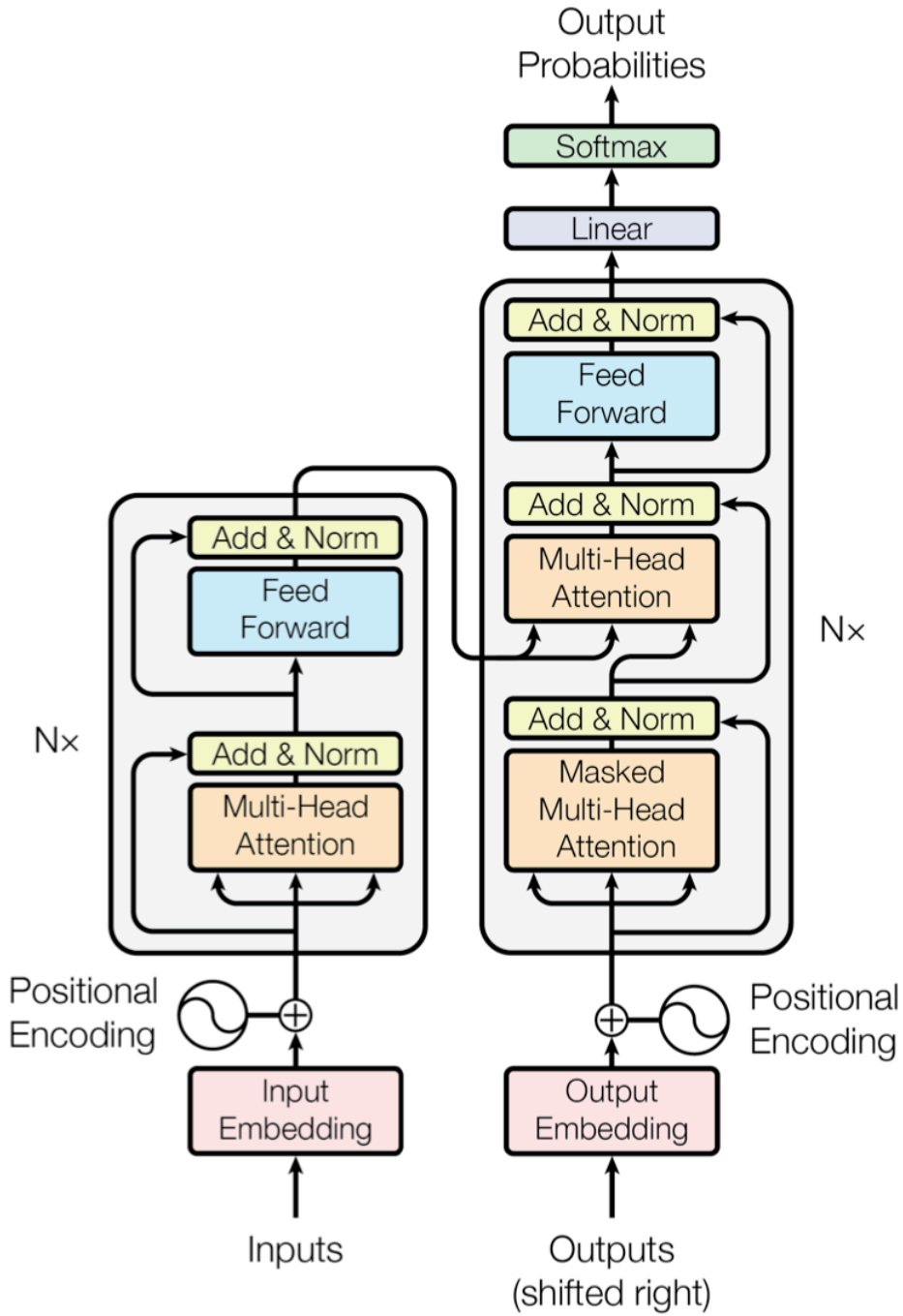
$$y = \text{Transformer}(x)$$

$x,\ y$: sentence

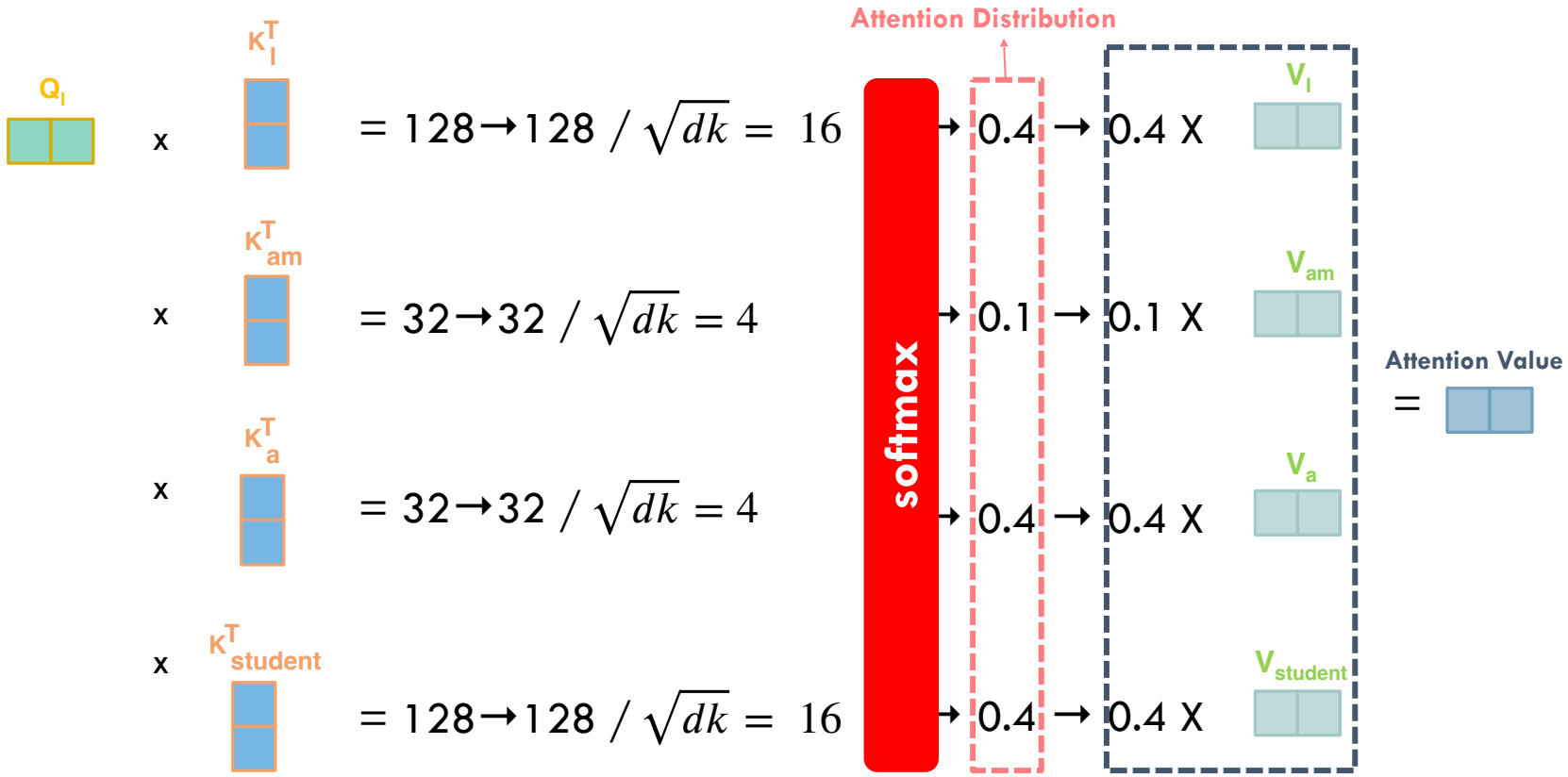
TRANSFORMER의 개괄적인 구조

- Seq-to-Seq




Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)

ATTENTION 개념



ATTENTION 개념

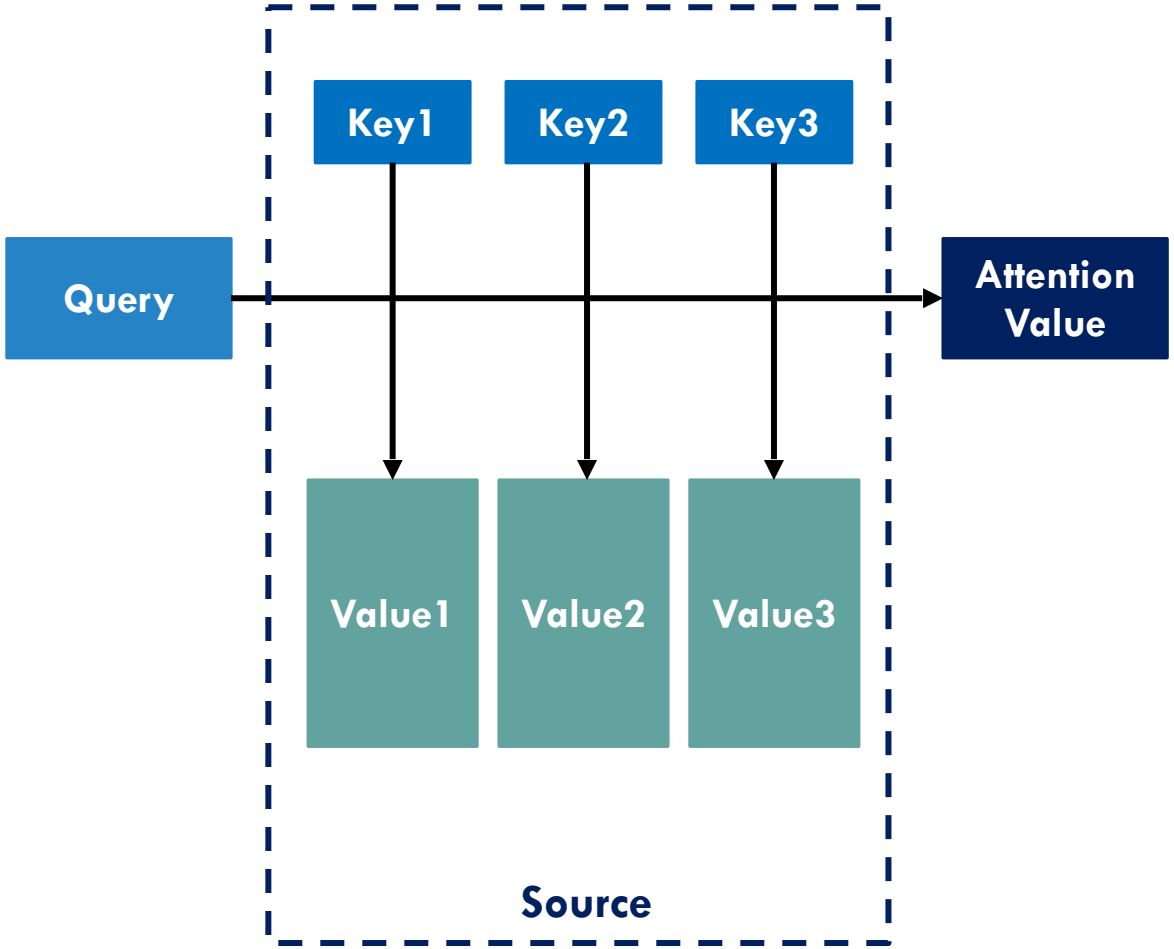
This ~~article is amazing~~

This article ~~is amazing~~

This article is ~~amazing~~

This article is amazing

$$\begin{bmatrix} 0 & -\infty & -\infty & -\infty \\ 0 & 0 & -\infty & -\infty \\ 0 & 0 & 0 & -\infty \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



ATTENTION 개념

$Q = X \times W_Q$

$K = X \times W_k$

$V = X \times W_V$

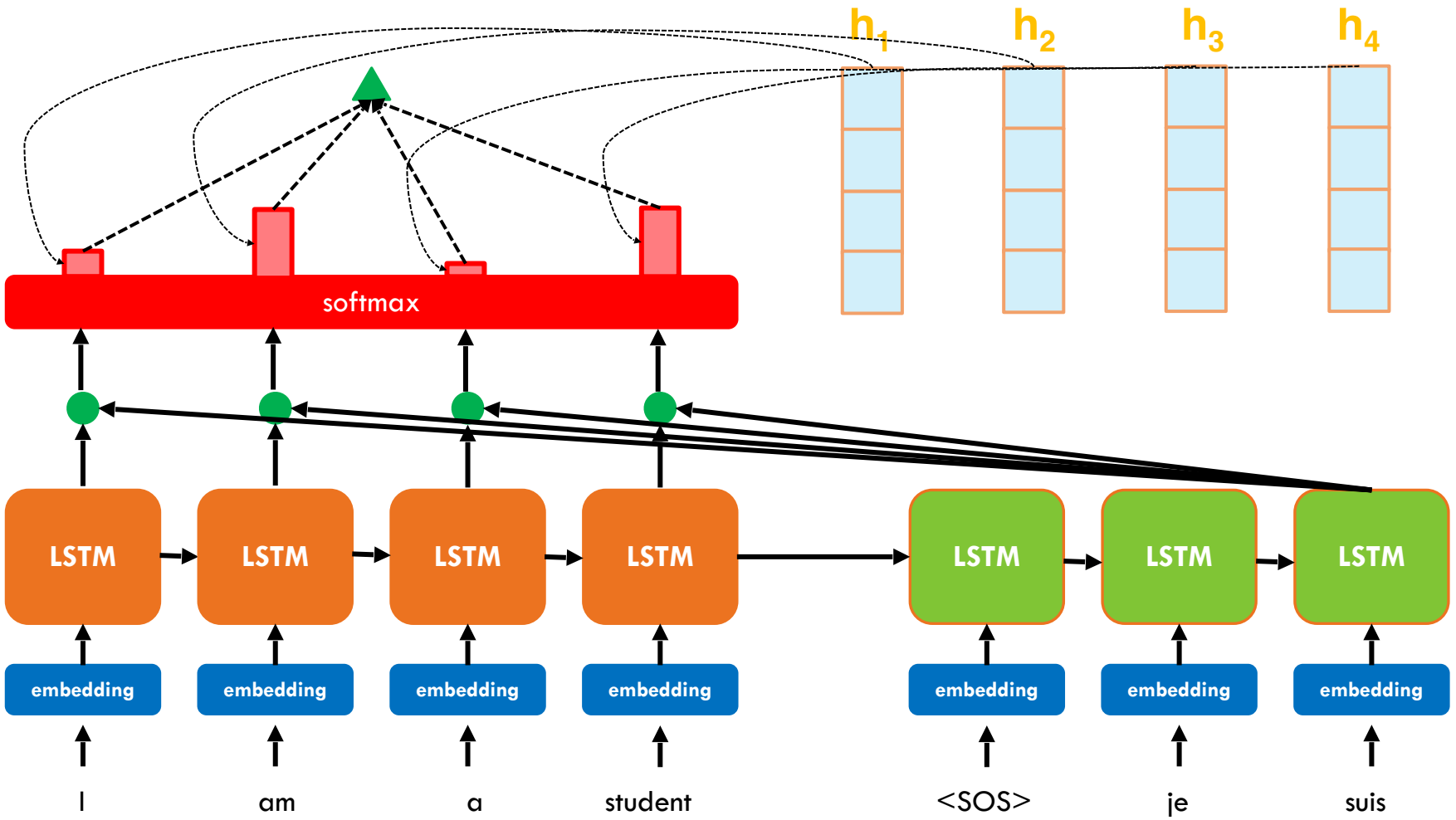
$$\begin{bmatrix} 1 & 0 & 2 \end{bmatrix} \times \begin{bmatrix} 0 & 4 & 2 \\ 1 & 4 & 3 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 4 \end{bmatrix}$$

$$\text{softmax}(xi) = \frac{\exp(xi)}{\sum_j xj}$$

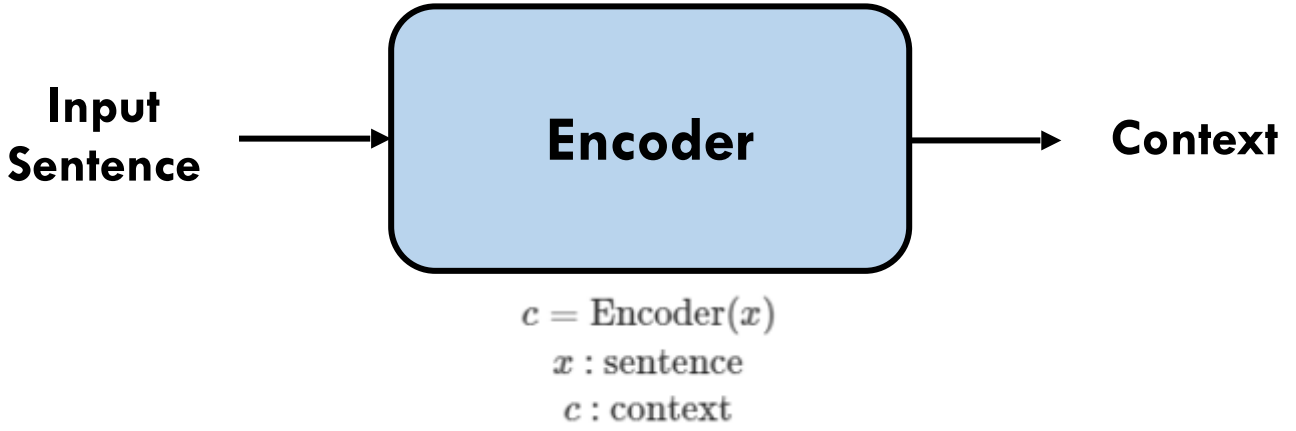
$$\text{Attention}(Q,K,V) = \text{softmax}(\frac{QK}{\sqrt{dk}})V$$

ATTENTION 개념

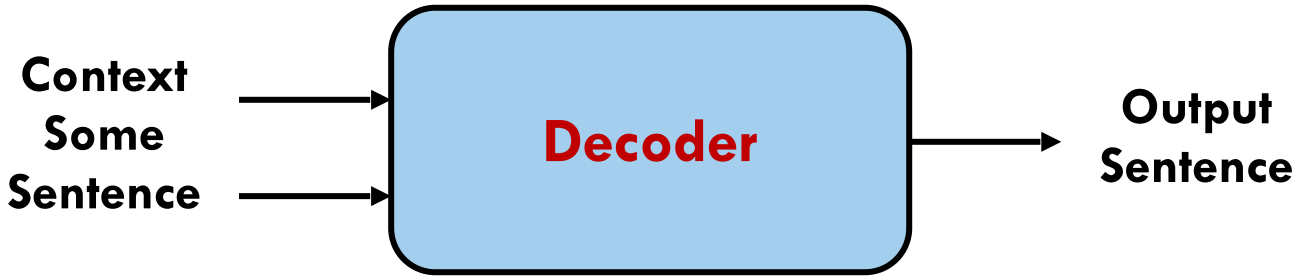
Attention Value a_t



TRANSFORMER 개념

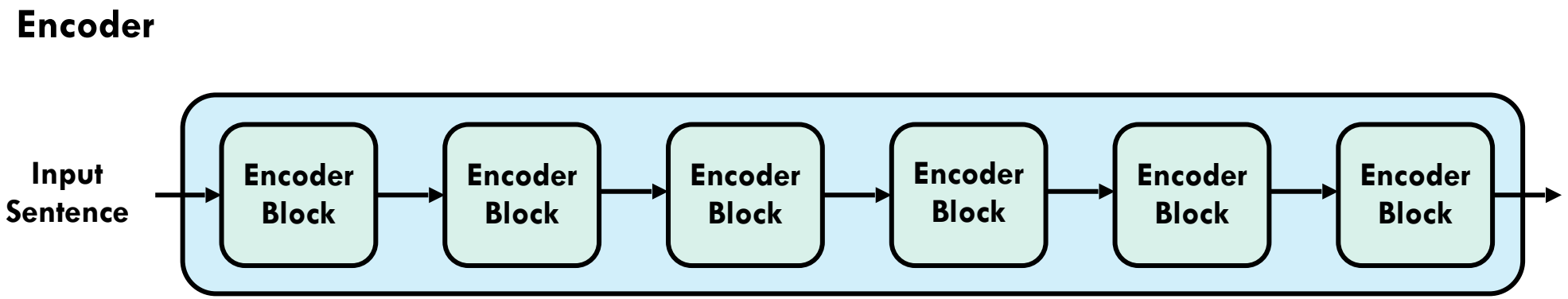


$c = \text{Encoder}(x)$
x : sentence
c : context

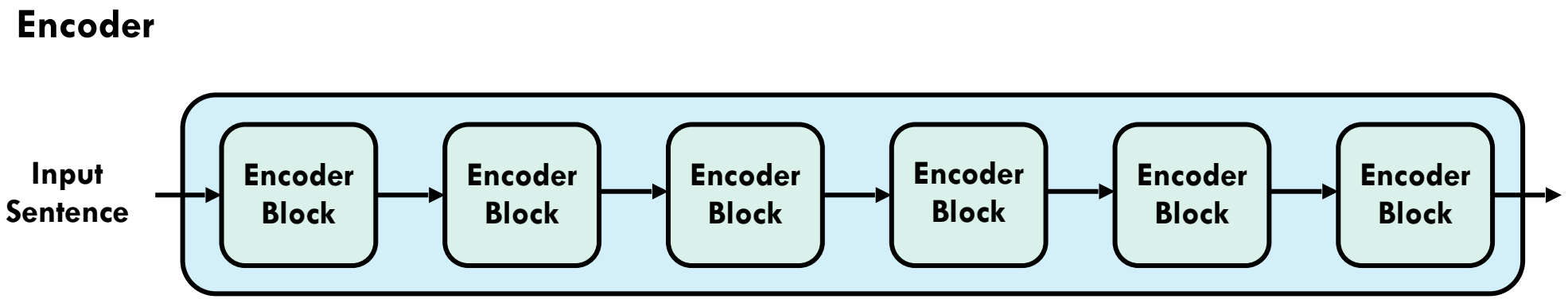


$y = \text{Decoder}(c, z)$
$y,\ z$: sentence
c : context

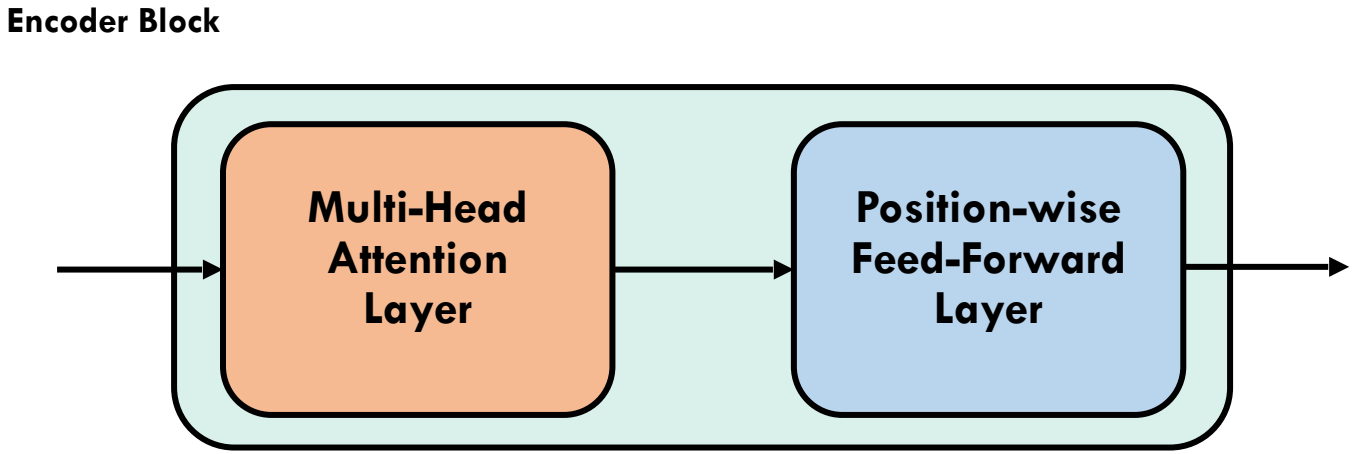
TRANSFORMER 개념



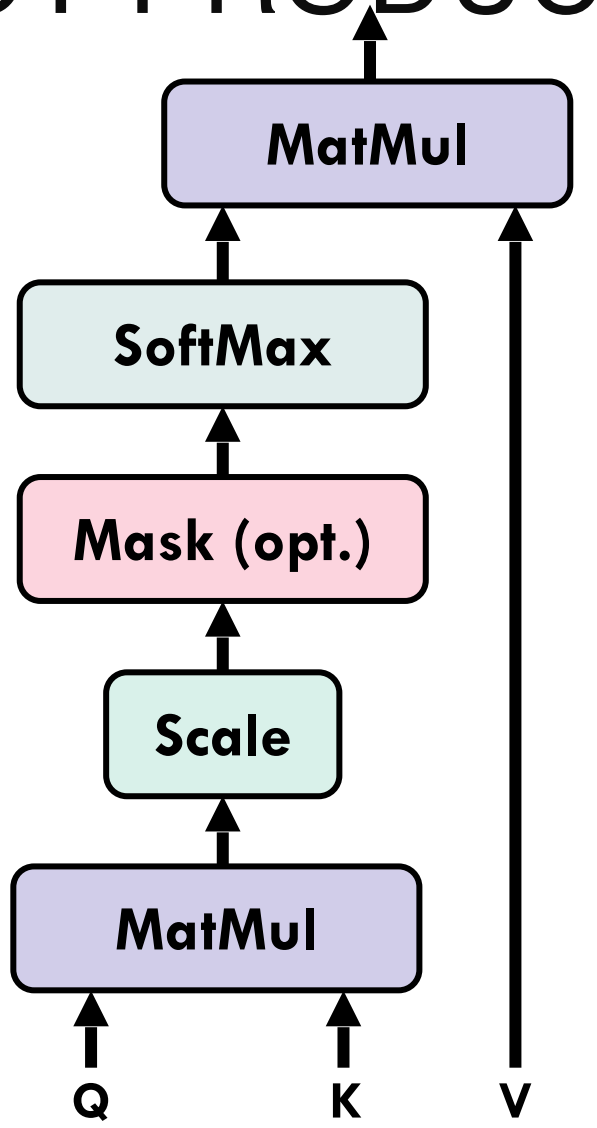
TRANSFORMER 개념



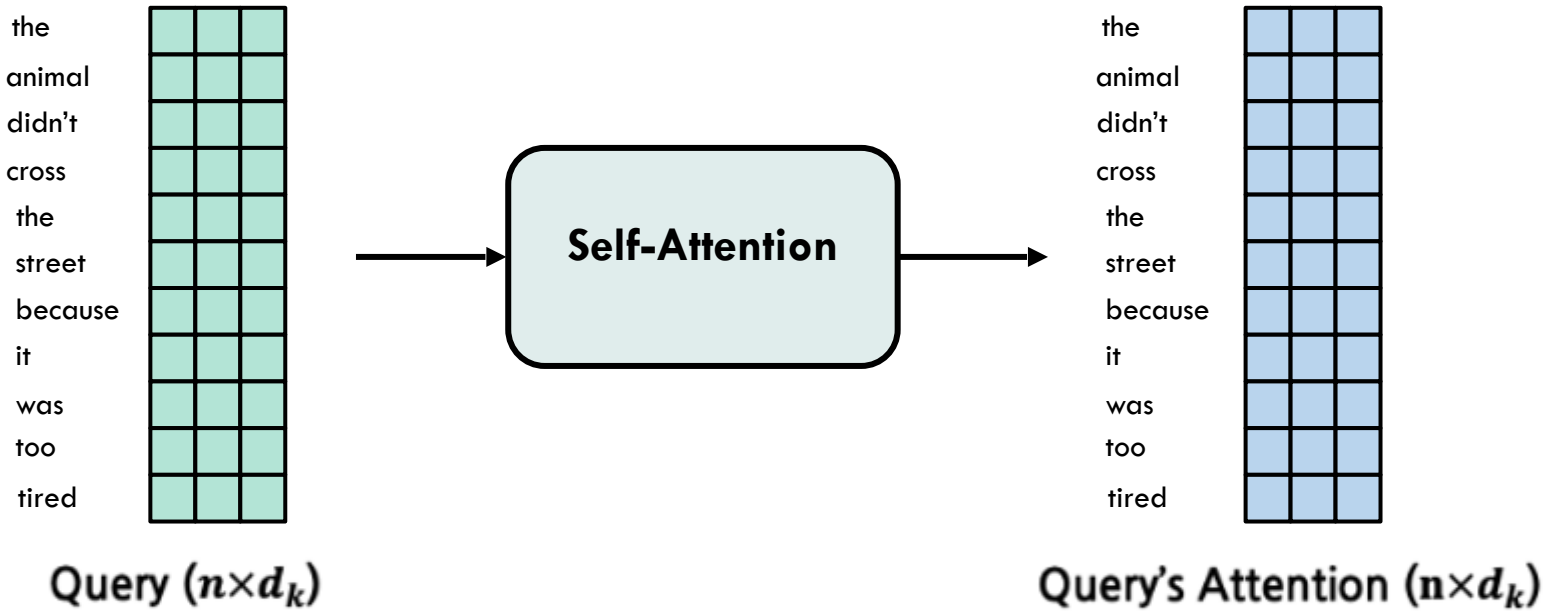
ENCODER BLOCK



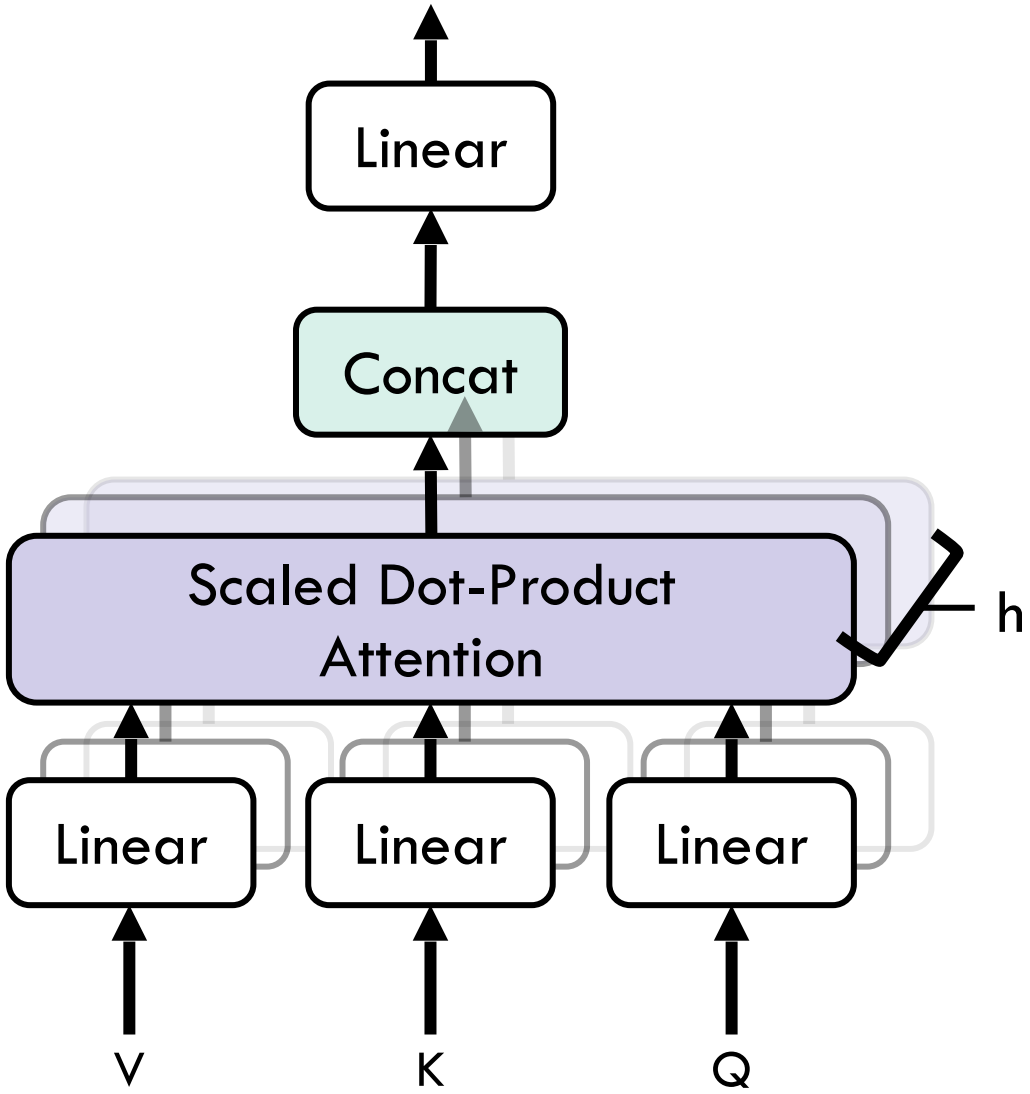
SCALED DOT-PRODUCT ATTENTION



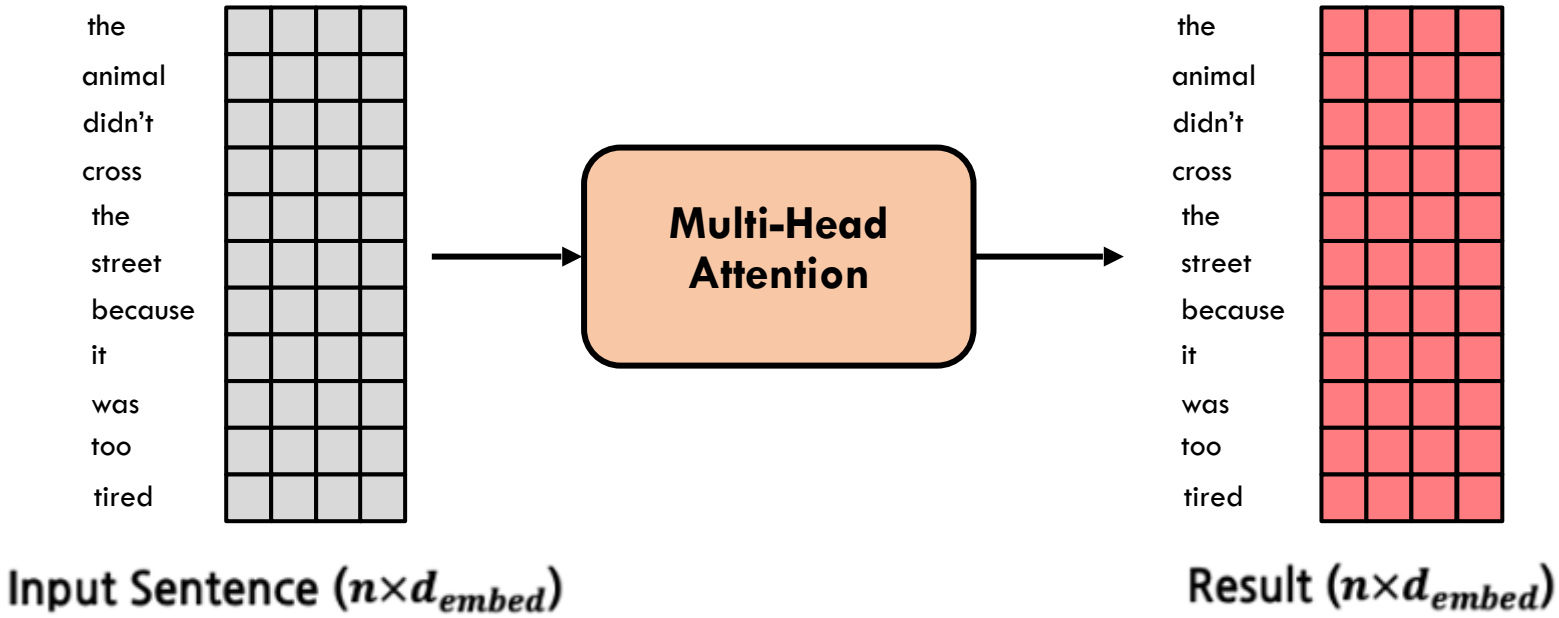
SELF ATTENTION



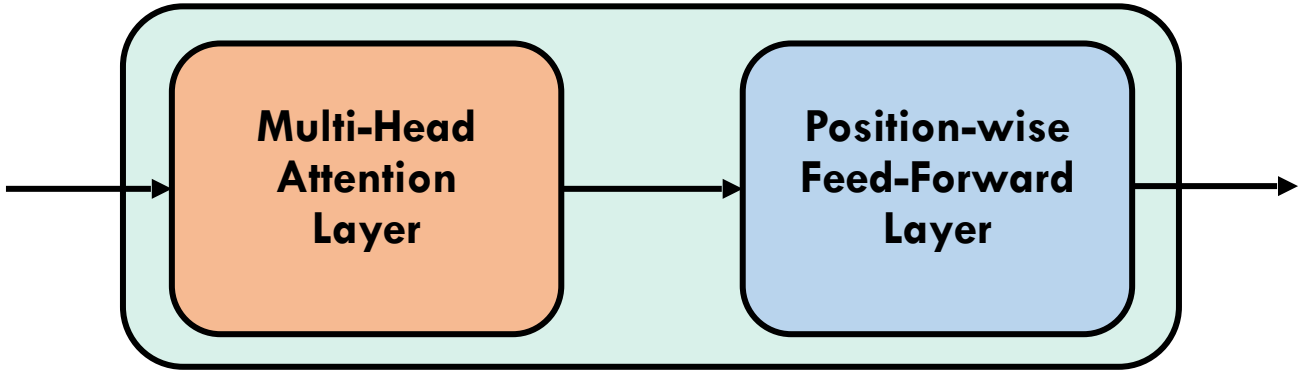
MULTI-HEAD ATTENTION LAYER



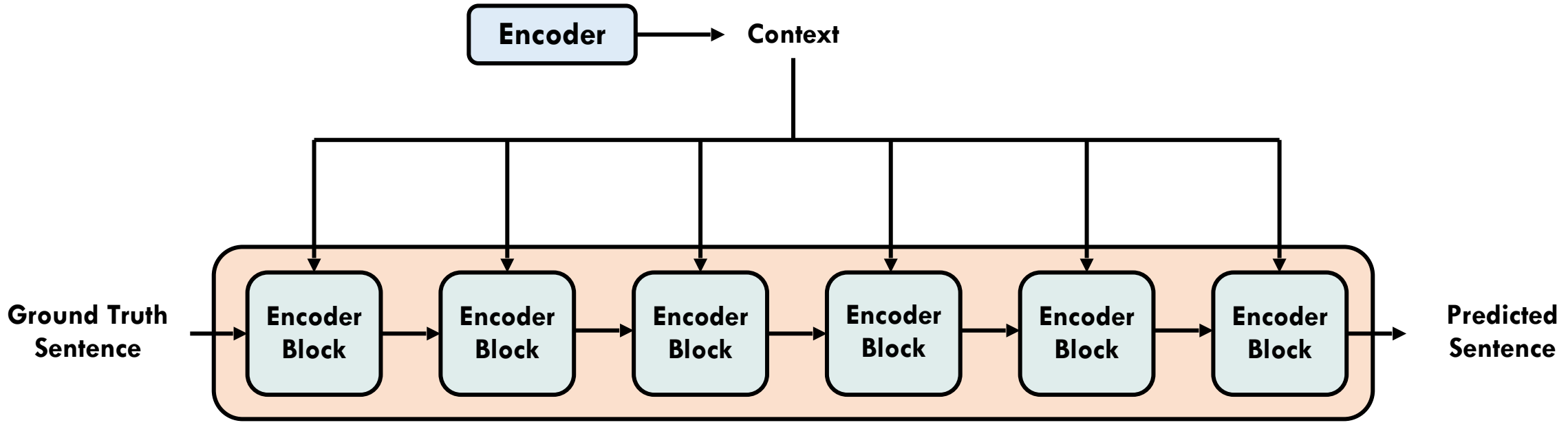
MULTI-HEAD ATTENTION LAYER



RESIDUAL CONNECTION LAYER



DECODER PART



DECODER PART