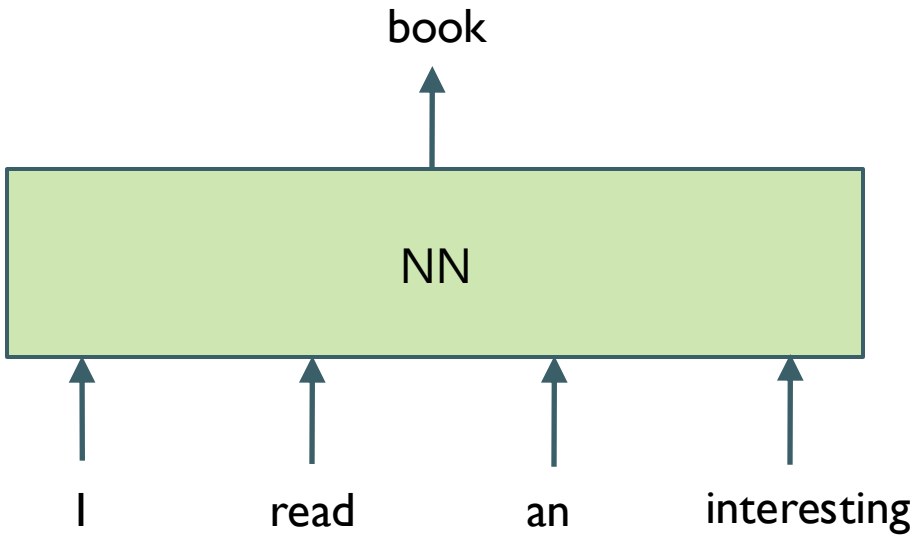
NLP + Transformer

성균관대학교 소프트웨어학과
이 지 형

Next Word Prediction

- **Example**

I read an interesting → book



Next Word Prediction

- **Training Data**

Crawled Text

I read an interesting book. It was about the life of a famous scientist who made groundbreaking discoveries in the field of biology. The book provided a detailed account of his experiments, challenges, and triumphs. I was particularly fascinated by how his persistence and dedication led to significant advancements in science. It was an inspiring read, and I learned a lot about the scientific process and the importance of never giving up.

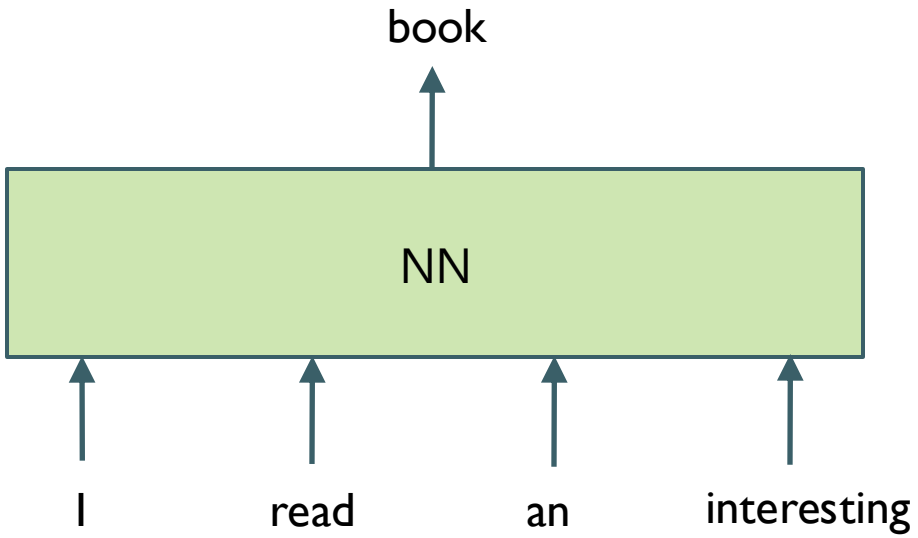
Training Data

(NA NA NA I → read)
(NA NA I read → an)
(NA I read an → interesting)
(I read an interesting → book)
(read an interesting book →.)
(an interesting book. → It)
(interesting book. It → was)
(book. It was → about)
…

Next Word Prediction

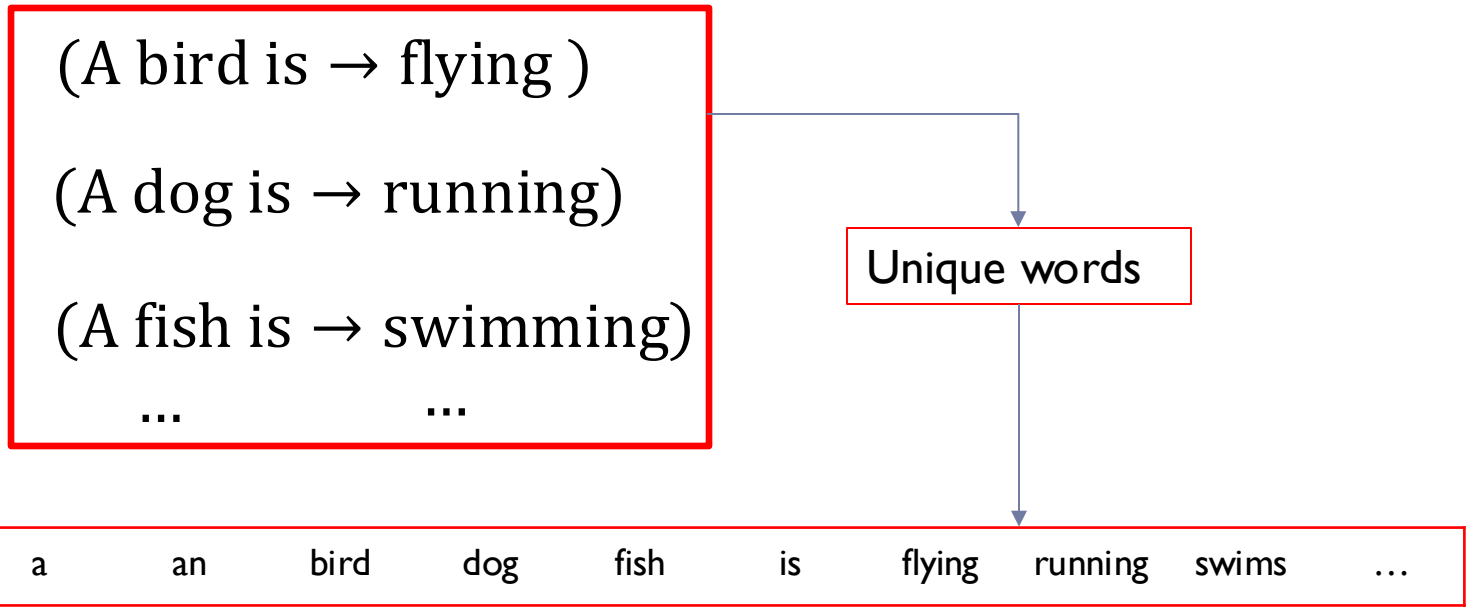
- **How to Handle Texts**

I read an interesting → book



Next Word Prediction

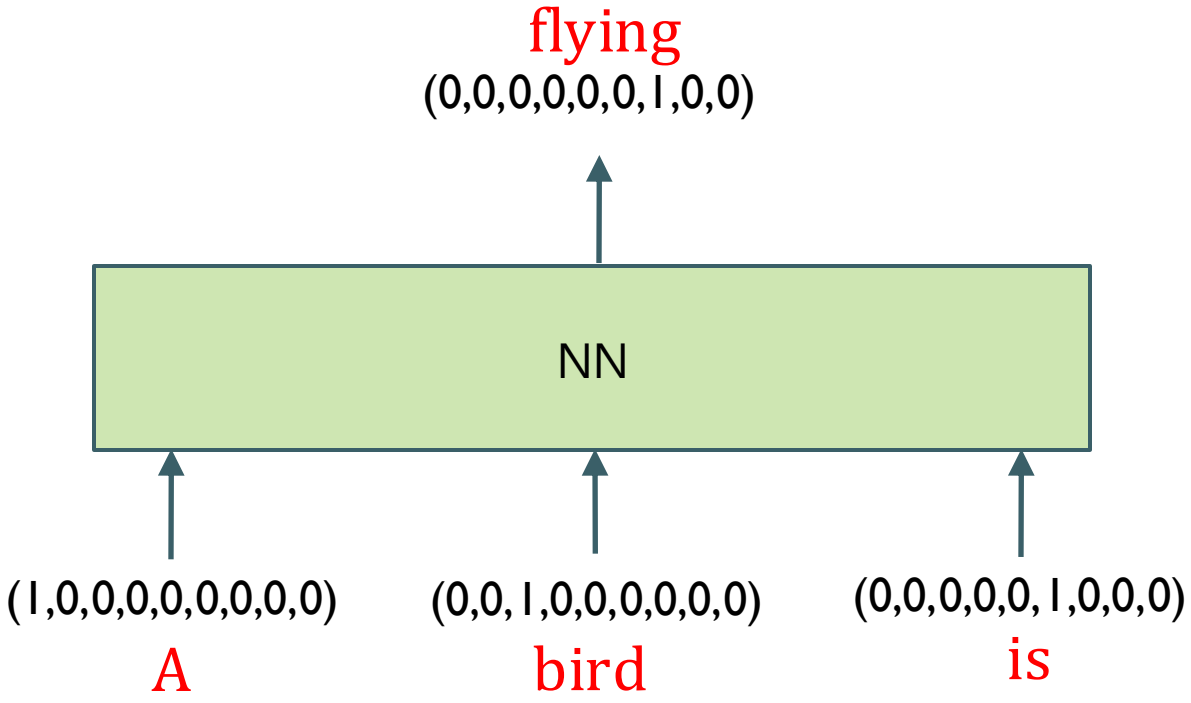
- **How to Handle Texts**



A	(	1	,	0	,	0	,	0	,	0	,	…)
bird	(	0	,	0	,	1	,	0	,	0	,	…)
is	(	0	,	0	,	0	,	0	,	0	,	…)

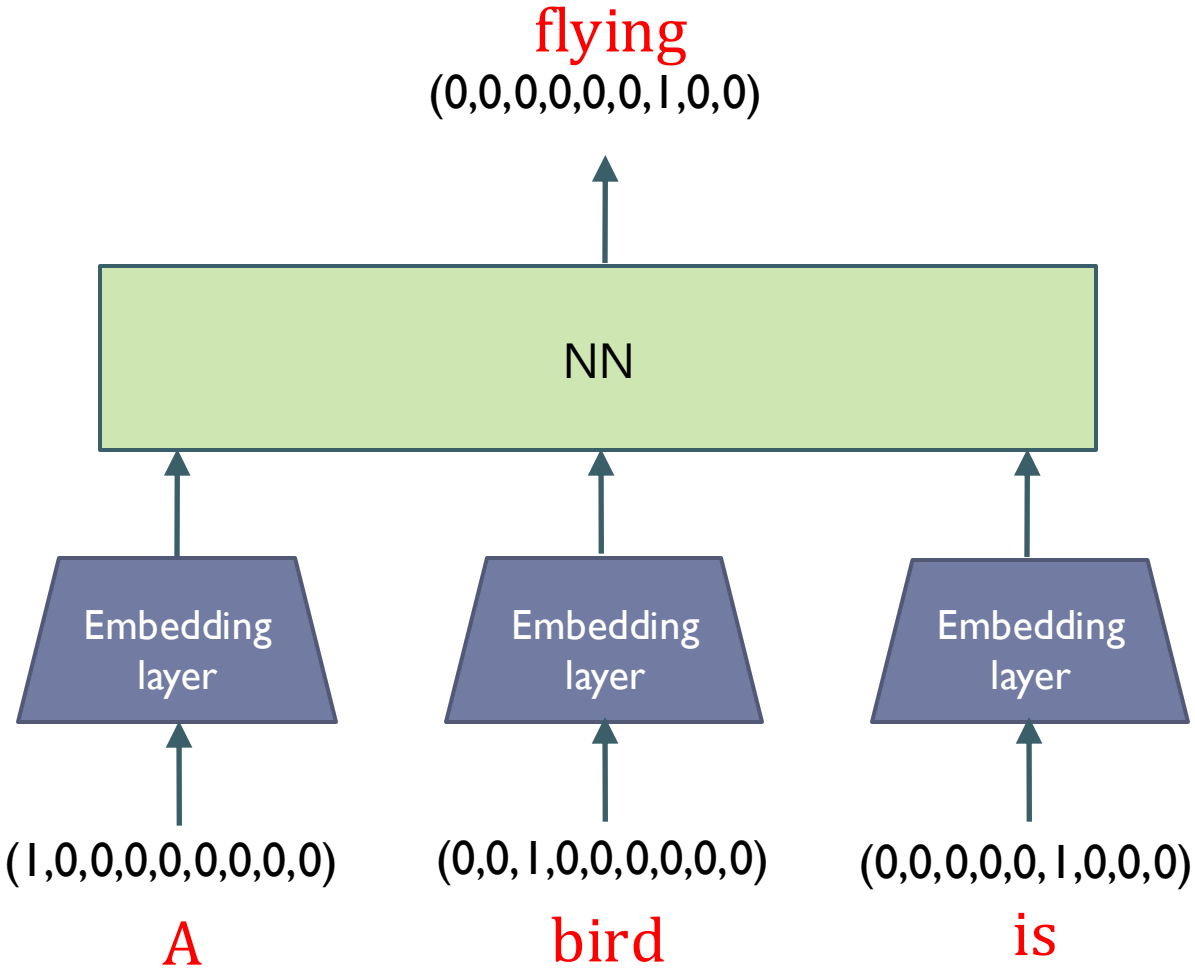
Next Word Prediction

- How to Handle Texts



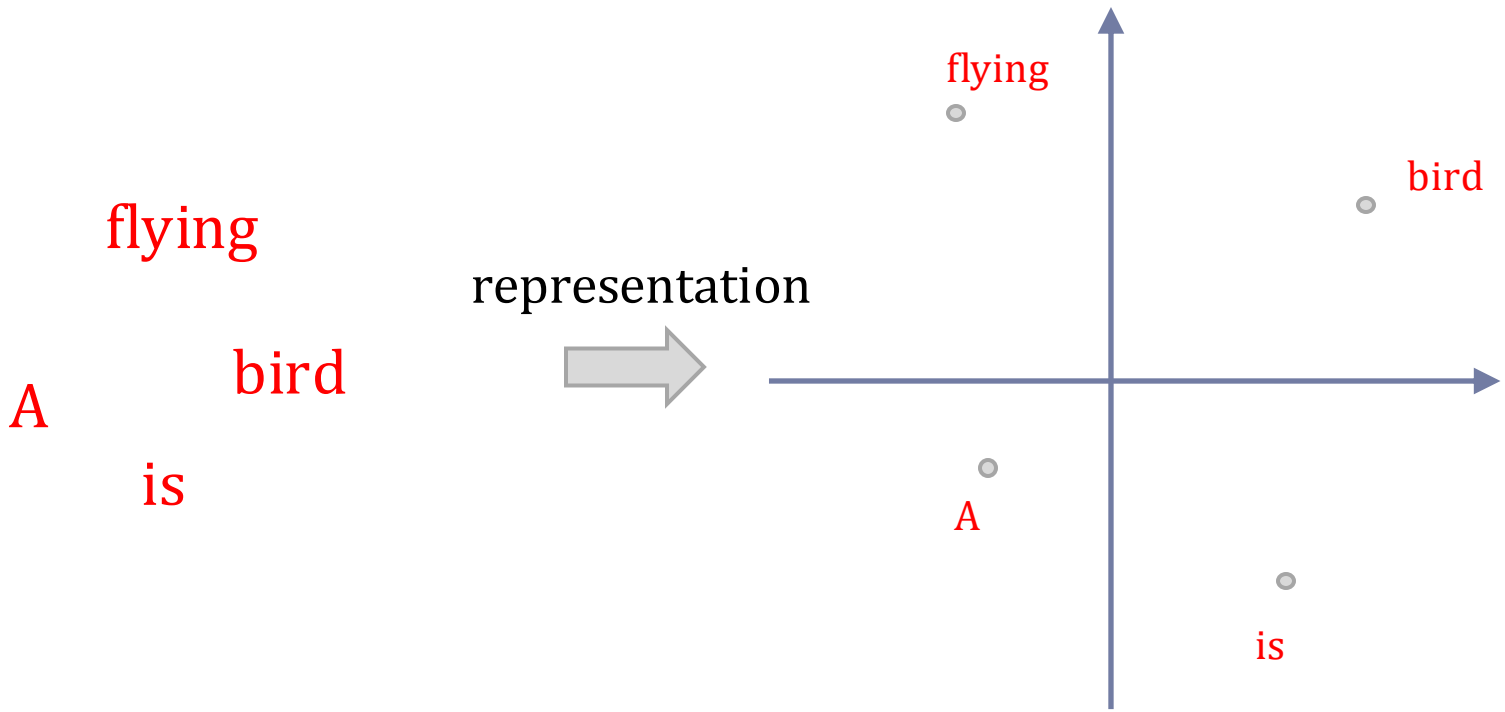
Next Word Prediction

- How to Handle Texts



Next Word Prediction

- **Representation, Embedding**



Values in non-Euclidean space

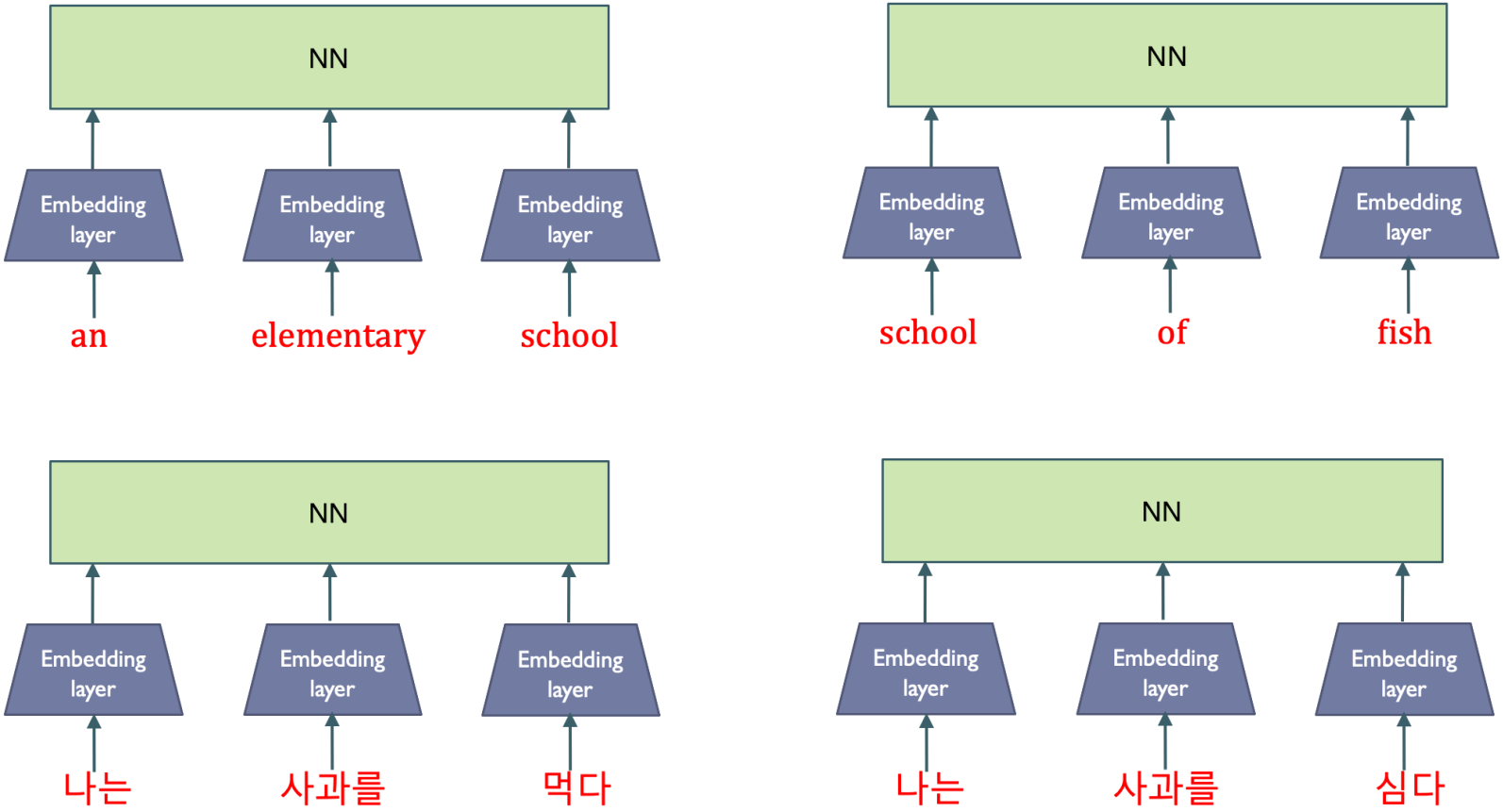
Values in Euclidean space

Attention Mechanism

성균관대학교 소프트웨어학과
이 지 형

Word Representation

- Enough? Embedding Layer



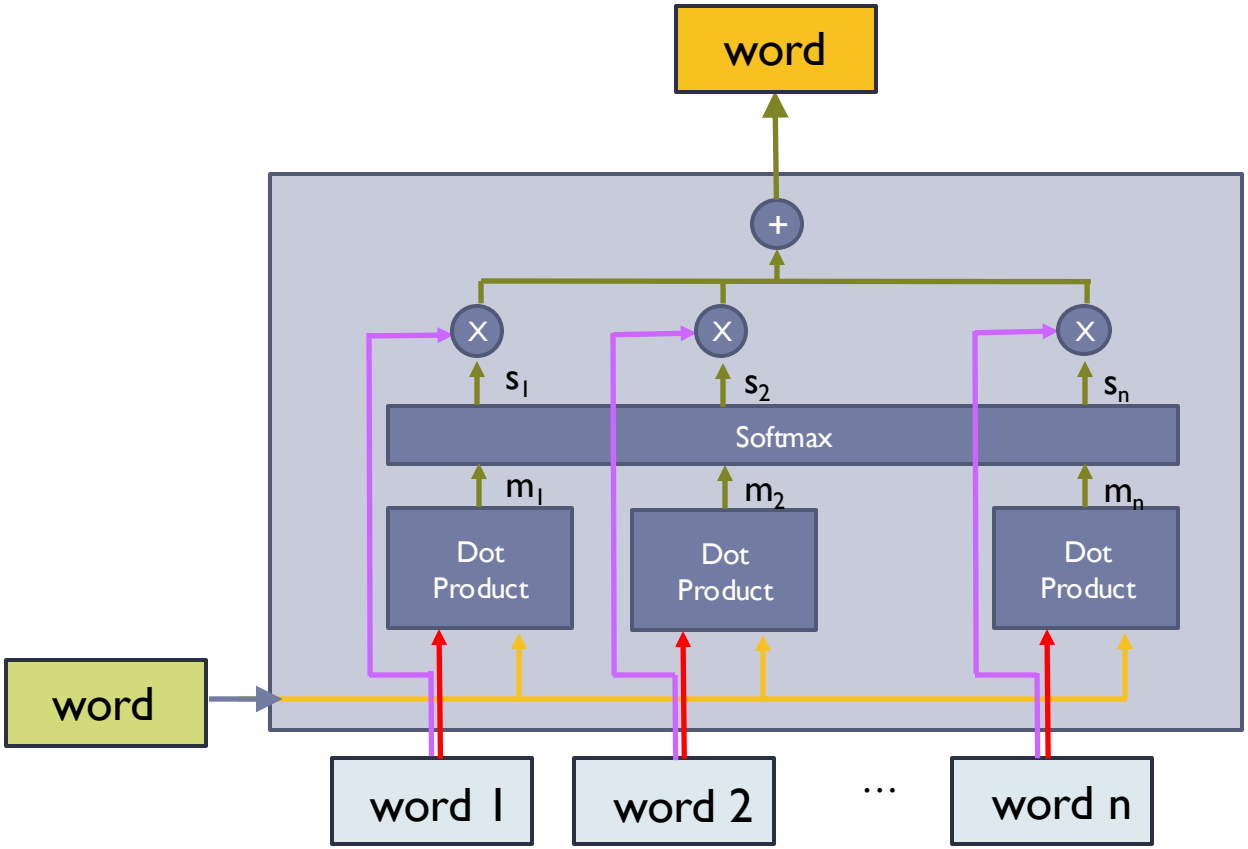
Word Representation

- Words have their own meanings independently
- But in a sentence, they often take on different, more specific meanings depending on context.
- Let's represent a word by blending it with other words in the sentence

→ This leads to the concept of Attention

Attention Mechanism

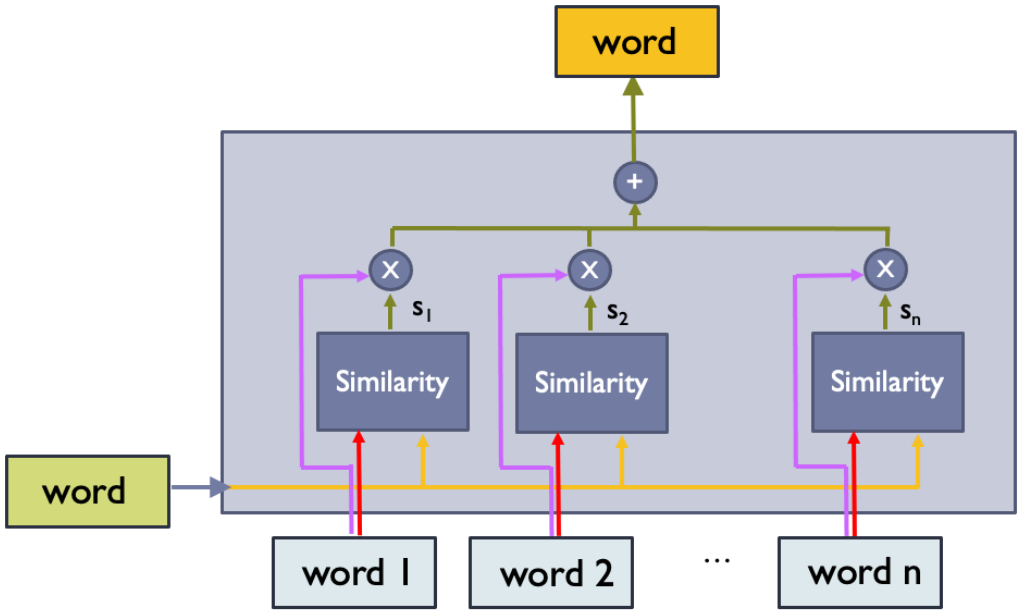
- Overview



Context Consideration

- **Generating Word Representation with Context**

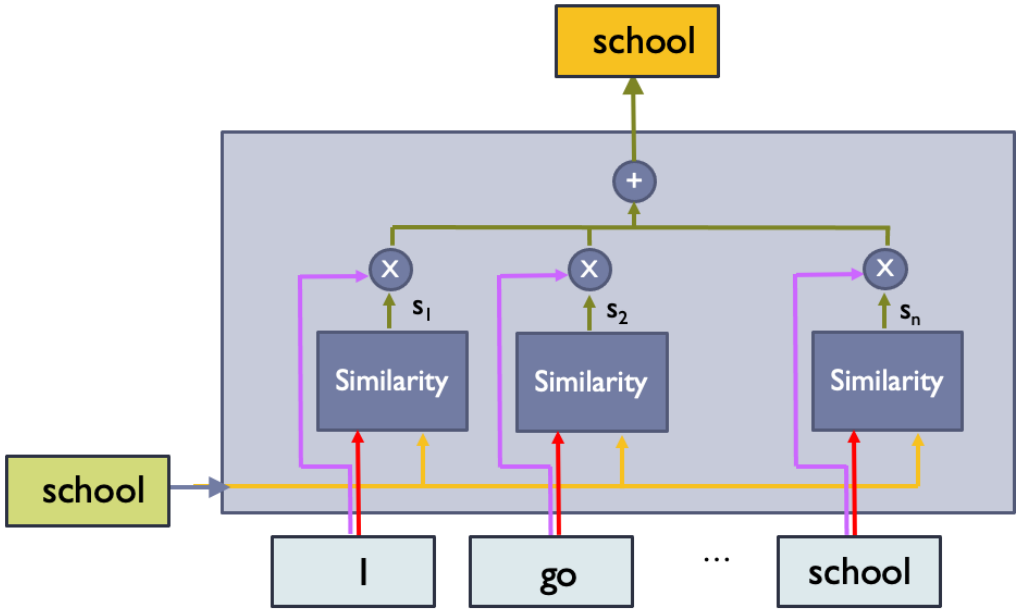
$$\text{word} = s_1 \times \text{word 1} + s_2 \times \text{word 2} + \cdots + s_n \times \text{word n}$$



Context Consideration

- Generating Word Representation with Context
  - Example

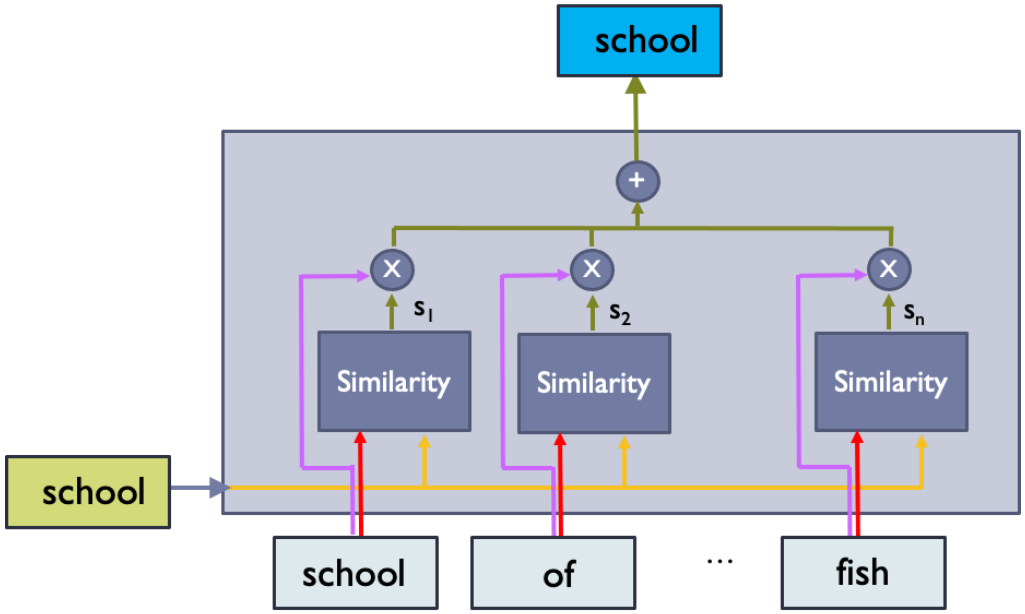
$$\text{school} = s_1 \times \text{I} + s_2 \times \text{go} + \cdots + s_n \times \text{school}$$



Context Consideration

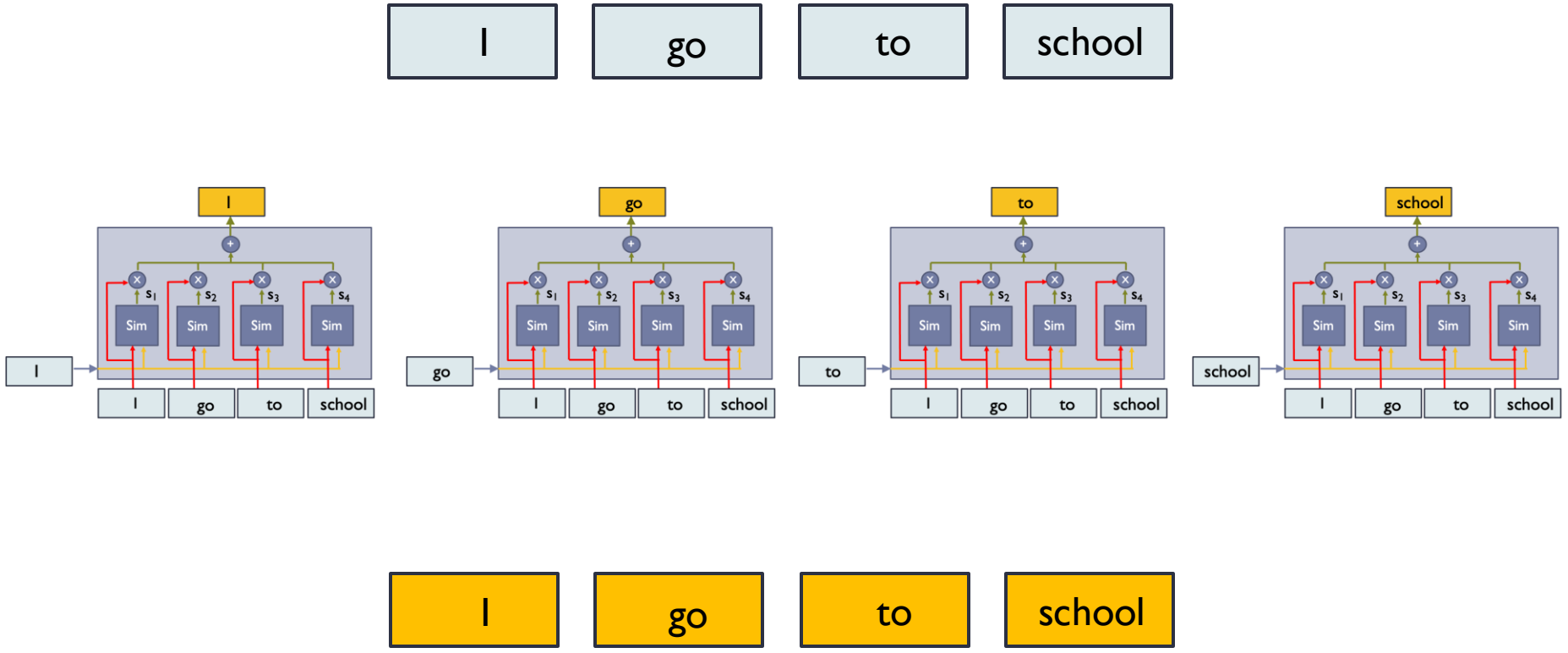
- Generating Word Representation with Context
  - Example

$$\text{school} = s_1 \times \text{school} + s_2 \times \text{of} + \cdots + s_n \times \text{fish}$$



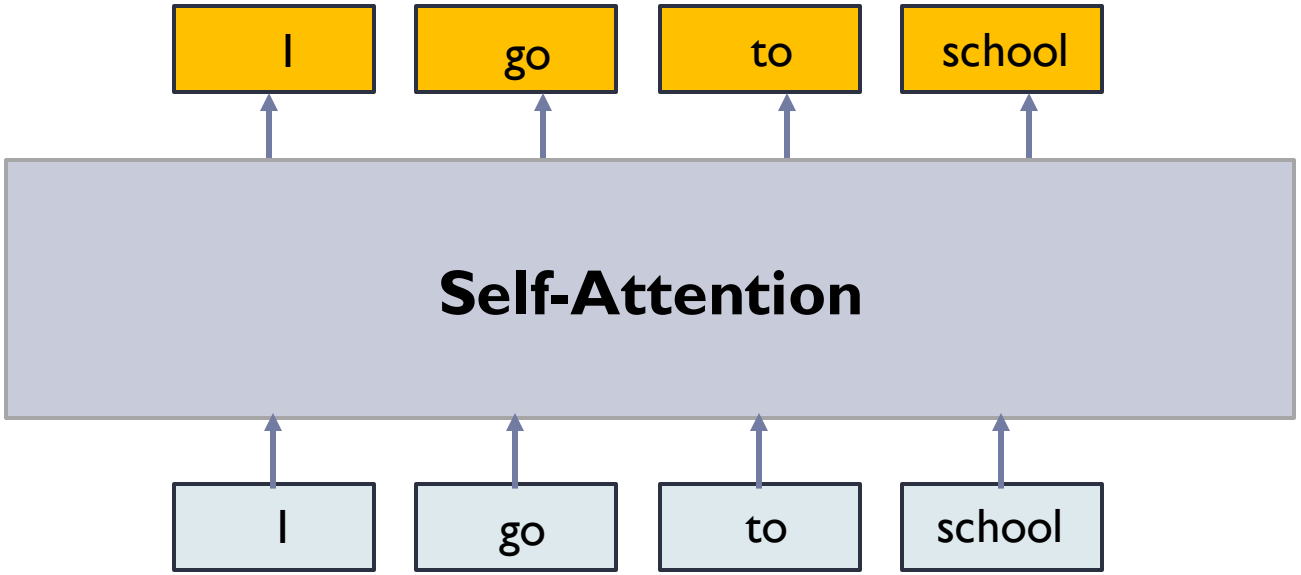
Self-Attention

- Context: The words in the same sentence



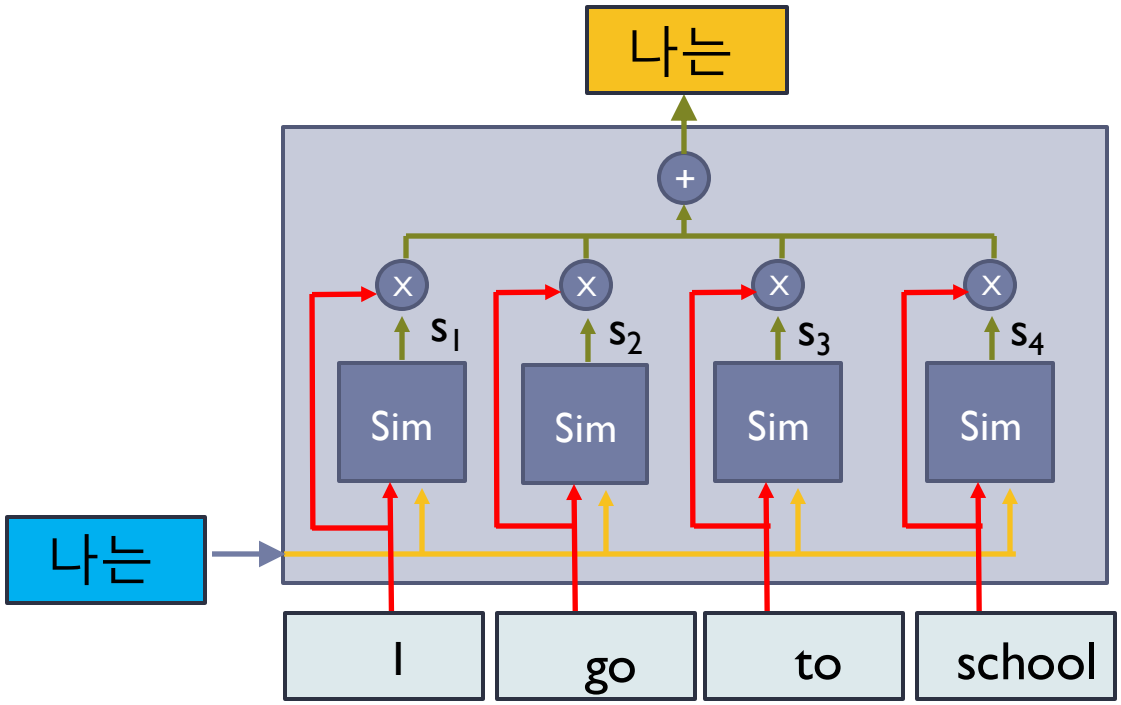
Self-Attention

- Context: The words in the same sentence



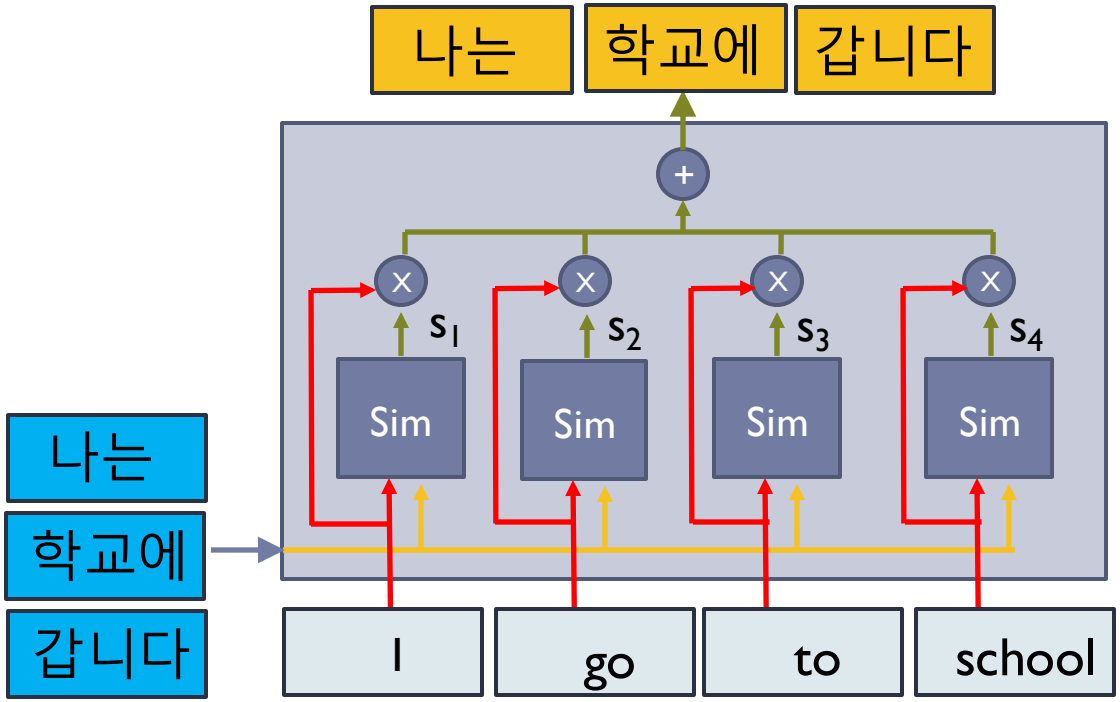
Cross-Attention

- Context: The words in the other sentence

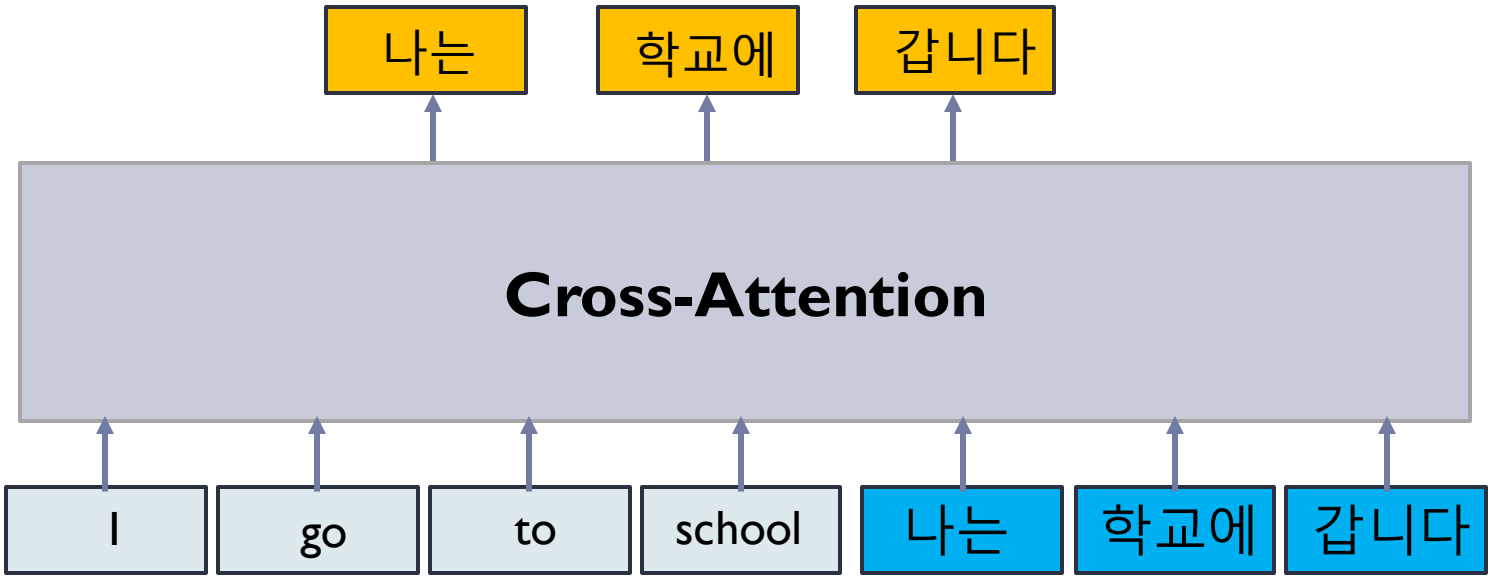


Cross-Attention

- Context: The words in the other sentence



Cross-Attention

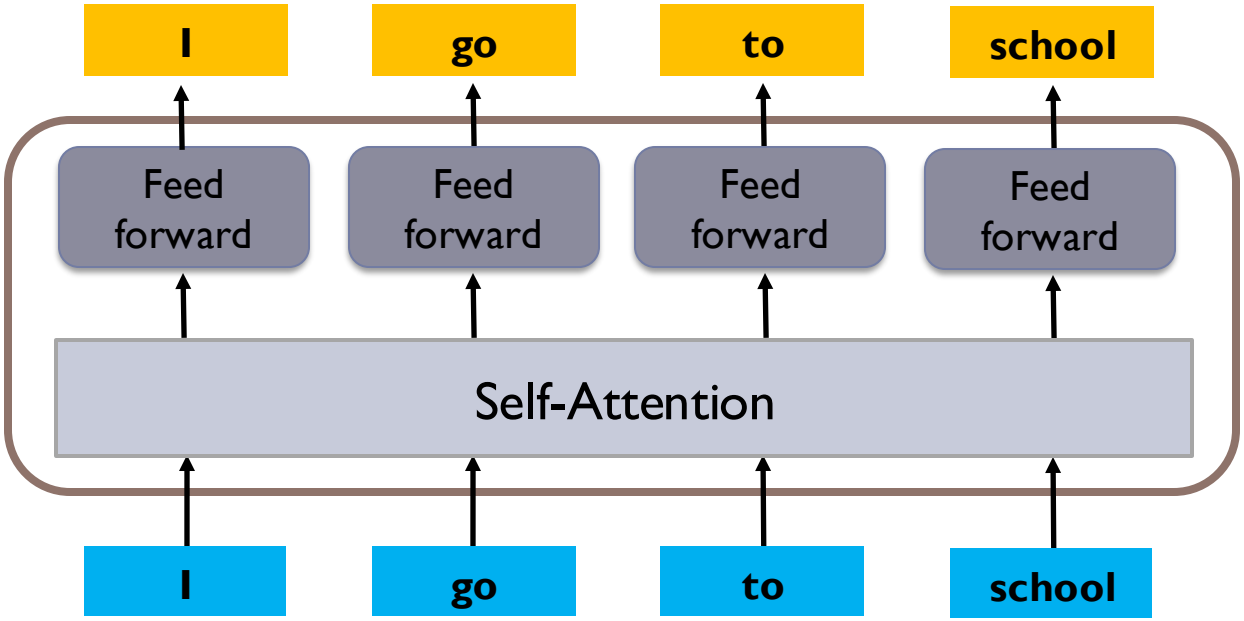


Transformer, BERT, and GPT

성균관대학교 소프트웨어학과
이 지 형

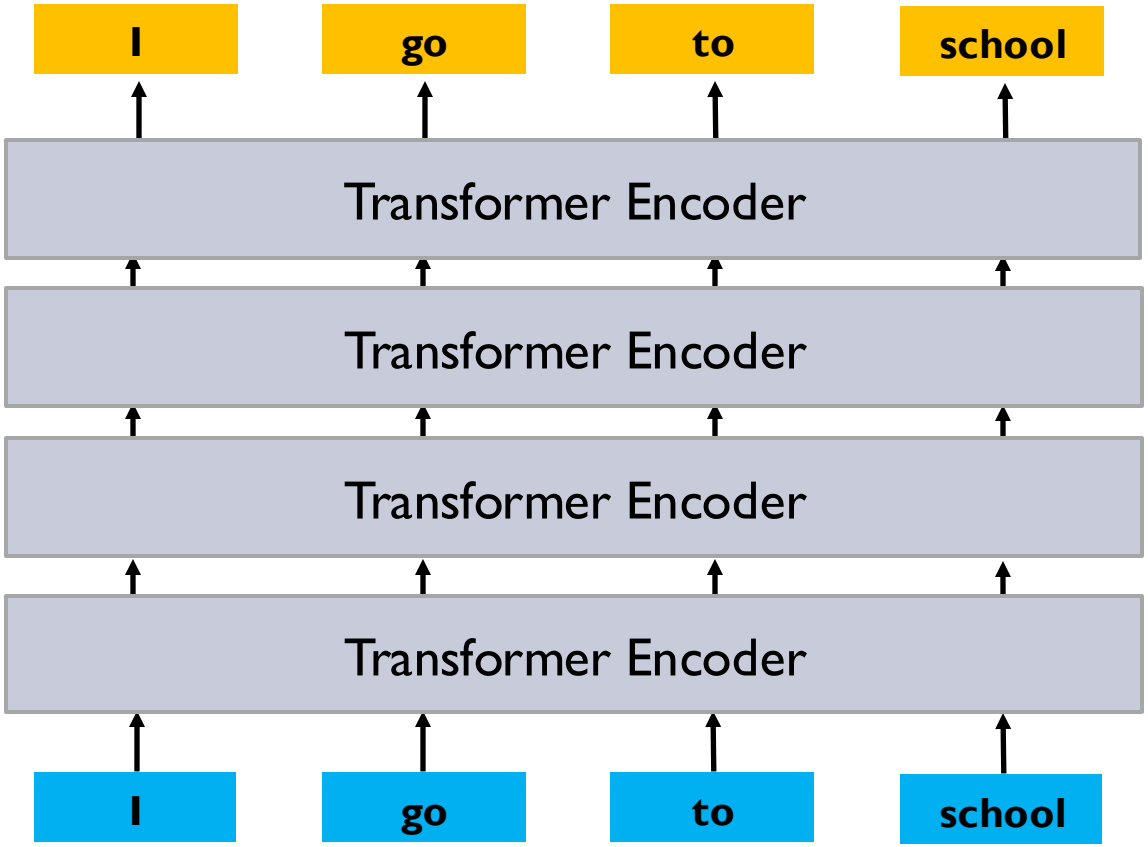
Transformer Encoder

- **Structure**
  - Self-Attention + Multi-Head + Position Encoding + Feed forwad NN layer
- **More Sophisticated Semantic Processing**



Transformer Encoder

- **Word Representation Reflecting Sentence-Level Semantic**

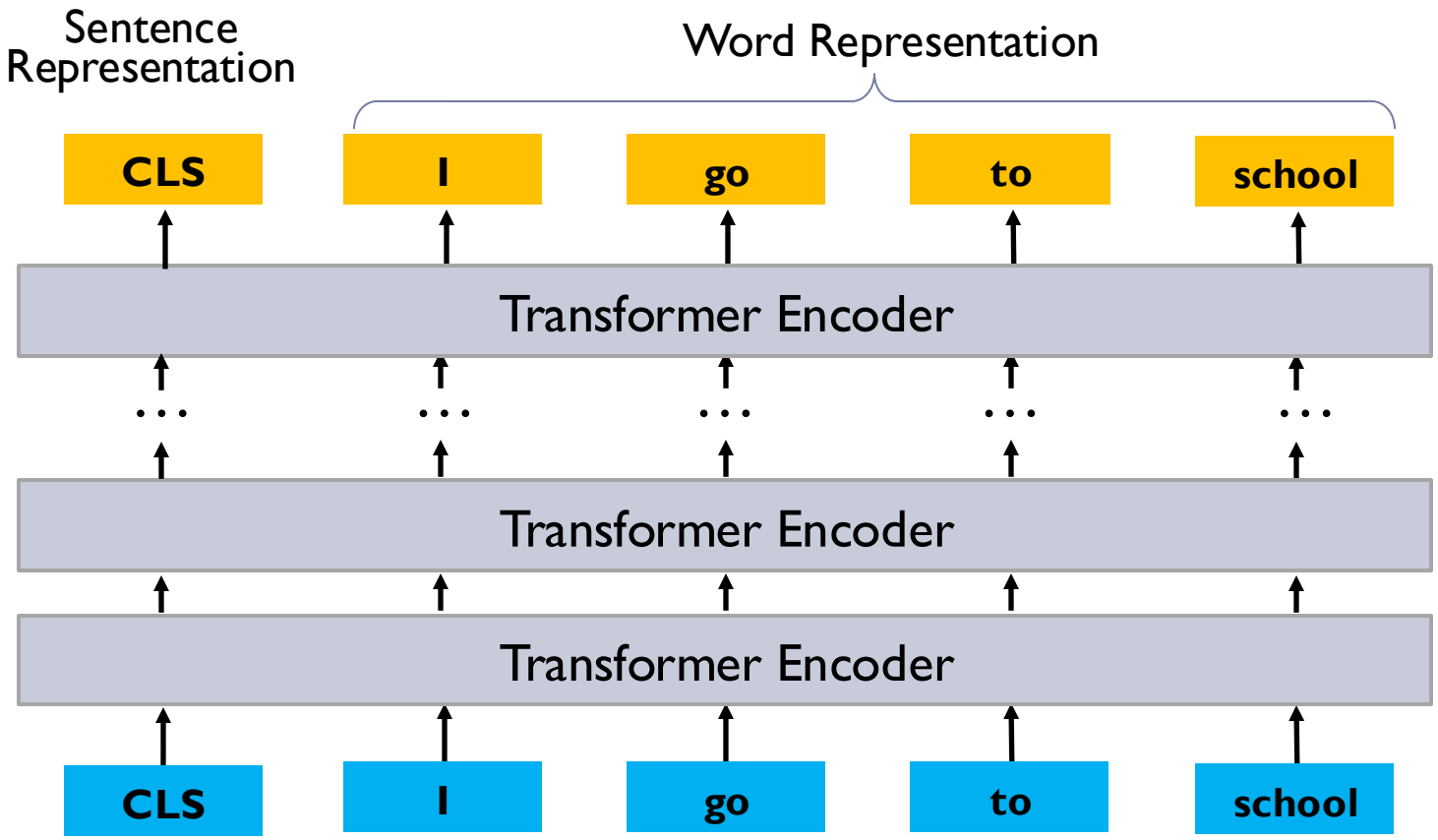


BERT Model

- BERT (Bidirectional Encoder Representations from Transformers)
- A powerful language understanding model
- Developed by Google in 2018
- Generating
  - Word Representations Highly Processed considering Context
  - Sentence Representation
- Pretrained Models
  - KoBERT (by SKT): 한국어 위키백과 + 뉴스 기사로 학습됨
  - KorBERT (by ETRI)

BERT Model

- **Structure: CLS token + Transformer Encoders**



BERT Model

- **Usage of CLS token**
  - **Text classification**
    - Spam detection
    - News categorization
  - **Sentiment analysis**
    - "I absolutely loved this product!" → Positive
    - "The service was terrible and slow." → Negative
  - **Semantic similarity**
    - Two sentences have similar meaning or not

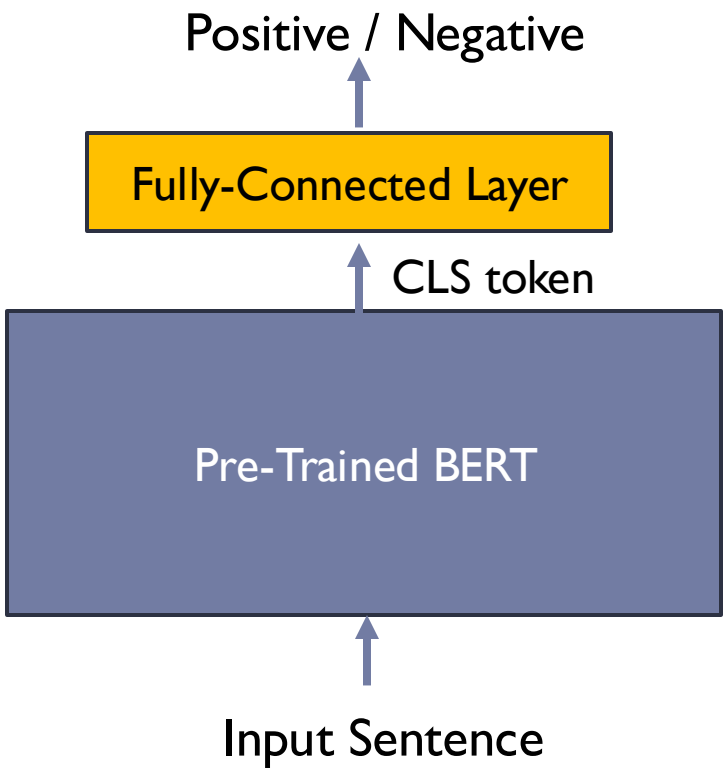
BERT Model

- **Usage of Word Representation**
  - **Named Entity Recognition (NER): 고유명사 찾기**
    - "Apple released a new product."→ O, X, X, X, X
  - **Part-of-Speech Tagging: 품사 찾기**
    - "Apple released a new product" → noun, verb, article, adjective, noun
  - **Semantic Role Labeling: 문법적 역할 찾기**
    - "Apple released a new product" → subject, action, object

BERT Model

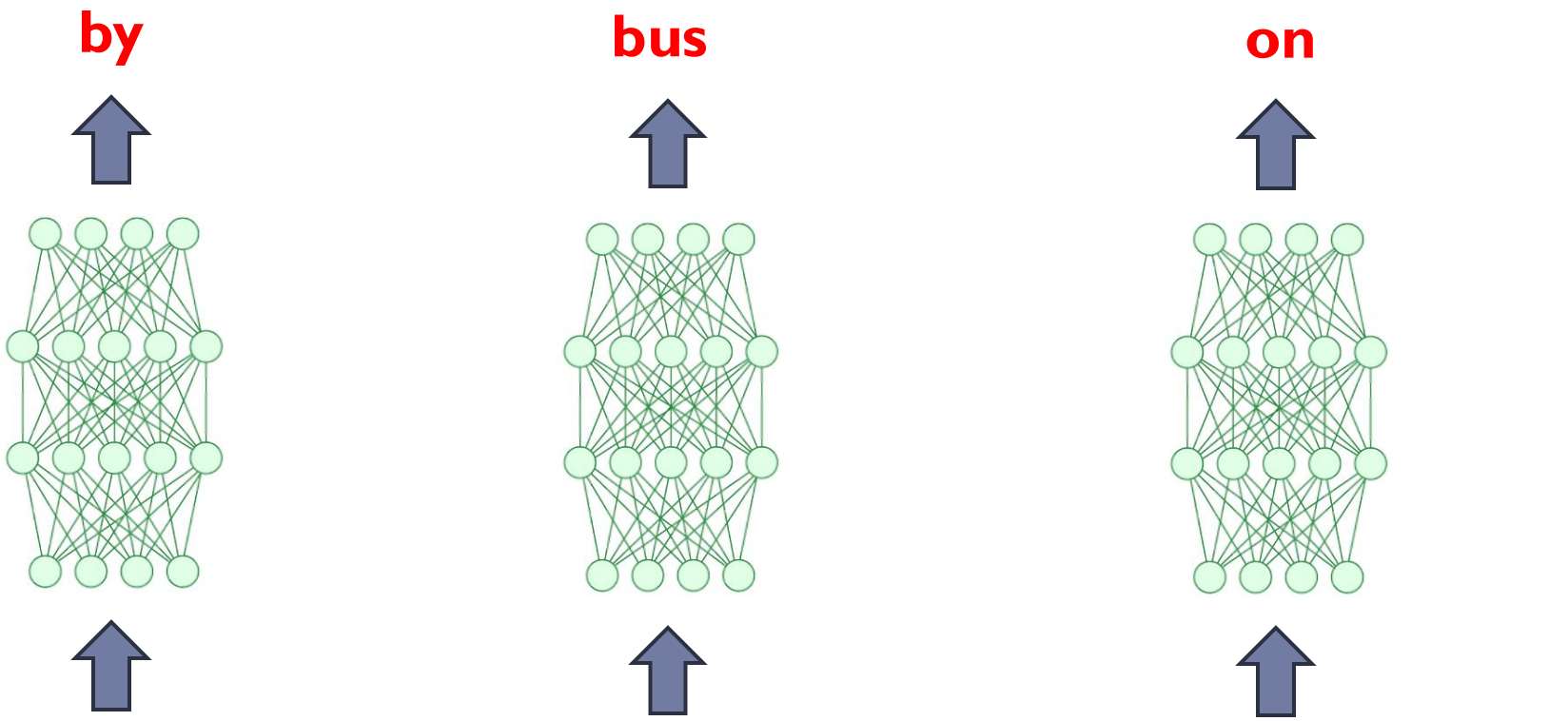
- **Example: Sentiment Analysis**
  - CLS 토근에 NN layer를 붙인 후에
  - 데이터를 이용하여 Fine-Tuning

이 제품 정말 좋아요. →	긍정
배송이 너무 느려요 →	부정
생각보다 괜찮았어요 →	긍정
기대 이하입니다 →	부정
가성비 최고 →	긍정
디자인 진짜 별로 →	부정



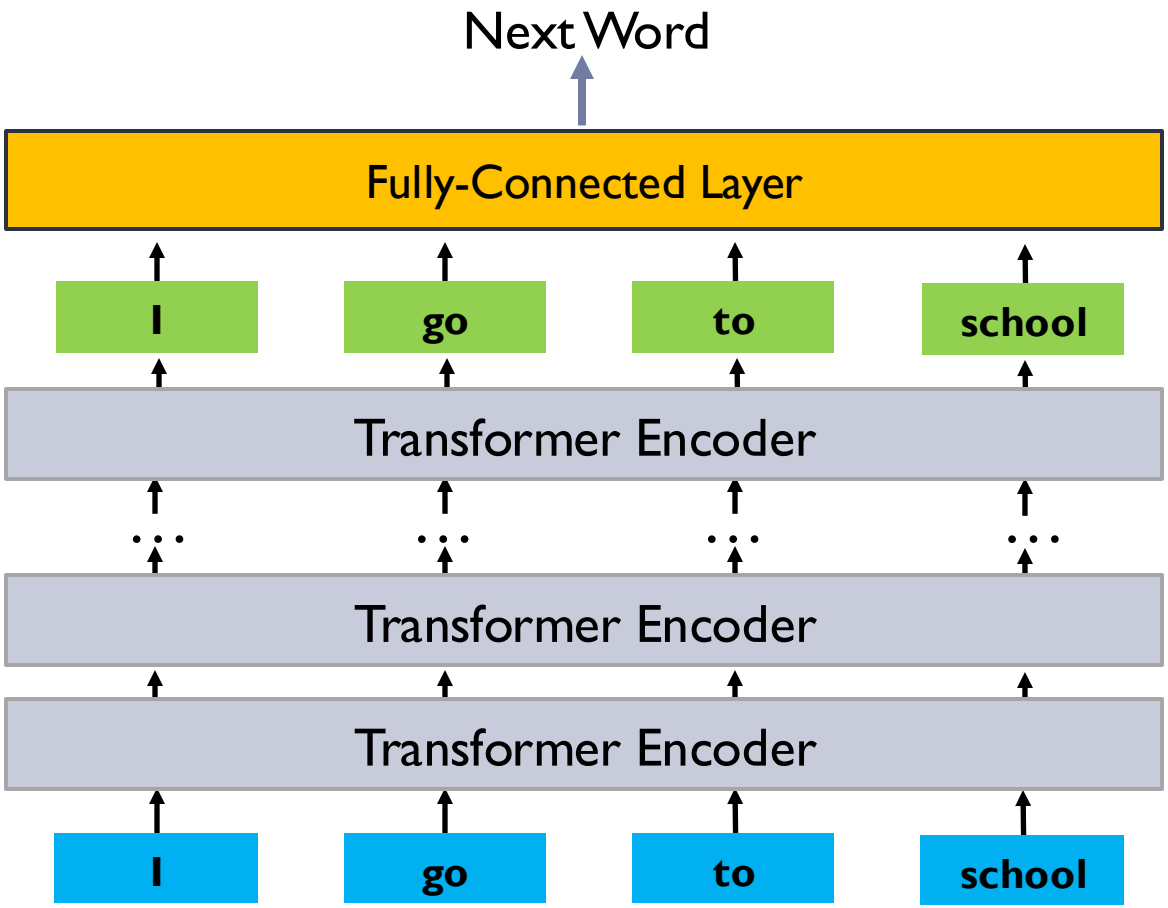
GPT Model

- Word Generation



GPT Model

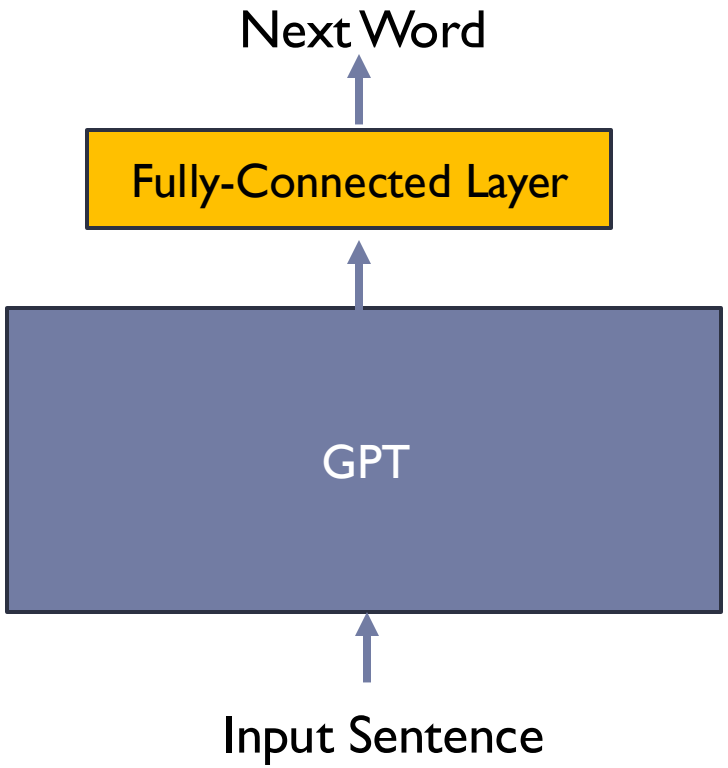
- Transformer Encoder + Fully-Connected Layer



GPT Model

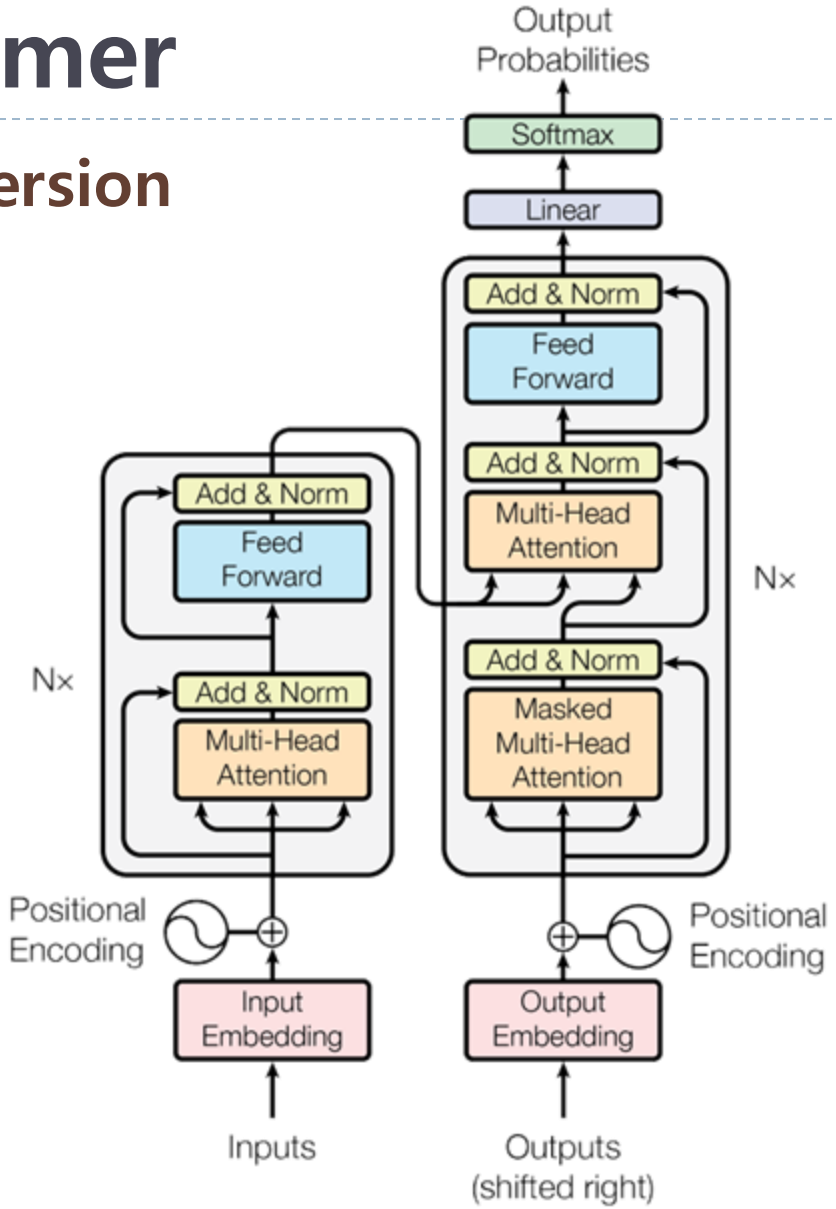
- **Example: English to Korean Translation**

I go to school → 나는
I to to school 나는 → 학교에
I to to school 나는 학교에 → 갑니다
A bird is flying → 새가
A bird is flying 새가 → 날고
A bird is flying 새가 날고 → 있어요



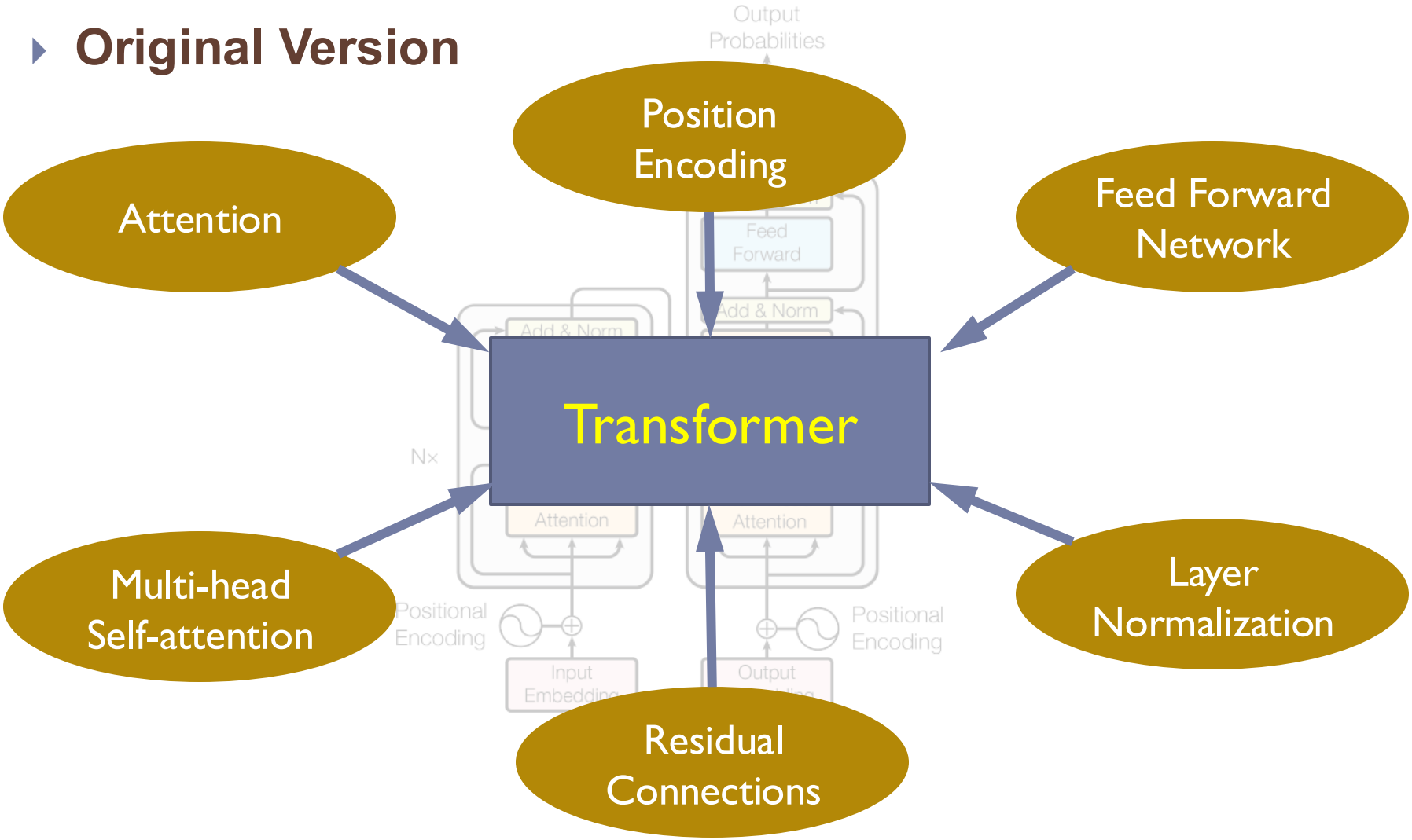
Transformer

- **Original Version**



Transformer

- **Original Version**



Question and Answer