

<https://arxiv.org/abs/1810.04805>

BERT (Bidirectional Encoder Representations from Transformers) : Pre-training of Deep Bidirectional Transformers for Language Understanding

KUBIG 15기 남정재

BERT

간단히 말해서...

Bidirectional Encoder Representations from Transformers

>> Transformer를 Bidirection한 Encoder의 형태로!

- Bidirection : 양방향
- Encoder : 입력 값을 숫자 형태로 바꾸는 모듈

즉, 문맥을 양방향을 위해 숫자의 형태로 바꿔주는 딥러닝 모델

양방향으로 문맥을 이해하는 식으로 진행되는 Language Model



BERT



BERT

BERT의 핵심은...

1) 전이학습 모델

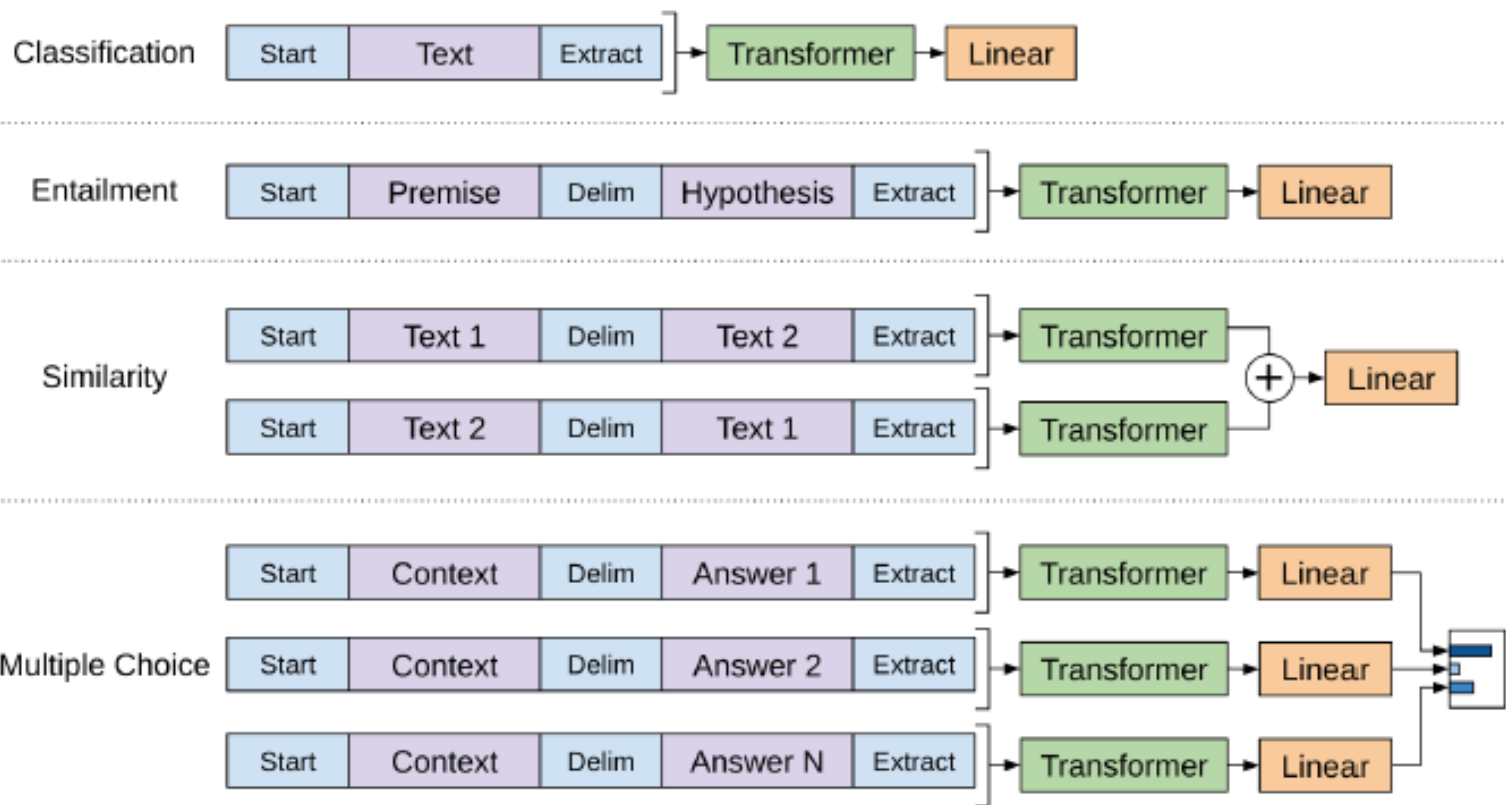
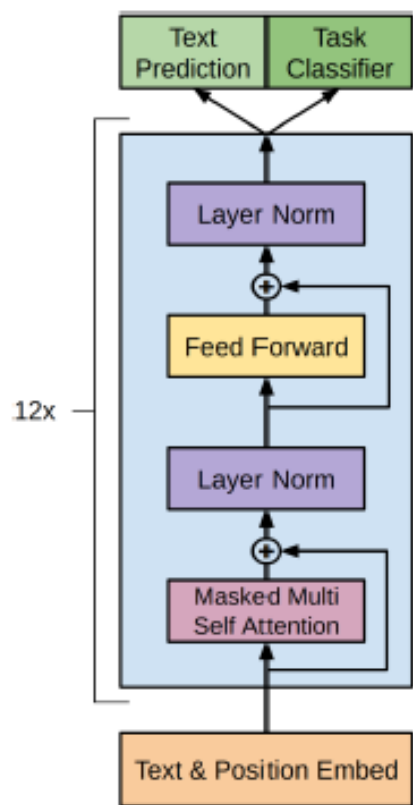
- 사전 학습된 대용량의 레이블링 되지 않은 데이터를 이용하여 언어 모델을 학습
- 이를 토대로 목적에 맞는 작업을 위한 신경망을 추가하는 전이 학습 방법

2) 사전학습 모델

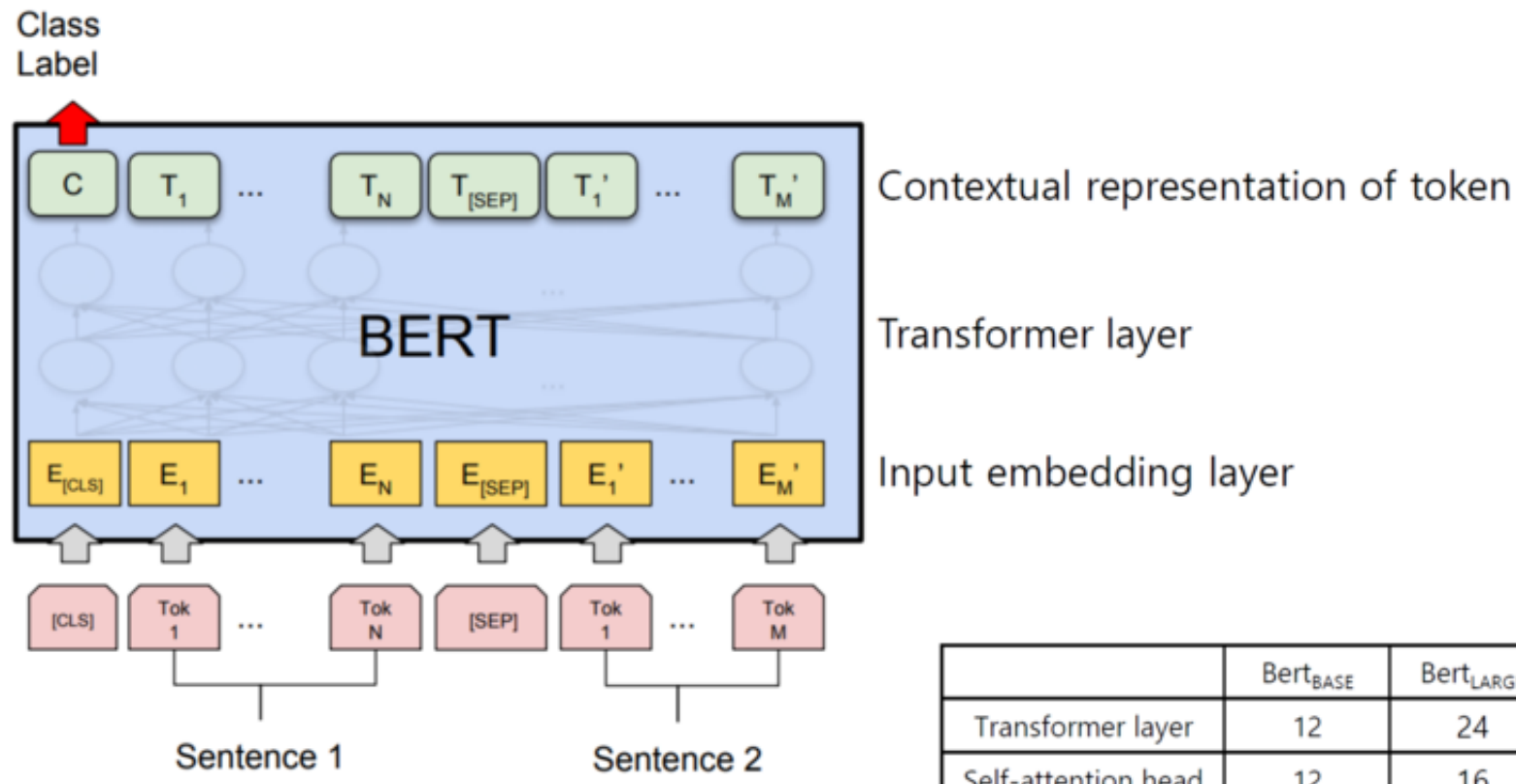
- 기본적으로 대량의 단어 임베딩 등에 대해 사전 학습이 되어 있는 모델 제공
- 상대적으로 적은 자원 만으로도 finetuning을 통해 다양한 작업 수행 가능



GPT-1

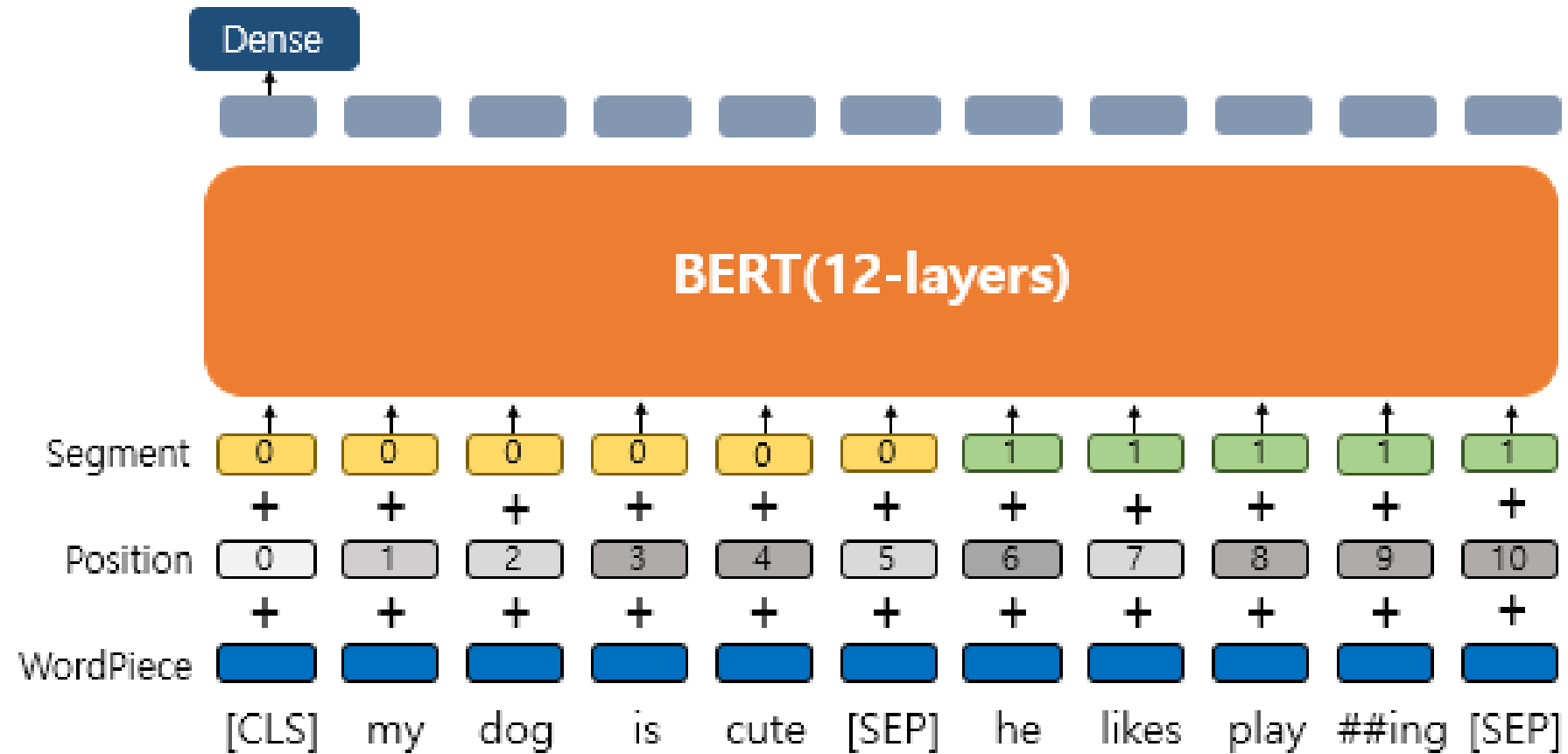


BERT

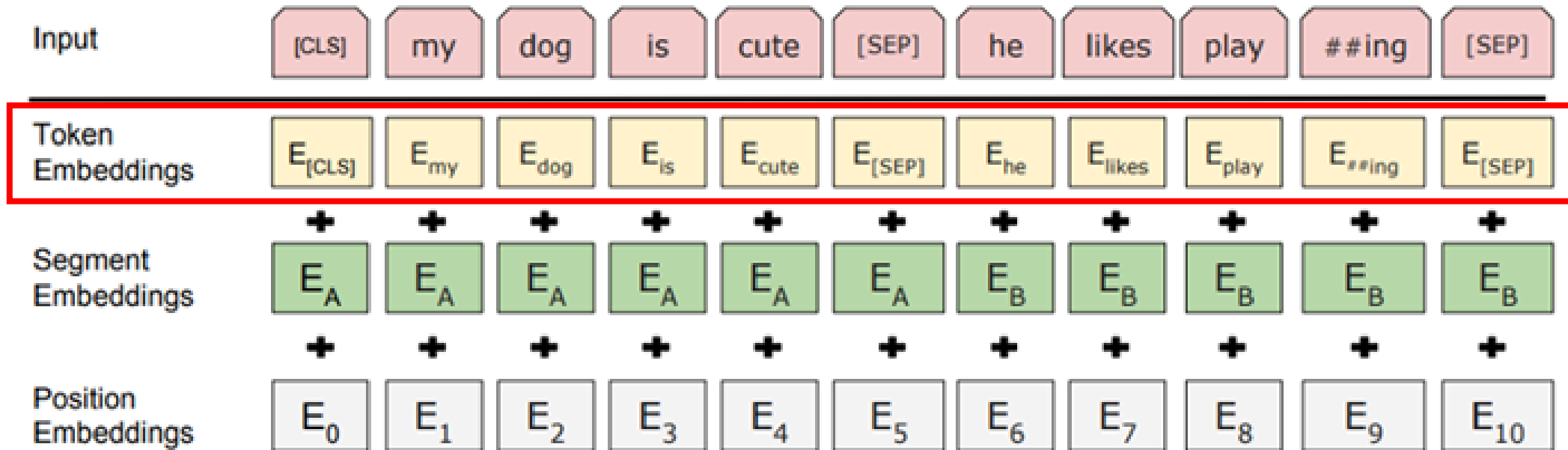


	Bert _{BASE}	Bert _{LARGE}
Transformer layer	12	24
Self-attention head	12	16
Total	110M	340M

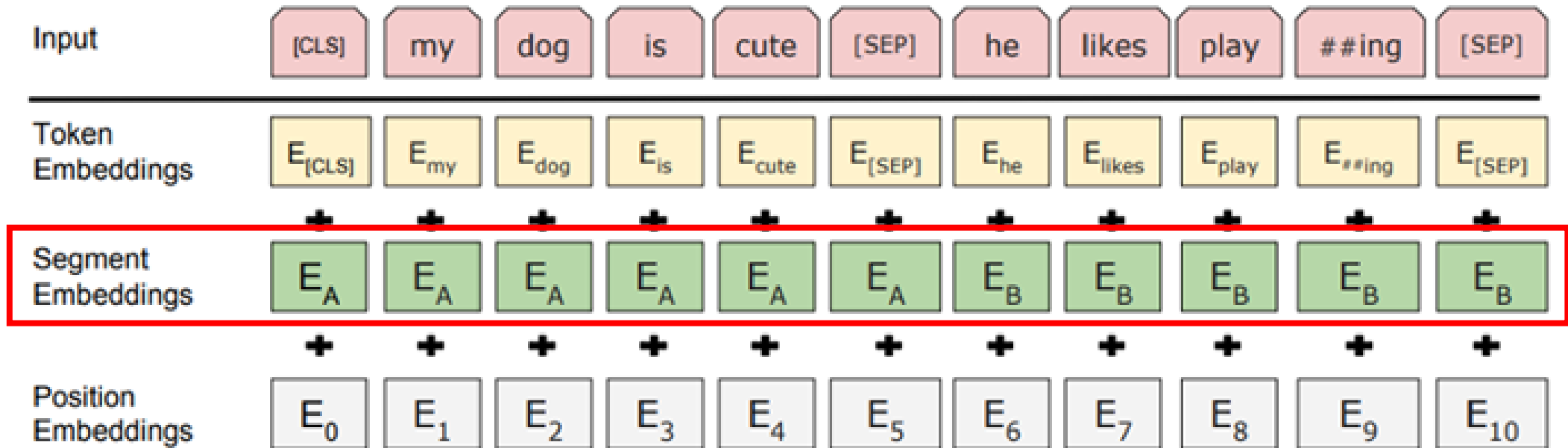
BERT



BERT



BERT



BERT

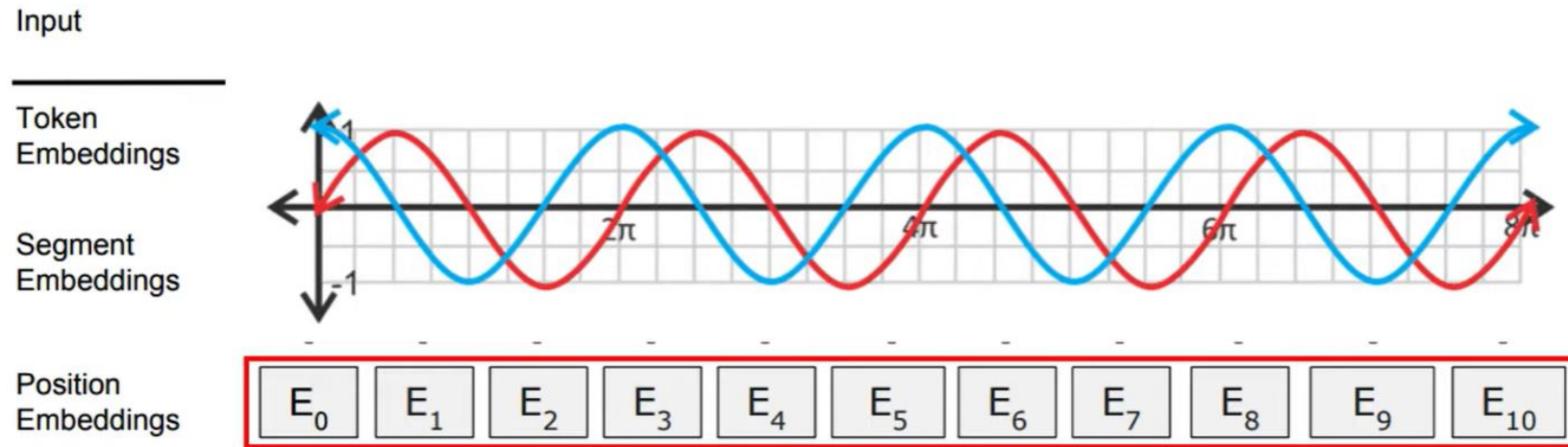
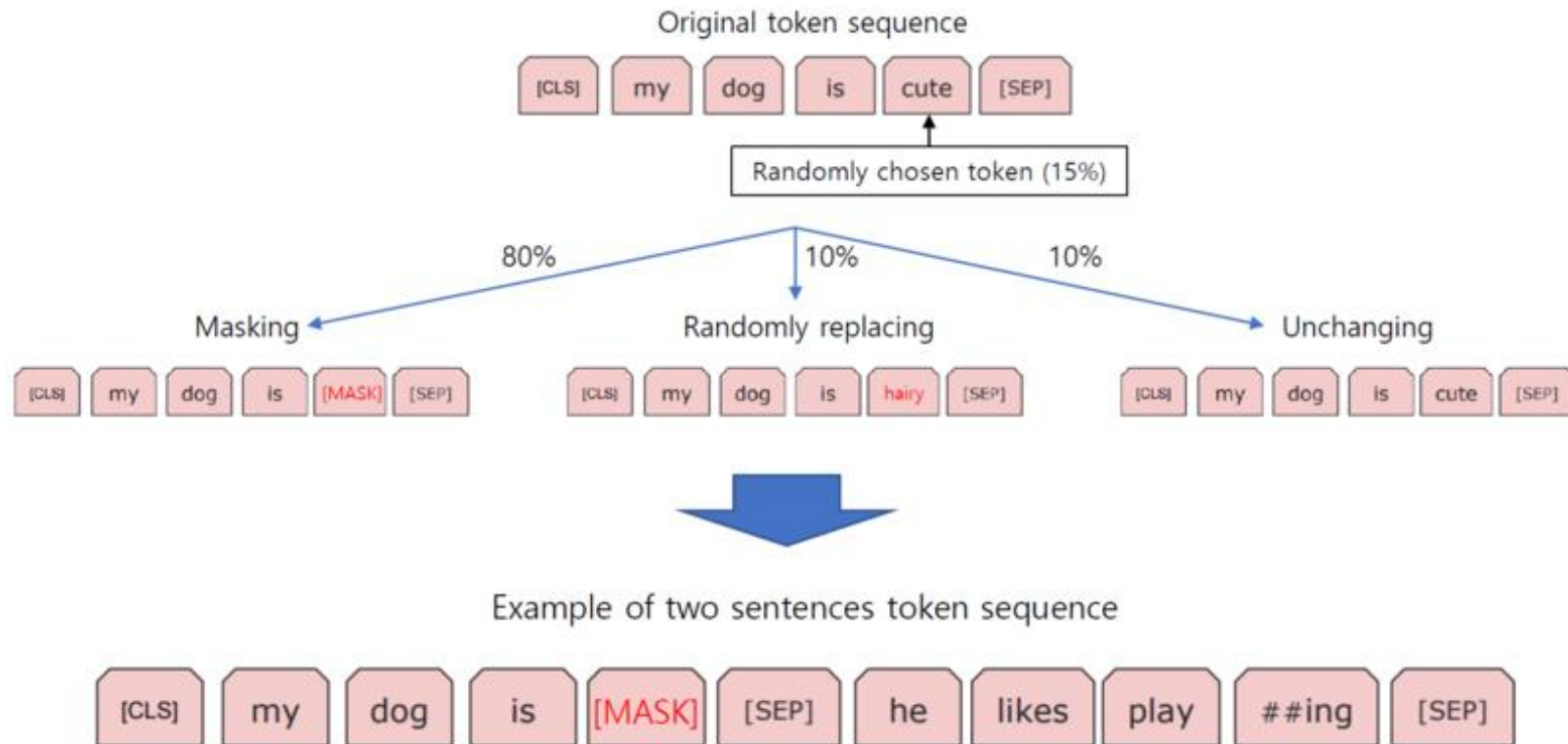
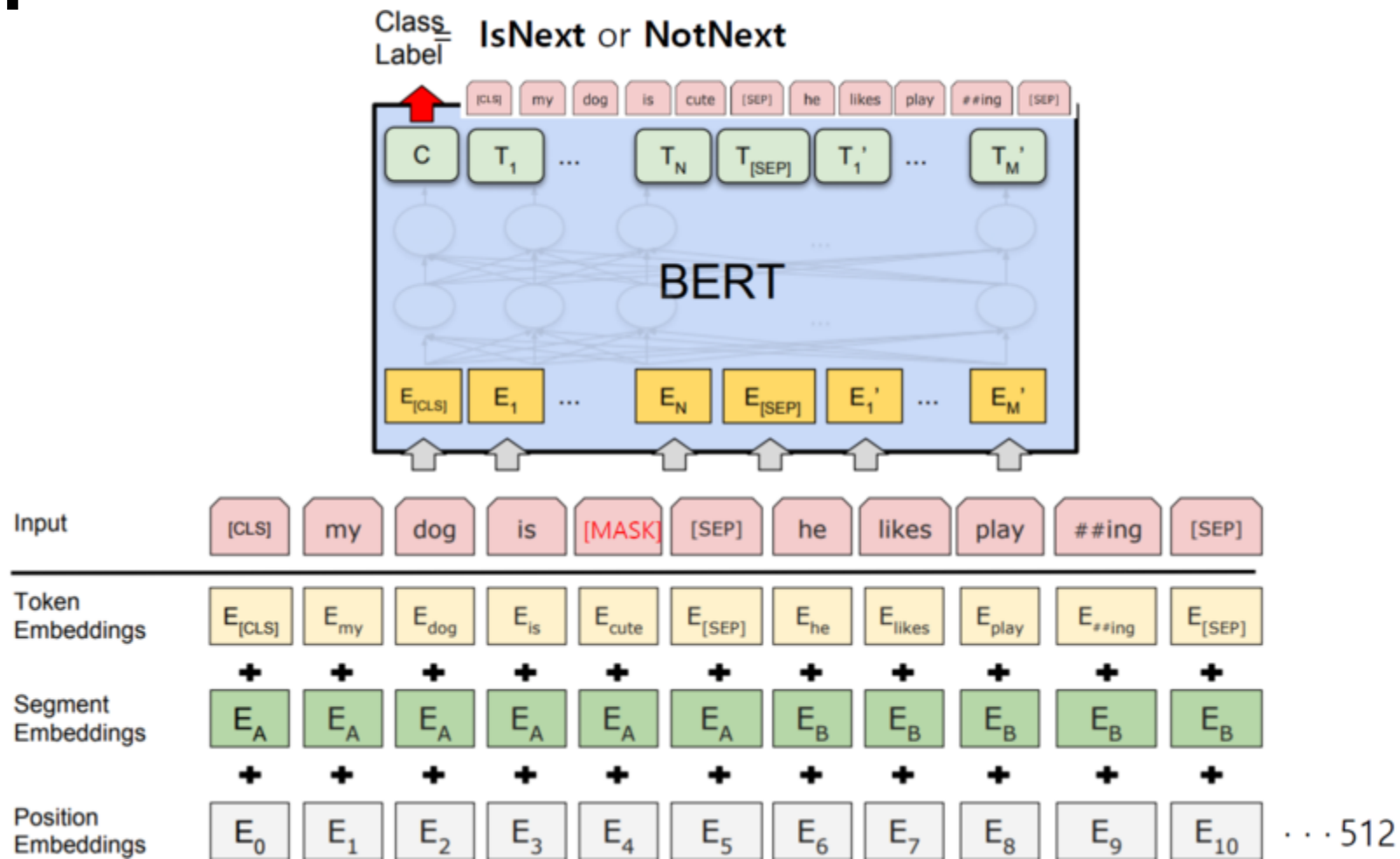


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

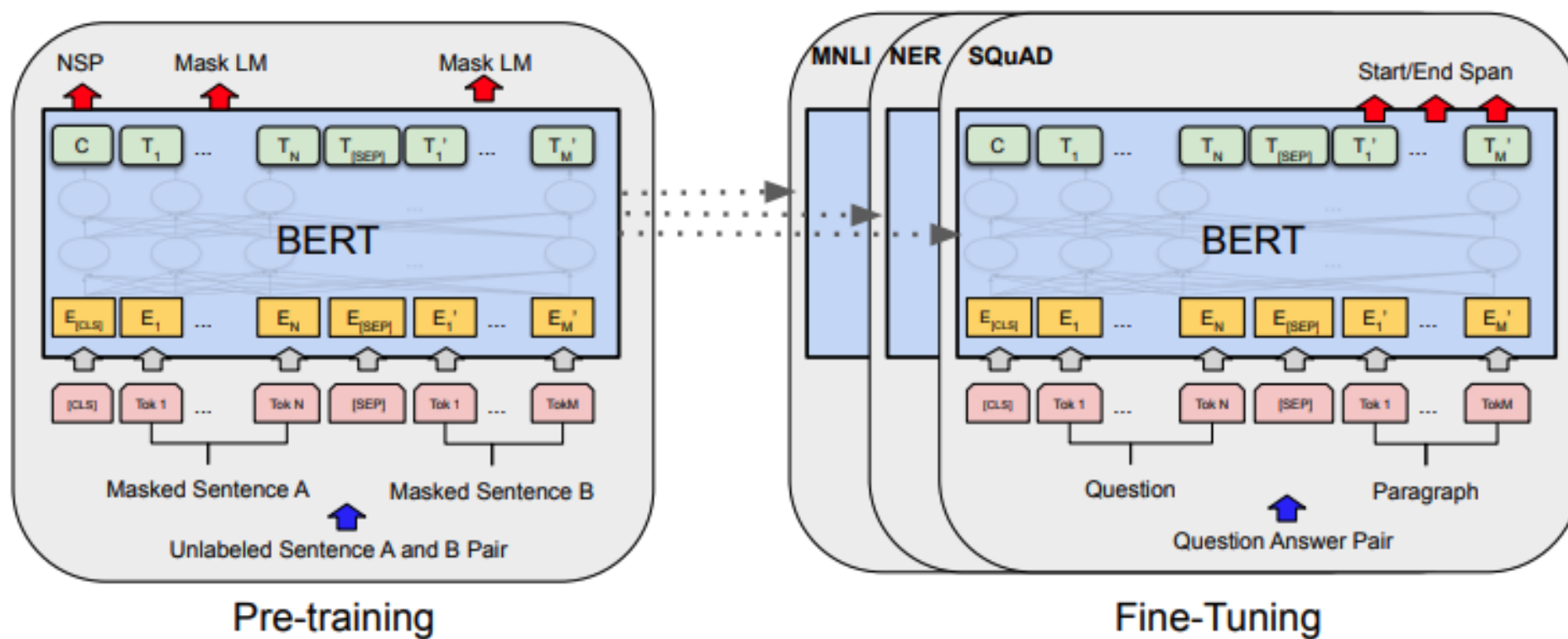
BERT



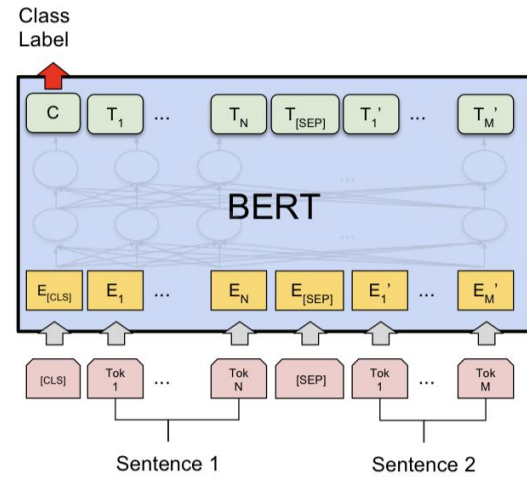
BERT



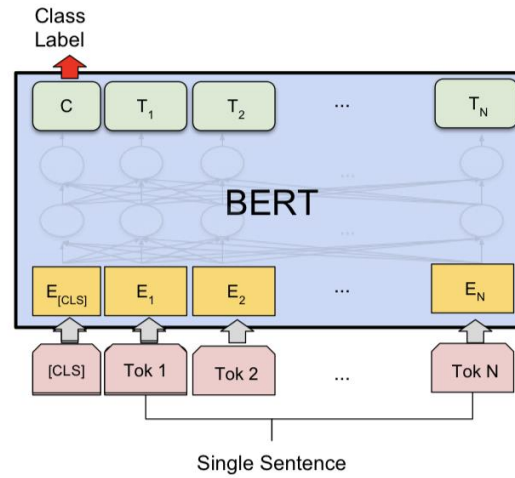
BERT



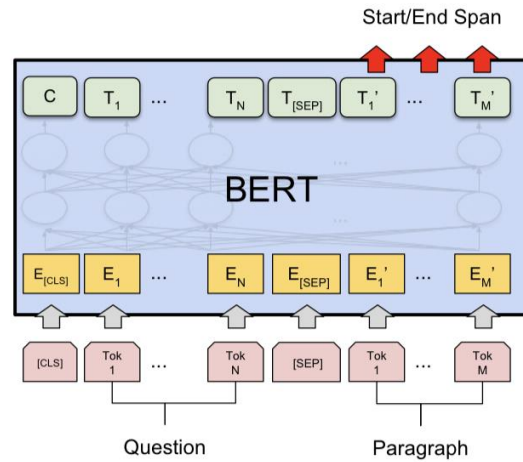
BERT



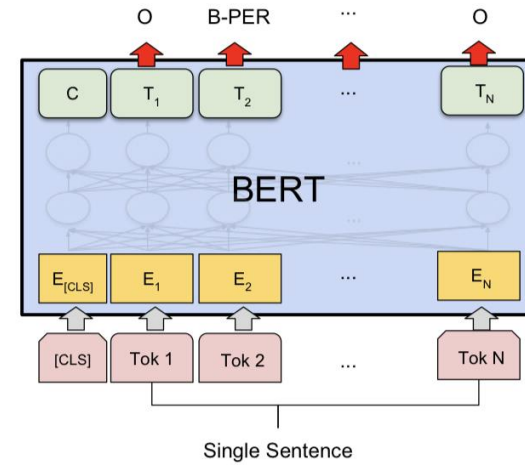
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



(b) Single Sentence Classification Tasks:
SST-2, CoLA



(c) Question Answering Tasks:
SQuAD v1.1



(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

BERT
- The end -