

Automatic Post Editing in NMT

2021/02/03

문현석



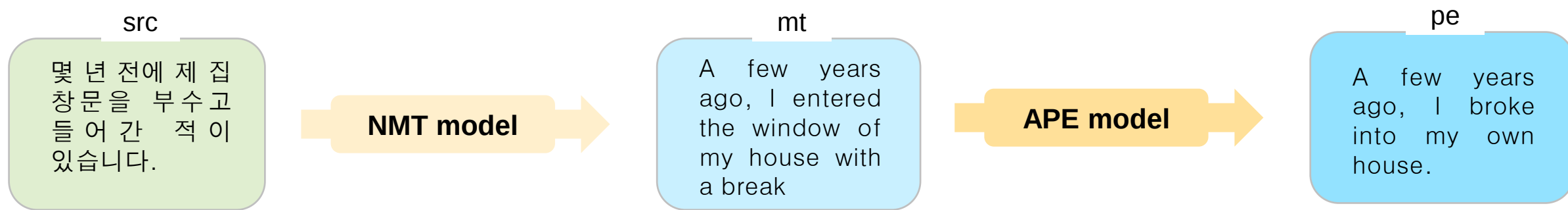
고려대학교
KOREA UNIVERSITY

Contents

- APE
- 기존 APE의 접근방법
- 현재 APE SOTA모델
- 앞으로의 연구 계획

Automatic Post Editing (APE)

APE : 기계번역 사후 교정



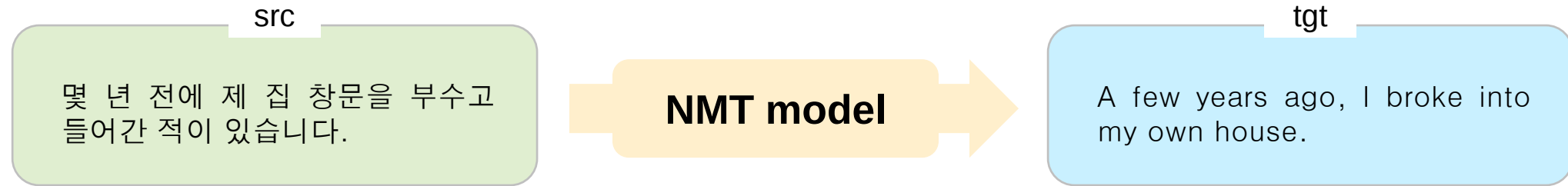
- 번역 결과문의 오류를 수정하기 위한 또 하나의 번역기
- 사람의 직접적인 수정작업 없이 더 뛰어난 번역결과물 생성
- 일반적인 도메인에서 생성된 번역문을 특정 도메인에 적합한 번역문으로 변경 (domain adaptation)

Automatic Post Editing (APE)

NMT: (src, tgt)로 이루어진 병렬 코퍼스를 통해 학습

src: source문장(입력)

tgt: target문장

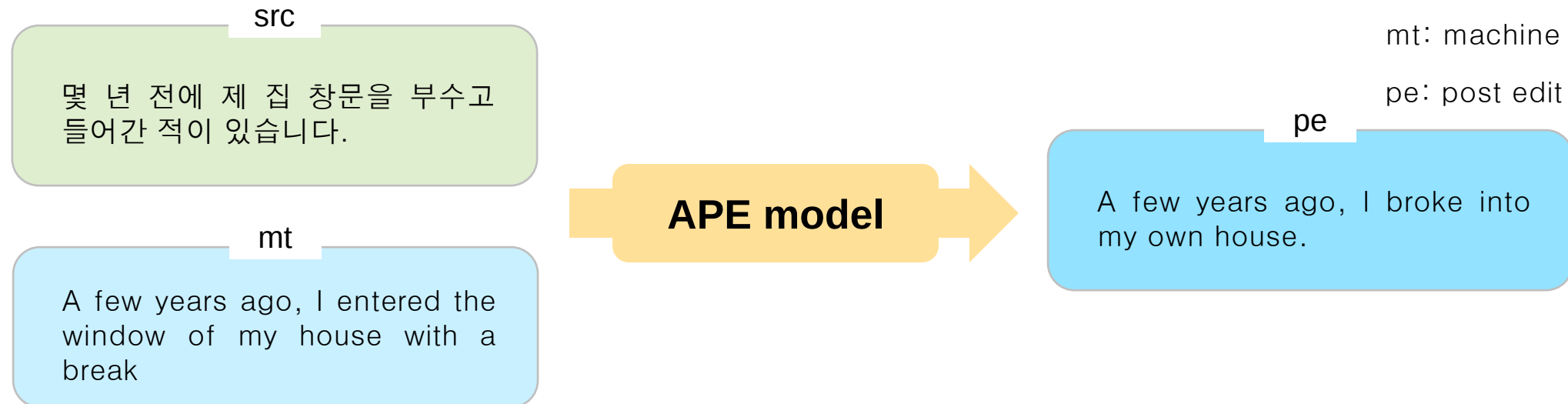


APE: (src, mt, pe)로 이루어진 코퍼스를 통해 학습

src: source문장(입력)

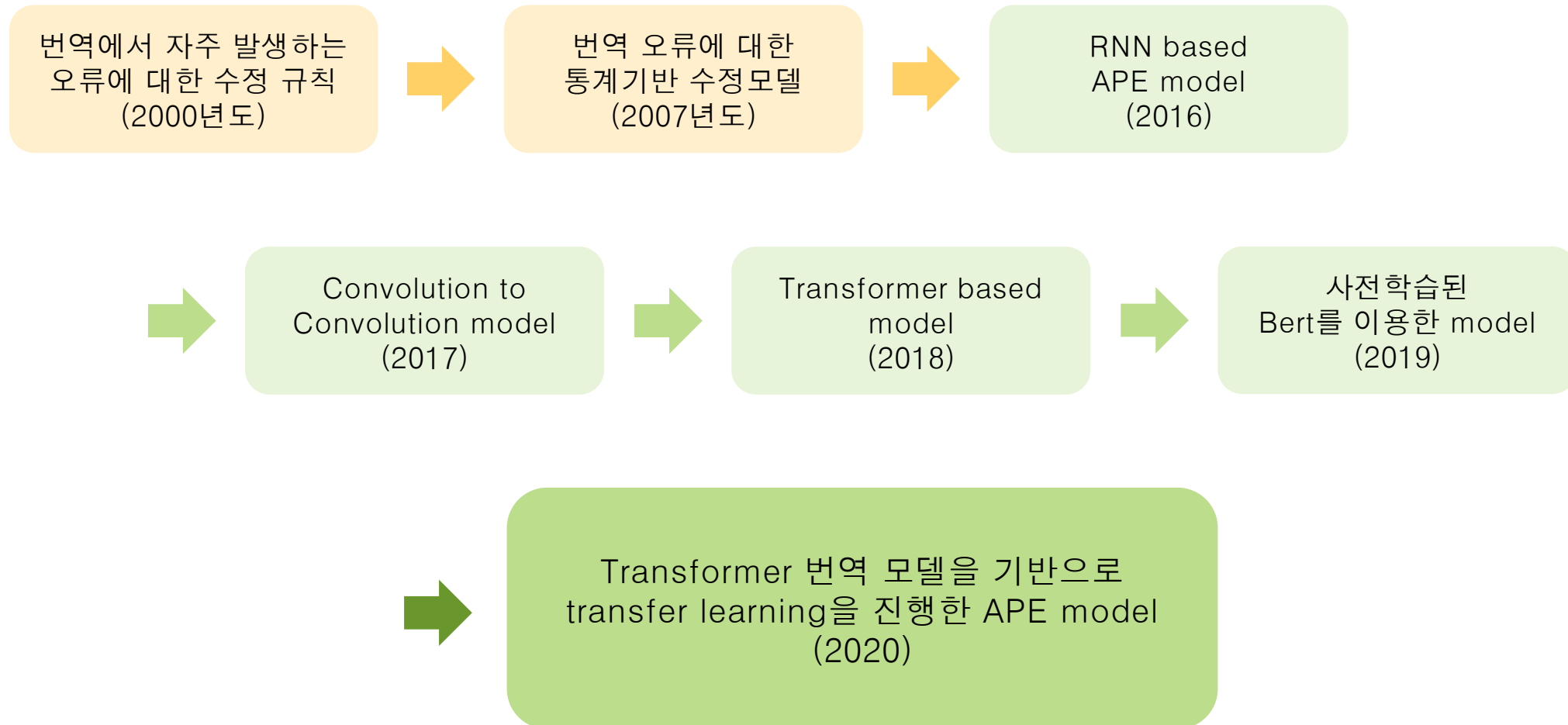
mt: machine translation

pe: post edit



Automatic Post Editing (APE)

- APE연구 발전단계



Automatic Post Editing (APE)

- 번역 된 문장에 추가적인 처리를 가하여 더 좋은 문장을 만들기
- (src, mt, pe) triplet으로 구성된 Dataset이 요구됨
- TER (Translation Error Rate), BLEU score로 성능 평가

$$* \text{TER} = \frac{\# \text{ of required edits}}{\text{average \# of tokens in pe}} \cdot \text{mt에 있는 token들 중, 교정이 필요한 token의 비율}$$

- WMT2020 En-De, En-Zh
일본어 한국어 APE데이터는 있지만, 한국어가 source문장인 데이터는 존재하지 않음

Automatic Post Editing (APE)

WMT2019 APE En-De

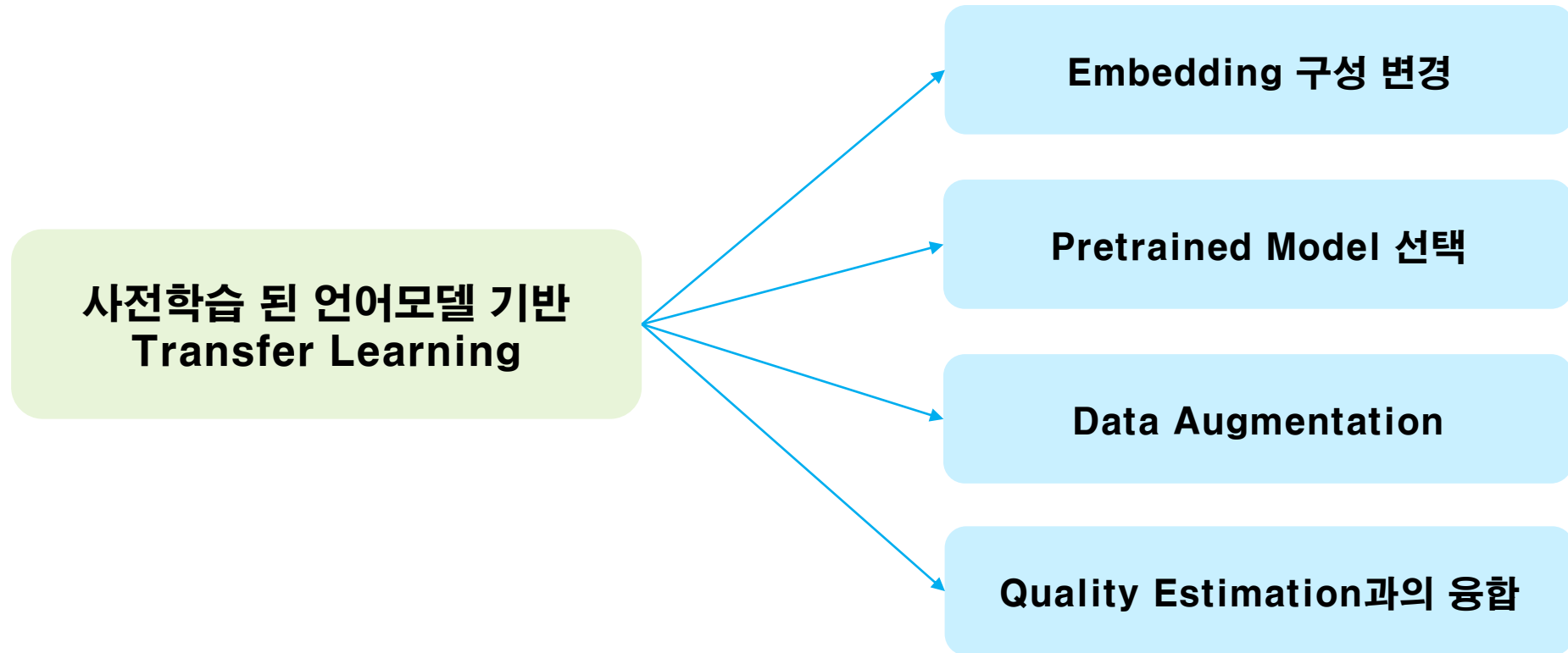
ID	TER (pe)	BLEU (pe)	TER (ref)	BLEU (ref)	TER (pe+ref)	BLEU (pe+ref)
UNBABEL Primary	16.06*	75.96	41.66	44.95	15.58	78.1
POSTECH Primary	16.11*	76.22	42.04	44.57	15.68	78.08
POSTECH Contrastive (var2Ens8)	16.13*	76.21	42.09	44.53	15.73	78.05
USAAR_DFKI Primary	16.15*	75.75	41.84	44.65	15.69	77.84
POSTECH Contrastive (top1Ens4)	16.17*	76.15	42.09	44.52	15.74	78.01
UNBABEL Contrastive (2)	16.21*	75.7	41.59	45.08	15.72	77.98
UNBABEL Contrastive (1)	16.24*	75.7	41.62	45.01	15.76	77.97
FBK Primary	16.37*	75.71	42.18	44.39	15.90	77.54
FBK Contrastive	16.61†	75.28	42.12	44.49	16.1	77.43
UDS Primary	16.77†	75.03	42.64	43.78	16.34	76.83
IC_USFD Contrastive	16.78†	74.88	42.45	44.01	16.31	76.82
UDS Contrastive (Gaus)	16.79†	75.03	42.55	44.0	16.33	76.87
UDS Contrastive (Uni)	16.80†	75.03	42.66	43.79	16.37	76.85
IC_USFD Primary	16.84†	74.8†	42.58	43.86	16.41	76.68
Baseline	16.84	74.73	42.24	44.2	16.27	76.83
ADAPT_DCU Contrastive (SMT)	17.07	74.3	42.40	44.14	16.54	76.36
ADAPT_DCU Primary	17.29	74.29	42.41	44.09	16.81	76.51
USAAR_DFKI Contrastive	17.31	73.97	42.45	43.71	16.87	76.06
ADAPT_DCU Contrastive (LEN)	17.41	74.01	42.44	44.01	16.91	76.2

WMT2020 APE En-De

	TER	BLEU
en-de		
HW-TSC_DIRECT_CONTRASTIVE.pe	20.21	66.89
HW-TSC_CONCAT_PRIMARY.pe	20.52	66.16
MinD-mem_enc_dec_post-CONTRASTIVE	26.99	55.77
POSTECH-ETRI_XLM-Top4Ens_CONTRASTIVE	27.02	56.37
MinD-mem_enc_dec-PRIMARY	27.03	55.58
POSTECH-ETRI_XLM-Top3Ens_PRIMARY	27.37	55.83
BeringLab_model1_PRIMARY	27.61	54.71
BeringLab_model2_CONTRASTIVE	27.96	54.60
POSTECH_TERNNoise-nFold-Ens8_CONTRASTIVE	28.22	54.51
POSTECH_TERNNoise-Ops-Ens8_PRIMARY	28.41	54.22
Baseline	31.56	50.21
KAISTxPAPAGO_EMT_PRIMARY	32.00	49.21

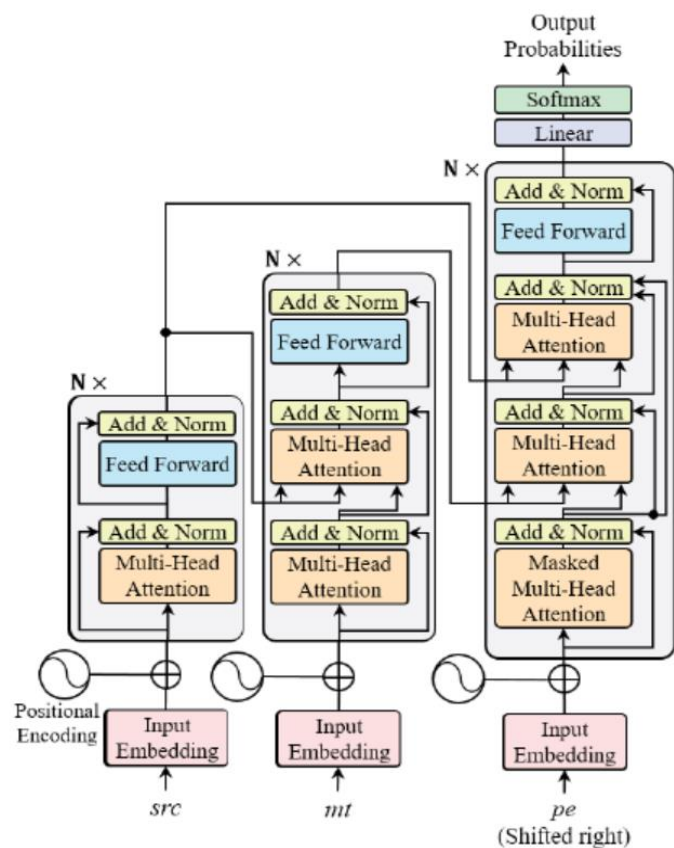
Automatic Post Editing (APE)

- APE 성능개선을 위한 연구 방향



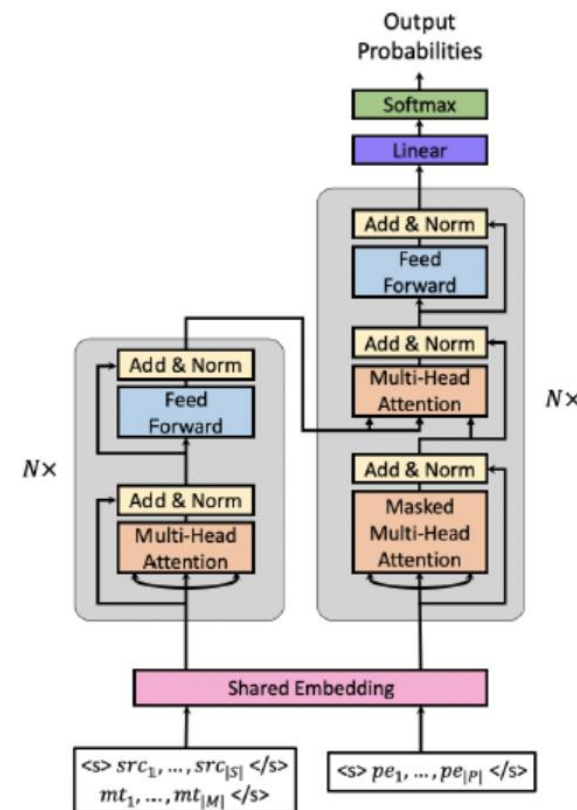
Input of APE

src와 mt를 처리하는 encoder를 따로 구성



Transformer-based Automatic Post-Editing Model with Joint Encoder and Multi-source Attention of Decoder (ACL 2019)

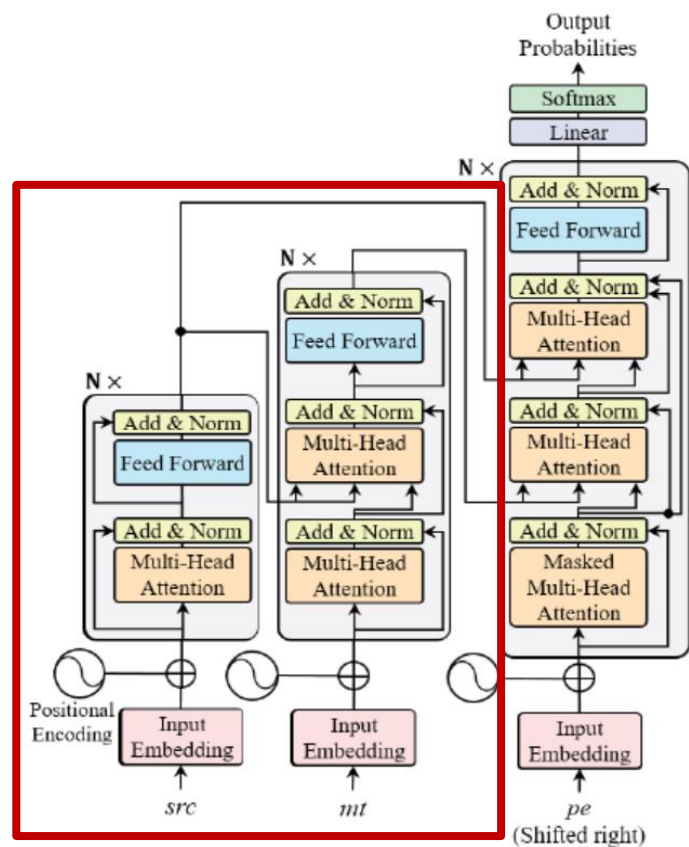
src와 mt를 한 문장으로 연결하여 입력을 구성



Cross-Lingual Transformers for Neural Automatic Post-Editing (WMT2020)

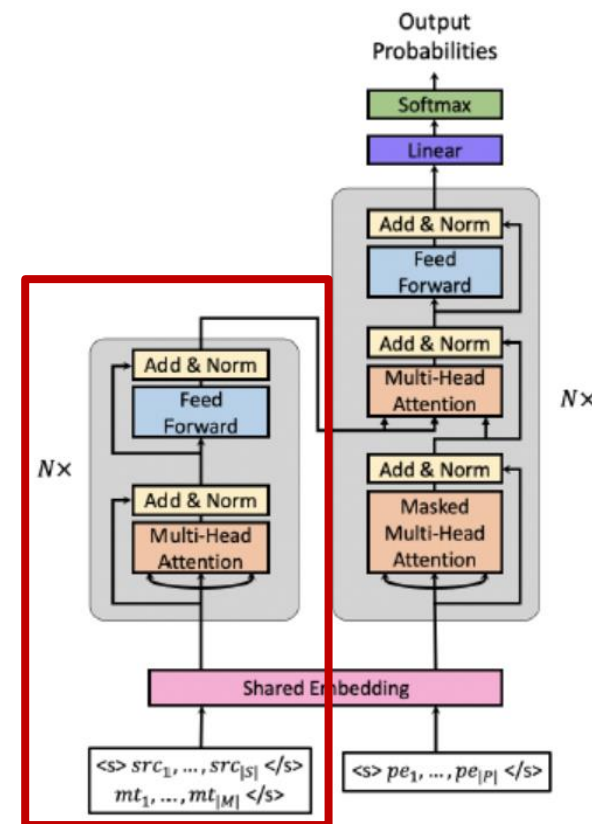
Input of APE

src와 mt를 처리하는 encoder를 따로 구성



Transformer-based Automatic Post-Editing Model with Joint Encoder and Multi-source Attention of Decoder (ACL 2019)

src와 mt를 한 문장으로 연결하여 입력을 구성



Cross-Lingual Transformers for Neural Automatic Post-Editing (WMT2020)

Transfer Learning

Language model 기반 사전학습 모델

- 같은 언어에 대한 denoising이 학습된 모델
- 기계가 번역한 결과물(MT)을 noise가 있는 데이터로 간주. Noise가 없는 원본 데이터(PE)로 복원

MT를 중심으로 PE를 생성 (SRC를 참고)

NMT model 기반 사전학습 모델

- 다른 언어에 대한 번역이 학습된 모델
- 다른 모델에서 번역한 결과물(MT)을 참고하여 SRC를 PE로 번역

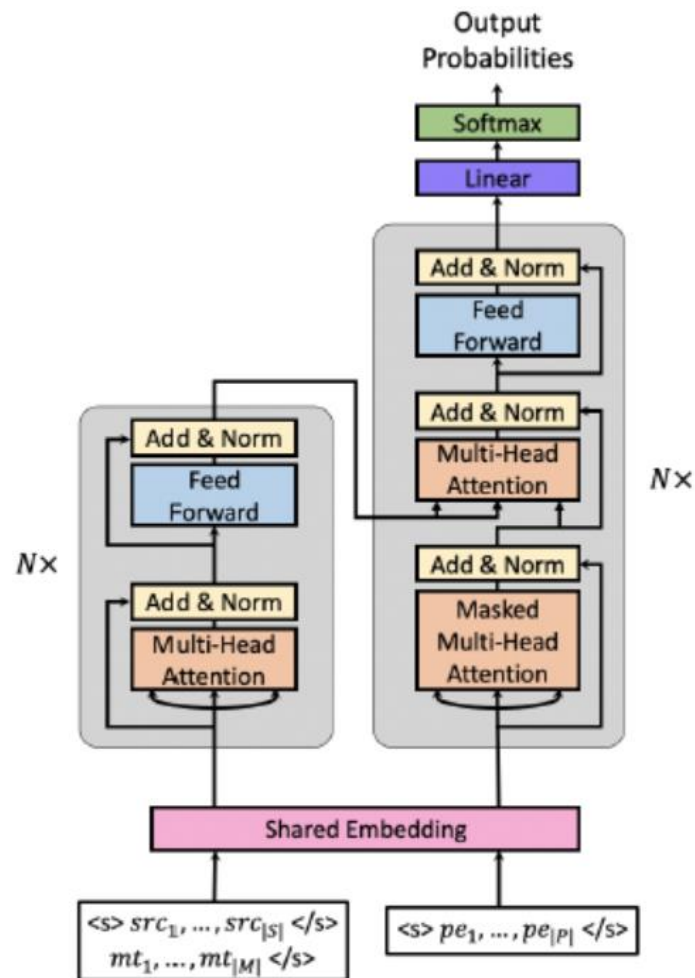
SRC를 중심으로 PE를 생성 (MT를 참고)

Transfer Learning

LM based APE

1)

BERT-base 모델을
encoder 구조에만 활용
decoder는 random initialize



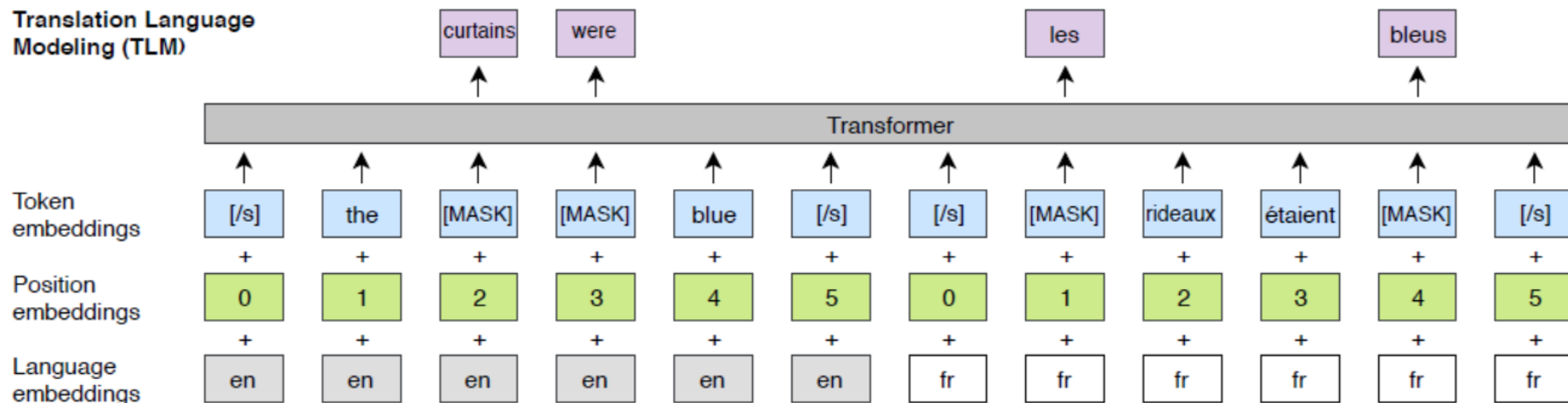
2)

BERT-base 모델을
encoder 구조 뿐 아니라
decoder 구조에도 활용

- 2019년 WMT에서는 mBERT를 encoder, decoder에 적용시킨 모델이 가장 우수한 성능
- 2020년 WMT에서는 decoder를 random initialize
- 현재까지 mBERT, XLM, XLM-R를 사전 학습 모델로 활용

Transfer Learning

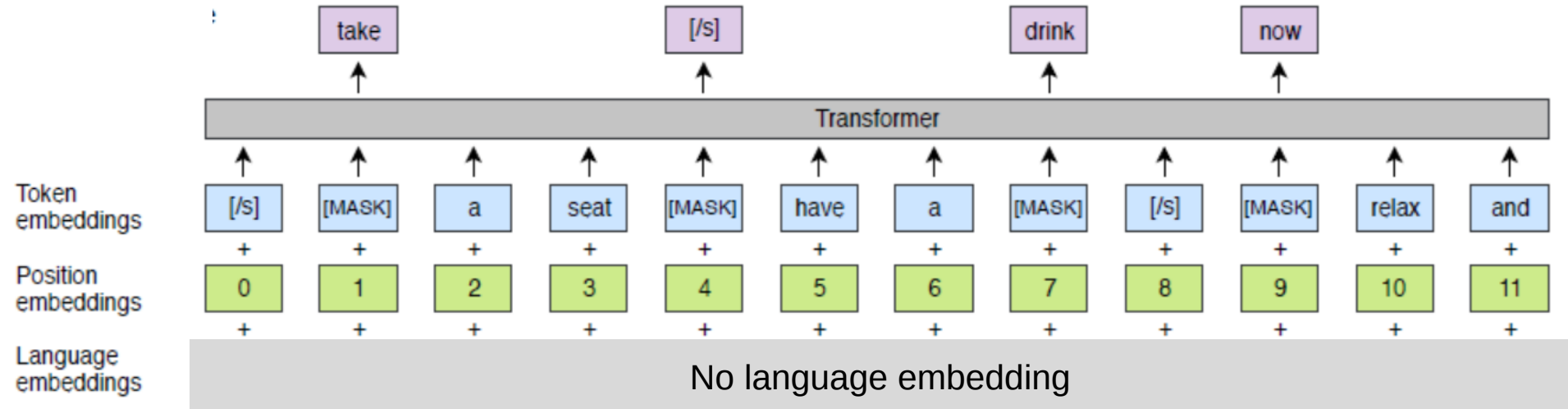
XLM



- Parallel corpus를 통해서 모델을 사전학습하는 방법론 제시
- src와 tgt를 한 문장으로 연결하여 모델의 입력을 구성함
- MLM과 마찬가지로, 전체 문장의 일부를 [MASK] token으로 치환
- src의 [MASK]를 위해서 tgt의 정보를 이용
- tgt의 [MASK]를 위해서 src의 정보를 이용

Transfer Learning

XLM-R



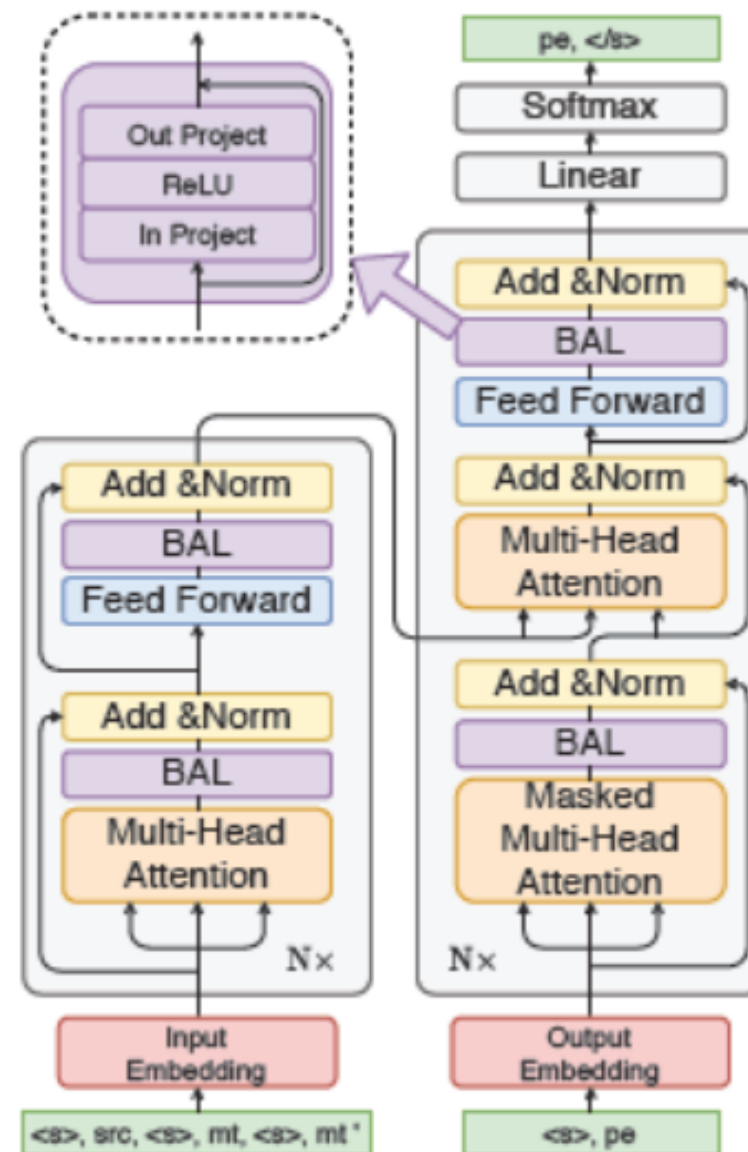
- 100개 언어의 mono corpus들로 MLM기반의 사전학습 진행
- RoBERTa 학습방법
- Common crawl 데이터로 학습
- Language embedding을 사용하지 않음

Transfer Learning

NMT based APE

HW-TSC's Participation at WMT 2020 Automatic Post Editing Shared Task (WMT2020)

- 사전 학습된 transformer NMT모델에 APE를 fine tuning
- 전체 모델을 fine tuning하지 않고, BAL(bottleneck adapter layer)구조를 도입하여 해당 부분만을 tuning
- 데이터 증강기법을 적용하지 않고, 주어진 APE training data만으로 훈련 진행
- src와 mt를 한 문장으로 연결하는 것 이상으로, src와 pe를 통해 자체적으로 생성한 mt' 을 추가적으로 연결하여 입력을 구성



Data Augmentation of APE

- WMT2020에서 제공되는 APE훈련 데이터는 7,000개
- 성능 향상을 위하여 대부분의 논문에서 데이터 증강기법 채용

eSCAPE

병렬 코퍼스(src, tgt)를 통해
자체적인 번역 모델 개발.
번역 모델을 통해 src를 mt로 번역

pe를 tgt로 하여 triplet(src, mt, pe)
구축

**eSCAPE: a Large-scale Synthetic
Corpus for Automatic Post-Editing**

Denoise

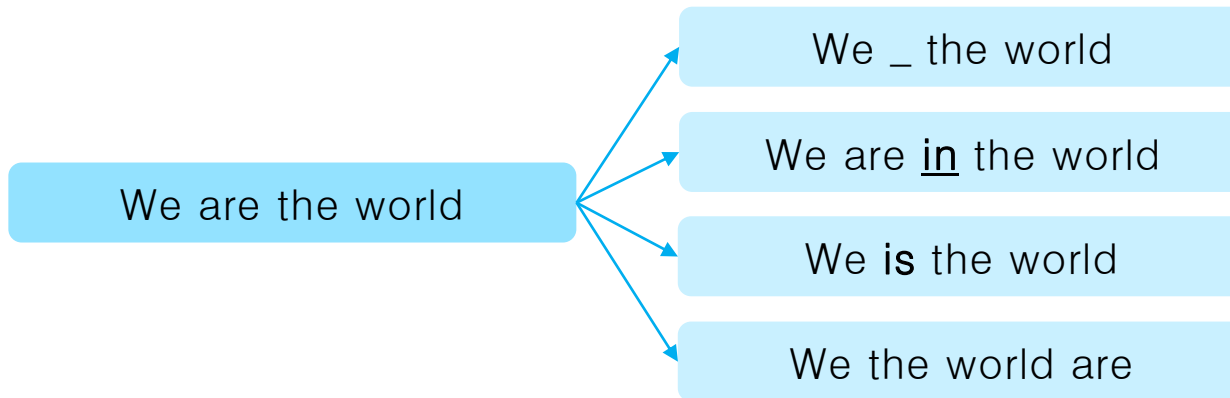
(src, tgt)로 구성된 병렬 코퍼스의
tgt에 noise를 추가하여 mt를 생성

pe를 tgt로 하여 triplet(src, mt, pe)
구축

**Noising Scheme for Data
Augmentation in Automatic
Post-Editing**

Data Augmentation of APE

- 충분한 데이터가 확보되지 않은 언어쌍의 경우, eSCAPE방법론 적용하기 어려움
- Parallel corpus가 NMT모델을 만들만큼 충분히 확보되지 않은 언어에서도 데이터 증강 가능



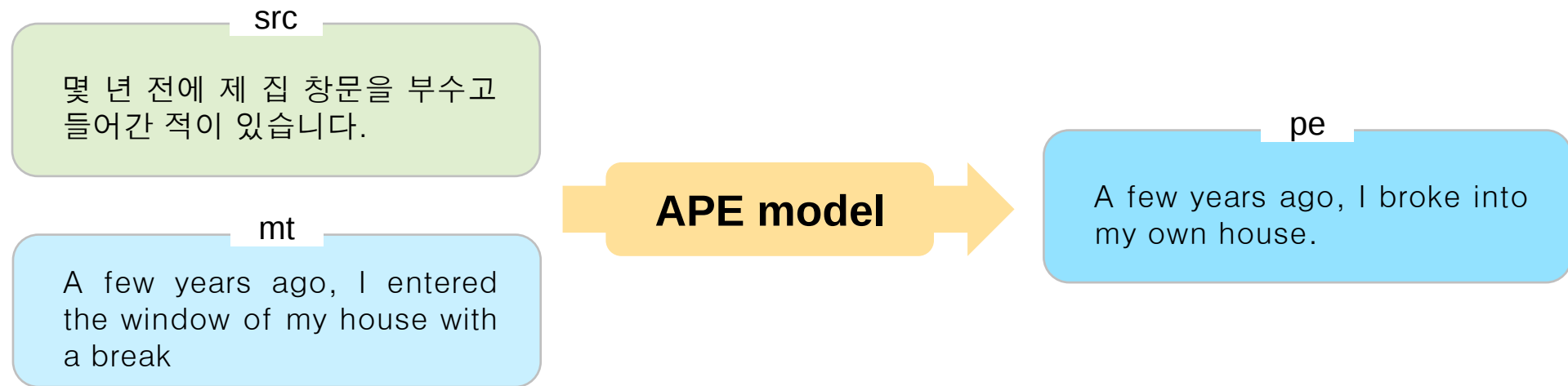
Denoise

(src, tgt)로 구성된 병렬 코퍼스의 tgt에 noise 추가.

Noising Scheme for Data Augmentation in Automatic Post-Editing

QE in APE

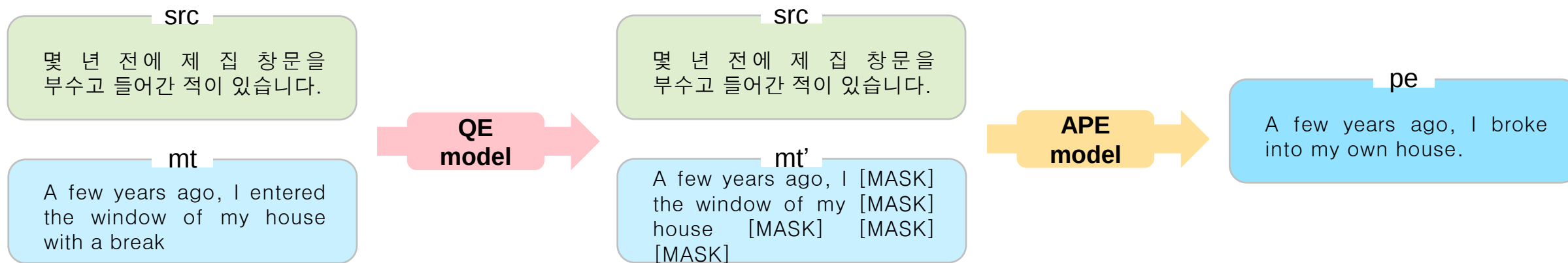
- QE: 번역된 문장의 품질을 예측하는 모델 (단어수준, 문장수준, 문단수준)



QE in APE

- QE: 번역된 문장의 품질을 예측하는 모델 (단어수준)
- mt의 token들 중, 교정이 필요한 부분을 [MASK]로 치환

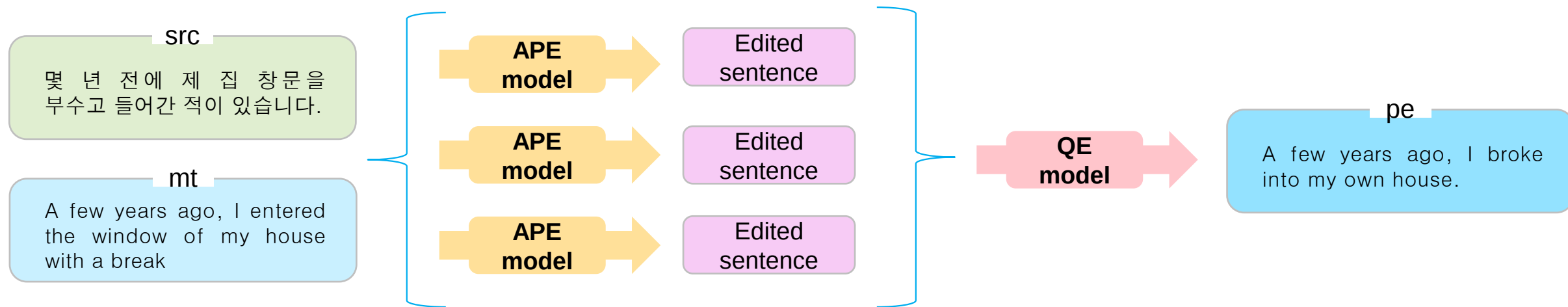
Cross-Lingual Transformers for Neural Automatic Post-Editing (WMT2020)



QE in APE

- QE: 번역된 문장의 품질을 예측하는 모델 (문장수준)
- 여러 APE모델을 합친 ensemble 모델에서 최고 품질의 번역 문장을 선정하는 용도

Cross-Lingual Transformers for Neural Automatic Post-Editing (WMT2020)



QE in APE

Systems	TER↓	BLEU↑
Baseline (MT Output)	31.37	50.37
APE Transformer	29.52	52.70
APE Transformer + Sentence-QE	29.10	52.99
Word-QE + Sentence-QE	30.27	50.83
APE Transformer + Word-QE + Sentence-QE	28.83	52.80
+ Ensemble	28.47	53.82

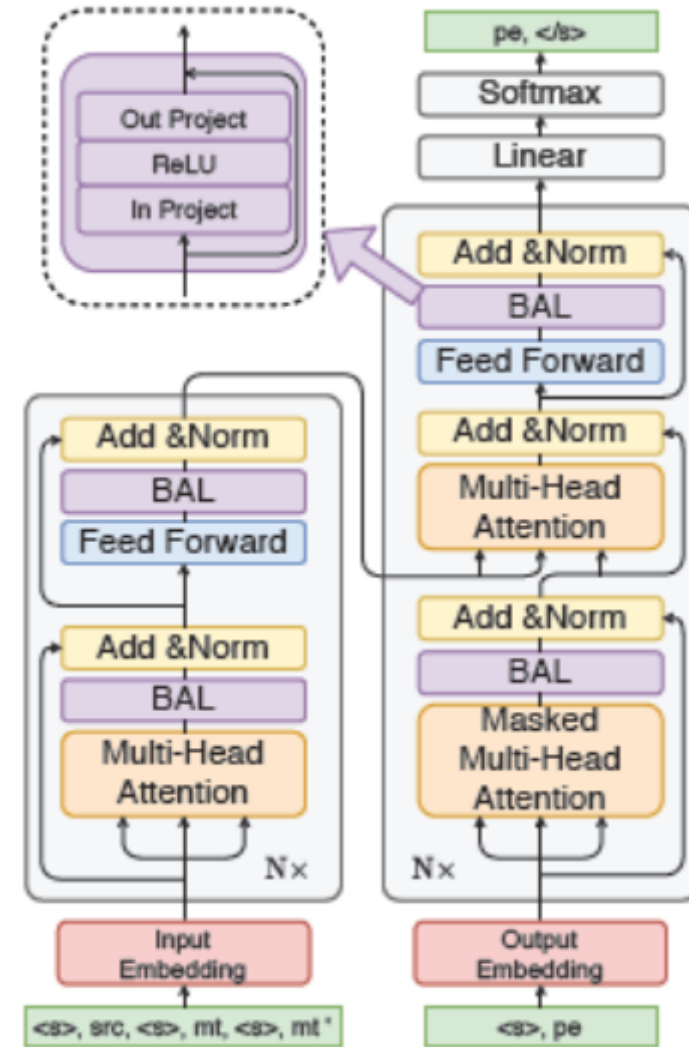
- QE를 적용한 Ensemble 모델을 통해 더 좋은 교정 성능을 기대할 수 있음
- 단, APE모델과 별개로 QE데이터를 통해 QE모델을 훈련해야 함

APE Summary

- 입력문장 구성방법: separate src, mt / combined src, mt
- 사전학습 모델: LM based / NMT based
- 데이터 증강: eSCAPE / denoise / No augmentation
- QE의 적용여부

APE SOTA

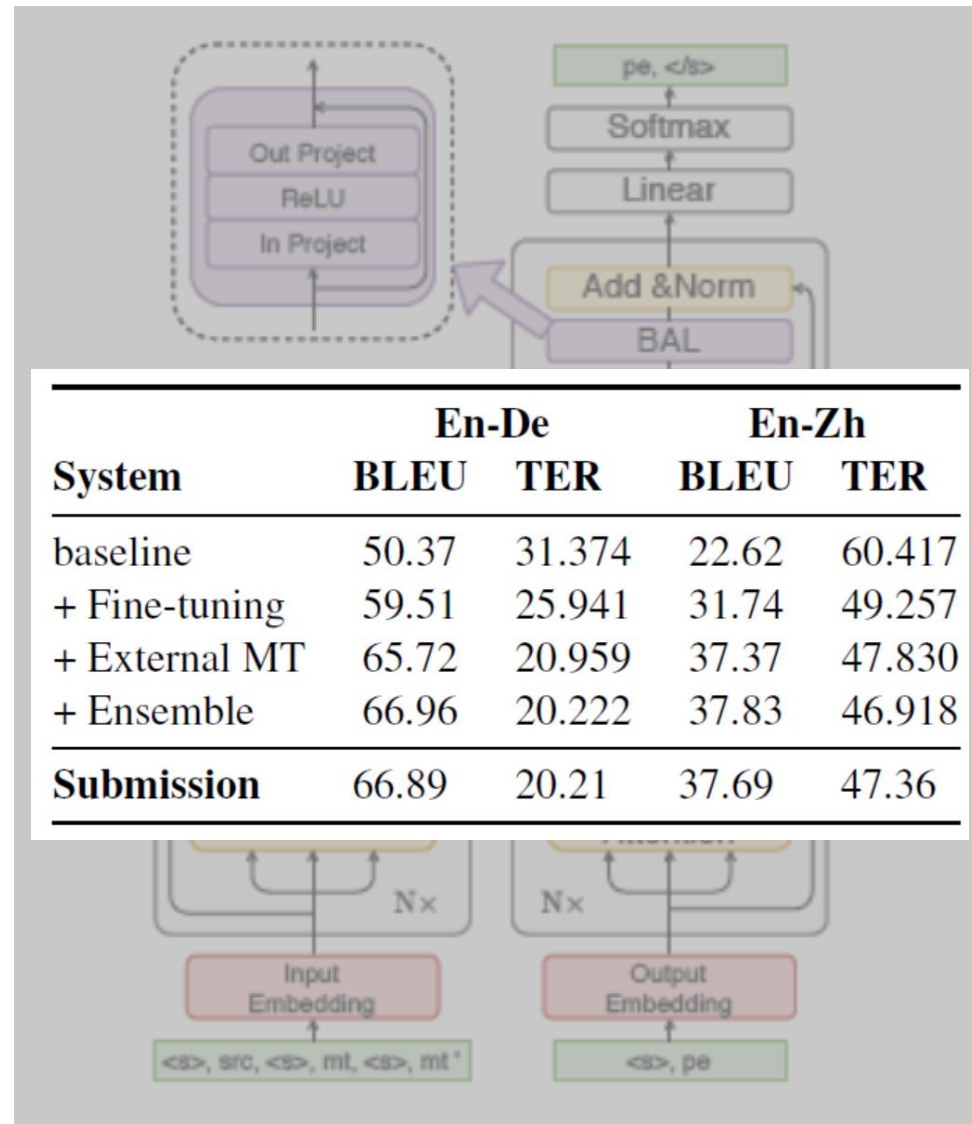
1. LM based 모델이 아닌, NMT 모델을 사전 학습 모델로 활용함
2. Fine tuning 시 BAL 구조 도입
3. Data augmentation을 진행하지 않고, 주어진 훈련 데이터만으로 학습
4. 입력을 구성할 때, 자체적으로 생성한 mt' 을 추가적으로 연결
→ 입력 문장은 <s> src <s> mt <s> mt'
5. QE 모델을 활용한 ensemble 모델



HW-TSC's Participation at WMT 2020 Automatic Post Editing Shared Task

APE SOTA

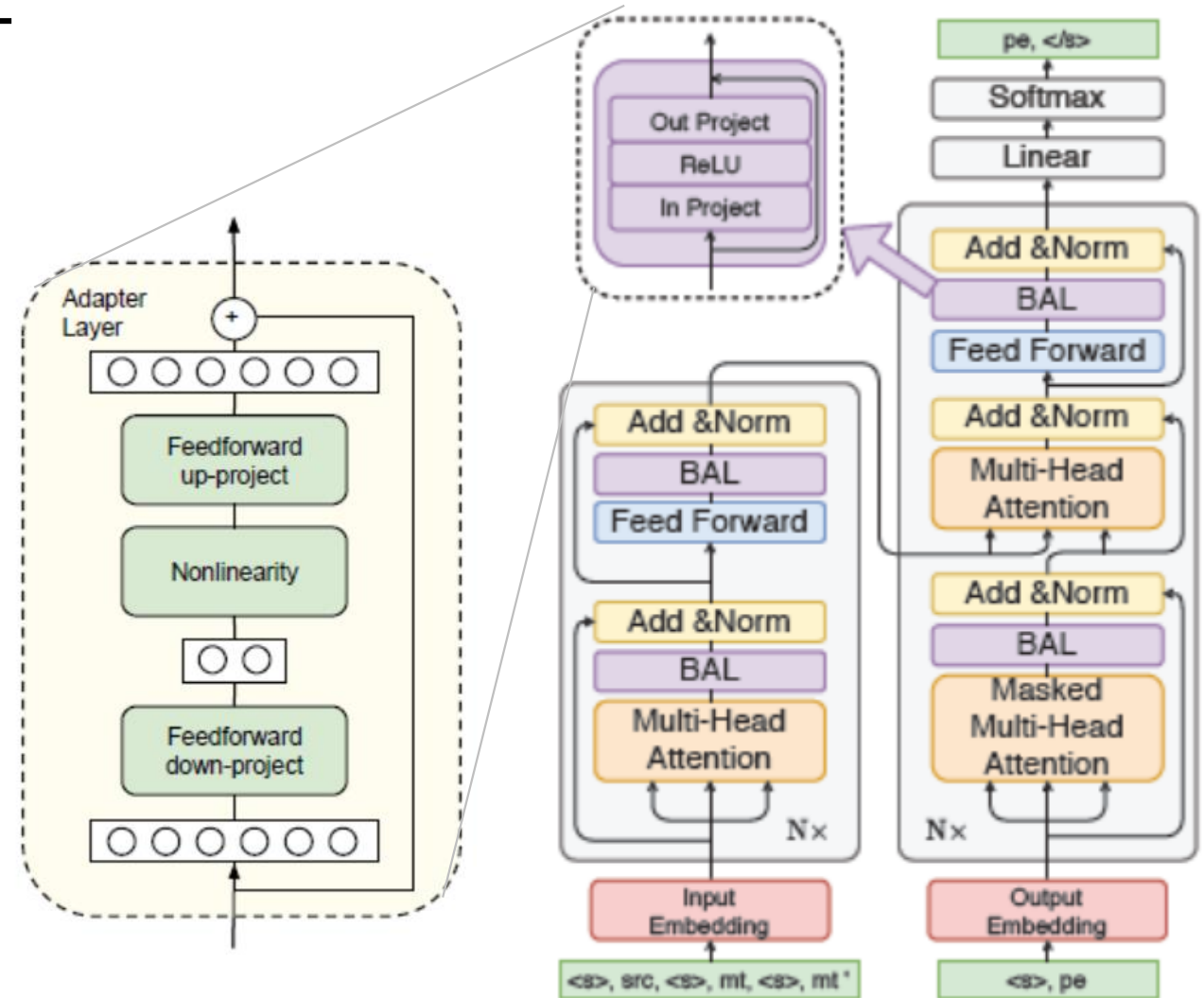
1. LM based 모델이 아닌, NMT 모델을 사전 학습 모델로 활용함
2. Fine tuning 시 BAL 구조 도입
3. Data augmentation을 진행하지 않고, 주어진 훈련 데이터만으로 학습
4. 입력을 구성할 때, 자체적으로 생성한 mt' 을 추가적으로 연결
→ 입력 문장은 <s> src <s> mt <s> mt'
5. QE 모델을 활용한 ensemble 모델



HW-TSC's Participation at WMT 2020 Automatic Post Editing Shared Task

Bottleneck Adapter Layer

- Feed forward network와 같이 Dense layer 2개와 activation function으로 구성됨
- Pretrained model은 학습 과정 중 freeze되고, finetuning 학습은 BAL에서만 진행됨
- Adapter layer에 skip-connection 구조를 도입
→ 만약 Adapter layer가 0에 가까운 값으로 학습되는 경우, identity block과 같은 역할
- Bottleneck layer의 dimension은 hidden size의 2배로 설정



Parameter-Efficient Transfer Learning for NLP

앞으로의 연구방향

1. LM based모델이 아닌, NMT 모델을 사전 학습모델로 활용함

2. Fine tuning시 BAL구조 도입

3. Data augmentation을 진행하지 않고, 주어진 훈련 데이터만으로 학습

4. 입력을 구성할 때, 자체적으로 생성한 mt' 을 추가적으로 연결

5. QE모델을 활용한 ensemble모델



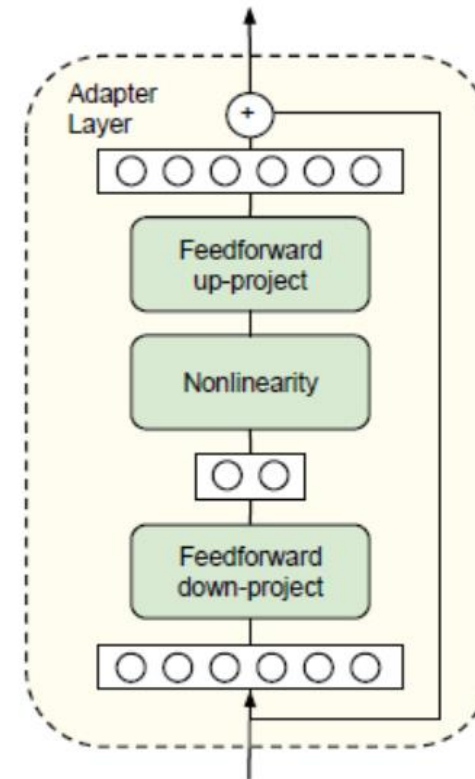
- En-De 데이터로 학습시킨 APE모델을 한국어로 전이 학습



- NMT 이후, 자체적으로 APE를 진행하는 모델

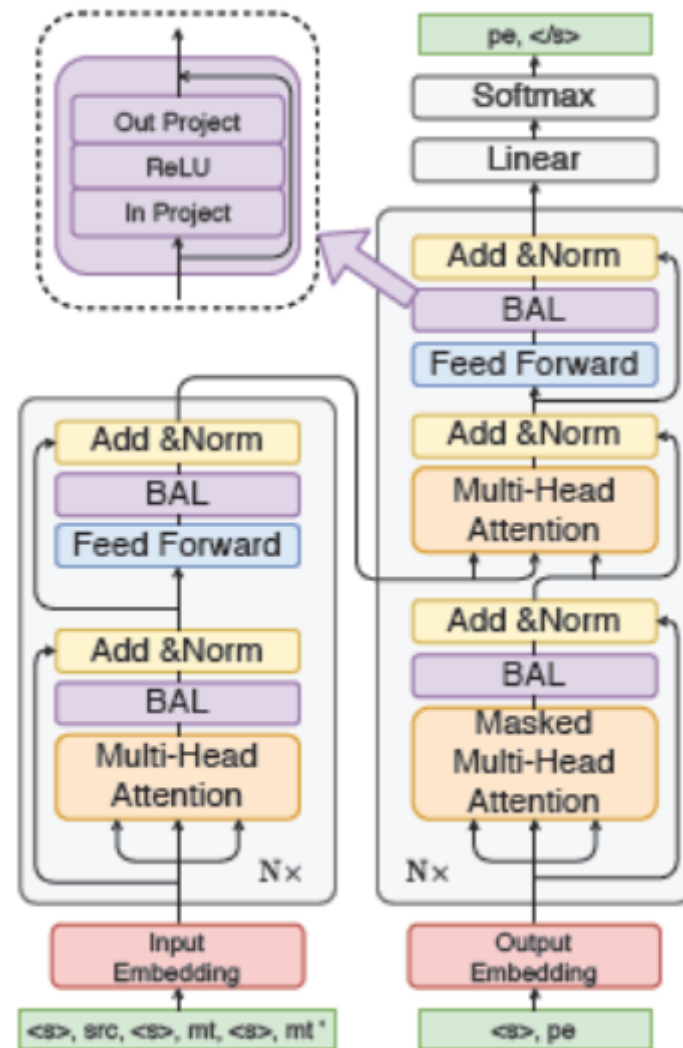
Model Transfer

- Bottleneck adapter layer(BAL)는 사전학습된 모델을 변경하지 않고, 일부 구조를 추가함으로써 쉽게 fine tuning할 수 있다는 장점이 있음
- BAL은 fine tuning 뿐 아니라, 다른 언어로의 전이학습에도 이용될 수 있음. 데이터가 풍부한 언어쌍을 통해 번역을 학습시켜 general NMT모델을 만들고, 저자원언어를 통해 BAL을 학습시켜 번역에서 높은 성과를 보임
(2019, Simple, Scalable Adaptation for Neural Machine Translation)
- APE에서 언어 전이학습은 아직 연구된 바가 없으며, 한국어 APE모델 및 데이터셋도 공개된 바 없음



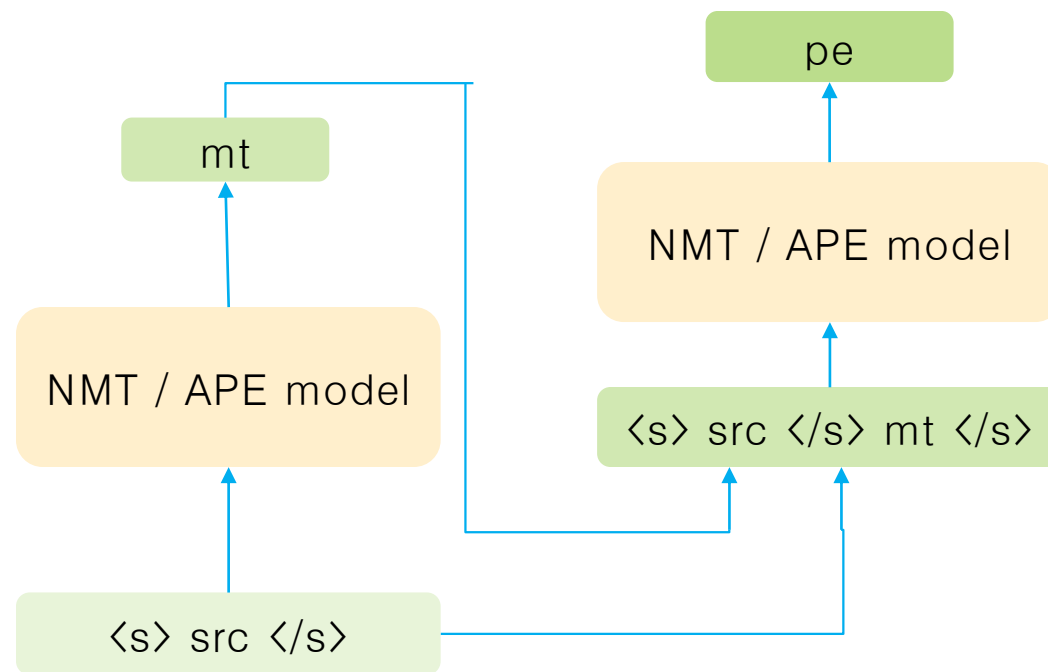
Self APE

- APE SoTA모델을 NMT모델로 해석
- 입력을 src, mt, mt' 으로 구성했을 시, src와 mt를 입력으로 넣어 번역했을 때보다 더 좋은 번역성능을 얻음
- 번역모델에서 mt를 입력에 추가한다면 더 좋은 성능을 얻을 수 있을 것으로 예상



Self APE

- 번역모델이 생성한 mt를 다시 번역모델의 입력으로 구성하는 방법론 구상
- src만을 입력으로 받는 NMT모델이 수정하지 못했던 오류를 수정할 수 있을 것으로 기대함



감사합니다

Q&A