XLM

CROSS-LINGUAL LANGUAGE MODEL PRETRAINING

Moon Sungwon

2020. 07. 02

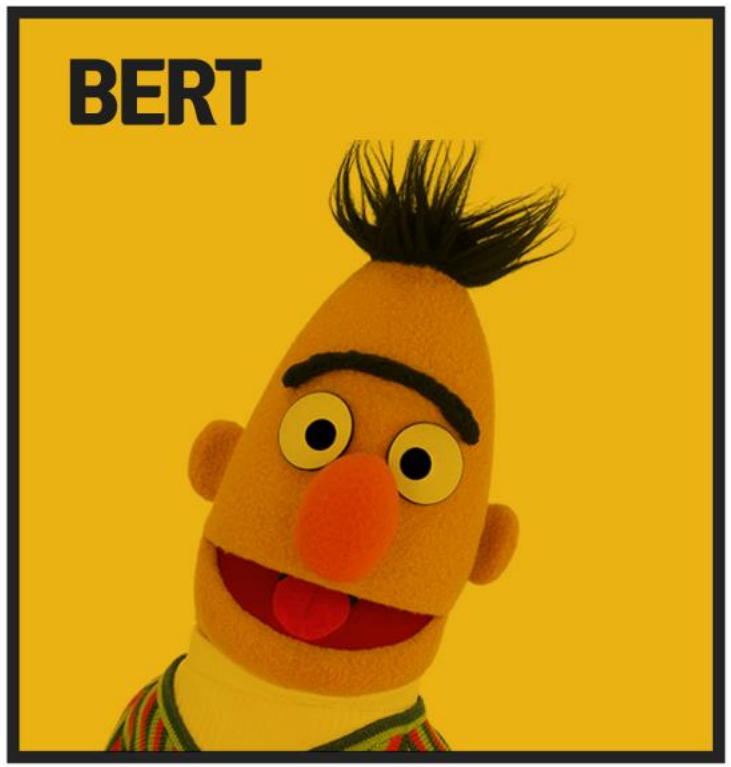
INTRO

최근엔
마스크
토큰
을
사용하는
언어 모델
이
유행입니다.

→

최근엔
마스크
[MASK]
을
[MASK]
언어 모델
이
유행입니다.

→

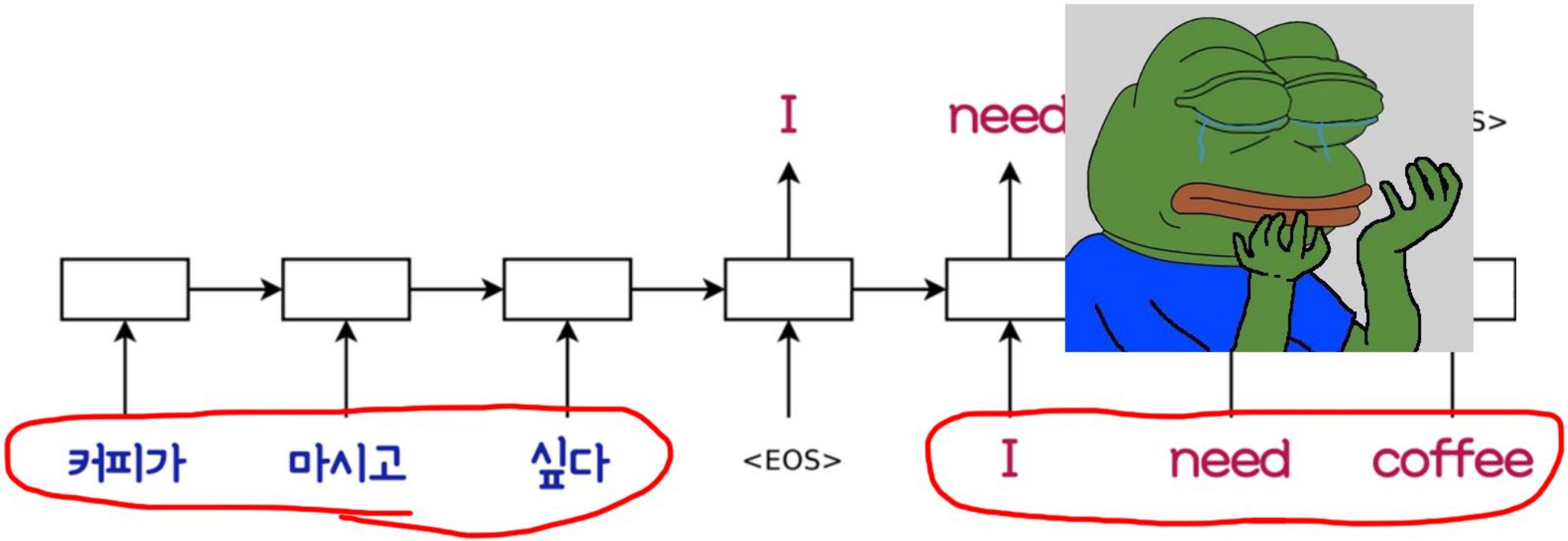


→

최근엔
마스크
토큰
을
활용하는
언어 모델
이
유행입니다.

Masked Language Model (MLM)

INTRO



요것만 있어도 MLM이 되는데

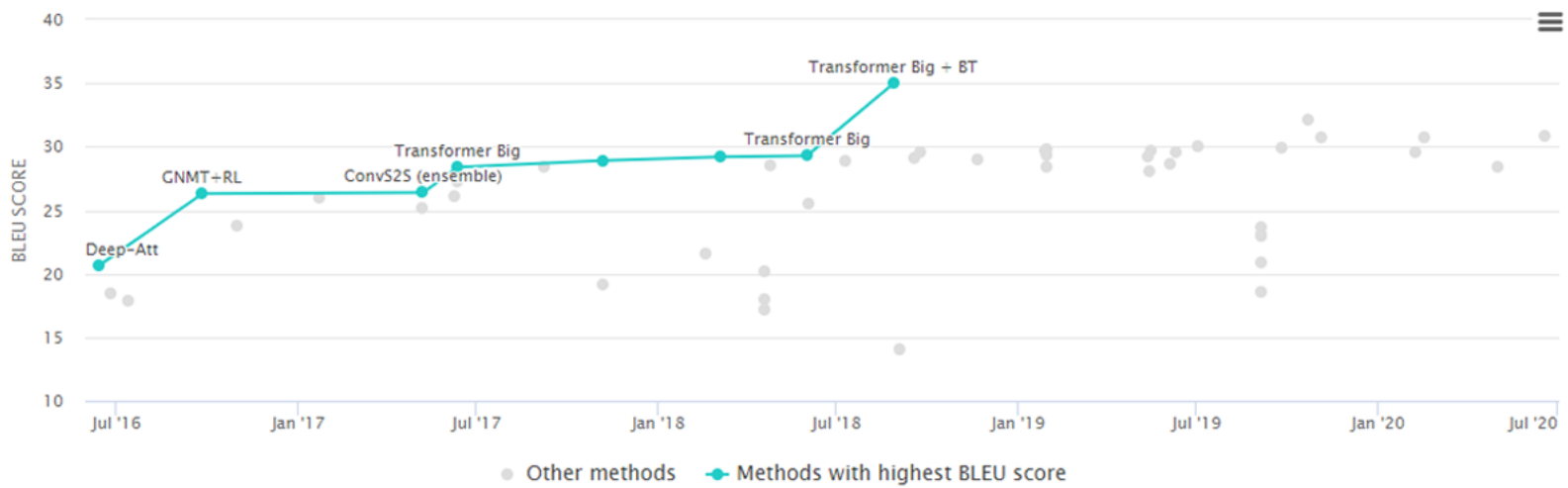
이거까지 모아야되는 기계 번역..?

기계 번역과 MLM...?

INTRO

Browse > Natural Language Processing > Machine Translation > WMT2014 English-German dataset

Machine Translation on WMT2014 English-German



View BLEU score All methods Edit

RANK	METHOD	BLEU SCORE	SACREBLEU	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	Transformer Big + BT	35.0	33.8	✓	Understanding Back-Translation at Scale			2018
2	T5-11B	32.1		✓	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer			2019
3	MAT	30.8		×	Multi-branch Attentive Transformer			2020
4	BERT-fused NMT	30.75		×	Incorporating BERT into Neural Machine Translation			2020

심지어 Transformer는 2017년도 논문

INTRO

SQuAD2.0

The Stanford Question Answering Dataset

Dataset (SQuAD) is a
t, consisting of questions
et of Wikipedia articles,
estion is a segment of text,
g reading passage, or the
le.

000 questions in SQuAD1.1
e questions written
to look similar to
SQuAD2.0, systems must
en possible, but also
supported by the paragraph

d model predictions

urkar & Jia et al. '18)

on of the SQuAD dataset,
nswer pairs on 500+

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Apr 06, 2020	SA-Net on Albert (ensemble) *QIANXIN*	**90.724**	**93.011**
2 May 05, 2020	SA-Net-V2 (ensemble) *QIANXIN*	90.679	92.948
2 Apr 05, 2020	Retro-Reader (ensemble) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694v2	90.578	92.978
3 May 04, 2020	ELECTRA+ALBERT+EntitySpanFocus (ensemble) *SRCB_DML*	90.442	92.839
4 Jun 21, 2020	ELECTRA+ALBERT+EntitySpanFocus (ensemble) *SRCB_DML*	90.420	92.799

So XLM!

“

MLM을 다중 언어로 확장해보자!

”

Causal Language Modeling (CLM)

= 인과 언어 모델

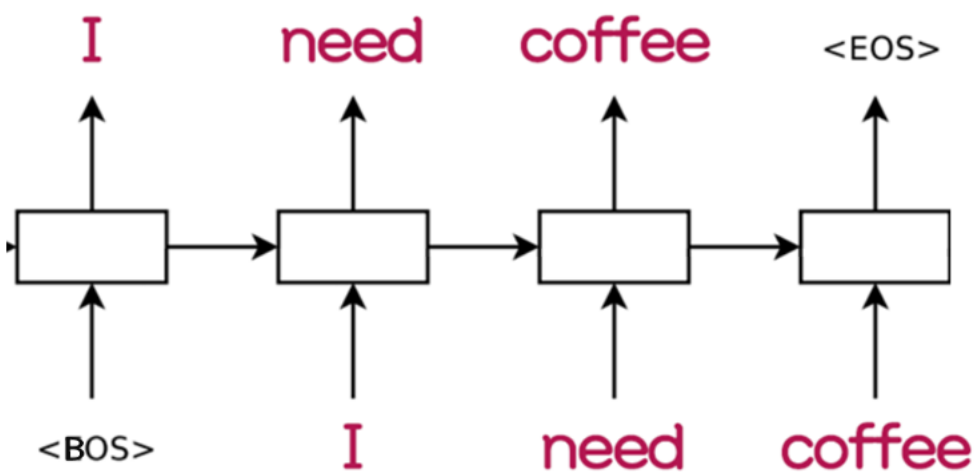
갈색 줄무늬를 가진 작고 털이 보송한 아기 [ ]

Because of...

고양이

강아지

호랑이



Masked Language Modeling (MLM)

최근엔 마스크 토큰 을 사용하는 언어 모델 이 유행입니다.

→

최근엔 마스크 [MASK] 을 [MASK] 언어 모델 이 유행입니다.

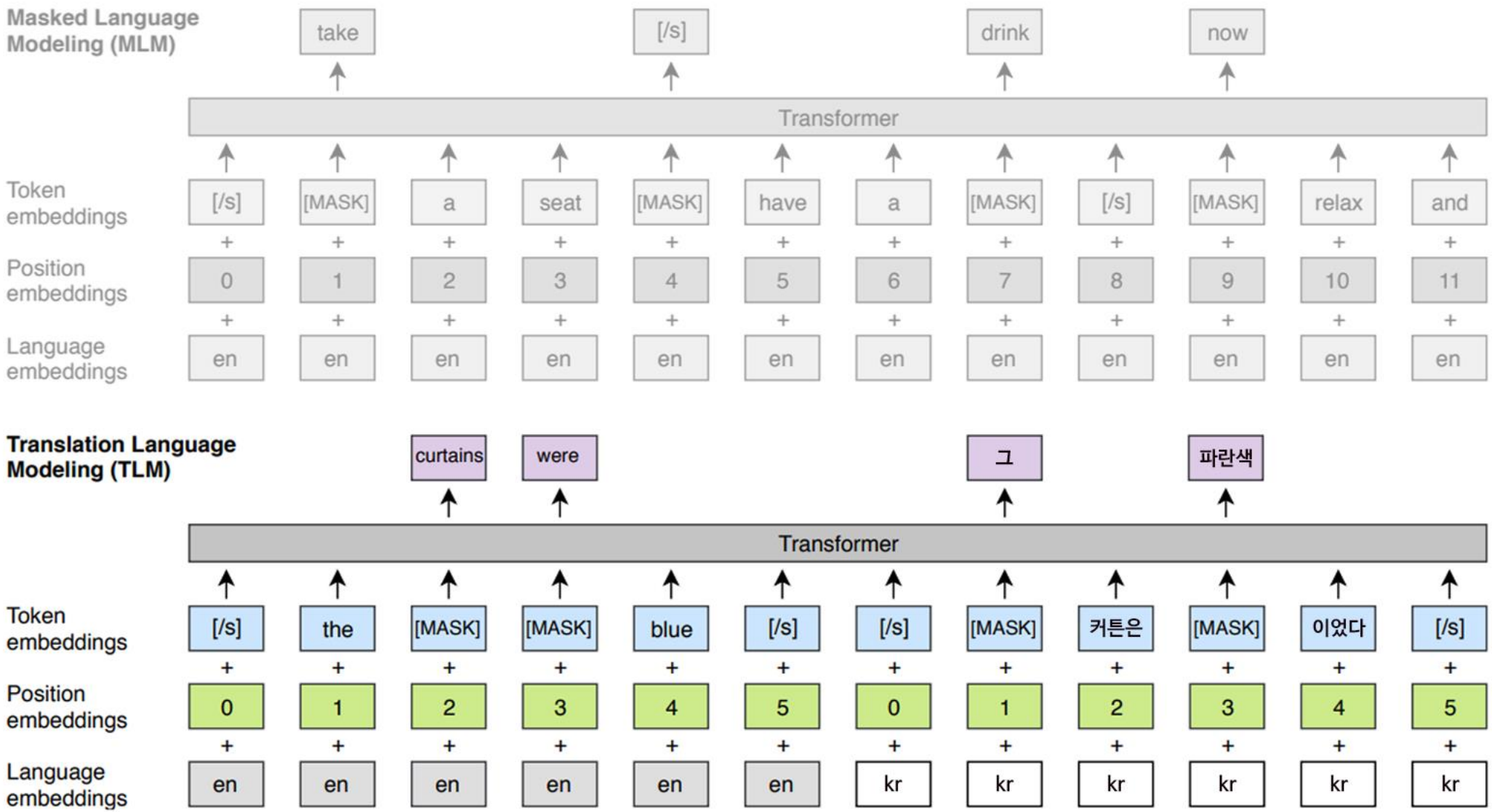
→

→

최근엔 마스크 토큰 을 활용하는 언어 모델 이 유행입니다.

CLM & MLM = Unsupervised !

Translation Language Modeling (TLM)



XNLI: The Cross-Lingual NLI Corpus

Language	Premise / Hypothesis	Genre	Label
English	You don't have to stay there. You can leave.	Face-To-Face	Entailment
French	La figure 4 montre la courbe d'offre des services de partage de travaux. Les services de partage de travaux ont une offre variable.	Government	Entailment
Spanish	Y se estremeció con el recuerdo. El pensamiento sobre el acontecimiento hizo su estremecimiento.	Fiction	Entailment
German	Während der Depression war es die ärmste Gegend, kurz vor dem Hungertod. Die Weltwirtschaftskrise dauerte mehr als zehn Jahre an.	Travel	Neutral
Swahili	Ni silaha ya plastiki ya moja kwa moja inayopiga risasi. Inadumu zaidi kuliko silaha ya chuma.	Telephone	Neutral
Russian	И мы занимаемся этим уже на протяжении 85 лет. Мы только начали этим заниматься.	Letters	Contradiction
Chinese	让我告诉你，美国人最终如何看待你作为独立顾问的表现。 美国人完全不知道您是独立律师。	Slate	Contradiction

Total 15 Languages

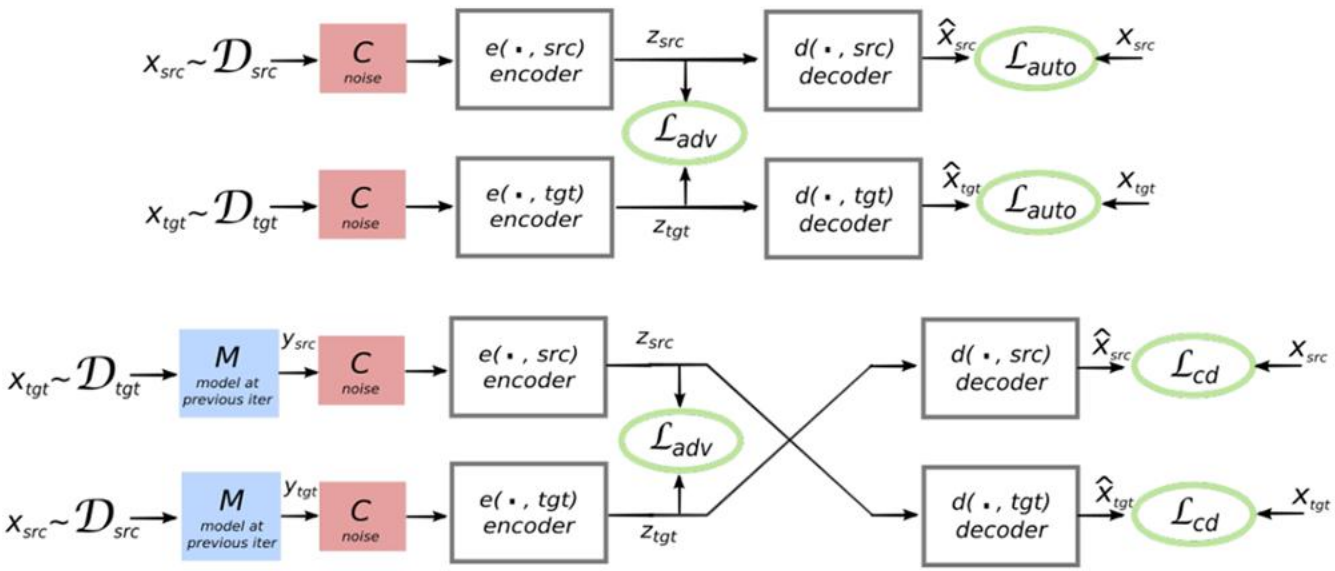
다국어 병렬 데이터를
TLM으로 학습한 후,
영어 NLI로 Fine Tuning

= Zero-shot

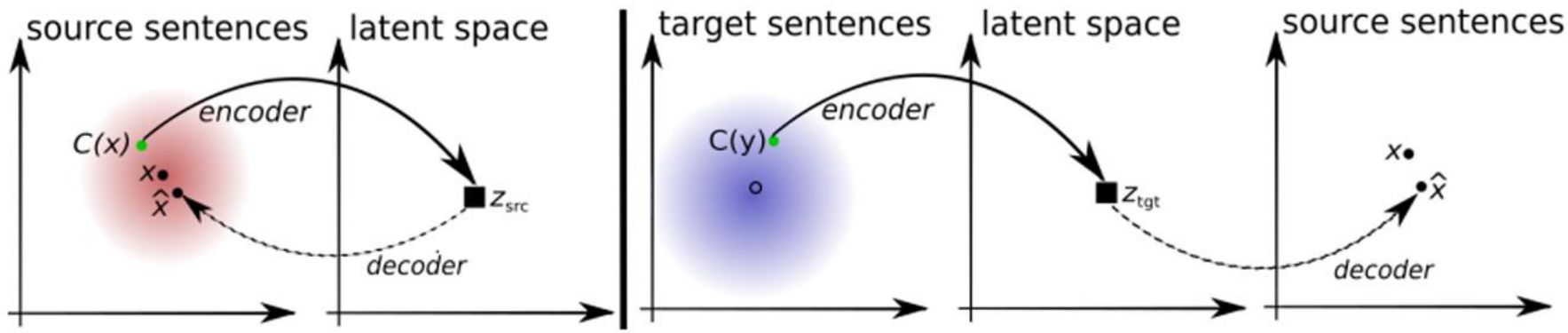
XNLI: The Cross-Lingual NLI Corpus

	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	Δ
Machine translation baselines (TRANSLATE-TRAIN)																
Devlin et al. (2018)	81.9	-	77.8	75.9	-	-	-	-	70.7	-	-	76.6	-	-	61.6	-
XLM (MLM+TLM)	85.0	80.2	80.8	80.3	78.1	79.3	78.1	74.7	76.5	76.6	75.5	78.6	72.3	70.9	63.2	76.7
Machine translation baselines (TRANSLATE-TEST)																
Devlin et al. (2018)	81.4	-	74.9	74.4	-	-	-	-	70.4	-	-	70.1	-	-	62.1	-
XLM (MLM+TLM)	85.0	79.0	79.5	78.1	77.8	77.6	75.5	73.7	73.7	70.8	70.4	73.6	69.0	64.7	65.1	74.2
Evaluation of cross-lingual sentence encoders																
Conneau et al. (2018b)	73.7	67.7	68.7	67.7	68.9	67.9	65.4	64.2	64.8	66.4	64.1	65.8	64.1	55.7	58.4	65.6
Devlin et al. (2018)	81.4	-	74.3	70.5	-	-	-	-	62.1	-	-	63.8	-	-	58.3	-
Artetxe and Schwenk (2018)	73.9	71.9	72.9	72.6	73.1	74.2	71.5	69.7	71.4	72.0	69.2	71.4	65.5	62.2	61.0	70.2
~~XLM (MLM)~~	~~83.2~~	~~76.5~~	~~76.3~~	~~74.2~~	~~73.1~~	~~74.0~~	~~73.1~~	~~67.8~~	~~68.5~~	~~71.2~~	~~69.2~~	~~71.9~~	~~65.7~~	~~64.6~~	~~63.4~~	~~71.5~~
XLM (MLM+TLM)	**85.0**	**78.7**	**78.9**	**77.8**	**76.6**	**77.4**	**75.3**	**72.5**	**73.1**	**76.1**	**73.2**	**76.5**	**69.6**	**68.4**	**67.3**	**75.1**

Unsupervised Machine Translation

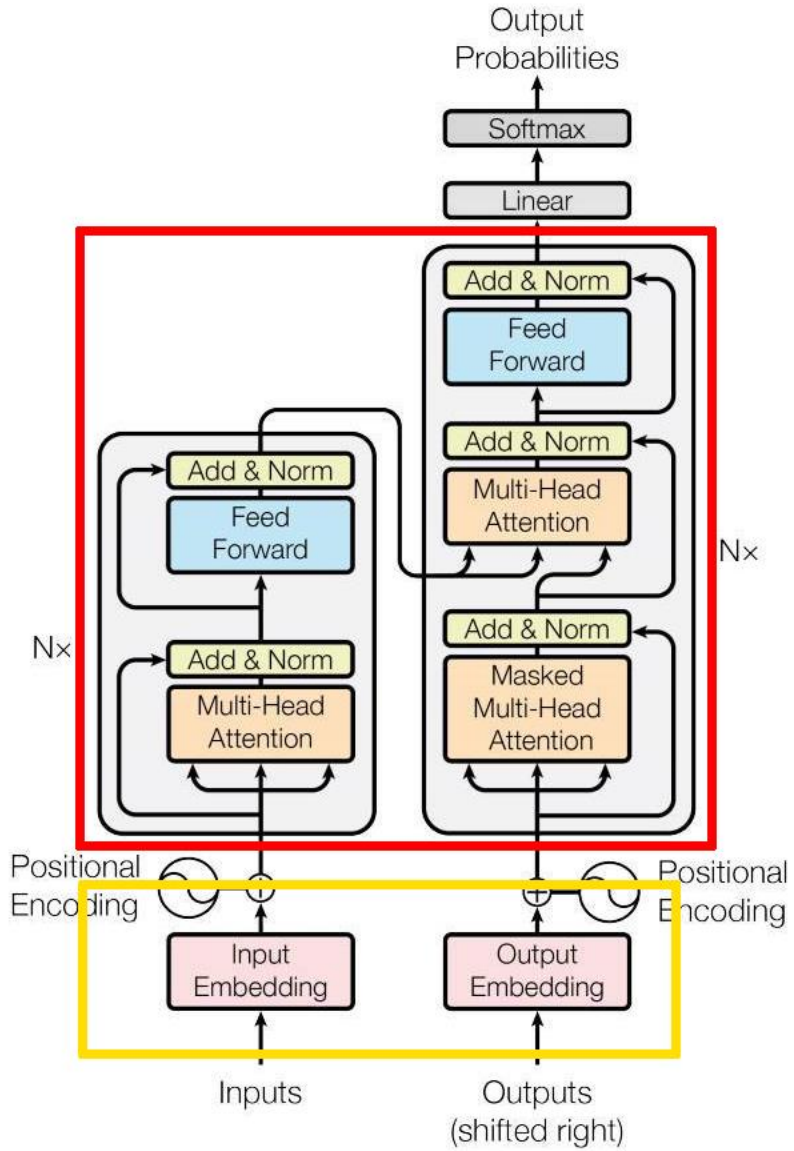


⟨ UNSUPERVISED MACHINE TRANSLATION USING MONOLINGUAL CORPORA ONLY ⟩



Unsupervised Machine Translation

		en-fr	fr-en	en-de	de-en	en-ro	ro-en
Previous state-of-the-art - Lample et al. (2018b)							
NMT		25.1	24.2	17.2	21.0	21.2	19.4
PBSMT		28.1	27.2	17.8	22.7	21.3	23.0
PBSMT + NMT		27.6	27.7	20.2	25.2	25.1	23.9
Our results for different encoder and decoder initializations							
FastText		29.4	29.4	21.3	27.3	27.5	26.6
-	-	13.0	15.8	6.7	15.3	18.9	18.3
-	CLM	25.3	26.4	19.2	26.0	25.7	24.6
-	MLM	29.2	29.1	21.6	28.6	28.2	27.3
CLM	-	28.7	28.2	24.4	30.3	29.2	28.0
CLM	CLM	30.4	30.0	22.7	30.5	29.0	27.8
CLM	MLM	32.3	31.6	24.3	32.5	31.6	29.8
MLM	-	31.6	32.1	**27.0**	33.2	31.8	30.5
MLM	CLM	**33.4**	32.3	24.9	32.9	31.7	30.4
MLM	MLM	**33.4**	**33.3**	26.4	**34.3**	**33.3**	**31.8**

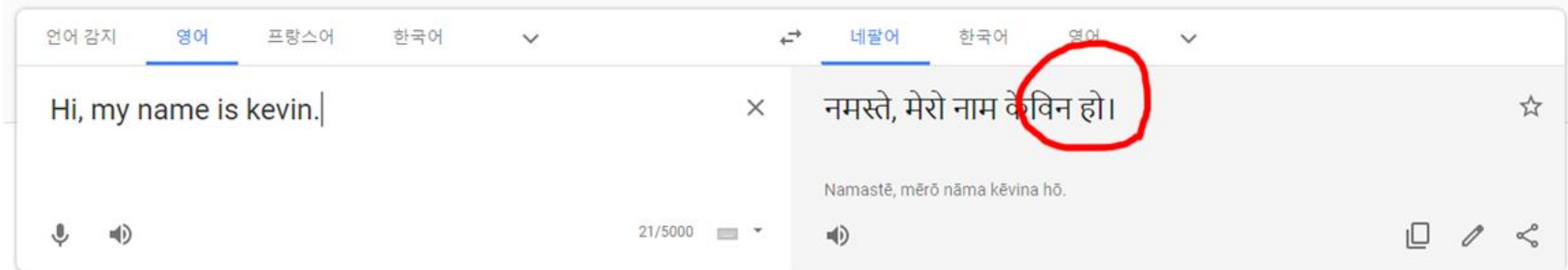


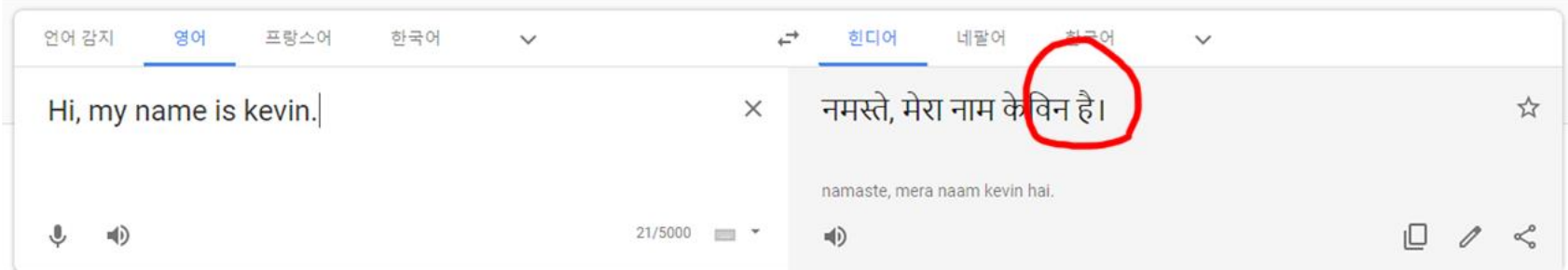
Supervised Machine Translation

Pretraining	-	CLM	MLM
Sennrich et al. (2016)	33.9	-	-
ro $\rightarrow$ en	28.4	31.5	35.3
ro $\leftrightarrow$ en	28.5	31.5	35.6
ro $\leftrightarrow$ en + BT	34.4	37.0	**38.5**

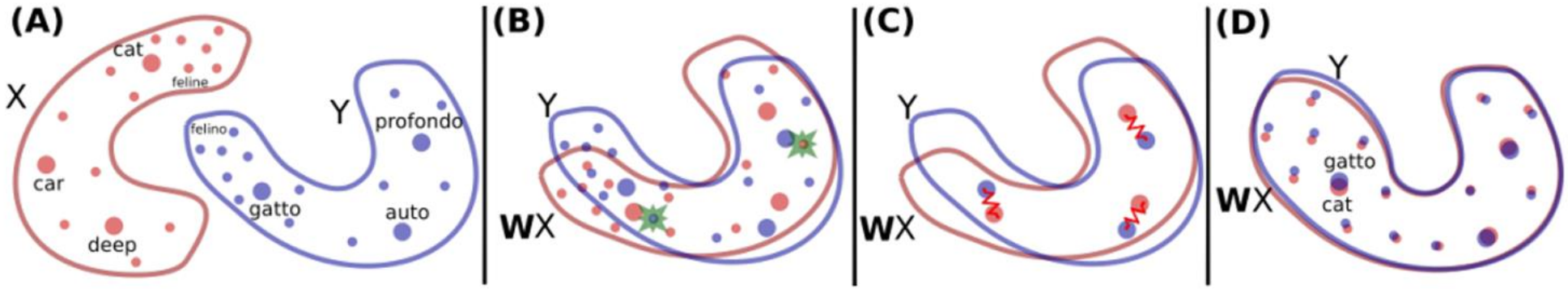
Low-resource Language Modeling

Training languages	Nepali perplexity
Nepali	157.2
Nepali + English	140.1
Nepali + Hindi	115.6
Nepali + English + Hindi	**109.3**





Unsupervised Cross-lingual Word Embeddings



〈 WORD TRANSLATION WITHOUT PARALLEL DATA 〉

	Cosine sim.	L2 dist.	SemEval'17
MUSE	0.38	5.13	0.65
Concat	0.36	4.89	0.52
XLM	**0.55**	**2.64**	**0.69**

감사합니다