


찰떡같이 알아듣는 번역기를 만들어 보자

- 한-영 관용구 기계번역 연구 -

Judy Choi

Judy Choi (최민주)

- Careers

- 2012 ~ 2019 악성코드 분석가
- 2018 프랑스 워킹홀리데이
- 2020 ~ 강원대학교 석사과정
 - Intelligence Software Lab
 - 주요 연구분야 :
NLP(자연어처리) – 기계번역 

- SNS

- <https://www.facebook.com/minjoo.choi.562/>
- <https://github.com/Judy-Choi>



Neural Machine Translation & 관용구

- NMT 기반 기계번역은 훌륭한 번역 성능을 보이나, 종종 오역이 발생함
 - 예) 관용구 번역
- 관용-구(慣用句)
 - 두 개 이상의 단어로 이루어져 있으면서 그 단어들의 의미만으로는 전체의 의미를 알 수 없는, 특수한 의미를 나타내는 어구(語句)
 - 예) '눈이 높다' = '눈' + '높다'"
 - 뜻 : '좋은 것을 찾다', '안목이 높다'

Quiz : 다음 문장을 영어로 번역해 봅시다 😊

- 나는 남자 보는 눈이 높아요
- 학부모만 발을 동동 구르는 상황이다

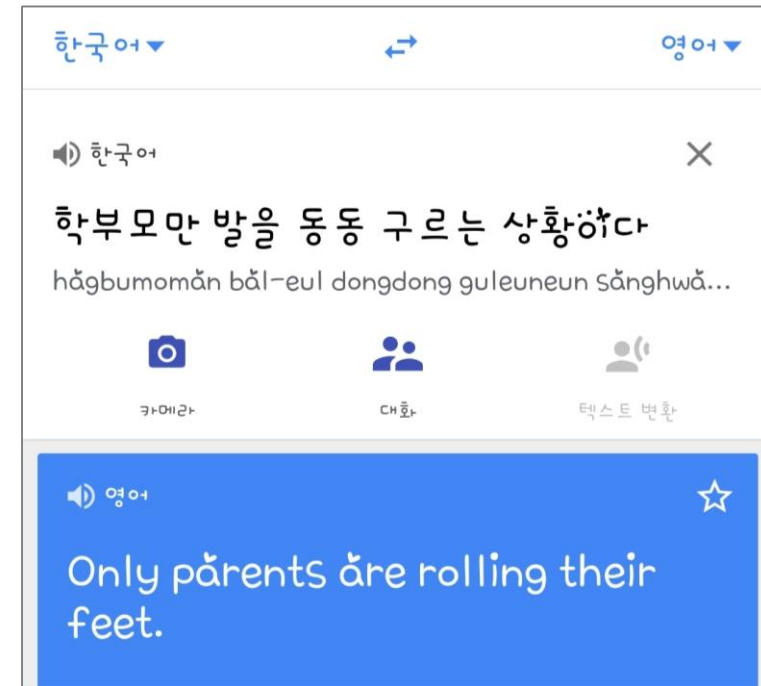
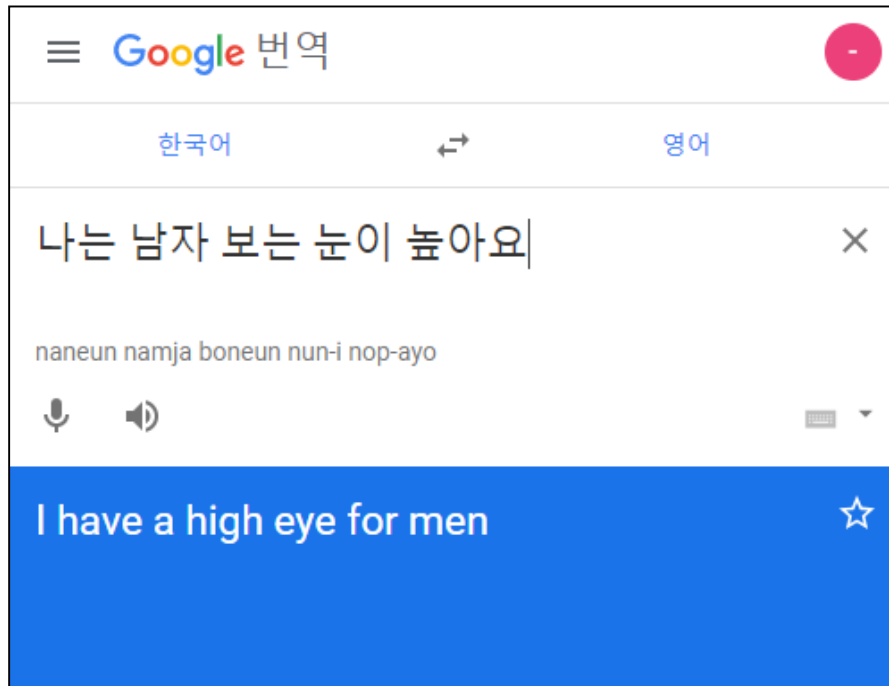
Quiz : 다음 문장을 영어로 번역해 봅시다 😊

- 나는 남자 보는 눈이 높아요
 - I have high standards for man
- 학부모만 발을 동동 구르는 상황이다
 - Only the parents are anxious

Quiz : 다음 문장을 영어로 번역해 봅시다 😊



- 나는 남자 보는 눈이 높아요
 - I have high standards for man
- 학부모만 발을 동동 구르는 상황이다
 - Only the parents are anxious



관용구 기계번역의 문제점

- 기계번역은 함축적인 의미를 지닌 관용구를 정확하게 번역할 수 없음
- 기계번역 모델이 관용구를 효과적으로 학습하고 번역 결과를 평가할 수 있도록 관용구 번역에 특화된 연구가 필요함.

- 관용구 기계번역 연구내용

1. 데이터셋 구축 (KISS)
2. 평가방법 제안 및 평가지표 구축 (Black List)
3. 번역 모델 학습방법 제안 (Special Token Tagging)

관용구 기계번역을 위한 한-영 데이터셋 구축 및 평가 방법

최민주^o

국민대학교 컴퓨터공학과
mjchoi0831@gmail.com

KCC 2020

한-영 관용구 기계번역을 위한 NMT 학습 방법

최민주^o, 이창기

강원대학교 컴퓨터공학과
mjchoi0831@gmail.com, leeck@kangwon.ac.kr

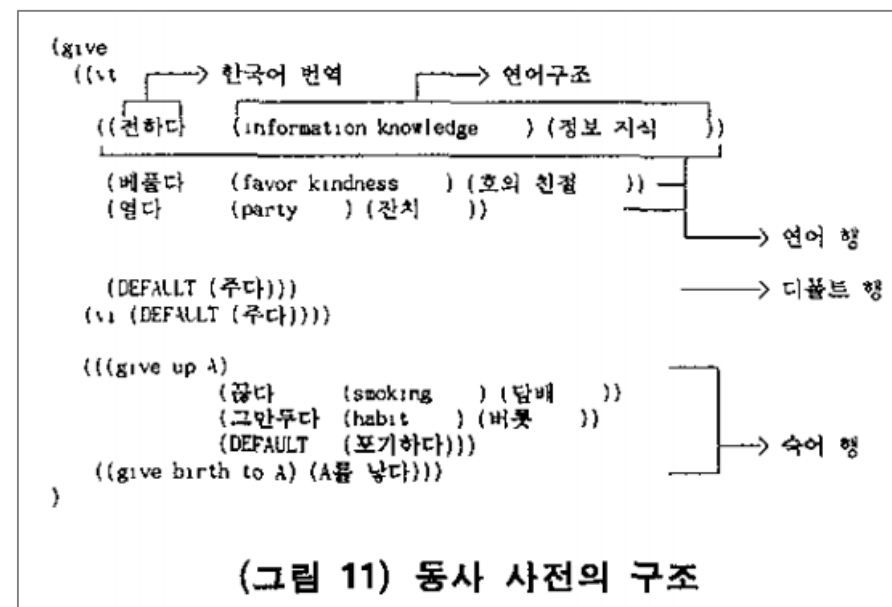
HCLT 2020

라떼는 말이야 (Latte is Horse...)



http://philamilal.org/index.php?mid=board7&document_srl=10902

- NMT 이전에는...
 - 구-대-구 방식으로 관용구의 구조 정보를 이용하여 번역
- 사전 기반 방식
 - 문장 내에 관용구 사전에 존재하는 관용구가 포함되어 있으면 관용구로 번역
 - 단점
 - 사전 구축에 많은 시간과 노력 필요
 - 문장 단위로 학습하고 번역하는 NMT에는 적합하지 않음



[1] 이호석, 김영택. 영어-한국어 기계번역을 위한 언어와 숙어 트랜스퍼 사전.
(구) 정보과학회논문지. 20.7: 976-987. 1993.

NMT에서 관용구를 학습/평가하려면 뭐가 필요할까?

1. 관용구가 포함된 다량의 문장 데이터셋

- 한-영 번역 쌍 말뭉치로부터 관용구 포함 번역 쌍 추출
⇒ 관용구 학습 데이터셋 구축 (KISS)

2. 관용구 번역에 특화된 번역 평가 지표

- 관용구 학습 데이터셋을 이용
⇒ 블랙리스트(평가 지표) 생성

3. 관용구를 효과적으로 학습시키는 방법

- 데이터셋 내의 관용구 앞/뒤에 특수한 토큰을 붙여 학습
⇒ 토큰 부착 전/후 번역 품질을 비교하여 학습 효과 검증

1. KISS : 관용구 포함 한-영 번역 쌍 데이터셋 구축

표준국어대사전
수록 관용어

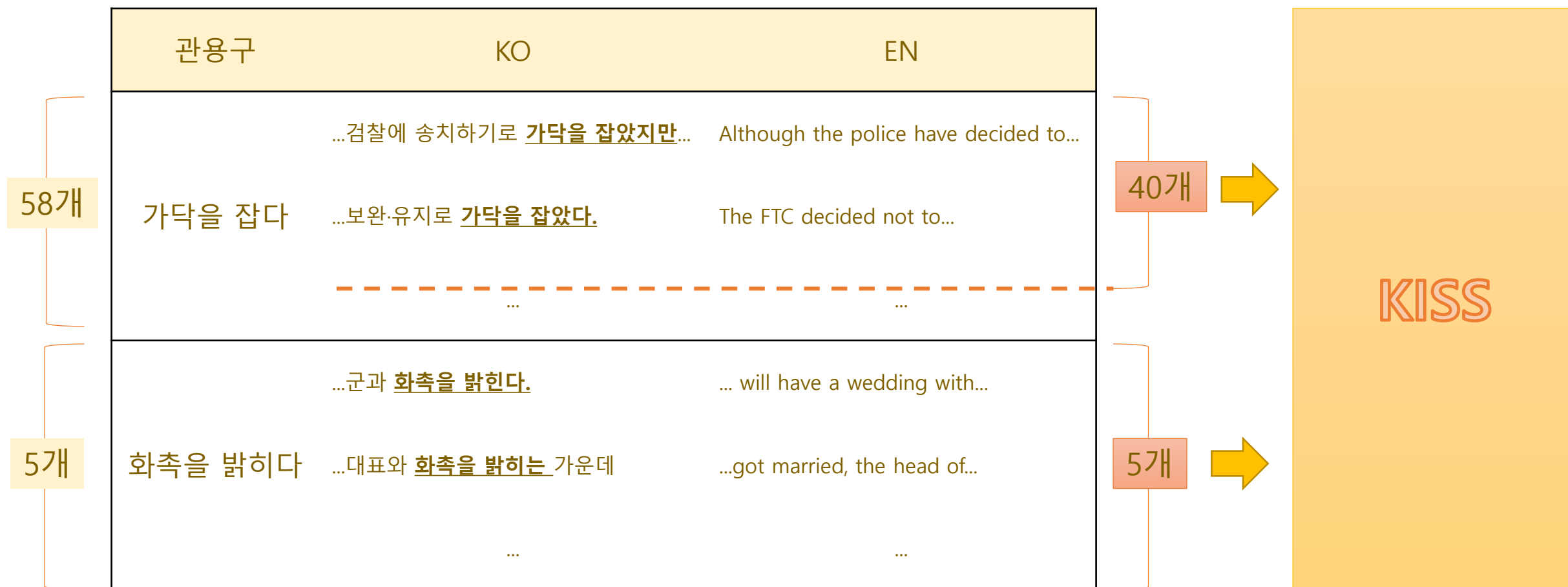
...
눈이 높다
마침표를 찍다
총대를 메다
화촉을 밝히다
...



AI Hub 한국어-영어 번역(병렬) 말뭉치

KO	EN
나는 여자 보는 <u>눈이 높아요</u>	I have high standards for woman.
아버지의 명예회복을 위한 김지훈의 기나긴 여정이 마침내 <u>마침표를 찍었다</u> .	Kim Ji-hoon's long journey ... has finally come to an end.
기획재정부가 혁신성장 관련 규제 완화에 <u>총대를 멘다</u> .	The Ministry of Economy and Finance takes charge of ...
중국 출신 할리우드 스타 장쯔이 ...와 내년 <u>화촉을 밝힌다</u> .	Chinese Hollywood star Zhang Ziyi (28) will marry ...
...	...

1. KISS : 관용구 포함 한-영 번역 쌍 데이터셋 구축



1. KISS : 관용구 포함 한-영 번역 쌍 데이터셋 구축

- 관용구 수집
 - 표준국어대사전 온라인 사이트 → 3,887개 관용구 목록 다운로드
- 한-영 문장 번역 쌍 추출
 - AI Hub 한국어-영어 번역(병렬) 말뭉치
 - 430개 한국어 관용구 포함 18,808개 번역 쌍 추출
 - 동일한 관용구 포함 문장 4~40개로 제한
 - 420개 한국어 관용구 포함 7500개 번역 쌍 추출 => KISS 구축
- KISS : Korean-english Idioms in Sentences dataSet



github.com/Judy-Choi/KISS

KISS : Korean-english Idioms in Sentences dataSet

관용구	한국어 원문	영어 번역 쌍
눈이 높다	나는 여자 보는 <u>눈이 높아요.</u>	I have <u>high standards</u> for woman.
마침표를 찍다	아버지의 명예회복을 위한 김지훈의 기나긴 여정이 마침내 <u>마침표를 찍었다.</u>	Kim Ji-hoon's long journey to restore his father's honor has finally <u>come to an end.</u>
총대를 메다	기획재정부가 혁신성장 관련 규제 완화에 <u>총대를 멘다.</u>	The Ministry of Economy and Finance <u>takes charge of</u> the deregulation related to innovative growth.
화촉을 밝히다	중국 출신 할리우드 스타 장쯔이(章子怡, 28)가 현 남자친구인 미국의 억만장자 비비 네보(42) 와 내년 <u>화촉을 밝힌다.</u>	Chinese Hollywood star Zhang Ziyi (28) will <u>marry</u> her current boyfriend, U.S. billionaire Aviv Nevo (42), next year.

Problem...

Correct

관용구	눈이 높다
한국어 원문	나는 여자 보는 <u>눈이 높아요</u> .
영어 번역 쌍	I have <u>high standards</u> for woman.
관용구	마침표를 찍다
한국어 원문	아버지의 명예회복을 위한 김지훈의 기나긴 여정이 마침내 <u>마침표를 찍었다</u> .
영어 번역 쌍	Kim Ji-hoon's long journey to restore his father's honor has finally <u>come to an end</u> .

Error



관용구	운을 떴다
한국어 원문	정부가 노인 연령 기준을 높이는 방안에 대해 <u>운을 뒀다</u> .
영어 번역 쌍	The government has been <u>lucky</u> about ways to raise the criteria of age for senior citizens.
관용구	유명을 달리하다
한국어 원문	오키나와 현지사인 오나가 다케시가 지난 8일 췌장암으로 <u>유명을 달리했다</u> .
영어 번역 쌍	Takeshi Onaga, an incumbent Governor of Okinawa, became <u>famous</u> for pancreatic cancer on the 8th.

Problem...

- 구축한 데이터셋에 다량의 오역 쌍 존재
 - 데이터셋에 포함된 다량의 오역을 제거할 수 있는 방법 필요
- 관용구 번역 시 다량의 오역 발생
 - 기계번역 오류를 탐지할 수 있는 품질 평가 지표 필요

Problem...

- 구축한 데이터셋에 다량의 오역 쌍 존재
 - 데이터셋에 포함된 다량의 오역을 제거할 수 있는 방법 필요
- 관용구 번역 시 다량의 오역 발생
 - 기계번역 오류를 탐지할 수 있는 품질 평가 지표 필요

1 shot 2 kill solution

Blacklist Method

2. 블랙리스트를 이용한 평가 방법

- 직역으로 인한 번역 오류 탐지
- 단어-대-단어 구조의 관용구를 번역한 결과로부터 오역 여부를 판별
- 원리
 - 관용구를 한-영 번역한 결과에 블랙리스트 단어가 1개 이상 포함되어 있으면 오역으로 간주

관용구	블랙리스트
꼬집어 말하다	nip pinch twitch
눈 높다	eye
운을 떴다	lucky
유명을 달리하다	famous

2. 블랙리스트를 이용한 평가 방법

Correct

관용구	눈이 높다
블랙리스트	eye
한국어 원문	나는 여자 보는 <u>눈이 높아요</u> .
영어 번역 쌍	I have <u>high standards</u> for woman.

Error

관용구	눈이 높다
블랙리스트	eye
한국어 원문	나는 여자 보는 <u>눈이 높아요</u> .
영어 번역 쌍	I have <u>high</u> eye for woman.

블랙리스트를 이용한 오역 탐지

2. 블랙리스트를 이용한 평가 방법

- 블랙리스트 구축
- 단어의 의미를 이용한 블랙리스트 단어 추출

관용구	<u>꼬집어</u> 말하다
블랙리스트	nip pinch twitch
관용구	<u>눈</u> 높다
블랙리스트	eye

- 오역결과를 이용한 블랙리스트 단어 추출

관용구	운을 떴다
한국어 원문	정부가 노인 연령 기준을 높이는 방안에 대해 <u>운을 땀다</u> .
영어 번역 쌍	The government has been <u>lucky</u> about ways to raise the criteria of age for senior citizens.
블랙리스트	lucky
관용구	유명을 달리하다
한국어 원문	오키나와 현지사인 오나가 다케시가 지난 8일 췌장암으로 <u>유명을 달리했다</u> .
영어 번역 쌍	Takeshi Onaga, an incumbent Governor of Okinawa, became <u>famous</u> for pancreatic cancer on the 8th.
블랙리스트	famous

2. 블랙리스트를 이용한 평가 방법

- 블랙리스트 구축
 - 420개 관용구 중 275개 관용구에 대한 블랙리스트 구축
 - 오역이 거의 없는 관용구 제외
 - 예) '그건 그렇고' → 'By the way' 로 대부분 정확하게 번역됨
 - 영어로 직역되는 관용구 제외
 - 예) '(사람의 마음을) 가지고 놀다' → 'play with' 로 직역되므로 블랙리스트 구축 불가
- 블랙리스트를 이용한 오역 없는 번역 쌍 추출
 - KISS 로부터 275개 관용구 포함 3,461 개 번역 쌍 추출

2. 블랙리스트를 이용한 평가 방법

- 오역 제외한 번역 쌍을 이용한 기계번역 서비스 품질 평가

- 관용구 275개

- 번역 쌍 3,461개



	Google 번역	Naver Papago	Kakao i 번역
블랙리스트 탐지	1,179	1,093	1,049
블랙리스트 미탐지	2,282	2,368	2,412
* 번역 정확도 (%)	65.93	68.41	69.69
평균 BLEU 점수	30.04	13.47	33.83

* 번역 정확도 : 전체 문장 중 블랙리스트 미탐지된 문장 수

2. 블랙리스트를 이용한 평가 방법

- 번역 품질과 BLEU 점수가 비례하지 않는 예시

관용구	가닥을 잡다
한국어 원문	당초 강 위원장은 국정감사에서 이 문제가 불거지자 “감사원 감사를 받겠다” 며 버텼지만 즉각 사퇴로 <u>가닥을 잡았다</u> .
영어 번역 쌍	At the beginning, Representative Kang said, “I will be audited by the auditor,” when the matter was raised in the state audit.
Naver Papago 번역 결과	Initially, Kang endured the issue during a parliamentary audit, saying he would undergo an audit by the Board of Audit and Inspection, but he immediately <u>decided to</u> step down.
BLEU 점수	4.7930e-76

3. NMT 모델 학습 방법

- NMT 모델이 관용구를 더욱 정확하게 번역하도록 학습시키려면?
 - 문장 내 관용구의 위치를 표기해 모델에 학습
 - 관용구 시작 위치에 <idm> token 부착
 - Unigram Precision, Word-level Accuracy 증가

We perform a German→English experiment by providing the model with additional input features. The additional features indicate whether a source sentence contains an idiom and are implemented as a special extra token <idm> that is prepended to each source sentence containing an idiom. This is a simple approach that can be applied to any sequence-to-sequence architecture.

Model	WMT test sets 2008-2016		Idiom test set		
	BLEU	BLEU	Unigram Precision	Word-level Accuracy	
PBMT Baseline	20.2	19.7	57.7	71.6	
NMT Baseline	26.9	24.8	53.2	67.8	
NMT <idm> token on source	25.2	22.5	64.1	73.2	

Table 5: Translation performance on German idiom translation test set. *Word-level Idiom Accuracy* and *Unigram Precision* are computed only on the idiom phrase and its corresponding translation in the sentence.

3. NMT 모델 학습 방법

- NMT 모델이 관용구를 더욱 정확하게 번역하도록 학습시키려면?
 - 문장 내 관용구의 위치를 표기해 모델에 학습
 - 관용구 시작 위치에 `<start_fig>`,
관용구 끝 위치에 `<end_fig>` token 부착
 - BLEU score 증가

We use BLEU to measure the quality of translations. From the result presented in Table 18, we can see that the baseline achieves a BLEU score of 17.2, which is lower than the performance of previously reported models on the standard test set (WMT 2008-2016) [Sennrich et al., 2016]. This suggests that it is much harder to translate sentences containing idioms. Further, simply appending the `< fig >` token to indicate the usage of idiom gets a BLEU score of 16.6, which is slightly lower than the baseline model; **using the `< start_fig >` and `< end_fig >` tokens** outperforms the baseline by 2.3 BLEU. This suggests that the usage information and the position information of the idiom can help boost the performance of neural machine translation models on idioms.

Table 18: The performance on English-to-German idiom translation test set.

Model	BLEU
NMT Baseline	17.2
with <code>< fig ></code> token	16.6
with <code>< start_fig ></code> <code>< end_fig ></code> token	19.5

3. NMT 모델 학습 방법

- 3.1 데이터셋 구축

- AI Hub 한국어-영어 번역(병렬) 말뭉치 AI 데이터 이용 (160만개)
- 이 중 420개 관용구 포함 문장 쌍 7,500개로 KISS 구축
- KISS 에 블랙리스트 적용 → 275개 관용구 포함 3,461개 문장 쌍 추출

✓ 수작업 검증 ⇒ 최종적으로 총 3,377 개 관용구 포함 문장 쌍 추출

3. NMT 모델 학습 방법

- 3.2 데이터셋 분할

- 관용구가 포함된 문장 / 관용구가 포함되지 않은 일반 문장

	관용구 포함 문장 쌍 (KISS – 오역 제거)	관용구 미포함 문장 쌍 (AI Hub – KISS)
개발셋	500	5,000
학습셋	1,877	1,577,426
평가셋	1,000	10,000
합계	3,377	1,592,426

3. NMT 모델 학습 방법

- 3.3 실험 조건별 학습 데이터셋 구축
 - 1. 관용구가 포함되지 않은 문장으로만 이루어진 학습셋
 - 2. 관용구를 포함하는 문장과 관용구가 포함되지 않은 문장으로 이루어진 학습셋
 - 발을 동동 구르는 상황이다.
 - 3. 2에서 문장 내 관용구의 시작 위치를 <idm> 으로 표기한 학습셋
 - <idm>발을 동동 구르는 상황이다.
 - 4. 2에서 문장 내 관용구의 시작과 끝 위치를 각각 <idm>, </idm> 으로 표기한 학습셋
 - <idm>발을 동동 구르는</idm> 상황이다.

3. NMT 모델 학습 방법

- 4가지 데이터셋을 총 4가지 NMT 모델에 학습
- NMT 모델별 Hyperparameter

	2 Layer LSTM (OpenNMT Default)	4 Layer LSTM	4 Layer / 6 Layer Transformer
Encoder hidden state size	500	1,000	512
Batch size	64	100	2,048
Dropout	0.3	0.1	0.1
Train steps	100,000		

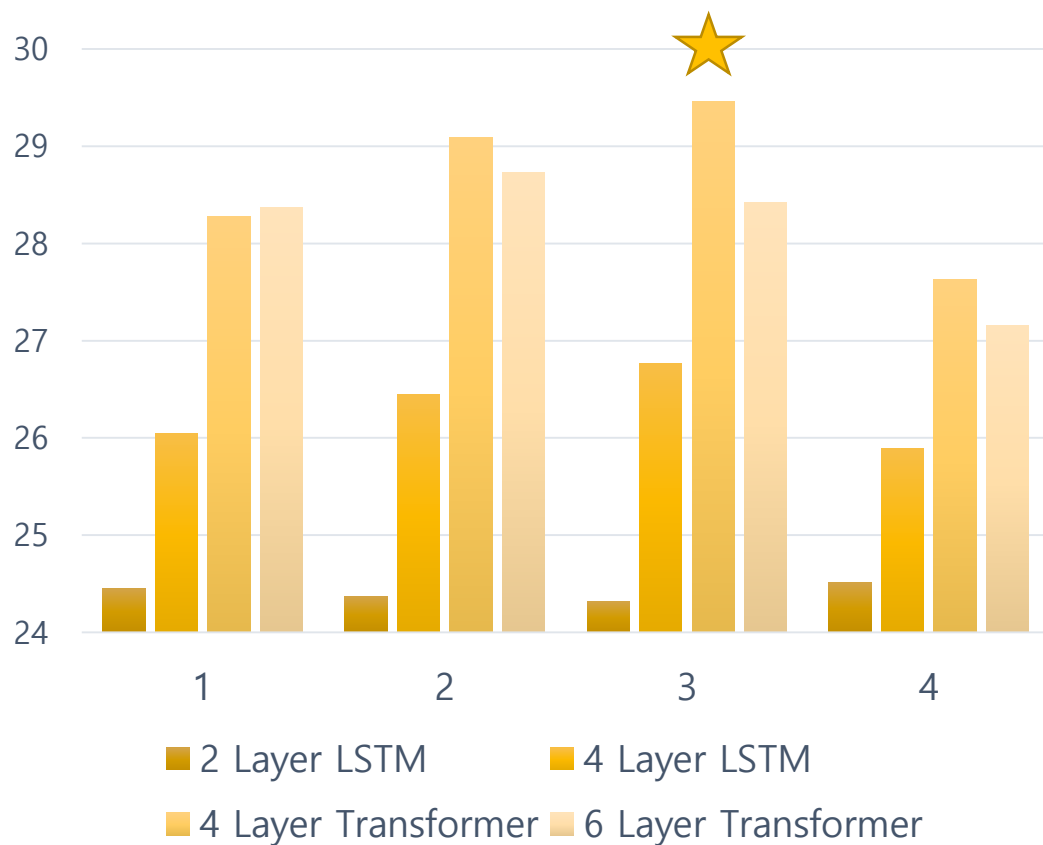
3. NMT 모델 학습 방법

- 실험 결과 BLEU 점수 비교

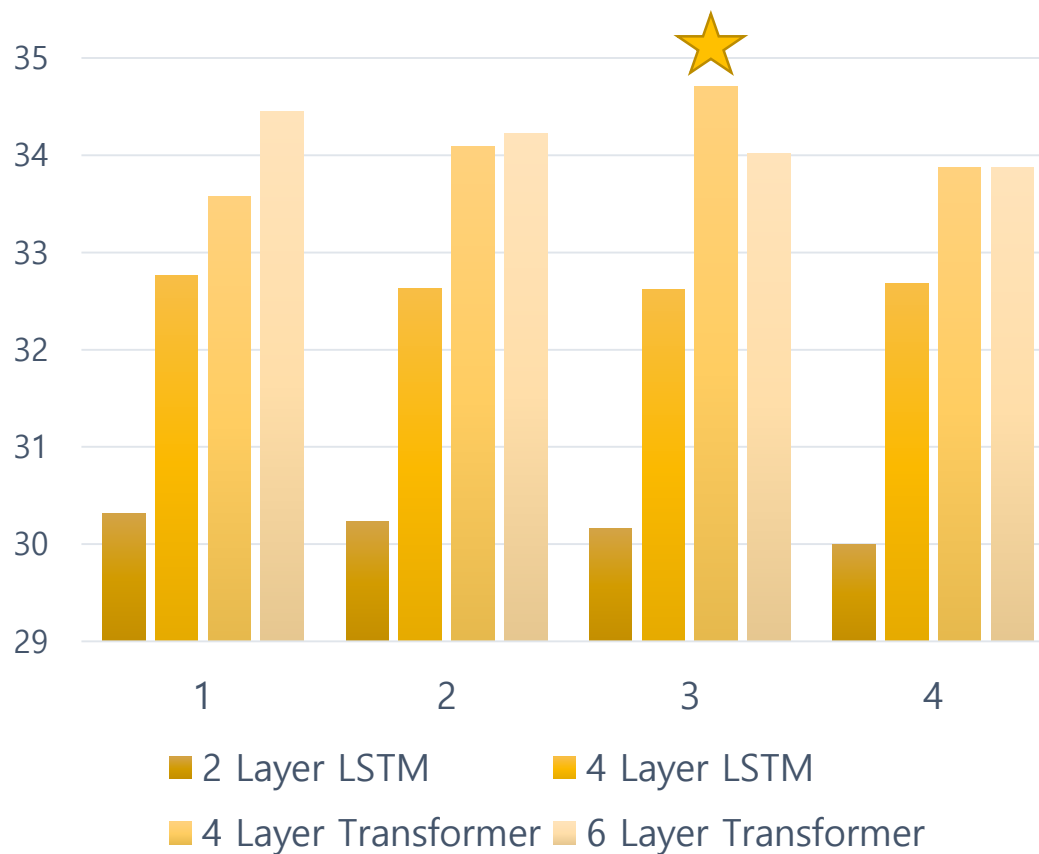
NMT 모델		2 Layer LSTM (OpenNMT Default)		4 Layer LSTM		4 Layer Transformer		6 Layer Transformer	
데이터셋		관용구 포함 평가셋	관용구 미포함 평가셋	관용구 포함 평가셋	관용구 미포함 평가셋	관용구 포함 평가셋	관용구 미포함 평가셋	관용구 포함 평가셋	관용구 미포함 평가셋
1	관용구 미포함 학습 (Baseline)	24.45	30.32	26.05	32.76	28.28	33.58	28.37	34.45
2	관용구 포함 학습	24.37 (-0.08)	30.23 (-0.09)	26.45 (+0.40)	32.63 (-0.13)	29.09 (-0.81)	34.09 (+0.51)	28.73 (+0.36)	34.22 (-0.23)
3	관용구 포함 학습 + <idm> 태그	24.32 (-0.13)	30.16 (-0.16)	26.76 (+0.71)	32.62 (-0.14)	29.46 (+1.18)	34.71 (+1.13)	28.42 (+0.05)	34.02 (-0.43)
4	관용구 포함 학습 + <idm> </idm> 태그	24.51 (+0.06)	30.00 (-0.32)	25.89 (-0.16)	32.68 (-0.08)	27.63 (-0.65)	33.88 (+0.30)	27.16 (-1.21)	33.88 (-0.57)

3. NMT 모델 학습 방법

관용구가 포함된 문장의 BLEU 점수 비교



관용구가 미포함된 문장의 BLEU 점수 비교



3. NMT 모델 학습 방법

- 관용구 학습 방법에 따른 번역 결과 비교

	관용구 학습 방법	번역 결과
원문	학부모만 <u>발을 동동 구르는</u> 상황이다.	only the parents are <u>anxious</u> .
1	관용구 미포함 학습 (Baseline)	only parents are <u>rolling their feet</u> .
2	관용구 포함 학습	only parents are <u>rolling their feet</u> .
3	관용구 포함 학습 + <idm> 태그	only parents are <u>struggling</u> .
4	관용구 포함 학습 + <idm> </idm> 태그	only parents are <u>jumping</u> .

3. NMT 모델 학습 방법

- NMT 학습 시 관용구 포함 문장을 함께 학습시킬 경우 성능 향상
 - 문장 내에 토큰을 삽입하여 관용구 위치를 표기하였을 때
관용구가 포함된 문장의 번역 성능이 대부분의 NMT 모델에서 향상
 - 관용구의 시작 위치를 특수한 토큰(<idm>)으로 표기한 경우 성능 향상
- ✓ Best case : <idm> 관용구 + 4 Layer Transformer
- ※ 관용구가 포함되지 않은 일반 문장의 번역 성능은 반대로 저하

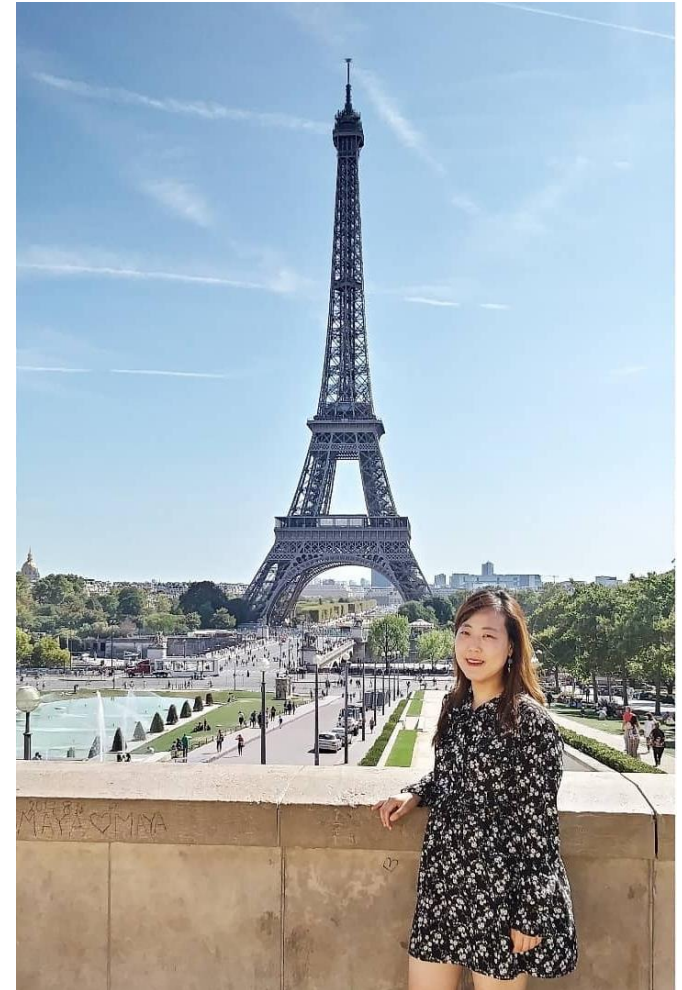
결론

- 관용구가 포함된 다량의 문장 데이터셋 구축
 - KISS : 420개 관용구, 7,500개 한-영 번역 쌍 데이터셋
- 관용구 번역에 특화된 번역 평가 지표 생성
 - 275개 관용구, 3,461개 번역 쌍에 대한 블랙리스트 구축
- 블랙리스트를 이용한 기계번역 서비스의 번역 정확도 측정 방법 제안
- NMT 학습 시 관용구가 포함된 문장을 함께 학습시킬 경우 번역 품질 상승함을 증명
 - 특수한 토큰을 사용하여 문장 내 관용구의 위치를 표기한 경우 번역 품질 상승
 - 한국어의 경우 관용구의 시작 위치만을 표기할 경우 번역 품질이 가장 크게 상승

Q & A

<https://www.facebook.com/minjoo.choi.562/>

<https://github.com/Judy-Choi>





감사합니다 – Thank you