JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 27 No. 3, pp. 13-23, March 2022
https://doi.org/10.9708/jksci.2022.27.03.013

Probing Sentence Embeddings in L2 Learners' LSTM Neural Language Models Using Adaptation Learning

Euhee Kim*

*Professor, Dept. of Computer Science & Engineering, Shinhan University, Gyeonggi-do, Korea

[Abstract]

In this study we leveraged a probing method to evaluate how a pre-trained L2 LSTM language model represents sentences with relative and coordinate clauses. The probing experiment employed adapted models based on the pre-trained L2 language models to trace the syntactic properties of sentence embedding vector representations. The dataset for probing was automatically generated using several templates related to different sentence structures. To classify the syntactic properties of sentences for each probing task, we measured the adaptation effects of the language models using syntactic priming. We performed linear mixed-effects model analyses to analyze the relation between adaptation effects in a complex statistical manner and reveal how the L2 language models represent syntactic features for English sentences. When the L2 language models were compared with the baseline L1 Gulordava language models, the analogous results were found for each probing task. In addition, it was confirmed that the L2 language models contain syntactic features of relative and coordinate clauses hierarchically in the sentence embedding representations.

▸ **Key words**: LSTM, (L2) language model, probing method, syntactic priming, adaptation effect

[요 약]

Prasad et al.는 사전학습(pre-trained)한 신경망 L1 글로다바(Gulordava) 언어모델을 여러 유형의 영어 관계절과 등위절 문장들로 적응 학습(adaptation learning)시켜 문장 간 유사성(sentence similarity)을 평가할 수 있는 통사 프라이밍(syntactic priming)-기반 프로빙 방법((probing method)을 제안했다. 본 논문에서는 한국인 영어학습자가 배우는 영어 자료를 바탕으로 훈련된 L2 LSTM 신경망 언어 모델의 영어 관계절 혹은 등위절 구조의 문장들에 대한 임베딩 표현 방식을 평가하기 위하여 프로빙 방법을 적용한다. 프로빙 실험은 사전 학습한 LSTM 언어 모델을 기반으로 추가로 적응 학습을 시킨 LSTM 언어 모델을 사용하여 문장 임베딩 벡터 표현의 통사적 속성을 추적한다. 이 프로빙 실험을 위한 데이터셋은 문장의 통사 구조를 생성하는 템플릿을 사용하여 자동으로 구축했다. 특히, 프로빙 과제별 문장의 통사적 속성을 분류하기 위해 통사 프라이밍을 이용한 언어 모델의 적응 효과(adaptation effect)를 측정했다. 영어 문장에 대한 언어 모델의 적응 효과와 통사적 속성 관계를 복합적으로 통계분석하기 위해 선형 혼합효과 모형(linear mixed-effects model) 분석을 수행했다. 제안한 L2 LSTM 언어 모델이 베이스라인 L1 글로다바 언어 모델과 비교했을 때, 프로빙 과제별 동일한 양상을 공유함을 확인했다. 또한 L2 LSTM 언어 모델은 다양한 관계절 혹은 등위절이 있는 문장들을 임베딩 표현할 때 관계절 혹은 등위절 세부 유형별로 통사적 속성에 따라 계층 구조로 구분하고 있음을 확인했다.

▸ **주제어**: LSTM, 영어학습자 언어 모델, 프로빙 방법, 구조/통사 프라이밍, 적응 효과

• First Author: Euhee Kim, Corresponding Author: Euhee Kim
*Euhee Kim (euhkim@shinhan.ac.kr), Dept. of Computer Science & Engineering, Shinhan University
• Received: 2022. 02. 10, Revised: 2022. 02. 11, Accepted: 2022. 03. 18.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

인공 신경망(Artificial Neural Network, ANN)의 언어 모델링 성능이 향상됨에 따라 사람과 동등한 수준의 문장 이해 능력 보이는 신경망의 이해/처리 능력을 향상시키기 위한 연구가 증가하고 있다. 이 연구 영역에서 사용되는 ANN(종종 LSTM, 최근에는 트랜스포머 기반 Bert)는 일반적으로 대규모 텍스트 말뭉치를 사전학습하며, 그 다음 다양한 통사적 구조를 가지는 문장들로 구성된 비교적 적은 데이터를 갖고 추가학습(fine-tuning)을 한다. 문장의 통사적 구조에 대한 감독 학습(supervised learning)없이 비지도 학습(unsupervised learning)으로 훈련된 ANN은 문장의 통사적 구조에 대한 민감도를 요구하는 자연어처리(Natural Language Processing, NLP) 태스크에서 놀라운 성능을 보여주고 있다[1-2].

그러나 이 성공의 기반이 되는 사전 학습된 신경망의 문맥 임베딩 표현(contextual embedding representation) 공간을 해석하는데 상당한 어려움이 존재한다. 이 어려움을 해결하기 위해 신경망의 프로빙 태스크(probing task)나 시각화 분석(heat map) 기술을 사용하고 있다[3-4].

이러한 연구는 적어도 두 가지 이유에서 중요하다. 특별한 언어 학습 과정 없이 단순한 언어 모델링에서 학습할 수 있는 정보의 유형과 양을 밝힐 수 있다. 또한 사람의 일반적 보편 문법에 관한 논쟁에 기여하고, 언어 모델링에 뛰어난 성능을 보이는 신경망 모델이 실제로 사람이 인식하는 동일한 통사적 구조에 민감한지 조사하여 ANN 자체에 대한 진단 도구로 사용될 수 있다.

ANN 네트워크 중 하나인 순환 신경망(Recurrent Neural Network: RNN) 기반을 둔 언어 모델은 자연어와 같은 순차적인 데이터를 처리하는데 좋은 성과를 보인다. 그러나 RNN은 입력 문장이 길어질수록 정확한 예측이 어려운 장거리 의존성(long distance dependency) 문제가 존재했다. 이러한 문제를 해결하기 위해 장단기 메모리(Long Short-Term Memory: LSTM) 기반 언어 모델을 제안했다[2].

최근 LSTM 기반 사전학습 언어 모델이 적응 학습을 통해 사람처럼 어휘 항목뿐만 아니라 문장의 추상적 문장 구조에도 적응한다는 것을 보여주었다. 즉, 다양한 유형의 통사적 구조를 지닌 문장들에 대하여 통사 프라이밍 진단을 통하여, 언어 모델의 임베딩 표현 공간이 언어적 해석 방식에 의해 계층적으로 구성되어 있으며 문장의 추상적 구조 속성을 반영한다는 것을 보여주었다. Fig. 1은 문장의 다양한 통사 구조에 대한 계층적 구성을 표현한다[5-6].

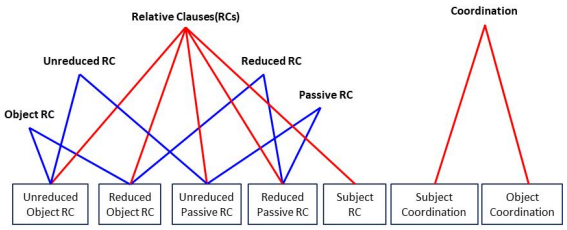

Fig. 1. The hierarchically syntactic structure in the LSTM language model's representation space

또한 최근 SOTA 성능을 보여주고 있는 Bert와 같은 언어 모델에 대해서도 문장들에 대한 이해 범위와 정도를 분석하기 위해 사전학습 언어 모델의 단어 임베딩 표현 공간에 내재하는 통사적 속성을 찾아내기 위해 통사 프로빙을 적용했다. 프로빙 태스크는 임베딩 표현에 담겨져 있는 문장이 가지는 언어적 속성을 진단하는 문제인데, 다른 복잡한 다운스트림 태스크(downstream task)와 비교할 때 실험이 명확하고 간단해서 결과의 해석이 쉽다[7-10].

본 논문에서는 영어를 외국어(L2)로 학습하며 한국인 영어학습자가 사용하는 영어 교재의 말뭉치로 심층 학습한 LSTM 신경망 언어 모델의 문장 표현 공간에서 영어 문장에 담겨진 통사적 구조의 위상을 평가하기 위해 통사 프라이밍-기반 프로빙 태스크를 적용한다.

본 논문의 구성은 다음과 같다. 2장에서는 신경망 언어 모델의 통사 프라이밍-기반 프로빙 태스크 관련 선행연구를 살펴보고, 3장에서는 영어 학습자가 사용하는 문장들로 학습한 LSTM 신경망 언어 모델에 적용할 통사 프라이밍-기반 프로빙 태스크를 설계한다. 4장에서는 실험에 활용한 언어 모델 및 실험 결과를 살펴보고, 마지막으로 5장에서 결론을 맺는다.

II. Related works

본 장에서는 신경망 언어 모델의 통사 프라이밍-기반 프로빙 태스크 관련 선행연구를 살펴본다. 심리언어학에서 사용해 온 사람 학습자를 대상으로 한 영어 문장들의 통사 구조 유사성을 진단하는 통사 프라이밍, 신경망 언어 모델의 문장 처리, 신경망 언어 모델을 대상으로 한 영어 문장들의 통사 구조 유사성을 진단하는 통사 프라이밍, 그리고 신경망 적응 언어 모델을 고찰한다.

2.1 Syntactic priming in humans

사람들은 (a)문장과 (c)문장이 'that' 관계절(Relative clauses; RCs)을 포함하기 때문에 구조적으로 유사하고,

(b)문장이 관계절이 없으니 다르다고 판단한다:

(a) The secretary that managed the plan submitted the project. (주어 관계절)
(b) The secretary managed the plan and submitted the project. (등위절)
(c) The nurse that reviewed the method corrected the error. (주어 관계절)

심리언어학에서는 (a)문장과 통사 구조가 다른 영어 문장들이 (a)문장 앞에 올 때보다 구조가 같은 다른 문장들이 앞에 올 때 (a)문장에 대한 사람의 이해 정도가 빠른 현상을 '통사 프라이밍(syntactic priming)' 혹은 '통사 점화'라 말한다[8].

Well et al.는 통사 프라이밍 효과를 이용하여 문장 구조의 유사성을 연구했다. 첫 번째 실험에서 학습자가 (c)문장과 같은 주어 관계절 문장들을 먼저 읽은 후 동일한 구조를 공유하는 다른 문장 (a)을 나중에 읽었을 때 반응 시간(reaction time)을 측정했다. 두 번째 실험에서는 (b)문장과 같은 등위절(coordination) 문장들을 먼저 읽은 후 관계절 구조를 가지는 (a)문장을 나중에 읽었을 때 반응 시간을 측정했다. 첫 번째 실험의 프라이밍이 두 번째 실험의 프라이밍보다 크게 나타났다. 그들은 통사 프라이밍을 이용하여 통사 구조 공유 여부에 따라 문장들 간의 유사성 여부를 규명했다[11].

2.2 Syntactic predictions in LSTM neural LMs

NLP 분야에서 신경망 언어 모델의 문장 처리 관련 연구가 최근 활발히 진행되고 있다.

Gulordava et al.연구에서는 영어 문장의 복잡한 주어와 관계하는 동사의 단수형 또는 복수형을 결정하는 문법성 판단 과제와 관련 사람의 문법성 판단력과 비교할 때 LSTM 신경 언어 모델(LSTM neural LMs)도 통계적으로 유의미한 판단 예측을 할 수 있다고 규명했다.

그리고 Gulordava et al.은 LSTM 신경망은 언어 정보를 추출할 뿐만 아니라 통사 구조 예측 능력에 있어서도 상당한 수준에 도달한다는 가설을 검증했다.

또한 Gulordava et al.이 제안한 LSTM 언어 모델은 실제 Google에서 제안한 신경망 언어 모델보다 성능이 우수했다[7].

2.3 Syntactic priming in Adaptive LSTM LMs

Hewitt et al.는 신경망 언어 모델의 단어 표현 공간에 담고 있는 통사적 구조에 대한 프로빙 태스크를 수행하여 언어 모델이 통사 구조를 어떻게 이해하는지를 연구했다[9].

그러나 새로운 도메인에 특화된 신경망 언어 모델을 구축하기 위해 매번 대용량 훈련 코퍼스를 이용하여 언어 모델을 훈련하기에는 과다한 시간과 노력이 필요하다. 이런 문제를 완화하기 위해 사전학습 언어 모델에 적용하여 추가학습(fine-tuning)이나 적응학습(adaptation learning)을 통해 새로운 체제의 신경망 언어 모델을 제안하고 있다[5-6].

언어 모델의 적응학습이란 기존에 사용하고 있는 언어 모델에 추가로 퓨-샷 학습(few-shot learning)을 설정하여 새로운 도메인에 속한 문장에 적응하도록 빠른 시간에 새로운 언어 모델을 만드는 방법이다. 적응 언어 모델이 필요한 이유는 모든 도메인에서 해당되는 큰 규모의 언어 모델을 사용하는 것보다는 작은 규모의 적응 언어 모델을 사용하는 것이 효율적이기 때문이다[12-15].

Prasad et al.는 글로다바(Gulordava) 언어 모델에 통사 프라이밍 패러다임을 적용하여 문장 임베딩 표현에 담겨진 통사구조 속성에 대해 진단했다. 즉, Prasad et al.는 사전학습 글로다바 언어모델에 여러 유형의 관계절과 등위절 문장들로 적응 학습시켜 문장 간 유사성(sentence similarity)을 평가할 수 있는 통사 프라이밍-기반 프로빙 테스크를 제안했다. 이들이 제안한 문장 간 유사성 측정은 LSTM 언어 모델을 통해 문장들의 통사 구조에 대한 상대적 예외성 확률을 추출하고, 적응 학습에 따른 적응 효과를 계산하여 문장들의 유사성을 평가했다. 문장 간 유사성 평가 방법에 의해 적응 언어 모델은 언어적 해석 방식으로 관계절 구조 정보를 표현 공간에 계층적으로 구성하고 있음을 규명했다[6].

본 논문에서 제안하는 신경망 언어 모델을 이용한 관계절 세부 유형 통사 구조를 찾는 통사 프라이밍-기반 프로빙 태스크는 Prasad et al.가 제안한 방법을 한국인 영어학습자(L2 learner; L2er)가 학습하는 영어 자료로 사전 학습된 L2 LSTM 언어 모델과 퓨-샷 학습 방법에 적용했다.

본 논문에서는 앞으로 Prasad et al.가 설계한 LSTM 언어 모델을 글로다바 언어 모델로 지칭하고, 본 논문에서 제안하는 L2 언어 모델과 비교할 때 베이스라인 언어 모델로 정한다.

통사 프라이밍-기반 프로빙 태스크 실험 수행 과정에서 Prasad et el.이 사용한 언어 모델 종류를 본 논문에서 제안한 언어 모델과 비교하여 Table 1에 정리했다.

Table 1. Language models(LMs) for probing tasks

Parameter		Gulordava model	L2 model
pre-trained LMs	RNN	LSTM	LSTM
	Hidden layers	2	2
	Hidden units	100, 200, 400, 800, 1600	100, 200, 400
	Tokens size	2M, 10M, 20M	7M, 9.8M, 13M
Pre-trained LMs		75	9
Adaptive LMs		375	45

III. Methods

본 장에서는 Prasad et al.의 연구 문제가 L2 LSTM 언어 모델의 문장 임베딩 표현에 어떻게 포착되는지를 진단하는 통사 프라이밍-기반 프로빙 태스크를 소개한다.

프로빙 태스크에 대한 전체 구성도은 Fig. 2—3과 같다. 실험에서 사용한 하드웨어는 Table 2에 정리했다.

Table 2. Hardware configuration

Name	Version
GPU	RTX A5000
CPU	i9-10980XE
RAM	32G
HDD	2TB
SSD	1TB

제안한 언어 모델의 구현은 Linux Ubuntu 운영체제와 Visual Studio Code 개발도구 환경에서 Python언어와 Pytorch 라이브러리를 이용하여 수행했다.

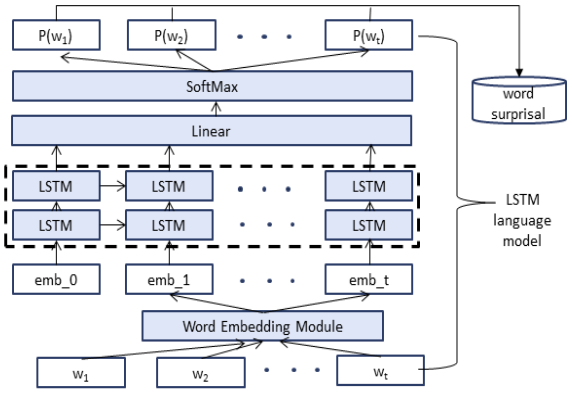

Fig. 2. L2 LSTM language model architecture

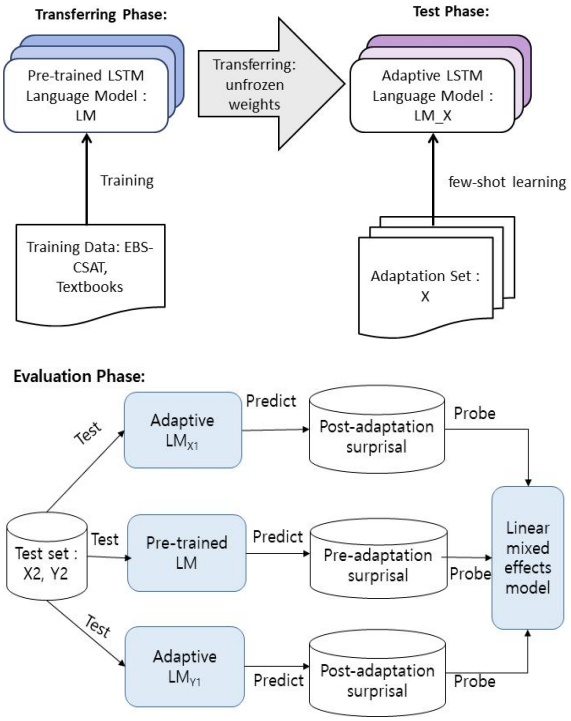

Fig. 3. A probing procedure using syntactic priming

3.1 Experimental data for adaptive LSTM LMs

통사 프라이밍-기반 프로빙 태스크에서 진단한 통사적 구조는 Prasad et al.이 사용했던 실험 데이터를 참조하여 관계절 5종류와 등위절 2종류를 포함한 템플릿 7종을 사용하여 문장들을 생성했다[6].

7종 탬플릿은 능동형-주어-관계절(active subject RC; SRC), 목적어-관계절(unreduced object RC; ORC), 축약-목적어-관계절(Reduced object RC; ORRC), 축약-수동형-관계절(reduced passive RC; PRRC), 주어-등위절(subject coordination; SCONT), 목적어-등위절(object coordination; OCONT)로 구성된다.

템플릿 문장들은 동사 223개, 명사 164개, 부사 24개, 형용사 78개로 슬롯을 채워 명사, 동사, 부사, 형용사를 조합하여 의미적으로 타당하게 만든다.

7종 템플릿을 사용하여 5개의 실험 목록(experiment list)을 생성하고 각 목록은 적응 세트(adaptation set)와 테스트 세트(test set)의 쌍으로 구성되어 있으며 기능 단어(that, and)와 일부 수식어만 공유하고 어휘 중복을 최소화했다. 각 적응 세트는 20개 적응 문장들을 포함하며, 테스트 세트는 50개 테스트 문장들을 포함한다.

Table 3의 7가지 변형으로 생성된 예문들은 거의 동일한 어휘들로 구성된 것을 알 수 있다.

Table 3. Examples of sentences with 7 templates

Structure	Sentence
ORC	The conspiracy **that the employee welcomed** divided the beautiful country.
ORRC	The conspiracy **the employee welcomed** divided the beautiful country.
PRC	The conspiracy **that was welcomed by the employee** divided the beautiful country.
PRRC	The conspiracy **welcomed by the employee** divided the beautiful country.
SRC	The employee **that welcomed** the conspiracy quickly searched the buildings.
OCONT	The conspiracy welcomed **the employee and** divided the beautiful **country**.
SCONT	The employee **welcomed** the conspiracy **and** quickly **searched** the buildings.

3.2 L2 Pre-trained LSTM LMs (L2PLMs)

통사 프라이밍-기반 프로빙 태스크에 사용한 L2 사전학습 LSTM 신경망 언어 모델(L2PLMs)은 Kim et al.이 제안한 언어 모델을 사용했다[16-17]. 제안한 L2PLMs 모델의 아키텍처는 Fig. 2와 같다. 일반적으로 언어 모델의 기능은 문장 $w = (w_1, \cdots, w_t)$이 입력계층에 순차적으로 입력되었을 때 앞서 입력된 단어 배열을 통해 문맥을 학습하고, 출력계층에서 단어사전의 모든 단어에 대해 각 단어가 문장 w의 다음에 나올 단어 w_{t+1}로 예측되는 확률 $P(w_{t+1})$을 출력한다.

Table 1의 L2PLMs 모델은 공통으로 입력계층과 출력계층 사이에 임베딩 계층(embedding layer), 2개 은닉 계층(hidden layer), 완전 연결 계층(fully connected layer)로 사용한 선형 계층(linear layer)으로 구성된다. 입력 계층과 출력 계층의 유닛수는 사전의 크기(vocabulary size)로 설정하고, 임베딩 계층은 입력 계층에서 전달받은 단어 배열을 사전학습 모델 word2vec의 임베딩 공간으로 인코딩하여 각 단어를 256차원의 단어 임베딩 벡터(word embedding vector)로 표현한다. 은닉 계층은 2개의 동일한 차원의 계층을 사용하며 각각의 은닉 유닛은 LSTM 블록으로 구성된다.

Fig. 2의 L2PLMs 모델에 사용된 학습 코퍼스는 2016년부터 2018년까지 출판된 EBS-CSAT English Prep Books와 2001년 그리고 2009년에 한국에서 출판된 중학교 및 고등학교 영어 교과서에서 수집한 영어 문장들을 사용했다.

통사 프라이밍-기반 프로빙 태스크에서는 임베딩 계층과 은닉 계층의 차원을 휴리스틱 방법으로 설정하여 9개의 모델을 구축했다. Prasad et al.은 글로다바 모델을 사용하여 75개의 사전학습 모델을 구축했다.

3.3 L2 Adaptive LSTM LMs (L2ALMs)

본 절에서는 Table 3에서 제시한 7종류의 통사 구조가 L2 언어 모델의 문장 임베딩 표현에 Fig. 1과 같이 계층적으로 구성되는지를 진단하기 위해 통사 프라이밍-기반 프로빙 절차를 소개하며 절차 과정은 Fig. 3과 같다.

통사 프라이밍-기반 프로빙 절차는 L2PLMs 모델의 전이 단계, 테스트 단계, 그리고 평가 단계로 구성된다. Fig. 3의 전이 단계에서는 9개 L2PLMs 모델을 각각 사용했다.

Fig. 3의 테스트 단계에서는 L2PLMs 모델의 전이 단계에서 추출한 가중치를 사용하여 적응 학습을 추가 진행한다. 적응 학습은 L2PMs 모델에 적응 세트의 적응 문장을 하나씩 입력하면서 목표 문장을 예측할 때 교차 엔트로피 손실(cross-entropy loss)을 계산하여 언어 모델의 가중치를 업데이트한다. 업데이트된 가중치를 사용하여 다음 적응 문장을 예측한다.

실험에서는 원-샷 학습을 수행하기 위해 1절의 관계사절 문장들로 구성된 실험 목록의 적응 세트를 대상으로 테스트를 수행하면서 L2PLMs 모델의 가중치들을 조정하고 모든 적응 문장들에 업데이트된 새로운 총 45개 L2 LSTM 적응 언어 모델(L2ALMs)을 저장했다. Fig. 4는 이 과정에 대한 도식도이다.

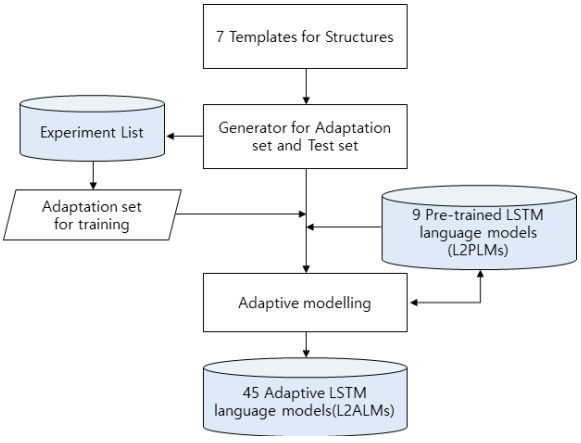

Fig. 4. Flowchart for Adaptive Language Model

Fig. 4는 9개의 L2PLMs 모델을 이용하여 적응 세트로 적응 모델링하는 과정을 나타낸다. 본 실험에서 사용한 적응 알고리즘은 다음과 같다:

1. Test on a sentence
2. Update weights based on that sentence
3. Repeat on remaining sentences

Fig. 3의 평가 단계에서는 문장 임베딩이 문장의 통사 구조 속성을 잘 적응 학습하고 있는지를 평가하기 위해 적응 효과와 통사적 속성 관계에 대한 통계 분석 실험을 진

행한다. 두 언어 모델 L2PLMs와 L2ALMs을 이용하여 테스트 문장의 통사 프라이밍 적응 효과를 측정한다. 그리고 L2ALMs 모델의 표현 공간에 관계절 및 등위절 통사 구조들이 계층적으로 구성됨을 진단하기 위해 선형 혼합효과 모형 분석을 수행한다.

3.4 Surprisal(s) for LSTM LM

심리언어학이나 NLP 분야에서는 문장 속의 각 단어를 처리하는데 요구되는 인지 노력을 측정하는 방법으로 Hale이 제안한 놀라움(surprisal) 정보를 일반적으로 사용한다. 놀라움은 문맥 속 단어들의 상대적 예외성을 측정하여 현재 관찰 단어 이전까지 주어진 문맥에 이어서 목표 단어가 나타날 확률을 음의 로그로 계산한다[18].

Fig. 2의 LSTM 언어 모델(LM)을 활용하여 목표 단어(w_i)에 대한 놀라움($Surp_{LM}(w_i)$)을 구하는 수식은 다음과 같다.

$$Surp_{LM}(w_i) = -\log_2(P(w_i|h_1h_2\cdots h_{i-1})) \quad (1)$$

h_i는 w_i로 수렴되기 이전 은닉 상태이며, 확률은 소프트맥스 활성화 함수를 사용하여 계산한다. 단어의 출현 확률이 낮으면 놀라움이 높고, 출현 확률이 높으면 놀라움이 낮다. 단어를 처리하는 데 필요한 인지 노력은 그 놀라움에 비례한다.

LM 모델의 문장을 처리하는데 요구되는 인지노력에 대한 측정은 다음과 같이 정의한다. 수식 (1)을 이용하여 문장(S)에 대한 놀라움($Surp_{LM}(S)$)을 문장을 구성하는 각 단어에 대한 놀라움을 구하여 합을 계산한다.

$$Surp_{LM}(S) = \sum_{1}^{n} Surp_{LM}(w_i) \quad (2)$$

복문을 포함하는 집합(X)에 대한 놀라움($Surp_{LM}(X)$)는 수식 (2)을 이용하여 X에 속하는 모든 문장에 대한 $Surp_{LM}(S)$을 구하여 평균을 계산한다.

Fig. 3의 평가 단계에서는 각 L2PLMs 모델(LM) 과 적응 세트(X)로 적응시킨 L2ALMs 모델(LM_X)을 사용하여 테스트 세트(Y)에 대한 적응 전-놀라움($Surp_{LM}(Y)$)과 적응 후-놀라움($Surp_{LM_X}(Y)$)을 계산한다. 그리고 적응 전후의 놀라움의 차이를 $Adapt(Y \mid X)$로 표시한다.

$$Adapt(Y|X) = Surp_{LM}(Y) - Surp_{LM_X}(Y) \quad (3)$$

3.5 The adaptation effect ($AE(Y \mid X)$)

L2ALMs 모델의 표현에 관계절 구조 표현 정보를 담고 있는지에 대한 통사 프라이밍-기반 프로빙 태스크는 Fig. 3의 평가 단계에서 문장의 적응 효과(adaptation effect; AE)을 계산하여 서로 다른 통사 구조들 간 유사성 측정을 수행한다.

평가 단계에서 적응 세트(X)가 주어진 경우 테스트 세트(Y)에 대한 적응 효과를 측정할 때 L2ALMs 모델의 적응 효과($AE(Y \mid X)$)는 Prasad et al.가 제안한 수식을 사용했다[6]:

$$AE(Y|X) = Adapt(Y|X) - Surp_{LM}(Y) \quad (4)$$

수식 (4)에서 $Surp_{LM}(Y)$와 $Adapt(Y \mid X)$사이에 유의미한 양의 상관관계가 존재함($p < 2.2e-16$)을 Fig. 5에서 보여준다.

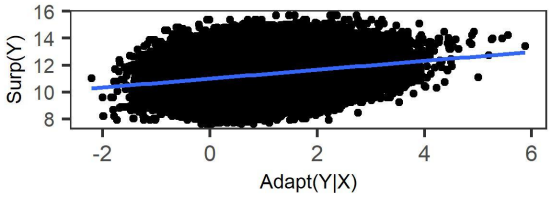

Fig. 5. The relation between $Surp_{LM}(Y)$ and $Adapt(Y \mid X)$

만일 서로 다른 테스트 세트 Y와 Z가 주어진 경우 LM_X 모델의 적응 효과가 $AE(Y \mid X) > AE(Z \mid X)$을 만족한다면, 적응 세트 X는 Z 보다 Y와 유사함을 의미한다. 그 이유는 수식 (3)-(4)을 이용하여 다음 결과를 도출할 수 있기 때문이다: $Surp_{LM_X}(Y) < Surp_{LM_X}(Z)$. 그리고 양의 상관관계에 의해 $Adapt(Y|X) < Adapt(Z|X)$을 만족한다.

IV. Results

통사 프라이밍-기반 프로빙 태스크 실험은 Fig. 3의 총 9개 L2PLMs 모델과 45개 L2ALMs 모델을 사용했다. 이 실험은 통사 프라이밍-기반 프로빙 태스크별로 각 테스트 문장을 각각 L2PLMs 모델과 L2ALMs 모델에 입력하여 적응 효과 $AE(Test|Adaptation)$을 평가했다.

L2ALMs 모델의 적응 효과와 Table 3의 관계절 구조를 이용하여 언어적 속성을 복합적으로 분석하기 위해 혼합효과 모형 분석을 수행했다. 혼합효과 모형 분석은 피험자별 평균 및 항목별 평균, 실험에서 조절하는 다양한 독립변수별 평균을 이용하여 데이터를 분석하는 방법이다. 따라서 통계적으로 유의미한 효과가 나타나면, 언어 모델의

문장 임베딩 표현에 문장이 가지고 있던 통사 구조의 속성이 잘 담겨져 있다고 검증했다.

L2LMs 모델에서 얻은 실험 결과들은 선행 연구 Prasad et al의 실험 결과들과 동일한 양상을 공유함을 확인했다[6].

4.1 Similarity between sentences by AE

첫 번째 통사 프라이밍-기반 프로빙 태스크는 Table 3의 통사 구조에 대해 적응 효과 $AE(Test|Adaptation)$을 계산하여 L2ALMs 모델이 영어 문장의 통사 구조 속성을 잘 담고 있는지를 평가하기 위한 실험이다.

적응 효과를 측정하기 위해 적응 세트($X1$)과 테스트 세트($X2$)는 Table 3에서 제시한 통사 구조를 공유하지만 어휘적으로 겹치지 않는 문장들을 포함한다고 가정했다. 테스트 세트($Y2$)는 $X1$과 다른 통사 구조를 가지고 어휘적으로는 $X2$와 일치하는 문장들을 포함할 경우 수식 (4)을 이용하여 적응 효과를 측정하였을 때 다음과 같다:

$$AE(X2|X1) > AE(Y2|X1) \quad (5)$$

5절의 적응 효과의 정의에 의해 $X1$는 $Y2$보다 $X2$와 유사함을 의미한다.

Fig. 6은 위의 적응 효과 관계를 막대그래프로 시각화한 그래프이다.

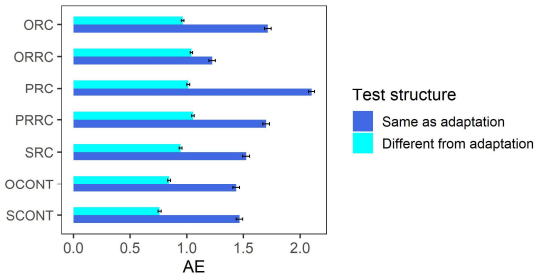

Fig. 6. The AE averaged across all 9 models

위 그래프는 Table 3의 통사 구조(Y축)에 따라 $X1$과 동일한 구조(아래쪽 막대그래프)를 공유하는 $X2$와 다른 구조(위쪽 막대그래프)를 가지는 $Y2$에 대해 AE를 9개 L2ALMs 모델에서 측정하여 평균을 낸 결과이다.

Prasad et al. 연구 결과와 마찬가지로 L2ALMs 모델도 적응 문장과 테스트 문장이 동일한 통사 구조를 공유할 경우 문장 간 유사성이 더 높음을 확인했다 (즉, 이 경우 막대그래프의 길이가 더 길다).

위의 결과를 통계적 분석을 통하여 검증하기 위해 적응 효과와 통사적 구조 간의 요인 관계를 R을 이용하여 혼합효과 모형 분석을 했다. 실험 목록(experimentList)에서 Table 3의 통사 구조 유형별 적응 효과 AE가 유의미 효과가 있는지 분석하기 위해 설계한 혼합효과 모형은 다음과 같다:

$$AE \sim structure + (1|experimentList) \quad (6)$$

여기서 고정 효과(structure)는 테스트 통사 구조가 적응 통사 구조와 같으면 1로, 다르면 -1로 코딩되는 범주형 변수이다. 랜덤 효과(experimentList)는 적응과 테스트 세트가 쌍으로 구성된 목록이다.

혼합효과 모형 분석 결과에 따라 AE는 통사 구조 유형별 유의미한 관계가 있음을 Table 4에 정리했다. 즉, 동일한 구조를 공유하는 문장들은 어휘만 일치하는 다른 구조를 갖는 문장들보다 L2ALMs 모델의 표현 공간에서 서로 유사하게 표현된다는 것을 규명했다.

Table 4. Mixed-effects model with AE

Structure	Estimate	Std	Pr(>|t|)
ORC	0.376	0.008	0.000***
ORRC	0.092	0.008	0.000***
PRC	0.545	0.008	0.000***
PRRC	0.323	0.008	0.000***
SRC	0.289	0.008	0.000***
OCONT	0.296	0.08	0.000***
SCONT	0.352	0.009	0.000***

4.2 Similarity between sentences with CONTs

두 번째 통사 프라이밍-기반 프로빙 태스크는 Table 3의 등위절을 중심으로 $AE(Test|Adaptation)$을 측정하여 L2ALMs 모델이 영어 문장의 등위절 속성을 잘 담고 있는지를 평가하기 위한 실험이다. Table 3에 제시된 등위절의 2가지 유형을 살펴보면 OCONT와 SCONT의 예문들은 다른 통사 구조 속성을 가지며 의미가 다른 문장들이다.

등위절과 관계절 관련 예문들을 비교해보면 OCONT와 ORC는 서로 다른 구조를 가지며 동일한 어휘들로 구성된 문장들이다. SCONT는 SRC와 다른 구조를 가지면서 동일한 어휘로 구성된 문장들이다. 따라서 등위절을 가지는 문장들을 적응 문장과 테스트 문장으로 나눠서 적응 효과(AE)를 구하여 등위절 문장들의 유사성을 진단했다.

우선, 적응 세트($Coord_{X1}$)와 테스트 세트($Coord_{X2}$)는 서로 다른 유형의 등위절 구조를 가지며 어휘적으로 거의 겹치지 않는 문장들을 포함한다. 테스트 세트(RCs)는 $Coord_{X1}$와 다른 통사 구조를 가지고 어휘적으로 $Coord_{X2}$와 일치하는 문장들을 포함한다.

수식 (2)을 이용하여 적응 효과를 측정하였을 때 다음과 같다:

$$AE(Coord_{X2}|Coord_{X1}) > AE(RCs|Coord_{X1}) \quad (7)$$

Fig. 7의 상단 그래프는 위의 적응 효과 관계를 막대 그래프로 시각화한 그래프이다.

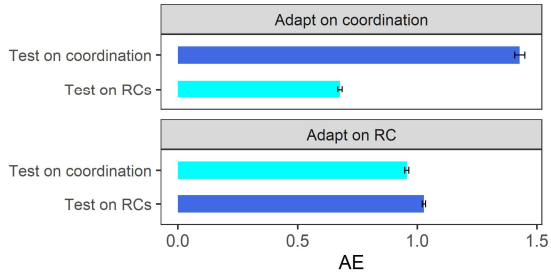

Fig. 7. Similarities between sentences with different types of coordination(upper panel) and RCs(lower panel)

위의 결과를 통계적으로 검증하기 위해 주어진 적응 효과(AE)에 적합하여 적응 효과에 영향을 미치는 등위절 구조에 대해 R을 이용한 혼합효과 모형 분석을 수행했다. 여기서, 고정 효과(testRC)는 관계절 구조로 테스트 할 경우 1로 코딩하고 다른 유형의 등위절로 테스트할 경우 -1로 코딩되는 범주형 변수이다.

$$AE \sim testRC + (1|adaptlist) \quad (8)$$

혼합효과 모형 분석 결과를 중심으로 적응 효과에 고정 효과 변인의 영향이 통계적으로 유의미한 효과가 있음을 Table 5에 정리했다.

Table 5. Mixed-effects model with AE

Adaptation	testRC	Estimate	Std	Pr(>|t|)
Coordination	Intercept	1.053	0.070	0.00***
	TestRC1	-0.375	0.005	0.00***

Prasad et al.의 연구 결과와 마찬가지로, 등위절이 있는 문장들로 적응된 L2ALMs 모델을 다른 유형의 등위절이 있는 문장들로 테스트할 경우 관계절 문장들로 테스트했을 때보다 적응 효과가 상당히 컸다(위 막대그래프의 길이가 아래 막대그래프의 길이보다 상당히 길다). 이것은 등위절이 있는 문장들이 서로 다른 구조를 포함하고 있어도 L2AMs 모델은 서로 유사한 문장들로 진단했다.

4.3 Similarity between sentences with RCs

세 번째 통사 프라이밍-기반 프로빙 태스크는 Table 3에 제시된 다양한 유형의 관계절을 중심으로 적응 효과 $AE(Test|Adaptation)$을 측정하여 L2ALMs 모델이 관계절이 가지는 대표적 통사 구조 속성 중의 하나인 '공백(gap)'을 추적하는지를 평가하기 위한 실험이다. 공백이란 관계절을 수식하는 관계절 선행사와 연관되어 관계절 내부에서 생략된 주어나 목적어를 의미한다.

Table 3의 주어-관계절(SRC), 목적어-관계절(ORC), 축약 관계절(ORRC, SPRRC)은 공백을 공유한다. 관계절이 있는 영어 문장들은 항상 공백을 공유하지만 등위절 문장은 공백을 갖지 않는다.

본 실험에서는 L2ALMs 모델이 관계절의 공백 속성을 추적할 수 있는지를 진단하기 위해 특정 유형의 관계절이 있는 문장들로 구성된 적응 세트(RC)와 나머지 다른 유형의 관계절들이 있는 문장들로 구성된 테스트 세트(RCs), RCs와 어휘적으로 겹치면서 등위절이 있는 문장들로 구성된 테스트 세트($Coordination$)을 선택했다.

수식 (4)을 이용하여 적응 효과를 측정하였을 때 다음과 같다:

$$AE(RCs|RC) > AE(Coordination|RC) \quad (9)$$

Fig. 7의 하단 그래프는 위의 적응 효과 관계를 막대그래프로 시각화한 그래프이다. L2ALMs 모델은 다른 유형의 관계절 문장들을 테스트할 경우 등위절 문장들을 테스트했을 때보다 적응 효과가 약간 큰 것으로 나타났다.

위의 결과를 통계적으로 검증하기 위해 혼합효과 모형을 주어진 언어 모델의 적응 효과에 적합하여 적응 효과에 영향을 미치는 테스트 문장의 구문 유형별에 대해서 통계 분석했다.

$$AE \sim testRC + (1|experimentList) \quad (10)$$

여기서 고정 효과(testRC)는 적응 문장의 구조와 다른 유형의 관계절로 테스트할 경우 1로 코딩하고, 등위절로 테스트할 경우는 -1로 코딩했다.

Table 6의 혼합효과 모형 분석 결과를 중심으로 적응 효과와 다양한 유형의 관계절 간 유의미한 효과가 나타났다.

Table 6. Mixed-effects model with AE

Adaptation	testRC	Estimate	Std	Pr(>|t|)
RC	Intercept	0.992	0.056	0.00***
	TestRC	0.035	0.002	0.00***

Prasad et al.의 연구 결과와 마찬가지로, L2LMs 모델도 관계절의 공백 속성을 추적할 수 있으며 등위절 문장과 관계절 문장을 다르게 처리함을 진단했다.

4.4 Similarity between sentences in sub-classes of RCs

네 번째 통사 프라이밍-기반 프로빙 태스크는 Table 3의 5 종류의 관계절을 하위 범주인 축약(reduction)과 수

동형(passivity)로 나눠서 $AE(Test|Adaptation)$을 측정하고 L2ALMs 모델이 하위범주 속성들을 추적하는지를 평가하기 위한 실험이다.

Table 3을 보면 기능어 'that'가 생략된 축약 속성을 공통으로 갖는 관계절은 ORRC와 PRRC이고, 수동형 속성을 공통으로 갖는 관계절은 PRC와 PRRC가 된다. 축약 또는 수동형 속성 하나만 갖는 관계절은 ORRC, PRC이고, 축약과 수동형 속성 둘 다 갖는 관계절은 PRRC가 된다. 축약과 수동형 속성 둘 다 갖지 않는 관계절은 SRC가 된다.

L2ALMs 모델이 위의 관계절의 하위 범주 속성을 추적할 수 있는지를 진단하기 위해 우선 수동형 적응 세트(RCs_{Pa})을 PRC와 PRRC 관계절이 있는 문장들로 구성하고, 축약형 적응 세트(RC_{Re})는 PRC와 ORC 관계절 문장들로 구성했다. 수동형 적응 모델의 테스트 세트는 RC_{Pa}와 동일하게 선택했다. 마찬가지로 축약형 적응 모델의 테스트 세트는 RC_{Re}와 동일하게 선택했다.

수식 (4)을 이용하여 적응 효과를 측정하였을 때 다음 결과를 얻었다:

$$AE(Match \mid RC_{Pa}) > AE(Mismatch|RC_{Pa}) \quad (11)$$

$$AE(Match \mid RC_{Re}) > AE(Mismatch|RC_{Re}) \quad (12)$$

$$AE(Match \mid RC_{Pa\cap Re}) > AE(Mismatch \mid RC_{Pa\cap Re}) \quad (13)$$

Fig. 8은 수식 (11) - (13)의 적응 효과 관계를 막대그래프로 시각화한 그래프이다.

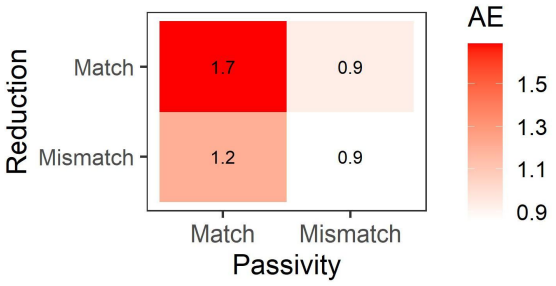

Fig. 8. Similarity between sentences in sub-classes of RCs

수식 (11)은 수동형 속성에서 적응 문장과 테스트 문장의 구조가 불일치할 때의 적응 효과(AE = 0.9+0.9)는 일치할 때의 적응 효과(AE = 1.7+1.2)보다 적은 것을 알 수 있다. 수식 (12)은 축약 속성에서 적응 문장과 테스트 문장의 구조가 일치할 때의 적응 효과(AE = 1.7+0.9)가 불일치할 때의 적응 효과(AE = 1.2+0.9)보다 큰 것을 알 수 있다. 수식 (13)은 수동형 속성과 축약 속성이 둘 다 일치한 관계절은 두 가지 속성 중 하나만 일치하거나 두 속성이 모두 불일치한 경우보다 적응 효과(AE =1.7)가 크게 나타났다.

종합적으로 L2LMs 모델은 문장의 수동형 속성을 추적한다고 진단할 수 있다. Prasad et al. 의 연구결과와 마찬가지로 L2LMs 모델의 표현 공간에서도 수동형 구조가 축약형 구조보다 문장 간의 유사성에 더 많이 영향을 미치는 것을 알 수 있다. 또한 관계사 'that' 생략 구조의 문장들이 관계사 생략 구조가 없는 문장들보다 적응 언어 모델의 표현 공간에서 훨씬 더 유사도가 높음을 진단할 수 있다.

Fig. 8의 L2LMs 모델의 적응 효과에 적합한 혼합효과 모형은 수식 (10)과 같다. 고정 효과(testRC)은 4가지 수준을 갖는 범주형 변수가 된다. 4가지 수준에 대한 코딩은 Table 7에 정리되어 있다.

Table 7. TestRC as fixed-effect

	Reduced_ match	Both_ match	No_ match
Passive_match	0	0	0
Reduced_match	1	0	0
Both_match	0	1	0
No_match	0	0	1

L2ALMs 모델의 적응 효과에 영향을 미치는 고정 효과 요인으로서 관계절의 수동형 속성과 축약형 속성은 통계적으로 유의미한 효과가 있음을 Table 8에서 확인할 수 있다.

Table 8. Mixed-effects model with AE

testRC/Contrast	Estimate	Std	Pr(>|t|)
Reduced_match	-0.270	0.010	<2e-16***
Both_match	0.482	0.010	<2e-16***
No_match	-0.344	0.010	<2e-16***

4.5 Effect of model size and tokens size on the similarity between sentences with all RCs

본 절에서는 언어 모델 크기와 훈련 말뭉치 크기가 관계절이 있는 문장 간 유사성에 영향을 미치는지 검증하고자 한다.

평가 모델들의 표현 공간에서 문장 간 유사성을 비교하기 위해 적응 효과(AE)를 직접 사용할 수 없다. 그 이유는 더 많은 데이터로 훈련된 모델은 적은 데이터로 훈련된 모델보다 더 강한 언어적 속성을 포착할 가능성이 높으므로 추가 학습을 수행해도 적응 효과가 클 가능성이 낮다. 이 문제를 완화하기 위해 Prasad et al.가 제안한 거리 척도를 사용했다.

클래스(C)에 속하는 문장들(S_C)과 해당 클래스에 속하지 않는 문장들($S_{\neg C}$) 사이의 거리는 수식 (4)을 이용하여 적응 효과 AE의 비율로 다음과 같이 정의한다.

$$D(S_C, \neg S_C) = \frac{AE(X_2|X_1)}{AE(\neg X_2|X_1)} \quad (14)$$

본 실험에서는 언어적으로 해석이 가능한 3종류의 클래스를 정의하여 문장 간 거리를 계산했다. 3종류의 클래스는 Table 3에서 제시한 동일한 관계절 구조를 가지는 문장들의 클래스(Specific RCs), 축약 관계절을 가지는 문장들의 클래스(Reduced RC), 그리고 모든 관계절 구조를 가지는 문장들의 클래스(All RCs)가 된다.

관계절이 있는 문장 간 거리 $D(S_{RC}, S_{\neg RC})$ 계산은 Fig. 9를 이용하여 수행한다. Fig. 9에서 각 행의 검정색 사각형은 L2ALMs 모델의 적응 학습에 사용했던 특정 구조 X_1을 표시한다. 흰색 사각형은 X_1과 다른 구조를 표시한다. 그리고 파랑색 사각형은 위에서 정의한 3종류의 클래스에 속하지 않는 구조를 표시한다.

수식 (14)을 통해 $D(S_{RC}, S_{\neg RC})$거리는 먼저 각 행에 따라 주어진 흰색 사각형별로 적응 효과(AE)을 계산해서 그들의 평균을 구한다. 마찬가지로 파랑색 사각형별로 적응 효과(AE)을 계산해서 그들의 평균을 구한다. 두 평균값의 비율을 계산한다. 그다음 각 행별 구한 비율을 모두 합하여 평균을 구한 결과가 된다.

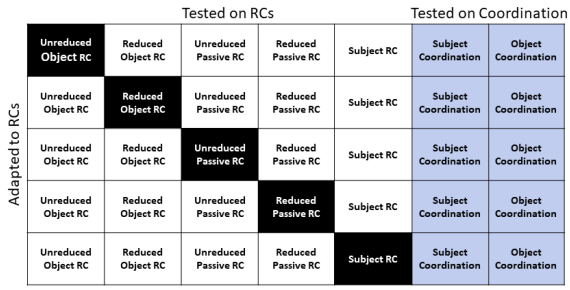

Fig. 9. A schematic of how $D(S_{RC}, S_{\neg RC})$ is calculated

Fig. 10은 L2LMs 모델의 은닉층 유닛의 개수와 토큰 말뭉치 크기를 달리했을 때 각 클래스별 문장 간 거리에 대한 결과 그래프이다. 각 클래스별 문장 간 거리는 $D(member,\ Non-member) \geq 1$을 만족함을 확인할 수 있다.

Fig. 10. Effect of model size and training corpus size on $D(member,\ Non-member)$

수식 (14)로부터 $D(S_C, \neg S_C) > 1$는 L2LMs 모델의 표현 공간에서 해당 클래스에 속하는 문장들은 그 클래스에 속하지 않는 문장들보다 서로 가깝게 표현된다는 것으로 해석할 수 있다.

주목할 점은 Fig. 10의 오른쪽 그림(All RCs)은 신경망 언어 모델과 학습 토큰의 크기와 무관하게 L2ALMs 모델의 문장 임베딩 표현에서 등위절 구조보다 관계절 구조가 서로 가깝게 표현된 것을 확인할 수 있다. Fig. 10의 왼쪽 그림(Specific RCs)은 동일한 관계절 구조를 갖는 문장 간 거리는 다른 클래스와 비교할 때 크게 표현된 것을 확인할 수 있다. 그러나 훈련 토큰의 개수가 증가할수록 문장 간 거리가 감소함을 확인할 수 있다.

위의 분석이 통계적으로 유의미한 결과인지를 검증하기 위해 혼합효과 모형 분석을 수행했다:

$$D(member, \neg memebr) \sim nhid^{*} csize + (1|adaptlist) \quad (15)$$

Table 9를 통해 첫 번째 클래스(Specific RCs)는 언어 모델 크기와 학습 토큰의 개수가 동일 관계절을 공유하는 문장 간 유사성에 통계적으로 유의미한 영향을 미치는 요인임을 확인할 수 있다. 두 번째 클래스(Reduced RC)는 언어 모델 크기만이 축약 관계절을 공유하는 문장 간 유사성에 통계적으로 유의미한 영향을 미치는 요인임을 확인할 수 있다. 세 번째 클래스(All RCs)는 학습 토큰의 개수가 모든 관계절 문장 간 유사성에 통계적으로 유의미한 영향을 미치는 요인임을 확인할 수 있다.

Table 9. Mixed-effects model with distance metric

Distance	Fixed-effect	Estimate	Std	p-value
D(RC, ¬RC)	nhid	-0.149	0.013	0.00***
	csize	-0.091	0.013	0.00***
	nhid:csize	0.021	0.013	0.123
D(Reduced, ¬Reduced)	nhid	-0.022	0.009	0.028*
	csize	0.013	0.009	0.190
	nhid:csize	0.000	0.010	0.925
D(RCs, ¬RCs)	nhid	-0.011	0.013	0.407
	csize	-0.035	0.013	0.010*
	nhid:csize	-0.032	0.013	0.021*

V. Conclusion

본 연구는 L2 영어 학습자로서 신경망 언어 모델의 표현 공간에서 영어 문장에 담겨진 언어적 속성을 추적할 수 있는지 평가하기 위해 통사 프라이밍-기반 프로빙 태스크를 적용했다. 프로빙 태스크 실험에서는 심리언어학에서 통사 구조 속성의 유사성 평가할 때 사용하는 통사 프라이밍과 사전 학습한 총 9개 L2 LSTM 언어 모델들을 기반으

로 관계절 문장들을 추가 적응 학습시킨 총 45개 L2 적응 언어 모델들을 사용하여 모델의 표현 공간에서 문장 간의 유사성을 통계적으로 규명했다.

본 연구에서 제안한 L2 LSTM 적응 언어 모델의 문장 임베딩 표현에 담고 있는 통사 구조 속성을 Prasad et al. 의 글로다바 언어 모델과 비교 분석했을 때 거의 동일한 양상을 공유함을 확인했다.

결과적으로 본 연구의 방법을 통하여 L2 LSTM 적응 언어 모델은 문장의 관계절 추상 통사 구조 속성을 추적할 수 있으며 특히 5종류의 관계절 문장들을 Fig. 1과 같이 계층적으로 재구성할 수 있음을 진단했다.

하지만 보다 더 근본적인 쟁점은 신경망 언어 모델이 문장 처리과정에서 반영하는 통사구조가 구문들 사이의 피상적인 구조에 기인하는 것인지 혹은 신경망 언어 모델이 탐침 하는 비명시적 내재적 속성에 기인하는 것인지의 문제이다. 본 연구가 이와 같은 쟁점을 연구하는 시발점이라는 점에서, 이 쟁점을 보다 심층적으로 연구하는 것은 후속 연구로 남겨둔다.

향후 연구계획은 본 연구의 진단 방법을 Bert나 GPT-2와 같은 트랜스포머 언어 모델에 적용하여 순환 신경망 LSTM 언어 모델의 연구 결과와 비교하고자 한다.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2020S1A5A2A01044957).

REFERENCES

[1] A. Vaswani et al., "Attention Is All You Need," June 2017, DOI: arXiv.1706.03762.

[2] S. Hochreiter and S. Jurgen, "Long short-term memory," Neural Computation, Vol. 9(8), pp. 1735-1780, Nov 1997.

[3] T. Linzen et al., "Assessing the ability of LSTMs to learn syntax-sensitive dependencies,", pp. 521-535, Nov 2016, DOI: arXiv.1611.01368.

[4] R. T. McCoy et al., "Revisiting the poverty of the stimulus: Hierarchical generalization without a hierarchical bias in recurrent neural networks," pp. 2093-2098, June 2018, DOI: arXiv.1802.09091.

[5] M. van Schijndel et al., "A neural model of adaptation in reading," pp. 4704-4710, Oct 2018, DOI: 10.18653/v1/D18-1499.

[6] G. Prasad et al., "Using priming to uncover the organization of syntactic representations in neural language models," pp. 66-76, Nov 2019, DOI: arXiv.1909.10579.

[7] K. Gulordava et al., "Colorless green recurrent networks dream hierarchically," Mar 2018, DOI: arXiv.1803.11138.

[8] H. P Branigan et al., "Syntactic priming across highly similar languages is not affected by language proficiency," Oct 2021, DOI: 10.1080/23273798.2021.1994620.

[9] J. Hewitt et al., "A Structural Probe for Finding Syntax in Word Representations," pp. 4129-4138, June 2019, DOI: 10.18653/v1/N19-1419.

[10] A. Conneau et al., "What you can cram into a single \$&!#* vector: Probing sentence embeddings for linguistic properties", June 2018, DOI: arXiv.1805.01070.

[11] J. B Wells et al., "Experience and sentence processing: Statistical learning and relative clause comprehension," Cognitive Psychology, 58, pp. 250-271, Mar 2009.

[12] Tianyu Gao et al., "Making Pre-trained Language Models Better Few-shot Learners," ACL, pp. 3816-3830, Aug 2021.

[13] J. R. Bellegarda, "An Overview of Statistical Language Model Adaptation," ITRW on Adaptation Methods for Speech Recognition, pp. 29-30, Aug 2001.

[14] Schick and Schütze, "It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners," NAACL, Apr 2021, DOI: arXiv.2009.07118.

[15] M. van Schijndel et al., "Quantity doesn't buy quality syntax with neural language models," Aug 2019, DOI: arXiv.1909.00111.

[16] E. Kim, "Sentence Comprehension with an LSTM Language Model," Journal of Digital Contents Society, Vol. 19(12), pp. 2393-2401, Dec 2018.

[17] E. Kim et al., L2ers' predictions of syntactic structure and reaction times during sentence processing. Linguistic Research 37, pp. 189-218, 2020.

[18] J. Hale et al., "Quantifying Structural and Non-structural Expectations in Relative Clause processing," Jan 2021, DOI: 10.1111/cogs.12927.

Authors

Euhee Kim received the M.S. degrees in Computer Engineering from Dongguk University, Korea, in 2002 and Ph.D. degrees in Mathematics from The University of Connecticut, U.S.A in 1995.

Euhee Kim is currently a Professor in the Department of Computer Science & Engineering at Shinhan University. She is interested in AI, NLP and Big Data computing.