Korean Journal of English Language and Linguistics
Vol. 20, December 2020, 881–903
10.15738/kjell.20..202012.881

코퍼스와 딥러닝 언어 모델을 활용한 문장 처리의 예측성과 행동 반응 시간과의 관계 연구

서혜진 (동국대학교)
신정아 (동국대학교)*

Seo, Hye-Jin and Jeong-Ah Shin 2020. Exploring the relationship between the predictability and the behavioral reaction time in sentence processing using corpus and deep-learning language models. *Korean Journal of English Language and Linguistics* 20, 881-903. This study examined whether the predictability is associated with the behavioral reaction times in sentence processing. The information complexity measures have been proposed to quantify the predictability for word-by-word human sentence processing. The most traditional information complexity measure is known as surprisal, which calculates relative unexpectedness at each word in a sentence (Hale 2001, Levy 2005, 2008). The most traditional information complexity measure is known as surprisal, which calculates relative unexpectedness at each word in a sentence (Hale 2001, Levy 2005, 2008), and some studies suggested that surprisal and reading times are positively correlated (Monsalve, Frank and Vigliocco 2012, Smith and Levy 2013). In order to calculate surprisal, the previous studies used one of two ways: Corpus based language models and deep learning based language models. This study, however, used both of them to analyze human reading times, comparing surprisal calculated from corpus-based language models with that calculated from deep-learning-based language models. Many studies partially investigated either of them. In this study, human reading times were analyzed by comparing surprisal calculated from corpus-based language models with that calculated from deep-learning-based language models. The results showed that surprisal calculated from corpus-based language models is more suitable to explain the behavioral reaction time data. Although the deep learning technology performs very well in the field of natural language processing, it does not seem to be human-like processing. Nonetheless, this study can contribute to the development of deep learning technology as well as computational psycholinguistic research in that it tried to compare the outcomes of corpus and deep learning technology with human behavioral responses.

Keywords: predictabililty, surprisal, corpus-based language model, deep-learning-based language model

* 제1저자: 서혜진; 교신저자: 신정아

1. 서론

사람들은 글을 읽을 때, 현재 주어진 정보를 토대로 다음에 어떤 단어가 나올지 예측하면서 읽는다. 심리언어학 연구에서는 여러 가지 실험을 활용하여 문맥 속 단어를 어떠한 방식으로 예측하면서 읽어나가는지 밝혀내고자 하였다(Clifton, Ferreira, Henderson, Inhoff, Liversedge, Reichle and Schotter 2016, Rayner 1998, Staub 2015). 문맥 속 어휘 예측성과 관련된 초기 실험 연구는 문장 속 목표 단어에 따라서 사람의 행동 반응이 어떻게 달라지는지를 관찰하였다. 다음 (1)의 문장은 문맥 속에서 예측이 쉬운 단어와 예측이 어려운 단어로 구성된 예이다.

(1) a. Since the wedding was today, the baker rushed the wedding cake to the reception.
b. Since the wedding was today, the baker rushed the wedding pies to the reception.

(1a)에서와 같이, 문맥 속에서 쉽게 예측 할 수 있는 단어 'cake'와 (1b)에서와 같이, 상대적으로 예측이 어려운 단어 'pies'가 주어졌을 때, 사람들은 예측이 쉬운 단어보다 예측이 어려운 단어를 더 자주, 더 오래 응시한다는 것이 안구 운동 추적 실험을 통해서 관찰되었다(Balota, Pollatsek and Rayner 1985, Choi, Lowder, Ferreira, Swaab and Henderson 2017). 또한, 이러한 효과는 뇌파측정 실험에서도 관찰되었는데, 의미적 불일치(semantic incongruity)일 때 나타나는 N400 성분이 예측이 어려운 단어보다 예측이 쉬운 단어에서 감소된 N400 효과(Reduced N400 effect)를 관찰할 수 있었다. 즉, 뇌파측정 실험으로도 예측이 쉬운 단어보다 예측이 어려운 단어에서 의미적 불일치가 더 컸다는 것을 확인할 수 있었다(Federmeier and Kutas 1999, Kutas and Hillyard 1984).

이러한 연구 결과는 인지 과학 분야에서 어휘 예측성이 사람의 인지정보처리(a cognitive information processing)에 어떠한 역할을 하는지 살펴볼 필요가 있다는 점을 시사한다(Clark 2013). 언어를 처리할 때, 사람의 뇌는 문맥 속 주어진 요소들을 활용하여 앞으로 전개될 어휘뿐만 아니라 문장의 통사적 요소까지도 미리 예측할 수 있다고 한다(Kuperberg and Jaeger 2016). 이때, 미리 예측했던 것과 실제로 전개되는 것이 일치한다면 문장 처리 촉진(processing facilitation) 효과가 나타나게 된다. 최근 심리언어학 분야에서는 문장 처리 촉진 효과와 통사적 구조 예측성의 상호관계를 연구하고 있다. 언어는 한정된 의미를 무한하게 표현할 수 있는 규칙이다(Chomsky 1965). 하지만 사람은 무한한 언어 표현을 모두 학습한 이후에나 언어를 활용할 수 있는 것이 아니라, 다양한 언어적 경험을 바탕으로 일반적인 언어적 지식을 구축해나간다. 따라서 언어를 이해할 때, 사람들은 현재까지 경험한 언어적 특징들을 토대로 비록 처음 접한

문장일지라도 이해하는 데 문제가 없으며, 문장 속 내포된 의미를 일반적으로 추론한다. 많은 연구에서 사람들은 인지적 자원(cognitive resource)인 의미, 통사, 화용의 언어 정보를 활용하며, 문장을 점진적으로 예측한다고 주장했다(Altmann and Kamide 1999, Federmeier 2007, Hale 2001, Levy 2008, Linzen and Jaeger 2015). 특히, 문장을 읽어나가며 현재까지 주어진 문장 속 정보를 토대로, 앞으로 이어질 문장의 통사적 구조를 '확률적으로' 미리 예측한다고 주장하였다.

구체적으로, Chomsky(1965)는 각 동사에 따라서 취할 수 있는 보어의 종류를 하위범주 자질(subcategorization feature)로 세분화하면서, 사람들이 이러한 동사의 하위범주 자질을 활용하여, 문장의 통사적 구조를 확률적으로 예측한다고 설명하였다. 예를 들어, 동사 *'accept'*는 명사구(Noun phrase; *accept a gift*)와 절 보충어(Sentential complement; *accept that you have lost*)를 보어로 취할 수 있다(그림 (1a)와 (1b)). 하지만 일상생활에서 동사 *'accept'*는 절 보충어보다는 명사구와 훨씬 더 빈도 높게 사용되기 때문에, 사람들이 동사 *'accept'*를 접했을 때 문장의 통사적 구조로 명사구가 뒤따를 것이라 예측한다(그림 (1a)). 좀 더 구체적으로, 사람들이 *'He accepted the proposal ...'*과 같은 문장을 읽어나가고 있는 중이라면, '*the proposal*'은 절 보충어의 주어로써 사용될 수도 있고 명사구로써도 사용될 수 있지만 일반적으로 '*the proposal*'을 명사구로 예측하며 문장을 이해해나간다(그림 (1c)). 이렇게 통사적 구조를 예측하면서 읽어나가다가 *He accepted the proposal was wrong*과 같이 문장이 이어진다면, 사람들은 재빠르게 문장의 통사적 구조를 다시 분석해야 한다(그림 (1d); Linzen and Jaeger 2016).

이러한 이유로, 절 보충어를 선호하는 동사(예, *prove*)보다 명사구를 더 선호하는 동사(예, *accept*)에 절 보충어가 이어진다면, *'was wrong'*과 같이 문장의 구조를 확실하게 파악할 수 있는 중의성이 해소되는 구역(disambiguating region)에서 문장 처리가 느려지는 경향이 발생하게 된다(Garnsey, Pearlmutter, Myers and Lotocky 1997, Trueswell, Tanenhaus and Kello 1993). 또한, 동사 *'forget'*은 동사 *'accept'*와 같이 절 보충어보다 명사구를 더 선호하지만 다양한 보어(*forgot about the party, forgot to buy groceries*)도 취할 수 있기 때문에, 각각의 동사를 접하고 앞으로 전개될 문장의 통사적 구조를 예측하는 것이 다를 것이다. 문장 처리 도중에 발생하는 통사적 구조의 예측 가능성(predictability)은 동사 *'forget'*보다 동사 *'accept'*가 더 높기 때문에 동사 '*accept*'보다 동사 '*forget*'이 포함된 문장을 처리하는 데 더 오랜 시간이 걸릴 가능성이 있다.

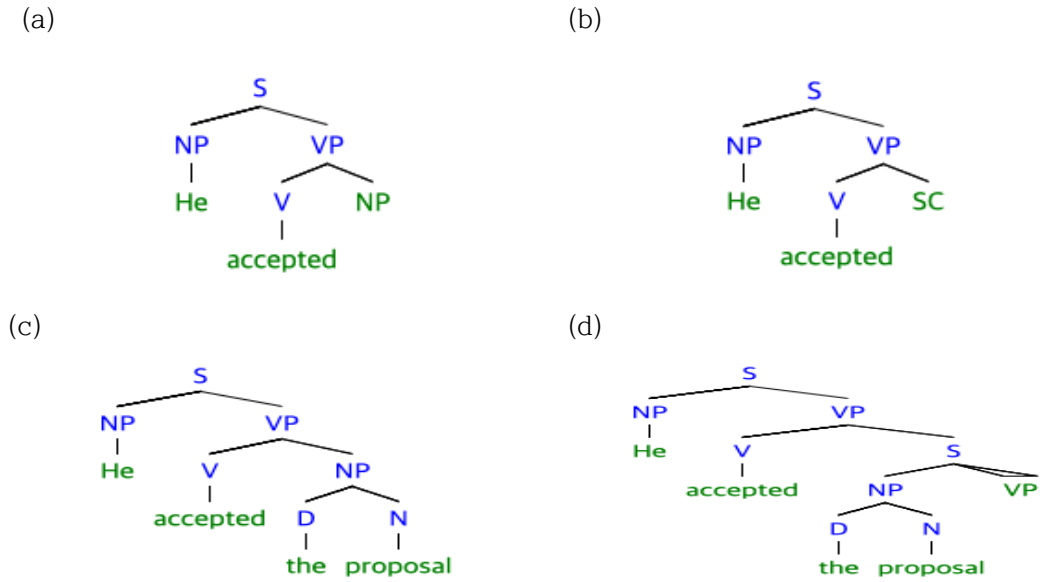


그림 1. *'he accepted the proposal was wrong'* 문장 처리의 실시간 점진적 예측 파서
(Linzen and Jaeger 2016)

이와 같이 문장의 예측성(predictability)과 문장 처리 난이도(processing difficulty)는 연관성이 있다(Boston, Hale, Kliegl, Patil and Vasishth 2008, Demberg and Keller 2008). 더 나아가 사람들이 앞으로 전개될 통사적 구조를 경험에 입각하여 확률적으로 예측하면서 문장을 처리한다는 것을 시사한다(DeLong, Urbach and Kutas 2005, Smith and Levy 2013).

오인 문장(garden-path sentence)은 사람들이 실시간 언어 처리를 할 때, 확률에 기반하여 가능성 있는 다양한 통사적 구조를 가정하는지, 아니면 가장 가능성 있는 하나의 통사적 구조만을 가정하는지를 모색하는 데 적합하기 때문에 심리언어학 분야에서 많이 다루어졌다. 오인 문장은 (2)와 같이 본동사와 축약 관계절(Reduced relative clause) 분석 사이에서 일시적인 구조적 중의성(temporary structural ambiguity)을 유발하는 문장이며, *'raced past the barn'* 구는 일시적인 중의성 구역에 해당한다. (2)와 같은 문장을 사람들이 읽어나가는 과정이라면 본동사인 *'fell'*을 접하기 전까지 *'raced past the barn'*을 본동사로도 분석할 수도 있고, 축약 관계절로 분석할 수도 있다. 일반적으로 사람들은 *'raced past the barn'*은 본동사로 분석하면서 문장 처리를 해나간다(그림 2a). 그 이후에 문장의 본동사인 *'fell'*을 접하게 된다면, 문장의 통사적 구조를 본동사 분석이 아닌 축약 관계절로 재분석하는 과정이 요구된다(그림 2b).

(2) The horse raced past the barn fell. (Bever 1970)

(a) 본동사 분석 (b) 축약 관계절 분석

그림 2. 통사 구조별 나무 그림 (Hale 2001)

본 연구에서는 오인 문장을 대상으로 크게 두 가지 방식으로 언어 처리 과정을 살펴보고자 한다. 첫 번째로, 다양한 통사적 구조를 확률로 변환한 확률적 문맥자유문법(PCFG, Probabilistic Context Free Grammar)으로 구축한 문장 처리 언어 모델(a language model)을 활용한다. 두 번째로는 딥러닝 기술로 구현한 다양한 언어 모델을 활용한다. 이러한 두 가지 방식으로 구현한 오인현상 문장의 언어 처리 방식과 실제 영어 원어민의 자기조절 읽기 시간(self-paced reading times)과의 연관 관계를 살펴보면서, 예측 모델들이 점진적으로 통사적 구조를 처리하는 사람의 언어처리 방식을 설명하는 데 활용될 수 있는지 살펴보고자 한다. 이러한 과정을 통해서, 사람의 언어처리 방식을 행동반응 시간뿐만 아니라 다양한 기재를 통해서보다 더 심도 있게 알아보고, 각 언어 모델이 사람의 언어 처리 방식과 얼마나 유사한지, 어떠한 한계점이 있는지도 자세히 보고자 한다.

2. 이론적 배경

2.1 문장 처리 계산 모델 써프라이절(Surprisal)

2.1.1 써프라이절의 이론적 정의

어휘 예측성과 관련된 선행 연구에서는 빈칸 채우기 예측성(cloze predictability)을 활용하면서, 문장 속 목표 단어 하나를 예측하기 쉬운 단어와 상대적으로 예측하기 어려운

단어로 구성하면서 문장 처리 방식의 본질을 파악하고자 하였다. 하지만 이러한 빈칸 채우기 실험은 실험 디자인에 따라서 매우 제한적으로 실험 문장들이 구성되므로, 사람들의 문장 처리 방식 본질을 파악하기에는 한계가 있다(Ferreira and Lowder 2016). 최근 연구 동향은 문장 속 목표 단어 하나가 아니라 문장의 모든 단어들의 예측성을 파악하는 문장처리 계산모델(computational models of sentence processing)을 활용하면서 사람들의 문장 처리 난이도를 설명하고자 한다(Futrell, Wilcox, Morita and Levy 2018, Hale 2006, Lizen and Jaeger 2016). 이러한 접근 방식은 기존의 정보 이론(Shannon 1948)과 최근 발전하고 있는 실시간(real time)으로 문장 속 각각 단어들의 정보 복합성(information complexity)을 추정하면서, 문장 처리 방식을 파악하는 계산 언어 모델(Hale 2001, Levy 2008)이 융합된 것이다.

연구가 가장 많이 된 문장 처리 계산 모델은 Hale(2001, 2016)이 제안한 써프라이절(surprisal)[1]이다. Hale(2001, 2016)은 문장 속 각각의 단어를 처리(process)하는 데 요구되는 인지 노력(cognitive effort)을 써프라이절로 측정 할 수 있다고 하였다. 써프라이절은 문맥 속 단어들의 상대적 예외성(unexpectedness)을 계산하며, 현재까지 주어진 문맥에 이어서 목표 단어(w_i)가 나타날 확률을 음의 로그 취한 것이다(수식1).

$$\text{Surprisal}(w_i) = -\log_2 P(w_i | w_1 \ldots w_{i-1})$$

수식 1. 써프라이절 수식

Hale(2001)은 문장을 실시간으로 처리할 때, 어휘 및 언어적 정보가 주어짐에 따라서 매순간 예측한 통사적 구문 분석(phrase-structural analyses)과 실제 주어진 언어적 정보가 얼마나 다른지를 점진적으로 계산하고자 하였다. 이러한 불일치(disconfirmation)는 맥락에 주어진 언어적 정보를 토대로 현재 주어진 어휘가 얼마나 나타나기 어려운지(빈도 낮은지)를 계산하는 것이 사람의 인지적 부하(cognitive load)를 반영한다고 가정한다. 문맥 속에서 빈도가 낮은 어휘는 써프라이절 값이 높고, 빈도가 높은 어휘의 경우엔 써프라이절 값이 낮다.

써프라이절은 주로 코퍼스를 이용하여 구현되었으나, 최근에는 딥러닝 기술을 활용하여 구현하고 있다. 다음 절부터는 코퍼스를 통해서 써프라이절을 구현하는 방법과 딥러닝 기술을 활용하여 구현하는 방법 및 특징들에 대해서 살펴보고자 한다.

[1] 써프라이절(Surprisal)은 정보 이론에서 섀넌의 정보량(Shannon information content)이라고 불리기도 한다.

2.1.2 코퍼스를 활용한 써프라이절 구현

써프라이절은 현재 단어에서 다음 단어로 이행할 때 발생하는 확률분포(probability distributions)와 관련이 있다. 이러한 분포를 우리는 흔히 언어 모델이라고 말한다. 언어 모델은 언어는 일종의 단어들의 집합(Chomsky 1956)이라고 가정하면서, 일련의 단어들에 확률을 부여한 것이다. 일반적으로 단어들의 나열에 확률을 부여하기 위해서 조건부 확률이 사용된다. 조건부 확률은 그 다음에 올 단어(w_i)의 확률은 n번째 이전부터 주어진 단어들(w_{i-1} w_{i-2}...$w_{i-(n-1)}$)과 연관이 있다고 말한다(수식2). Chomsky(1956)에서는 조건부 확률을 이용한 마르코프 모델(Markov model)[2]을 활용하여, 언어를 설명하고자 하였다. 하지만, 언어는 현재 주어진 단어와 가까운 단어뿐만 아니라 훨씬 더 이전에 제시된 단어와도 관련이 있을 수 있다. Chomsky는 사람들이 실제로 사용하는 주어-동사 수일치(subject-verb agreement) 및 대등접속구(coordinated phrase)와 같은 다양한 언어적 표현들을 마르코프 모델이 예측할 수 없다고 말한다. 문맥 속 의존(dependency) 관계를 설명하지 못하는 언어 모델은 자연 언어를 설명하는 데 한계가 있다(Hale 2006).

$$P(w_i | w_{i-1}\ w_{i-2}...w_{i-(n-1)})$$

수식 2. 조건부확률

이러한 언어 모델의 한계를 극복하기 위해서 Jelinek과 Lafferty(1991)는 확률문법(probabilistic grammar)을 제안하였으며, Hale(2001)과 Levy(2008a)는 코퍼스를 토대로 언어의 확률문법을 구현할 수 있다고 주장하였다. 확률문법이란 무한한 자연 언어를 유한한 확률적 규칙으로 정리한 것이며, 확률적 문맥자유문법(PCFG, Probabilistic Context Free Grammar)이라고 한다. 확률적 문맥자유문법은 코퍼스를 기반으로 추출한 계층적 구구조(a hierarchical phrase structure)를 확률로 계산함으로써, 현재 주어진 언어 정보를 토대로 앞으로 전개될 수 있는 통사적 구조들 및 실현 가능한 모든 통사적 구조들을 상대적 확률로 확인할 수 있다(Levy 2013). 이러한 확률적 문맥자유문법을 활용하여 써프라이절을 구한 것을 앞으로 코퍼스 언어 모델로 부르고, 그 예시는 <그림 3>과 같다. 구체적으로, 동사구(Verb Phrase)는 3가지 구조('VP → VBD PP', 'VP → VBN PP', 'VP → VBD')를 갖으며, 동사구가 'VP → VBN PP'로 상대적으로 빈도 높게 전개(약 75%)된다는 것을 파악 할 수 있다.

[2] 마르코프 모델은 어떠한 사건이 발생했을 때, 이전에 발생한 사건의 확률에 영향을 받는다는 확률 이론이다. Chomsky(1956)에서는 이러한 마르코프 모델을 활용하여, 언어를 설명하고자 하였다.

1.0	S	→	NP VP .
0.876404494831	NP	→	DT NN
0.123595505169	NP	→	NP VP
1.0	PP	→	IN NP
0.171428571172	VP	→	VBD PP
0.752380952552	VP	→	VBN PP
0.0761904762759	VP	→	VBD
1.0	DT	→	*the*
0.5	NN	→	*horse*
0.5	NN	→	*barn*
0.5	VBD	→	*fell*
0.5	VBD	→	*raced*
1.0	VBN	→	*raced*
1.0	IN	→	*past*

그림 3. 확률적 문맥 자유 문법 예시 (Hale 2001)

2.1.3 딥러닝 모델을 활용한 써프라이절 구현

최근 10년 동안 컴퓨터가 데이터를 토대로 스스로 학습하는 딥러닝 기술이 눈에 띄게 발달하면서 딥러닝 모델과 인간의 문장 처리 관계를 탐구하는 연구가 활발히 이루어지고 있다. 딥러닝 모델 중에서 가장 기본적인 순환신경망(Recurrent Neural Network; RNN) 모델은 자연언어 처리 분야뿐만 아니라 심리언어학 분야에서도 많이 활용되고 있다. 최근 연구에 따르면, 순환신경망이 사람의 문장 처리 방식을 반영하고 있는 실험 결과들이 보고되고 있다(Frank, Otten, Galli and Vigliocco 2019). 순환신경망은 사람의 읽기 시간(human reading times)과 읽기 실험 중 N400 뇌파 반응을 설명하기 위해서 활용되기도 했다(Monsalve et al. 2012, Goodkind and Bickneell 2018, Frank et al. 2015). 비교적 최근에 제시된 버트(BERT; Bidirectonal Encoder Representations from Transformers) 및 알버트(ALBERT; A Lite BERT for self-supervised learning of language representations) 모델은 다양한 자연언어 처리에서 순환신경망 모델보다 성능이 훨씬 좋다는 것이 입증되었다. 하지만 다양한 딥러닝 모델들이 사람의 언어 처리를 어느 정도 반영하는지에 대한 연구는 많이 진행되지 않았다.

딥러닝은 Frank Rosenblatt가 1957년 고안한 퍼셉트론(perceptron) 인공 신경망 알고리즘에서 시작되었는데, 퍼셉트론은 다양한 정보가 입력층(input layer)으로 들어가서 각각의 정보에 고유한 가중치(weight)를 부여하여 도출된 정보를 출력층(output layer)으로 보낸다. 딥러닝 학습은 입력층과 출력층 사이에 은닉층(hidden layer)이 2개 이상으로 구성되어있는 것이며, 입력층 및 출력층, 은닉층을 어떻게 구성하는지에 따라서 다양한 딥러닝 모델이 제시된다. 딥러닝 모델 중에서 가장 기본적인 형태인 순환신경망 모델을 먼저 살펴본다면, 순환신경망 모델은 데이터의 입력과 출력을 순차적인 시퀀스(sequence) 단위로 처리하며, 이전의 학습 정보를 현재 학습에 반영함으로써 시간/순서 정보까지 학습하는

모델이다. 좀 더 구체적으로 말하자면, 입력된 정보를 처리하여 은닉 상태(hidden state)에 저장한다. 그 다음에 뒤따르는 정보는 바로 직전의 층(the previous layer)에 저장된 은닉 상태의 정보와 현재 정보를 결합하여 현재 층(the current layer)의 은닉 상태에 저장한다. 각각의 층을 프로세스 하면서 시간/순서 정보까지 은닉 상태에 저장하기 때문에, 단어들 간의 순서가 있는 언어 데이터 처리를 할 때 유용하다. 이 모델은 입력된 정보를 시퀀스 단위로 바로바로 처리하기 때문에, 인간처럼(human-like) 굉장히 점진적인 언어 처리를 한다. 하지만 현재 입력된 정보를 바로 직전의 층에 저장된 은닉 상태의 정보와만 결합하며 처리(processing)하기 때문에, 시퀀스가 길어진다면 앞의 언어 정보가 뒤로 충분히 전달되지 않기 때문에 모델이 제대로 학습되지 않는 단점이 있다. 이러한 단점을 보완하기 위해서 제시된 모델은 순환신경망 모델의 일종인 장단기메모리 모델(LSTM; Long Short-Term Memory models)이다. 장단기메모리 모델은 전통적인 순환신경망 모델보다 은닉층을 더 세부적으로 나누었다. 장단기메모리 모델의 은닉층은 입력 게이트(input gate) 및 망각 게이트(forget gate), 출력 게이트(output gate)로 이루어져있다. 이렇게 함으로써 처리 도중에 불필요한 정보는 지우고, 기억해야할 정보들을 그 다음 처리까지 전달한다.

버트(BERT; Bidirectonal Encoder Representations from Transformers) 및 알버트(ALBERT; A Lite BERT for self-supervised learning of language representations) 모델은 비교적 최근에 제안된 딥러닝 모델이며, 위키피디아 및 북스코퍼스의 약 33억 개의 단어로 구성된 대량의 데이터를 사전 학습(pre-training)시킨 구글(Google)의 언어모델이다. 개인 연구자들이 이렇게 대량의 텍스트를 가지고 딥러닝 학습을 시키기 위해서는 상당한 지식을 필요로 하며, 학습시키기 위해서 엄청난 규모의 데이터 자료, 학습 파라미터 및 시간을 투자하여야 한다. 이러한 딥러닝 학습의 단점을 보완하기 위해서 개인 연구자들도 연구 목표에 따라서 손쉽게 파인튜닝(fine-turing)하여 활용할 수 있게 버트 모델이 고안되었다. 버트 모델은 문장 수준 임베딩(sentence embedding) 기법을 사용하면서 제시된 단어들의 정보뿐만 아니라 문맥의 정보까지도 활용 할 수 있는 양방향 모델이다(Devlin, Chang, Lee and Toutanova 2018). 임베딩은 자연어를 컴퓨터가 이해 할 수 있도록 단어나 문장을 벡터값으로 변환한다. 이렇게 단어 및 문장을 벡터값으로 표현함으로써, 컴퓨터가 사람처럼 단어 및 문장을 이해하고, 추론할 수 있게 된다. 임베딩은 대표적으로 단어 임베딩과 문장 임베딩이 있다. 2017년 이전에는 주로 단어 임베딩 기법을 활용하여 현재까지 제시된 단어들의 배열 정보를 가지고, 그 뒤에 뒤따를 단어가 무엇인지 예측하며 학습하는 방식이었다. 하지만 단어 임베딩 기법은 각 단어의 의미를 벡터로 표현할 수는 있지만 문맥에 따라서 달라지는 동음이의어를 구별하지 못하고 동일한 벡터값을 갖는다. 이러한 단점을 보완하기 위해서 문장 임베딩 기법이 제시 되었다. 문장 임베딩 기법은 문장 속 단어들을 토대로 각 단어의 의미를 추론하기 때문에, 동음이의어 단어들도 각각 문맥에 따라서 다르게 벡터값으로 변환될 수 있다. 현재 버트 모델을 토대로 다양한 언어 모델이 제시되고 있으며, 자연언어 처리 분야에서 상당히 좋은 성능을 내고 있다. 알버트는 버트

모델의 파라미터를 줄여서 학습 속도를 향상시킨 모델이다. 동일한 데이터 및 시간이 주어졌을 때, 데이터 학습 속도가 빠르고, 더 많은 학습 네트워크를 구성할 수 있기 때문에 버트라지 모델보다 성능이 더 좋다고 알려져 있다. 본 연구에서는 장단기메모리 및 버트, 알버트 모델을 다루며, 각 모델들과 사람의 문장 처리 연관성을 탐색해보고자 한다. 장단기메모리(LSTM) 모델은 Gulordava, Bojanowski, Grave, Linzen과 Baroni (2018)에서 영어 위키피디아의 약 9억 개의 단어로 구성된 데이터로 학습시킨 모델을 사용하고자 하며, 앞으로 글로다바 모델이라고 지칭하겠다. 또한, 버트베이스(bert-base) 모델, 버트라지(bert-large) 모델 및 알버트 모델을 사용하고자 한다. 버트라지 모델은 버트베이스 모델보다 학습을 더 많이 시켜서 단어 사이의 관계를 보다 더 잘 포착할 수 있다.

딥러닝을 활용하여 써프라이절을 구하는 수식은 <수식 3>과 같다. w_i는 현재 제시된 단어이고, h_{i-1}는 w_i로 수렴되기 전 은닉 상태(hidden state)이며, 확률은 소프트맥스 활성화함수(softmax activation)[3]를 사용하여 구한다.

$$\text{Surprisal}(w_i) = -\log_2 P(w_i | h_{i-1})$$

수식 3. 딥러닝 써프라이절 수식

2.1.4 코퍼스를 활용하는 방식과 딥러닝을 활용하는 방식의 장단점

코퍼스를 활용하여 써프라이절을 구현하는 방법은 직접적으로 전체 데이터 속 통사 구조의 확률을 사용하기 때문에, 언어 처리 과정에서 주어진 단어들에 따라 시시각각 변화하는 통사적 구조에 따른 인지적 부하를 간접적으로 유추할 수 있다. 하지만 코퍼스를 활용한 써프라이절을 구현하기 위해서는 대량의 언어 데이터가 필요하고, 연구 목적에 따라서 어떤 데이터를 사용할지를 선정해야 한다. 데이터 선정 후, 연구 목적에 따라서 데이터 핸들링(data handling) 작업이 필요하며, 연구 주제가 달라질 때마다 전체 코퍼스에서 확률적 문맥자유문법을 매번 새롭게 구축해야 한다[4]. 이러한 과정에서 상당한 시간과 자원이 필요하다. 반면에, 딥러닝 언어 모델을 활용하여 써프라이절을 구현하는 방법은 사전에

[3] 소프트맥스 활성화함수는 입력 받은 값을 0부터 1 사이의 값으로 모두 정규화하며, 이들의 총합을 1로 만들어주는 함수이다.

[4] 코퍼스를 토대로 구현하는 확률적 자유문맥규칙은 굉장히 많은 통사적 규칙으로 구성되어있다. 이러한 확률적 자유문맥규칙을 활용하여 써프라이절을 계산한다면, 컴퓨터는 거의 발생하지 않는 아주 희박한 통사적 규칙까지도 다 고려하기 때문에 간단한 문장의 써프라이절을 계산하는 데도 굉장히 많은 시간이 소요되며, 연산 오류가 발생할 가능성이 크다. 이러한 연유로, 코퍼스를 활용하여 써프라이절을 구할 때는 확률적 자유문맥규칙을 다소 간소화하여 아주 희박한 통사적 파생(syntactic derivation)은 배제한 후, 실현가능성이 있는 통사적 파생에 의거하여 써프라이절을 계산한다.

학습된 모델들을 가지고 써프라이절을 구하기 때문에 굉장히 빠른 시간 내에 구할 수 있다. 딥러닝 기술이 발달함에 따라서 언어학 연구의 중요성이 부각되고, 최근 언어 연구에서는 다양한 언어 현상을 딥러닝 모델 성능을 보다 더 정교하게 판단하는 척도로 사용하는 추세이다(Linzen 2019). 현재까지 주어-동사 일치(Linzen, Dupoux and Goldberg 2016), 오인 문장(Futrell, Wilcox, Morita and Levy 2018) 및 평서문-의문문(McCoy, Frank and Linzen 2018)과 같은 언어 현상에서 딥러닝 모델이 통사 구조 문법을 학습한 것과 같은 결과를 도출해냈다. 그러나 아직까지 딥러닝 모델이 실제로 통사적 지식까지 학습을 하는지는 뚜렷하게 밝혀지지 않았기 때문에 아직 더 많은 검증 절차가 필요하다.

2.2 선행 연구

최근 연구 중에는 순환신경망 모델이 계층 구조의 관계도 잘 처리 하는 것을 근거로 하여, 통사형태적인 지식까지 스스로 학습한다는 것을 보여주었다(Futrell et al. 2018, Linzen, Dupoux and Goldberg 2016, Gulordava, Bojanowsky, Grave, Linzen and Baroni 2018, Marvin and Linzen 2018). 특히, Futrell와 Wilcox, Morita, Levy(2018)는 순환신경망 모델이 자연언어 처리를 인간처럼(human-like) 점진적으로 처리하는지 살펴보기 위해서 다양한 오인 문장을 순환신경망 모델로 구한 써프라이절과 비교하여 분석하였다. 특히, (3)과 같은 전통적인 주절/생략관계절 중의성(main clause/reduced relative clause ambiguity) 문장을 탐구하면서 사람처럼 순환신경망 모델이 통사적 구조를 학습 할 수 있는지 알아보고자 하였다.

(3) a. The woman brought the sandwich from the kitchen tripped on the carpet.
b. The woman given the sandwich from the kitchen tripped on the carpet.
c. The woman who was brought the sandwich from the kitchen tripped on the carpet.
d. The woman who was given the sandwich from the tripped on the carpet.

(3a)에서는 초반에 'brought the sandwich from the kitchen'의 통사적 구조를 주절로 상정하다가 'tripped'를 처리하게 된다면, 통사적 구조를 다시 상정해야 된다. 이러한 오인현상은 (3b)와 같이 분사와 과거 동사의 형태가 다른 경우나 (3c)와 (3d)처럼 관계대명사 'who was'가 생략되지 않은 경우에서는 일어나지 않는다. (3)과 같은 문장에서의 중요한 구역(critical region)은 'tripped'와 같은 중의성 해소 구역(disambiguating region)이다. Futrell, et al. (2018)은 만약 (3a)의 중의성 해소 구역에서 다른 문장들보다 써프라이절이 높다면, 순환신경망 모델이 선호하는 통사적 분석이 있으며, 사람처럼 점진적인 자연언어 처리를 한다는 것을 내포한다고 가정하였다.

분석 결과, 사람처럼 순환신경망 모델의 써프라이절은 중의성 해소 구역에서 중의적인

분사 형태(brought)가 쓰인 문장이 중의적이지 않은 분사 형태(given)가 쓰인 문장보다 높았다(3a vs. 3b). 즉, 순환신경망 모델이 분사의 형태에 따라서 통사적 구조를 다르게 취하면서 자연언어 처리를 한다는 것을 의미한다. 하지만, 중의적이지 않은 분사가 쓰인 (3b)와 (3d)를 비교했을 때, 써프라이절이 'who was'가 생략되지 않았을 때보다 생략되었을 때 더 높았기 때문에 순환신경망 모델이 완벽하게 분사와 본동사(main verb)를 학습하지는 못한다고 말했다. Hale(2001)과 Levy(2008)를 포함한 다수의 연구에서 사람의 언어 처리 과정의 예측성과 관련된 써프라이절이 어떻게 심리언어학적인 언어 현상을 설명할 수 있는지 이론을 고안해냈지만, 실질적으로 써프라이절과 사람의 읽기 시간 데이터를 직접적으로 비교·분석하지 않았다.

이제부터는 써프라이절과 사람의 읽기 시간을 비교한 연구를 살펴보고자 한다. Linzen과 Jaeger(2016)는 펜트리뱅크(Penn Treebank) 코퍼스의 확률적 문맥자유문법으로 구한 써프라이절과 영어 원어민의 자기조절읽기 시간을 비교·분석하였다. 펜트리뱅크 코퍼스는 약 450만 영단어로 구성된 굉장히 큰 코퍼스이다(Marcus, Santorini and Marcinkiewick 1983). 의미적·어휘적 처리보다는 통사적 처리에 더 초점을 두기 위해서 분사와 동사를 제외한 모든 다른 어휘들은 품사 태깅(POS tagging; Part of Speech tagging)하였다. 이렇게 처리한 코퍼스를 바탕으로 확률적 문맥자유문법을 구축하였다. 이러한 방식을 취함으로써 각 동사 하위범주화의 특징에 따른 통사 구조 예측성을 더 잘 살펴볼 수 있다. 품사 태깅된 확률적 문맥자유문법으로 계산한 목표 문장의 단어별 써프라이절과 자기조절 읽기 시간을 비교하면서, 문장의 예측성과 사람의 행동반응시간의 관계성을 탐구하였다. 목표 문장(target sentence)은 총 32개의 세트(set)로 구성되었으며, 하나의 세트는 중의적 문장(ambiguous sentence)과 비중의적 문장(unambiguous sentence)으로 구성되어 있다. 보문소(complementizer) '*that*'이 생략되어있는 중의성 문장은 '*had been invaded*'를 처리할 때 문장의 통사적 구조를 재분석하는 과정이 필요하다.

(4) 문장 조건

a. The men discovered the island had been invaded by the enemy. (중의성 문장)

b. The men discovered that the island had been invaded by the enemy. (비중의성 문장)

목표 문장은 표 1과 같이 총 5개의 구역으로 나누었으며, 나머지 구역을 제외한 주어 및 동사, 중의성, 중의성 해소 구역을 통계 분석하였다. 그 결과, 중의성 해소 구역에서 써프라이절 효과가 있었으며, 써프라이절이 높아질수록 읽기 시간이 길어지는 통계적으로 유의미한 효과(significant effect)가 있었다(p = .03). 또한 써프라이절과 목표문장 조건의 상호작용에서 근소한 효과(marginal significant effect)가 있었다(p = .06). 비중의적 문장에서는 유의미한 써프라이절 효과가 없었지만(p > .2), 중의적 문장에서는 써프라이절

효과가 있었다(p = .007). 이러한 결과를 토대로, 써프라이절은 사람의 언어 현상을 설명하는 데 유용한 기재가 될 수 있다고 하였다.

표 1. 목표 문장의 단어별 구역

문장	The men	discovered	(that)	the island	had been invaded	by the enemy.
구역	주어	동사		중의성	중의성 해소	나머지

Frank(2009)는 간단한 순환신경망 모델(SRN; simple recurrent network)과 코퍼스 확률적 문맥자유문법에서 구현한 써프라이절을 읽기 시간과 비교·분석하였다. 던디 코퍼스(Dundee corpus; Kennedy, Hill and Pynte 2003, Kennedy et al. 2013)는 심리 언어학 분야에서 신뢰받는 코퍼스로 여겨지고, 안구운동 추적 실험(eye-tracking experiments) 결과와 함께 구성 되어있다. Frank(2009)는 던디 코퍼스에 구성된 안구운동 추적 실험 읽기 시간과 두 가지 종류의 써프라이절을 분석하면서, 써프라이절과 사람의 읽기 시간과의 상응관계를 알아보고자 하였다(Barrett, Agić and Søgaard 2015). 분석의 결과, 써프라이절과 읽기 시간은 굉장히 유의미한 양의 상관관계가 있다고 밝혔다. 간단한 순환신경망 모델과 코퍼스 언어모델로 구현한 써프라이절을 비교했을 때, 코퍼스 언어모델보다는 간단한 순환신경망 모델의 써프라이절이 읽기 시간과 더 강한 양의 상관관계가 있다고 하였다.

써프라이절과 읽기 시간과의 관계에 대하여, Kim, Park과 Seo(2020)에서는 원어민처럼 한국인 영어 학습자들이 동사 하위 범주화에 따라서 뒤따르는 통사 구조를 예측 할 수 있는지 코퍼스에서 추출한 써프라이절을 토대로 알아보았다. Linzen과 Jaeger(2016)의 실험 방식 및 실험 문장을 차용하여, 한국인 영어 학습자들을 대상으로 자기조절 읽기 실험을 진행하였고, 한국 중·고등학교에서 사용되는 영어 교과서 코퍼스를 활용하여 써프라이절을 계산하였다. 그 결과, 원어민들과 동일하게 중의성 해소 구역에서 유의미한 써프라이절 효과가 있었지만, 써프라이절과 문장 조건의 상호작용에서는 유의미한 효과가 없었다. 이러한 결과는 문장의 예측성과 관련된 써프라이절이 한국인 영어 학습자들의 행동 반응을 설명하는 데 좋은 기재가 될 수 있다는 것을 보여준다.

3. 연구 방법

본 연구에서는 Linzen과 Jaeger(2016)에서 128명의 영어 원어민을 대상으로 자기조절 읽기 행동반응을 실험한 데이터를 다양한 언어모델로 계산한 써프라이절을 토대로 재분석하였다. 펜트리뱅크 코퍼스에서 구축한 확률적 문맥자유문법을 활용한 방법과 딥러닝 테크닉을 이용하여 써프라이절을 계산하였다[5].

3.1 목표 문장(target sentence)

목표 문장은 Linzen과 Jaeger(2016)에서 사용한 32개 세트(set)이다. 한 세트당 중의적 문장('*the men discovered the island had been invaded by the enemy*')과 비중의적 문장('*the men discovered that the island had been invaded by the enemy*')으로 구성하였으며, 동사 및 중의성, 중의성 해소 구역을 통계 분석하였다. 특히, 중의적 문장에서 '*had been invaded*'를 처리하면서 재분석 과정이 요구되는 중의성 해소 구역을 중요한 구역으로 상정하였다.

3.2 실험 데이터 전처리 방법

Linzen과 Jaeger(2016)의 데이터 분석 방식을 최대한 따르며 재분석하였다. 한 단어의 읽기 시간(raw reading times)이 100ms 이하이거나 3,000ms 이상인 단어는 제거하였고, 모든 읽기 시간은 왜도(skewness)를 줄이기 위해서 로그값으로 변환(log-transform)[6] 하였다(Baayen and Milin 2010, Frank 2013). 로그로 변환한 행동반응시간의 평균으로부터 ±3 표준편차(sd; standard deviation)를 벗어난 값은 제거하였다. 그 이후에, 단어 길이에 따라서 읽기 시간에 영향을 줄 수 있기 때문에 로그로 변환한 읽기 시간의 잔차(residuals)를 구하였다(Jaeger 2009). 독립변수로써는 코퍼스 계산모델, 글로다바 모델, 버트베이스 모델, 버트라지 모델, 알버트 모델의 써프라이절이며, 모든 독립변수를 중심화(mean centering) 처리를 하였다(Jaeger 2009). 각 구역별 읽기 시간의 합에 로그로 변환한 읽기 시간 잔차를 선형 혼합효과 모형(Linear Mixed Effect Model)[7]의 종속변수로, 각 구역별 모든 모델들의 써프라이절의 평균값을 독립 변수로 설정하였다. 추가적으로, 문장의 조건(중의적 문장 vs. 비중의적 문장)이 독립 변수로 들어갔다.

---

[5] 감사하게도 Linzen과 Jaeger(2016)의 자기조절읽기시간 데이터 및 펜트리뱅크로 계산한 써프라이절 값을 저자들이 공유해주어서 재분석을 진행할 수 있었다.

[6] 많은 심리언어학 분야 읽기 시간 연구에서 영어 모국어 화자들의 경우 읽기 시간이 100ms 미만 또는 3,000ms 초과한 읽기 시간은 분석에서 제외한다. 반면에, 영어 학습자 화자들의 경우, 읽기 시간이 200ms 미만 또는 2,000ms 초과한 읽기 시간은 분석에서 제외한다. 이러한 과정들은 읽기 시간의 이상치(outlier)를 제거하고, 읽기 시간을 정규분포에 가깝게 만들기 위한 데이터 변형(transformation)에 해당한다(Shin 2019).

[7] 심리언어학 분야에서 선형 혼합효과 모형은 피험자(participant)별 평균 및 항목(item)별 평균, 언어 실험에서 조절(control)하는 다양한 독립 변수(independent variable)별 평균을 이용하여 복합적으로 언어 데이터를 분석하는 방법이다(Baayen, Davidson and Bates 2008). Clark(1973)이 언어를 고정효과(fixed-effect)로 분석했을 때의 오류와 관련된 논문을 발표한 뒤, 심리언어학 분야 실험 연구들은 선형 혼합효과 모형을 사용하며, 피험자 분석(participant analysis) 뿐만 아니라 항목 분석(item analysis)도 병행하게 되었다(Shin 2019).

4. 연구 결과

4.1 써프라이절 비교

<그림 4>와 <그림 5>는 코퍼스 및 딥러닝 모델(글로다바 및 버트베이스, 버트라지, 알버트)로 32개의 목표 문장들에서 구한 써프라이절의 조건별 평균 그래프이다. 다른 모델들과의 비교를 위해서 구역(region)별로 합한 써프라이절 값을 표준화하였다[8]. 문장의 정확한 구조를 파악할 수 있는 중의성 해소 구역에서 코퍼스 계산 모델과 글로다바 모델에서 비중의적 문장과 중의적 문장에서 써프라이절 값이 상당히 높은 것을 확인할 수 있었다. 버트베이스와 버트라지 모델에서는 비중의적 문장보다 중의적 문장에서 써프라이절 값이 상대적으로 더 높았지만 차이가 크지는 않았다. 반면에, 알버트 모델에서는 중의적 문장보다 비중의적 문장에서 써프라이절 값이 상당히 높은 것을 확인할 수 있었다. <그림 6>은 Linzen과 Jaeger(2016)의 자기조절 읽기 실험에서의 단어별 평균 읽기 시간을 그래프로 그린 것이다. 이 그래프로 확인할 수 있듯이, 일반적으로 사람들은 비중의적 문장보다 중의적 문장의 중의성 해소 구역에서 읽기 시간이 더 길다. 통계적으로 더 확실하게 비교를 해봐야하지만, 알버트 모델에서는 비중의적 문장에서 써프라이절이 더 컸기 때문에 사람의 언어 처리 과정과는 상당히 달라 보인다. 즉, 자연언어 처리 과정에서 더 좋은 모델이 딥러닝 처리의 퍼포먼스(performance)는 더 높을지 모를지라도 사람의 언어처리 과정과는 상이할 수 있다.

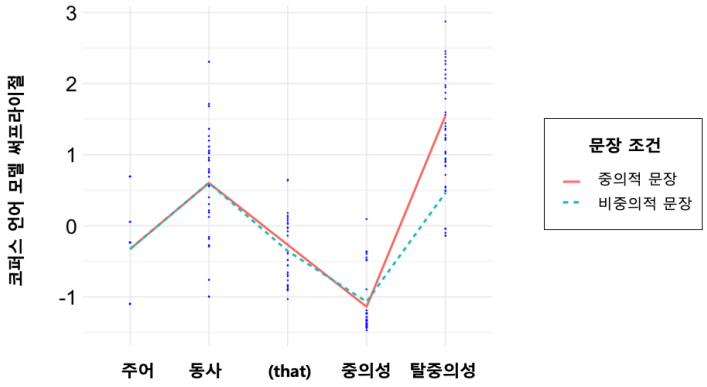
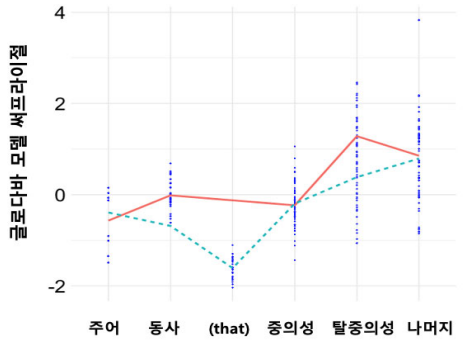


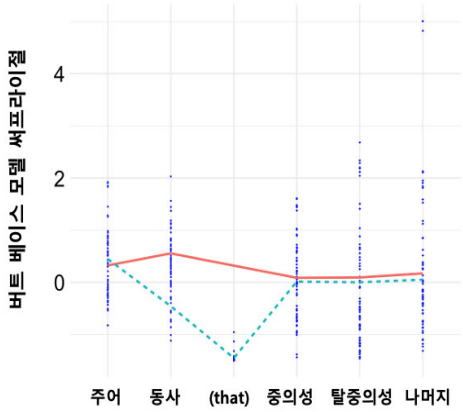
그림 4. 구역별 코퍼스 언어모델 써프라이절

[8] 딥러닝 모델과 달리 코퍼스를 활용하여 구한 써프라이절의 나머지 구역의 값은 없다. 한 문장의 써프라이절을 계산하는데 상당한 시간(약 50분)이 걸려서 통계 분석을 하지 않는 나머지 구역의 써프라이절 값은 구현하지 않았을 것이라 추측한다.

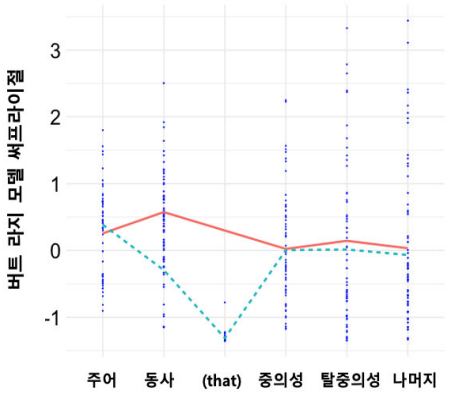
(a) 글로다바 모델



(b) 버트베이스 모델



(c) 버트라지 모델



(d) 알버트 모델

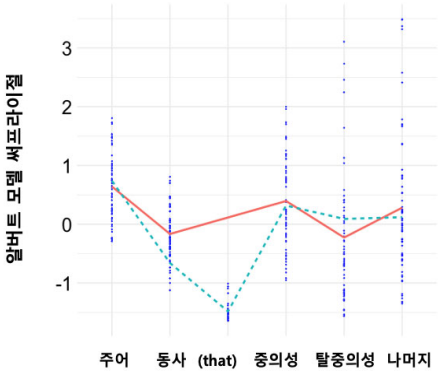


그림 5. 구역별 딥러닝 언어모델 써프라이절

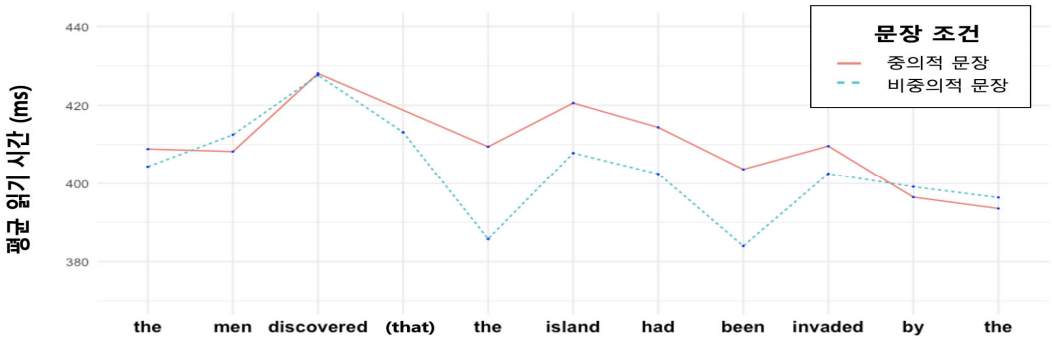


그림 6. 단어별 평균 읽기 시간

4.2 구역별 실험 결과

문장의 각 구역별 실험 결과는 다음과 같다.

1) 동사 구역: 코퍼스를 활용한 써프라이절 혼합효과모형에서만 유의미한 효과가 나타났다. 하지만 써프라이절 가정과는 달리 써프라이절 값이 커질수록 읽기 시간이 감소하였다(β = −0.051, SE = 0.02, t = −2.341, p < .05).

2) 중의성 구역: 코퍼스를 활용한 써프라이절 혼합효과모형을 제외한 딥러닝을 활용한 써프라이절 혼합효과모형에서 문장 조건별 효과가 나타났다. 비중의적 문장보다 중의적 문장일 때 읽기 시간이 더 오래 소요되었다(코퍼스 모델: β = −0.031, SE = 0.035, t = −0.898, p = .369; 글로다바 모델: β = −0.055, SE = 0.009, t = −5.966, p < .001; 버트베이스 모델: β = −0.046, SE = 0.007, t = −6.288, p < .001; 버트라지 모델: β = −0.042, SE = 0.009, t = −4.471, p < .001; 알버트 모델: β = −0.050, SE = 0.007, t = −6.383, p < .001)

3) 중의성 해소 구역: 버트베이스, 버트라지와 알버트 모델에서 문장 조건별 효과가 나타났다. 비중의적 문장보다 중의적 문장을 읽을 때 더 오랜 시간이 걸렸다(버트베이스 모델: β = −0.030, SE = 0.013, t = −2.327, p < .05; 버트라지 모델: β = −0.039, SE = 0.011, t = −3.584, p < .001; 알버트 모델: β = −0.027, SE = 0.009, t = −2.840, p < .01). 코퍼스 모델에서는 써프라이절 효과가 나타났으며, 써프라이절 값이 커질수록 읽기 시간이 늘어났다(β = −0.147, SE = 0.056, t = 2.620, p < .01). 하지만 딥러닝 모델에서는 유의미한 효과가 나타나지 않았다(글로다바 모델: β = −0.023, SE = 0.023, t = 0.977, p = .328; 버트베이스 모델: β = 0.025, SE = 0.023, t = 1.080, p = .281; 버트라지 모델: β = −0.015, SE = 0.015, t = −0.996, p = .319; 알버트 모델: β = 0.018, SE = 0.015, t = 1.240, p = .215).

다양한 언어 모델들이 사람의 언어 처리 과정을 모색 할 수 있는지 살펴본 결과, 통사적 재분석이 필요한 중의성 해소 구역에서 코퍼스 언어 모델에서만 써프라이절과 읽기 시간이 양의 상관관계를 갖는 것을 확인할 수 있었다.

5. 논의 및 결론

심리언어학 연구에서 써프라이절과 읽기 시간(reading times)이 양의 상관관계를 갖는다는 몇몇 연구가 있었지만(e.g., Monsalve et al. 2012, Smith and Levy 2013), 아직까지 많은 연구가 진행되지 않았으며, 코퍼스를 활용한 언어모델과 다양한 딥러닝을

활용한 언어모델을 통합적으로 비교·분석한 연구는 현재까지 진행되지 않았다. 본 연구에서는 동일한 실험 데이터를 가지고 코퍼스와 딥러닝 언어모델을 비교·분석 하면서 어떠한 언어 모델이 가장 사람의 언어 처리 방식을 잘 반영할 수 있는지 살펴보고자 하였다. 특히, 사람들이 문장을 읽어나가는 과정에서의 요구되는 인지적 노력을 써프라이절이란 값으로 측정하면서 문맥 속 단어들의 상대적인 예외성 또는 예측성을 언어 모델을 활용하여 구현하였다. 구체적으로, 영어 원어민들이 오인 문장을 처리하는 과정을 코퍼스 언어 모델이 예측한 써프라이절과 딥러닝 언어 모델이 예측한 써프라이절을 가지고 비교·분석하고자 하였다. 그 결과, 코퍼스를 활용한 써프라이절과 읽기 시간이 양의 상관관계를 갖는다는 것을 확인할 수 있었다. 특히, 오인 문장에서 중의적인 통사적 구조가 완전히 해소되는 중의성 해소 구역에서 이러한 현상을 살펴볼 수 있었다. 코퍼스를 활용한 써프라이절은 문장의 계층구조가 반영된 값인 반면에, 딥러닝을 활용한 써프라이절은 문장 속 단어들의 선형관계(linear relation)에서 스스로 계층구조까지 어느 정도 학습을 한다고 알려져 있다. 최근 딥러닝 기술이 비약적으로 발달하면서, 자연언어를 상당히 잘 처리하는 것이 입증되고 있지만, 아직까지 딥러닝 언어 모델이 인간처럼 언어를 처리한다고 주장하기에는 한계점이 보인다. 반면에, 언어적 정보가 보다 풍부하게 내포되어 있는 코퍼스 언어 모델은 어느 정도 인간의 언어 처리 과정을 반영하고 있는 듯하다.

현재 딥러닝 기술이 매우 빠른 기간에 비약적으로 발달되고 있지만, 딥러닝 기술 발달의 한계점이 있을 것이라 예측된다. 이와 같은 한계점을 해결하기 위해서는 사람의 언어 처리 과정과 딥러닝 자연언어 처리 과정을 면밀히 살펴보는 과정이 동반되어야 한다고 생각한다. 딥러닝은 파스 트리(parse tree)나 언어학 논리 구조식(logical formulas)과 같은 언어학적 자질을 바탕으로 한 입력 데이터(input data)가 아닌 대량의 미가공 데이터(raw data)를 통해서 학습이 이루어진다. 그럼에도 불구하고 딥러닝은 통사·형태적 및 의미론적으로 상당히 높은 정확도를 보였다(Radford, Wu, Child, Luan, Amodei, and Sutskever 2019). 이러한 현상은 언어학자들에게 '과연 언어 습득의 필수적인 메카니즘은 무엇일까?'와 같은 언어 습득과 연관된 주제를 제시해준다(Lasnik and Lidz 2017). Gulordava et al. (2018)에서는 장단기 메모리(LSTM) 모델이 주어/동사의 장거리 일치(long-distance agreement)를 확인해보았다. 장단기 메모리 모델은 *'the length of the forewings is...'*와 *'the length of the forewings are...'*를 비교했을 때, *'are'*이 포함된 문장보다 *'is'*가 포함된 문장의 확률이 더 높다고 잘 예측하였다. 또한 문법적으로 오류는 없지만 의미상으로 믿기 어려운(implausible) 문장인 *'the colorless green ideas near the duck (are/*is)...'*를 제시했을 때도 딥러닝 모델은 (예측성이 약간은 떨어졌지만) 잘 예측했다. 이러한 연구는 통사·의미적인 언어학적 자질 없이도 단어의 예측성을 학습시킴으로써 딥러닝 모델에 언어학적인 자질을 학습 시킬 수 있다는 점을 보여준다. Linzen과 Baroni(2020)에서는 딥러닝 모델의 성능을 다양한 언어학적인 현상을 토대로 검증하는 방법론은 이제 막 시작된 단계이며, 아직까지 딥러닝이 예측하는 것을 가지고 언어 현상을 타당하게 검증할 수는

없다고 하였다. 하지만, 이러한 방법론은 언어학 분야와 인지 과학 분야에 다양한 연구 주제를 제안해주며, 미래의 연구 분야가 될 것이라고 했다.

향후 과제로, 오인 문장 이외의 보다 더 다양한 언어 현상에서의 영어 원어민들의 행동 반응 시간과 코퍼스를 활용한 써프라이절, 딥러닝을 활용한 써프라이절을 비교·분석할 필요가 있다. 또한, 영어 원어민뿐만 아니라 한국인 영어 학습자들의 행동 반응 시간과 다양한 써프라이절을 비교·분석하면서, 영어 학습자들의 언어 처리 방식도 모색하고, 영어 원어민의 언어 처리 방식과 영어 학습자들의 언어 처리 방식의 차이점과 공통점을 알아보는 것이 필요할 것이다. 이러한 복합적인 언어 실험을 수행함으로써, 사람들이 어떻게 언어를 구사하고, 어떠한 과정을 거쳐서 언어를 습득하는지 알아볼 수 있을 것이라 기대한다.

사람은 문장을 예측하면서 이해해나간다. 예측하기 쉬운 단어는 굉장히 빨리 읽고, (예측하기 어려운 단어에 비해서 상대적으로) 주의 깊게 읽지 않는다. 반면에, 예측하기 어려운 단어는 다소 천천히 읽고, 해당 단어를 여러 번 읽는 경우가 많다. 이와 같이, 예측성은 사람의 언어 처리를 이해하는 데 굉장히 중요한 요소이다. 이러한 예측성을 토대로 구현하는 코퍼스 언어 모델 및 딥러닝 언어 모델을 연구하는 과정은 사람의 언어를 좀 더 과학적으로 입증할 수 있는 방안 중에 하나일 것이다. 만약 특정 언어 모델이 사람의 행동 반응을 전반적으로 잘 예측한다면, 우리는 해당 언어 모델이 잘 예측하지 못하는 사람의 행동 반응도 살펴볼 필요가 있다. 이와 같이, 사람의 행동 반응과 언어 모델이 예측하는 것 사이의 차이점을 연구하면서, 언어 모델 개발의 방향성을 제시할 수 있다(Linzen 2019). 실제 사람의 언어 처리 과정과 다양한 언어 모델들을 비교·분석하면서 심리언어학 분야에서는 사람의 언어 처리 방식을 좀 더 깊게 이해하고, 전산언어학 분야에서는 보다 더 나은 언어 모델 개발에 기여할 수 있을 것이라 기대한다.

References

신정아(Shin, J.). 2019. 혼합효과모형(Mixed-Effects Model)을 이용한 실험언어학 데이터 분석 방법 고찰: 자기조절읽기 실험 데이터를 중심으로(How to analyze experimental linguistic data using a mixed-effects model in R: Focusing on data from a self-paced reading experiment). ≪언어학≫ (*Korean Journal of Journal of English Language and Linguistics*) 19(1), 76-94.

Altmann, G. T. and Y. Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3), 247-264.

Baayen, R. H., D. J. Davidson and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390-412.

Barrett, M., Ž. Agić and A. Søgaard. 2015, January. The dundee treebank. In *The 14th International Workshop on Treebanks and Linguistic Theories (TLT 14)*.

Marcus, M., B. Santorini and M. A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19, 313-330.

Balota, D. A., A. Pollatsek and Rayner, K. 1985. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology* 17, 364−390.

Baayen, R. H. and P. Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2), 12−28.

Bever, T. G. 1970. The cognitive basis for linguistic structures. *Cognition and the Development of Language* 279, 1−61.

Boston, M., J. Hale, R. Kliegl, U. Patil and S. Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2(1), 1−12.

Chomsky, N. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2(3), 113−124.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge, MA: MITPress,

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3), 181−204.

Clifton, C. Jr., F. Ferreira, J. M. Henderson, A. W. Inhoff, S. P. Liversedge, E. D. Reichle and E. R. Schotter. 2016. Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language* 86, 1−19.

DeLong, K. A., T. P. Urbach and M. Kutas. 2005. Probabilistic word pre−activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8), 1117−1121.

Devlin, J., M−W. Chang, K. Lee and K. Toutanova. 2018. Bert: Pre−training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*

Federmeier, K. D. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44(4), 491−505.

Federmeier, K. D. and M. Kutas. 1999. A rose by any other name: Long−term memory structure and sentence processing. *Journal of Memory and Language* 41, 469−495.

Frank, S. 2009. Surprisal−based comparison between a symbolic and a connectionist

model of sentence processing. In *Proceedings of the annual meeting of the Cognitive Science Society* 31, 1139-1144.

Frank, S. L. 2013. Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science* 5(3), 475-494.

Frank, S. L., L. J. Otten, G. Galli and G. Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language* 140, 1-11.

Futrell, R., E. Wilcox, T. Morita and R. Levy. 2018. RNNs as psycholinguistic subjects: Syntactic state and grammatical dependency. *arXiv preprint arXiv:1809. 01329*.

Garnsey, S., N. Pearlmutter, E. Myers and M. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37(1), 58-93.

Goodkind, A., and K. Bicknell. 2018, January. Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th workshop on cognitive modeling and computational linguistics CMCL 2018,* 10-18.

Gulordava, K., P. Bojanowski, E. Grave, T. Linzen and M. Baroni. 2018. Colorless green recurrent networks dream hierarchically. *arXiv preprint arXiv:1803.11138*

Hale, J. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second of the North American Chapter of the Association for Computational Linguistics on Language*, 1-8, Stroudsburg, PA: Association for Computational Linguistics.

Hale, J. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30(4), 643-672.

Jaeger, F. 2009. Centering several variables. Retrieved from https://hlplab.wordpress.com/2009/04/27/centering-several-variables

Jelinek, F. and J. Lafferty. 1991. Computation of the probability of initial substring generation by stochastic context-free grammars. *Computational Linguistics* 17(3), 315-353.

Kennedy, A., R. Hill and J. Pynte. 2003. *The Dundee Corpus.* Paper presented at the 12th European Conference on Eye Movement, Dundee, Scotland.

Kennedy, A., J. Pynte, W. S. Murray and S.-A. Paul. 2013. Frequencyand predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology* 66, 601-618.

Kim, E., M-K. Park and H-J. Seo. 2020. L2ers' predictions of syntactic structure and reaction times during sentence processing. *Linguistic Research* 37, 189-218.

Kuperberg, G. R. and T. F. Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31, 32-59.

Kutas, M. and S. A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161-163.

Lasnik H, Lidz J. 2017. The argument from the poverty of the stimulus. In I Roberts, ed. *Oxford Handbook of Universal Grammar*, 221-248. Oxford: Oxford University Press.

Levy, R. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3), 1126-1177.

Linzen, T., & M. Baroni. 2020. Syntactic Structure from Deep Learning. *arXiv preprint arXiv:2004.10827.*

Linzen, T., E. Dupoux and Y. Goldberg. 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics* 4, 521-535.

Linzen, T. and T. F. Jaeger. 2016. Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science* 40(6), 1382-1411.

Lowder, M. W., W-I. Choi, F. Ferreira and J. M. Henderson. 2018. Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive science* 42, 1166-1183.

Marcus, M. P., B. Santorini and M. A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The PennTreebank. *Computational Linguistics* 19(2), 313-330.

Marvin, R. and T. Linzen. 2018. Targeted syntactic evaluation of language models. *arXiv preprint arXiv:1808.09031*

McCoy, R. T., R. Frank and T. Linzen. 2018. Revisiting the poverty of the stimulus: Hierarchical generalization without a hierarchical bias in recurrent neural networks. *arXiv preprint arXiv:1802.09091*

Monsalve, I. F., S. L. Frank and G. Vigliocco. 2012, April. Lexical surprisal as a general predictor of reading time. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* 398-408.

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog* 1(8), 9.

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 372-422.

Smith, N. J. and R. Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3), 302-319.

Staub, A. 2015. The effect of lexical predictability on eye movements in reading: Critical

review and theoretical interpretation. *Language and Linguistics Compass* 9, 311-327.
Trueswell, J., M. Tanenhaus and C. Kello. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(3), 528-553.

예시 언어(Examples in): 영어(English)
적용가능 언어(Applicable Languages): 한국어(Korean), 영어(English)
적용가능 수준(Applicable Level): 대학(Tertiary)

서혜진(Seo, Hye-Jin), 대학원생(Graduate Student)
동국대학교(Dongguk University)
영어영문학부
04620 서울특별시 중구 필동로 1길 30
Tel: 02) 2260-8705
E-mail: seohj0951@gmail.com

신정아(Shin, Jeong-Ah), 교수(Professor)
동국대학교(Dongguk University)
영어영문학부
04620 서울특별시 중구 필동로 1길 30
Tel: 02) 2260-3167
E-mail: jashin@dongguk.edu

논문 투고(Received): 2020년 11월 1일 (November 01, 2020)
논문 수정(Revised): 2020년 11월 30일 (November 30, 2020)
게재 확정(Accepted): 2020년 12월 15일 (December 15, 2020)

