

## LLM에서의 구조 점화 현상에 대한 소고: 한국어 수여 구문을 중심으로\*

허세문\*\*, 이상아†  
홍익대학교\*\*, 서울대학교†

Semoon Hoe, Sangah Lee. 2024. A short note on the structural priming in LLM: Focusing on dative constructions in Korean. *Language and Information* 28.3, 111-142. Recently, with the increasing diversity of research on neural language models, there has been growing discussion about whether the outcomes of purely probabilistic learning models reflect the so-called “linguistic knowledge” assumed in theoretical linguistics. This study investigates whether neural language models acquire structural cues of Korean dative constructions (DACs) so as to distinguish it from superficially similar constructions, in terms of structural priming. Using three large language models (LLMs)—LLaMA 3.1, Blllossom, and Polyglot-Ko—we calculate and compare the priming effects based on the probabilities, by examining various structural properties of our four target constructions (i.e., DAC, DAC\_Source, Causative, and Passive). The results reveal that at least the current LLMs employed in this study are capable of recognizing and distinguishing the abstract structures of our target constructions. Concomitantly, it is also observed that (i) cognitive concepts associated with spatial perception seem to have a stronger influence on structural priming than theta-roles, and (ii) the structural features of certain postposition/dative markers play a crucial role in differentiating the target sentences.

**Key words:** structural priming, prime effects, large language model, dative constructions, causative constructions, passive constructions

---

\* 이 연구는 서울대학교 신임교수 연구정착금으로 지원되는 연구비(100-20220148)에 의하여 수행되었음(교신저자)

\*\* 제1저자, (04066) 서울특별시 마포구 와우산로 94, 홍익대학교 영어영문학과, 초빙교수, E-mail: geisthoe@gmail.com

† 교신저자, (08826) 서울시 관악구 관악로 1, 서울대학교 언어학과, 조교수, E-mail: sanalee@snu.ac.kr

## 1. 서론

신경망 언어 모델(neural language model)에 대한 연구가 활발해지면서 순수 확률 기반 학습 모델의 결과에 이룬 언어학에서 논의되는 소위 “언어 지식”이 반영되어 있는지 살펴보는 연구도 다양하게 진행되고 있다(Hewitt & Manning 2018; Pater 2019; Warstadt & Bowman 2022 등). 특히, 최근 각광받는 트랜스포머(Transformer, Vaswani et al. 2017) 기반 신경망 언어 모델들은 일종의 블랙박스(black box)로 언어 지식의 반영 여부를 확인하기 위해서는 그 내부를 언어학적으로 탐침(linguistic probing)하기 위한 방법이 요구된다. 이에 따라, 특정 통사 현상이 반영된 문장의 문법성에 대한 평가 정확도를 통해 그 현상과 관련 통사 요인의 습득 여부를 확인할 수 있는 벤치마크(e.g., BLiMP; Warstadt et al. 2020)를 구축하고, 이를 통해 실제 신경망 언어 모델에 일치(agreement), 성분 통어(c-command), 구조적 위계(structural hierarchy) 판별 등 ‘구조 기반 통사 지식’이 반영되어 있는지 확인하는 식의 연구도 활발하게 이루어지고 있다(Linzen & Leonard 2018; Jawahar et al. 2019; Kann et al. 2019; Warstadt et al. 2020 등).

이는 인간 언어가 언어 능력(language competence)을 담당하는 뇌의 특정 기관(faculty of language)에 유전적으로 전해지는 생득적 정보(e.g., Universal Grammar)를 통해 습득되어야 하고(Chomsky 1986 *et seq.*), 이에 따라 확률 기반 학습만으로는 인간 언어의 특성을 포착할 수 없다는 결론(e.g., Katzir 2023; Lan et al. 2024 등)도 새로운 시각에서 재고해볼 필요가 있다는 주장으로 이어진다(Pater 2019; Linzen 2019; Piantadosi 2024 등). 물론 언어만을 위한 기관을 따로 가정하지 않고 인간의 ‘일반적인’ 인지 능력을 통해 언어의 습득이 이루어진다고 가정하는 인지 언어학(cognitive linguistics)기반 문법 이론이나 구문 문법(construction grammar; Goldberg 1995 등)들도 확률 중심 언어 모델과 어느 정도까지 양립 가능한지 쉽게 결론 내리기 어렵다는 점은 부정하기 어려울 것이다.

이러한 흐름에서, 인간 언어의 구조에 대한 언어 지식을 탐침하기 위해 고안된 방법을 신경망 언어 모델에 적용하려는 시도도 존재한다. 즉, 인간 언어와 유사한 방식으로 신경망 언어 모델의 “언어 능력”을 평가할 수 있는지 질문하고 그 해답을 찾으려는 것으로, 신경망 언어 모델이 구조의 형성이나 그 위계에 대한 지식을 실질적으로 습득하였는지 확인하는 것에서부터 인간 언어가 보이는 특성과 신경망 언어 모델 사이의 유사성을 파악하기 위한 시도까지 다양한 연구가 진행되고 있다. 그리고 이를 위한 분석 중 최근 주목 받는 방식의 하나로 구조 점화(structural priming)를 활용한 연구를 들 수 있다(Prasad et al. 2019; Sinclair et al. 2022; Jumelet et al. 2024; Prasad & Linzen 2024 등).

구조 점화는 심리 언어학(psycho-linguistics)에서 인간이 실제 통사 구조를 암묵적 지식(tacit knowledge)으로의 문법 요소로 인식하고, 이를 활용/구분하는지 확인하기 위해 고안된 실험 방식이다(Bock 1986, 1990; Branigan et al. 1995; Pickering & Ferreira 2008 등). 간단히 말해, 기 발화된 문장의 (추상적) 구조가 뒤따라 나오는 문장 형성에 직접적인 영향을 줄 수 있는지 확인하고, 영향이 있는 경우 그 영향이 “구조 형성”과 관련된 지식과 관련된 것인지 확인하는 방식이라 정리할 수 있다. 그러나 동사의 어휘 의미(lexical semantics)나 동사에 의해 선택된 논항(argument)들의 의미역(theta-role)과 같은 의미 요인에서부터 각 논항에 드러난 형태 표지(morphological marker)의 특성, 두 문장에 공유된 동일/유사 어휘 개수와 같은 다양한 요인들도 문장 구조의 형성에 영향을 준다고

볼 수 있다. 그러므로 형성된 구조 그 자체가 이들 요인과 완전히 독립적으로 점화 효과(priming effects)를 유발하는지, 아니면 다양한 요인들의 상호 작용까지 고려해야 하는지, 또 상호 작용이 존재한다면 그 영향력은 어느 정도인지 등 다양한 관점을 고려하면서 인간 언어의 특성을 탐구하는 것이 구조 점화 기반 연구의 주요 과제이다(Pickering & Ferreira 2008).<sup>1)</sup> 그리고 이를 대부분이 이미 신경망 언어 모델을 통해서 다양하게 연구되고 있다.

이러한 흐름에서 본 논문은 한국어의 여격 구문(dative constructions, 이하 DAC로 표기)을 중심으로 관련 구문의 구조를 구별할 수 있는 구조기반 지식이 신경망 언어 모델에 반영되어 있는지 구조 점화를 통해 확인해 보려 한다. 한국어의 경우에도 DAC 구문을 독립적인 구조로 구별하여 인식하는지, 또 그렇다면 이를 “외현적으로 유사해 보이는 관련 구문들”과 어떠한 구조적 요인을 통해 구분하는지 등 다양한 구조 점화 양상을 분석하고 확인하기 위한 실험을 구성하는 것도 주요 목적이라 할 수 있다.

이를 위한 논문의 구조는 다음과 같다. 우선 2장에서 구조 점화에 대해 간단히 정리하고, 이를 신경망 언어 모델에 적용하기 위해 본 논문에서 가정한 이론적 요인들에 대해 간략하게 살펴본다. 이 과정에서 한국어의 DAC에 대한 구조 점화 분석에서 비교/대조를 위해 사용될 세 가지 관련 구문의 특성에 대해서도 논의한다. 그리고 이러한 논의를 통해, 본 논문에서 수행하는 실험의 목표와 한계를 제시한다. 3장에서는 기본적인 실험 방법을 정리한다. 본 연구의 실험은 대규모 언어 모델(large language model, LLM)을 통해 이루어지는데, 실험을 위해 사용한 세 가지 언어 모델의 특성과 선택 이유를 간략하게 소개한다. 또한, 실험을 위해 어떠한 기준으로 동사를 선택하였는지 밝히고, 필요한 경우 각 동사들의 특성도 간략하게 정리한다. 4장은 각 실험을 소개하고 그 결과를 기술하고 분석한다. 5장에서는 이를 통해 LLM과 구조 점화와 관련해 드러난 몇 가지 논쟁 사항에 대해 더 자세하게 논의한다. 6장에서는 몇 가지 남은 문제를 논의하면서 논문을 마무리한다.

## 2. 구조 점화의 특성 정리

### 2.1. 구조 점화에 대한 이론적 배경<sup>2)</sup>

구조 점화는 기본적으로 다음과 같이 의미적으로 유사하여 그 사용에 있어서 경쟁이 일어날 수 있는 ‘구문’들을 비교하는 경우에 사용된다. 예를 들어 (1a)와 같은 DAC에 속하는 문장을 접하면 그 구

- 
- 1) 본 논문에서는 동사나 논항의 단순 어휘 의미는 최대한 배제하고 구조와 직접적으로 관련된 요인을 구분하여 실험을 전개할 것이다. 이에 따라, 의미역이나 구문 문법 관련 정보 등 몇 가지 변인은 의미 관련 요인으로 구분하기보다는, 구조 형성에 직접적으로 영향을 미치는 요인으로 보고 논의를 전개한다는 점에서 기존의 구조 점화 분석과 차이가 있을 수 있다. 이에 대해서는 2장과 3장에서 더 자세하게 논의한다.
  - 2) 구조 점화 연구에서는 후속 문장을 이해하는 실험(comprehension test)과 생성하는 실험(production test)의 차이에서 나타나는 다양한 변인들도 무시할 수 없다. 그러나 이러한 변인들이 신경망 언어 모델 기반 연구에 어떤 식으로 반영되는지 논의하는 것은 쉽지 않고, 이 구분 자체가 크게 영향을 미치지 않는다는 주장(Jumelet et al. 2024)도 있으므로, 본 논문에서는 더 자세하게 논의하지 않는다.

조직 특성에 영향을 받아 (1b)와 같은 관련 구문보다는 DAC에 속하는 문장을 반복하여 형성한다는 점을 보이고, 그 이유가 (1a)와 같은 문장을 인식하였을 때 그 구조적인 특성도 인식되어 후속 문장 형성에 직접적인 영향을 주기 때문이라 가정하는 것이다(Bock 1986 *et seq.*).

- (1) a. A teacher cooked a chicken for a worker. [DAC]  
 b. A teacher cooked a worker a chicken. [DOC]

이때, 후속 문장에 영향을 주는 문장을 접화 문장(prime sentence, 이하 PS로 기술)이라 부른다. 즉 PS의 구조적인 특성에 따라 DAC라는 구문과 동일한 구조를 가진 문장이 반복될 확률이 높아진다는 것이다. 이때, PS에 영향을 받는 후속 문장을 대상(target) 문장으로, PS가 대상 문장의 형성에 미치는 영향을 접화 효과(prime effects)라고 지칭한다.

당연하게도, 이러한 분석은 (1b)와 같은 이중 목적어 구문(double object constructions, 이하 DOC로 표기) 기반 PS가 주어지면 DAC보다는 DOC에 속하는 문장, 즉 전치사가 결합되고 두 내부 논항(internal argument)의 어순이 바뀐 형태의 문장이 형성될 것이라는 예측으로 이어진다. Bock(1986)은 이러한 예측도 사실이라는 것을 보이면서, 문장의 형성에서 적용되는 자연스러운 인지 작용이 진행되는 과정에서 바로 직전에 접한 문장의 “추상적 통사 표상(abstract syntactic representation)”을 인식하고, 바로 이 정보가 후속 문장 형성에 직접적으로 영향을 미친다는 것을 설득력 있게 제시하였다.

물론 이러한 주장이 타당성을 얻기 위해서는, 문장의 형성과 관련된 또 다른 요인이 아니라 “표상된 구조”만이 실제 직접적인 영향을 미치는지 확인해야 할 것이다. 이와 관련하여 Bock(1986)은 (1a)와 (2)의 경우 동일한 [DAC]이지만 PP를 이끄는 전치사가 ‘for’와 ‘to’로 서로 다름에도 불구하고 전치사 선택이나 전치사 사이의 의미 차이가 구조 접화와 그 결과에 유의미한 영향을 미치지 않는다는 점을 강조한다.

- (2) The guest threw the pot to the lady. [DAC]

물론 동사의 어휘 의미나 의미역 등 구조 형성에 영향을 주는 의미기반 요인이 더 중요한 역할을 하는 경우도 있다는 반박도 존재하지만, Bock & Loebell(1990)은 (3a)와 (3b)같은 PS 문장에 사용된 ‘to phrase’가 동일한 의미역을 부여받는다곤 볼 수 없음에도 불구하고(i.e., Goal vs. Recipient), 이 두 문장이 이야기하는 접화 효과에는 차이가 없다는 점을 보이고, 이를 포함한 다양한 논의를 통해 이러한 반박을 재반박하기도 하였다. 전치사의 형태나 그에 따른 의미 차이, 그리고 의미역이 구조 접화에 직접적인 영향을 미치지 않는다는 점을 들어 구조 접화의 정당성을 보장하려는 것이다.

- (3) a. The wealthy widow drove an old Mercedes to the church [DAC]  
 b. A rock climber sold some cocaine to an undercover agent [DAC]

그러나 여전히 문제는 남아 있다. 사실, 구조 접화에 대한 몇몇 분석에서 강조되는 구조적 특성이 설

령 동일하다 하여도, 이에 대한 이론적 배경까지 같지는 않다. 전통적인 구구조 문법(*phrase structure rules*)에서부터(see Bock et al. 1992) 구문 문법의 적용(see Ungerer 2023 for a brief summary)까지 다양한 방식이 활용되고 있는 것이다. 그런데, 특정 이론을 따르던 (3)의 두 예문은 다양한 의미 요인들을 고려하더라도 매우 유사한 구조라는 결론이 나올 수도 있다. 예를 들어 구문 문법의 측면에서 본다면 (3)의 두 예문은 모두 대격 논항이 주격 논항에서부터 여격 논항으로의 “공간 이동”의 대상이 되어야 한다는 점에서 충분히 동일한 인지적 특성을 공유하는 문장으로 볼 수도 있다. 즉 공간상의 이동이 존재하고 그 이동의 방향성이 같다는 두 측면에서 두 문장이 동일한 통사적 표상을 통해 실현되었다고 볼 여지도 있다는 것이다. 또한 의미역의 역할이 강조되는 이론에서도 Goal이나 Recipient에 대응되는 의미역을 각기 따로 설정하여 구분하지 않는 경우도 있고(e.g., Dowty 1991 등), 이에 따라 (3a)와 (3b)의 전치사구가 특별히 구조/의미적인 측면에서 구별되어야 하는 상이한 의미역을 가진다고 결론내리기 어려운 상황도 생각해 볼 수 있다.

이러한 문제는 구조 점화 관련 연구에서 지속적으로 제기되는 문제이고, 이를 하나하나 다 다루는 것은 본 논문의 범위를 넘어서는 일이 될 것이다. 무엇보다도 의미역과 같은 개념이나 인지/구문 문법에서의 공간 인지적 요인이 실제 신경망 언어 모델에 반영되어 있는지 확인하는 것부터 간단한 일이 아니다. 그러나, 이러한 정보가 특정 구조의 구성에 직접적인 영향을 주는 주요 요인일 수 있다는 점에서, 이를 단순히 주변 요인으로 보아 다루지 않는 것도 또 다른 문제를 불러 올 수 있다. 이 역시 구조 형성에 대한 배경 이론에 따라 그 양상이 달라지는데, 의미역과 통사 구조의 연관성을 다룬 논의 몇 개만 살펴봐도 이를 쉽게 확인할 수 있다. 이를 위해 먼저 구조 점화의 주요 동인인 통사 구조 이외에도 의미역을 도입하기 위한 독립적인 “의미 구조”가 존재하고 이 두 구조 사이의 상호작용(i.e., *linking*)을 통해 각 의미역의 구조적 위치가 결정된다는 이론(Jackendoff 1990 등)을 생각해 보자. 이러한 이론에서 의미역은 본질적으로 의미에 기반을 둔 개념이지만, 결과적으로 독립적으로 존재하는 통사 구조에 따라 그 값이 결정된다는 점에서 구조 형성과 무관한 의미 요인이라 단정 지을 수 없다. 이와 다르게, 의미역이라는 개념은 언어에 반영된 심리적 실체(*psychological reality*)로의 기초요소(*primitives*)가 아니라 통사 구조에 드러나는 “구조적 위계(*structural hierarchy*)”에 드러나는 의미적 특성 중 일부를 지칭하기 위한 편의적 개념일 뿐이라는 주장도 있다(Hale & Keyser 1993 등). 즉, 여기서 의미역은 특정 동사가 사용되었을 때 어떤 논항이 특정 통사 구조의 특정 위치에서 기저 생성되었다는 것을 알려주는 정보 그 이상도 그 이하도 아니라는 것이다. 이에 더해, 의미역 역시 공간과 사물을 인식하는 인간의 기본적인 인지 능력과 관련되어 형성된다는 점이 강조되는 인지 언어학적 입장도 고려해 보자. 예를 들어 구문 문법에서 가정하는 여러 인지적 요인(e.g., 공간 이동의 존재와 그 방향성 등)과 의미역 간의 상호작용 속에서 구조가 형성된다면, 앞선 논의와 또 다른 관점에서 의미역과 구조 형성 사이의 관계를 분석해야 하는 것이다. 즉, 이렇게 몇 가지 이론만 간단히 살펴본 것만으로도, 의미역이 추상적 구조와 무관하다고 보기 어렵고, 또 그 관계가 신경망 언어 모델에 어떻게 반영되어 있는지 세심하게 확인해야 할 필요가 있다는 것이 확인된다.

이러한 흐름에서 본 논문은 의미역에 대한 정보나 공간 관련 인지 요인들과 같은 관련 요인들 모두 구조 표상에 직접적으로 영향을 줄 수 있다고 보고, 이들이 구조적 요인으로 점화 효과에 영향을 미치는지 확인할 수 있는 방식을 활용하여 실험을 전개하려 한다. 즉, 구조 점화에 영향을 미치는 관련 요

인을 포괄적으로 고려하여 실험하려는 것으로, 이는 인간 언어와 신경망 언어 모델의 차이를 인정할 수밖에 없는 상황에서 지나치게 세밀하게 변인을 나누거나 아니면 아예 배제해 버리는 것 모두 부적절할 수 있다는 인식을 반영한 것이다. 그리고 이를 위해 구조 그 자체만이 아니라 관련 요인의 영향 및 상호 작용을 확인할 수 있도록 필요한 변인에 맞추어 비교에 사용되는 대조 구문의 수를 늘려 실험을 진행하려 한다. 물론 이는 일반적인 구조 점화 연구에의 기본 가정에서 약간 벗어난 방식일 수도 있다. 그러나 다양한 구문을 가정하고, 각 구문의 구조 형성에 영향을 미칠 수 있는 요인들을 최대한 고려하는 것이 신경망 언어 모델의 특성 그 자체를 이해하고 탐구하기에도 더 유리하다는 점도 무시할 수 없다. 본 연구에서는 인간 언어와 신경망 언어 모델에 대한 직접적인 비교까지 진행하지 않았지만, 언어 모델이 보이는 특성을 파악할 수 있는 언어학적 요인을 고려한 실험 설계는 필요한 것이다. 이에 따라, 의미역이나 논항 교체와 같은 지식이 신경망 언어 모델에 반영되어 있다는 기존 논의(Kann et al. 2019 등)에서부터 구문 문법의 형성에 작용하는 여러 의미 요인도 내제된 지식으로 신경망 언어 모델에서 인식될 수 있다는 가능성 등, 기존 연구의 결과나 실현 가능한 다양한 가능성을 최대한 고려하여 관련 요인을 다각화할 필요가 있는 것이다. 한 마디로, (추상적) 구조 그 자체만이 아니라 구조의 형성에 영향을 미치는 여러 요인들이 실제 신경망 언어 모델에 반영된 언어 지식으로 존재하는지 확인하는 과정이 필요하다는 점도 고려한 것이다.

## 2.2. 한국어 수여 구문과 구조 점화 실험을 위한 구조 유형 선택

1장에서 간략히 언급한 것처럼 본 논문은 한국어의 DAC가 가지는 구조적 특성을 구조 점화 기반 분석을 통해 확인하려 한다. 또한 2.1절에 언급한 것처럼 다양한 관련 요인들을 확인하기 위해 다음과 같은 네 가지의 유형을 구분하였다. 이들 유형들은 모두 격틀(case frame)과 어순을 동일하게 맞추어서 외현적으로 경쟁 대상이 될 수 있게 하였다. 이에 더해, 앞서 언급한 것처럼 의미역이나 인지/구문 문법의 공간 인식 요인이 구조 형성에 직접적인 영향을 줄 수 있다는 입장도 구문의 선택에 반영하였다. 이들 유형의 특성을 몇몇 예시들을 통해 살펴보면 다음과 같다.<sup>3)</sup>

### (4) DAC 유형

- a. 철수가 영희에게 책을 주었다.
- b. 철수가 영희에게 편지를 썼다.
- c. 철수가 영희에게 영어를 가르쳤다.

### (5) DAC\_Source 유형 (이하 DAC\_S 유형으로 표기)

- a. 철수가 영희에게 책을 받았다.
- b. 철수가 영희에게 차키를 뺏었다.
- c. 철수가 영희에게 피아노를 배웠다.

---

3) (4)에서 (7)까지의 유형에서 확인할 수 있듯이, 본 논문에서는 한국어의 DOC 구문을 의도적으로 배제하고 있다. 이러한 이유에 대해서는 6장에서 간략하게 다룬다.

- (6) 사동 유형
  - a. 철수가 영희에게 발표를 시켰다.
  - b. 철수가 영희에게 코트를 입혔다.
  - c. 철수가 영희에게 임무를 맡겼다.
- (7) 피동 유형
  - a. 철수가 영희에게 발을 밟혔다.
  - b. 철수가 영희에게 구타를 당했다.
  - c. 철수가 영희에게 기회를 빼앗겼다.

우선 (4)의 DAC 유형은 본 논문이 분석하려는 기본 유형이다. 의미역의 측면에서 주격 논항은 Agent를, 여격 논항은 Goal을, 대격 논항은 Theme으로의 의미역을 부여받는다고 볼 수 있다. 이밖에 인지/구문 문법적인 측면에서 이 유형은 주격 논항에서 여격 논항 쪽으로 대격 논항이 ‘공간 이동’을 한다는 특성을 보인다. 이때 대격 논항이 구체물이라면 실질적인 공간 이동이 드러나지만 (4c)와 같이 추상물인 경우에는 동일하다고 보기 어려울 수도 있다. 그러나 이와 관련해서 이미 많은 연구에서 추상물이 나오는 경우 말/내용 등의 추상적인 요소의 전달 역시 공간 이동으로 볼 여지가 있다고 지적하고 있다. 또한 물리적 공간을 추상화 시키면 일종의 상태로 간주할 수 있고, 이 상태라는 추상화된 공간에서의 일어나는 이동을 변화로 해석한다면 상태 변화와 공간 이동 역시 평행하게 분석할 수 있다는 점(Pinker 1989)을 받아들여, 여격 논항이 Goal에 대응된다고 충분히 인식될 수 있는 경우에는 구체/추상 구분 없이 이들 모두를 DAC 유형으로 분류하였다.<sup>4)</sup>

(5)의 DAC\_S 유형은 본질적으로 대격 논항의 공간 이동이 가정된다는 점에서 DAC 유형과 유사하다(Yi & Yun 2019).<sup>5)</sup> 물론 의미역의 측면에서 본다면 주격 논항이 Goal 혹은 Recipient에 대응되고, 여격 논항이 Source에 대응된다는 점에서 DAC\_S 유형은 DAC 유형과 명확하게 구분된다. 그러나 앞서 언급하였듯이, 의미역 자체가 구조와 관련된 특성이면서 인지/구문 문법적인 측면에서 본다면 공간 이동과 관련된 다양한 측면들과의 상호 작용을 한다고 볼 수 있다. 그런데 공간 이동이라는 인지/구문 문법적 속성에서 본다면 DAC\_S 유형과 DAC 유형의 차이는 ‘이동의 방향’ 밖에 없다는 점에서 매우 유사한 구문으로 인식될 수도 있다. 이러한 점에서 본 연구는 DAC 유형에 속하는 문장을 PS로 DAC\_S 유형 문장에 대한 점화 효과를, 혹은 DAC\_S 유형에 속하는 문장을 PS로 DAC 유

4) 감정의 변화도 동일하게 이동의 일종으로 볼 수 있다고 보아 DAC\_S 유형에 ‘느끼다’와 같은 동사도 포함시켰다. 그러나 양상(modal) 의미가 드러나는 소망(bouletic) 동사에 대해서는 쉽게 판단하기 어려워, ‘바라다’, ‘원하다’와 같은 동사는 제외하였다. DAC 유형과 DAC\_S 유형에서 대격 논항이 구체/추상의 구분이 실제 LLM에서 의미가 있는지에 대해서는 6장에서 다시 논의할 것이다.

5) 본 연구에서 DAC\_S 유형을 DAC 유형과 비교하는 것은 기본적으로 Yi & Yun(2019)에서 제시한 구조 점화에서 통사적 요인과 의미적 요인을 구분하려는 시도에 영향을 받은 것이다. 그러나 Yi & Yun(2019)의 분석을 신경망 언어 모델에 적용시키는 과정에서 어느 정도 수정이 이루어졌고, Yi & Yun(2019)과 달리 본 논문은 어순을 고려하지 않는 등 많은 차이도 존재한다. 이에 더해, 본 논문에서는 아직 인간 기반 실험과 신경망 언어 모델 기반 실험을 직접적으로 비교하지 않으므로, Yi & Yun(2019)의 분석에 대해서는 더 자세히 기술하지 않았다.

형에 속하는 문장에 대한 점화 효과를 확인하여 이들의 유사 혹은 구별 정도를 확인할 수 있다고 가정한다(see also Yi & Yun 2019). 즉, 구조 점화에 있어서 의미역과 공간 이동 중 어느 것이 더 직접적인 영향을 주는지도 확인할 수 있다고 보는 것이다. 만약 DAC와 DAC\_S 유형이 구조 점화에 있어서 서로 점화 효과를 준다면 구조적인 동일성이 있다는 것으로 해석되므로, 구조 점화에 있어서 의미역이라는 정보가 실질적으로 영향을 주지 않다고 확인할 수 있는 환경을 하나 제공한다고 볼 수 있다. 반대로, 이들이 점화 효과를 주지 않아 완전히 다른 구조로 인식된다면 공간 이동이라는 정보 자체보다는 그 방향성에 대한 정보보다 의미역이 더 직접적인 영향을 준다고 볼 수도 있을 것이다. 즉, 구조와 관련된 다양한 요인과 환경의 차이를 확인할 수 있는 기회를 얻을 수 있다는 가정에서, DAC\_S 유형을 분석 대상에 추가한 것이다.

(6)의 사동 유형은 동일한 격들을 유지한다는 점에서 이 유형에 속한 문장이 두 DAC 유형과 (외현적으로) 유사한 구조를 가졌다고 인식될 수 있어 추가하였다. 특히 한국어의 DAC는 “사동 의미”를 가지지 않는 경우가 더 많다고 지적하는 기존 논의도 고려하였다. Jung & Miyagawa (2004)에서는 (8)의 두 문장이 영어의 DAC와 DOC에 대응한다고 지적하고 있다(see also Shin & Christianson 2009).

- (8) a. 철수가 영희에게 책을 주었다. [DAC]  
 b. 철수가 영희를 책을 주었다. [DOC]

그러나 동시에 Jung & Miyagawa(2004)는 한국어의 (8b)와 같은 구문에서 대격 논항이 실현되는 동인은 영어의 DOC와 다를 수 있다고 지적한다. (9)와 같은 ‘보내다’ 동사는 영어의 ‘send’ 동사와 매우 유사한 의미를 가지지만 DOC의 형태를 허용하지 않는 것을 간과할 수 없다는 것이다. 이를 설명하기 위해 Jung & Miyagawa(2004)에서는 (8b)와 같은 문장에서 대격이 실현되기 위해서는 (10)과 같이 “사동을 드러내는 구조”가 DOC 동사구 내에 반영되어 있어야 한다고 주장한다.<sup>6)</sup>

- (9) a. 철수가 영희에게 책을 보냈다. (<sup>ok</sup>Chelswu sent a book to Yenghuy) [DAC]  
 b. \*철수가 영희를 책을 보냈다. (<sup>ok</sup>Chelswu sent Yenghuy a book) [DOC]  
 (10) a. 철수는 영희에게 심부름을 시켰다. [사동 구문 (with 사동 동사구 구조)]  
 b. 철수는 영희를 심부름을 시켰다. [사동 구문 (with 사동 동사구 구조)]

이러한 논의는 한국어의 (8b)와 같은 문장이 실제 영어의 DOC에 대응되는 문장이라 하여도 이를 위해서는 독립적으로 사동의 의미가 구조적으로 표상된 동사구가 실현되어야 한다는 가정으로 이어진

6) Jung & Miyagawa(2004)는 이러한 차이는 사동의 의미를 도입하는 특별한 기능핵(functional head)이 동사구 내에 존재하는지 여부를 통해 구분된다고 주장한다. 물론 이러한 최소 주의(Chomsky 1995 *et seq.*) 기반 설명은 구문 문법의 경우 사동성 탐지를 위한 독립적인 구문을 가정하는 것 등으로 대체될 수도 있을 것이다. 물론 이러한 구조적 차이가 신경망 언어 모델에서 어떻게 실질적으로 드러나는지 다루는 것은 본 연구의 범위를 벗어나는 것이므로, 이에 대해서는 더 자세하게 다루지 않는다.

다. 동시에 한국어의 DAC 유형에 속하는 동사들이 모두 사동성을 드러내는 특별한 구조의 동사구를 포함하는지 질문해 보아야 한다는 점도 드러난다. 즉, ‘주다’와 ‘보내다’는 동일한 DAC 유형 내에서도 사동성이라는 측면에서 서로 구분될 수 있다는 것이고, 만약 사동성이 구조에 반영된다면 두 동사가 서로 다른 구조의 동사구를 형성할 것이라는 점에서 구조 접화를 통한 분석에 문제를 일으킬 수도 있는 것이다.

다행히도 한국어의 경우 DAC 유형에 속하는, 즉 공간 이동에 기반을 둔 의미를 가진 동사 대부분은 여격 논항의 대격 논항으로의 교체를 허용하지 않는다. 이를 허용하는 ‘주다’나 ‘가르치다’와 같은 동사가 오히려 예외적인 것이라 볼 수도 있는 것이다. 그러나 이 문제를 완전히 무시할 수 없다는 점에서 (6)과 같은 사동 유형을 비교/대조 유형에 포함시켰다. 특히, 사동 유형에 배타적으로 포함될 수 있는 적절한 동사를 선택하기 위해 다음의 기준들을 추가적으로 적용하였다. 우선 여격 논항이 특정 행위를 실행하는 역할을 할 수 있는지 여부를 확인하였다. 사동 구문에서 여격 논항은 일반적으로 Patient로의 의미역을 부여받는다고 논의되지만, (6a)와 같은 상황에서 결과적으로 ‘영희’는 발표 행위를 실행해야 한다는 점은 변하지 않으므로, 이러한 특성을 사동 유형을 구분하는 주요 요인으로 본 것이다. 그런데, (6b)의 경우 ‘영희’가 어떤 행위를 하는지 확인하기 어렵다. ‘영희’는 특별한 행위를 하지 않고 ‘철수’가 ‘코트’를 ‘영희’의 몸에 입히는 행위가 있었다는 해석이 더 자연스러운 것이다. 그리고 이러한 경우 의미적으로 ‘철수’에서 ‘영희’로 ‘코트’가 공간 이동했다고 볼 수도 있다는 점에서 DAC 유형과 사동 유형의 구분은 쉽지 않다. 물론, (6b)는 ‘철수’는 행동 없이 명령만 하고, ‘영희’가 스스로 ‘코트’를 입었다고 볼 가능성이 아주 불가능한 것은 아니지만 이러한 해석은 보통 (11)과 같은 우언적 사동(periphrastic causation) 구문과 더 어울리는 해석이라고 분석되기도 한다.

(11) 철수가 영희에게 옷을 입게 했다.

이러한 점에서, 본 논문은 사동 유형에 속하는 동사의 특성을 “Goal이 드러내는 필수적인 구조적 특성과 구분되면서 대격 논항으로의 교체 가능한 여격 논항을 도입하는 것”이라 가정하였다. 이를 따르면 (6b)와 같은 동사는 사동 유형으로 분류되어야 할 것이다. 이에 더해, 한국어의 특정 동사에 어휘적 사동 표지가 결합되어 있고, 그 사동 표지를 제외한 동사가 한국어의 어휘부(lexicon)에 존재하는 경우 사동 유형에 속하는 동사로 분류하였다. 이는 (6c)와 같은 경우를 설명하기 위한 기준으로 이때의 여격 논항의 의미역을 특별히 특정하기는 어렵지만 이를 Goal로 보기는 어렵다는 점에서도 DAC 유형과 구분된다고 보았다.

지금까지 논의한 내용은 사동 유형 역시 의미역의 측면에서 DAC 유형과 구분되고, 이러한 차이가 구조에 영향을 미칠 것이라는 가정에서 시작한다. 또한 대격 논항의 공간 이동이 배제되어 있다는 점에서도 DAC 유형과 구분될 것이고, 일차적으로 본 논문은 이러한 특성이 구조 접화에 있어서 어떻

7) 물론 사동이라는 의미는 특정 사건을 야기하는 힘(force)의 이동으로 발현된다는 점(i.e., Force-dynamic framework; Talmy 2000, see also Jackendoff 1990)에서 일종의 공간 이동으로 해석될 여지는 존재한다. 특히나, 이러한 개념은 최소주의를 기반 통사 이론에서도 동사구 형성의 가장 중요한 요인으로 보는 등(Copley & Harley 2015) 다양하게 활용되고 있다는 점도 무시할 수 없다. 그러나 이동 대상으로서의 ‘힘’이 대격 논항으로

게 드러나는지 확인해 보려고 하는 것이다. 이에 더해져 필수적으로 고려해야 하는 요인이 하나 더 있다. 본 논문은 Jung & Miyagawa(2004)의 주장을 받아들여 실험 설계를 하였지만, 한국어 혹은 신경망 언어 모델 내의 문법의 경우 그들의 주장과 다르게 DAC 유형에 본질적으로 사동 의미와 관련된 구조가 반영되어 있을 수도 있는 것이다. 만약 사동 유형이 DAC 유형이나 DAC\_S 유형에 점화 효과를 보인다면 이러한 가능성에 대한 재고가 필요할 것이다.

마지막으로 (7)의 피동 유형도 비교/대조를 위한 유형으로 추가하였다. 사실 (7)의 유형은 DAC 유형과 완전히 구분되는 특성을 가진다는 예측된다는 점에서 중요하다.<sup>8)</sup> 우선 주격 논항은 Patient에 대응되고 여격 논항은 영어의 'by-phrase'에 대응된다는 점에서 Goal과의 유사점을 찾아볼 수 없다. 또한 타동사의 피동형에 대격 논항이 유지되는 것은 한국어의 피동 구문에서는 쉽게 찾을 수 있지만, 적어도 영어의 경우는 DOC 구문이 피동화 되는 경우 Theme에 대응되는 논항이 대격으로 유지되는 경우 정도를 제외하고는 쉽게 찾아보기 어려우므로 한국어에 국한된 문법적 특성도 드러낼 수 있다는 점에서 선택하였다. 즉, DAC 유형과의 유사점을 어느 정도 가정할 수 있는 (5)나 (6)의 유형과 다르게 (7)의 유형은 나머지 세 유형과 크게 구분되는 유형으로, 그 구조적 표상 역시 배타적인 것이라 가정한 것이다. 이러한 점에서 피동 유형에서는 여격 논항이 Agent(혹은 Causer)로 인식될 수 있는 환경을 야기하는 동사를 위주로 선택하였다.

### 2.3. 신경망 언어 모델에서 구조 점화 실험의 목적과 범위

2.2절에서 논의하였듯이, 본 논문은 비교/대조가 가능한 유형을 넷으로 구분하여 분석하고 있다. 물론, 구조 점화에 대한 초창기의 논문인 Bock(1986)에서도 능동문, 피동문, DAC, DOC, 이렇게 네 유형을 중심으로 분석하였다는 점에서 본 논문의 분석 대상 유형 설정이 크게 다를 것이 없어 보이기도 한다. 그러나 Bock(1986)이나 기존 논문들은 이러한 경우 보통 능동문-피동문, DAC-DOC 이렇게 두 부류로 구분하여 실험을 전개하였지만, 본 논문은 서로 다른 네 유형을 모두 직접적으로 대조한다는 점에서 차이가 있다.

이러한 대조 실험은 2.2절에서 언급하였듯이, 한국어에서 DAC 유형의 구조 형성에 영향을 주는 요인이 무엇이고, 또 비교 유형들과의 구조적 차이를 드러내는 요인이 무엇인지 조금 더 명확하게 구분하기 위해 고안되었다. 특히 이는 의미역이나 인지/구문 관련 정보 역시 구조를 구별하는 요인으로 간주하는 본 논문의 기본 가정과도 관련되어 있다. DAC 유형이 DAC\_S 유형과 실제 구조적인 차이가 있는지, 있다면 그 주요 요인은 무엇인지, 그리고 이러한 요인이 신경망 언어 모델에 반영되어 있는지

실현되지 않는다는 점에서 DAC 유형과는 확실한 차이가 있다고 보았다.

8) 영어의 경우 DAC가 DOC에서 '일종의' 피동화 과정을 통해 도출된다는 주장도 존재한다. Collins(2024)에서는 피동문에서 나타나는 'by-phrase'가 대응되는 능동문에서는 구조격을 받는 논항이라는 점을 고려한다면, DAC의 여격 논항도 대응되는 DOC에서는 구조격을 받는 논항이라는 점에서 유사점이 드러난다고 강조한다. 피동문과 DAC 모두 구조격을 받던 논항이 전치사와 결합하는 논항으로 바뀌게 되는 과정을 겪는데, 바로 이 과정이 일어나는 동기가 바로 피동화와 관련되어 있다는 것이다. 만약 이러한 분석을 받아들인다면 DAC 유형과 피동 유형도 구조적인 유사성을 가진다고 볼 수 있을 것이다. 이에 대한 더 자세한 논의는 추후 연구로 남긴다.

등 다양한 문제를 모두 살펴보아야 하고, 이러한 비교/대조는 DAC 유형과 나머지 두 유형 사이에서도 확인해 보아야 한다는 점에서, 다양한 요인을 효과적으로 살펴보기 위한 전략인 것이다.

물론 본 논문에서처럼 다양한 요인을 동시에 비교하는 것은 실험의 간결성을 저해하는 것으로 보일 수도 있다. 그러나 본 연구에서 사용하는 신경망 언어 모델이 구조 점화 실험에서 요구되는 언어 지식을 온전히 반영하고 있으리라고 보장할 수 없다는 점에서 이러한 시도는 불가피하다고 할 수 있다. 즉, 모델의 한계에 의해 실험의 결과가 기대와 다르게 왜곡되는 경우에도 그 실패 이유나 특징을 좀 더 정확하게 파악하고 보완할 수 있는 정보를 얻을 수 있게 실험을 설계한 것이다. 그리고 실제 인간을 대상으로 한 실험과 다르게 미세 조정(fine-tuning)을 통해서 실험에 필요한 문법 지식을 추가할 수 있고, 이러한 정보가 반영된 모델과 이전 모델도 쉽게 비교할 수 있다는 점도 고려하였다. 즉, 본 연구는 일종의 예비 실험적인 성격도 가진다고도 할 수 있는 것이다.

### 3. 실험 준비 및 모델 소개

#### 3.1. 모델 소개

본 연구에서 실험 대상으로 삼은 모델은 Meta에서 공개한 영어 기반 LLM인 LLaMA 3.1 8B,<sup>9)</sup> LLaMA 3.1 모델을 기반으로 하여 한국어 데이터로 추가 훈련한 한국어-영어 이중 언어 모델인 Bllossom 8B,<sup>10)</sup> EleutherAI에서 공개한 한국어 기반 모델인 Polyglot-Ko 12.8B<sup>11)</sup>이다. 세 가지 모델 모두 트랜스포머 구조에 기반한 생성형 언어 모델에 해당한다.

LLaMA 3.1 8B 모델은 80억 개 매개변수를 포함한 대규모 언어 모델로, 다양한 언어를 다룰 수 있는 다중 언어 모델이므로 한국어 역시 지원 가능하나 훈련 데이터의 많은 부분이 영어로 이루어져 있다. Bllossom 8B 모델은 이러한 LLaMA 3.1 8B 모델을 한국어 데이터를 이용해 추가 훈련 및 미세 조정된 것으로, LLaMA 3.1의 보편적인 언어 지식과 성능에 더해 한국어 처리 능력을 강화한 것으로 기대되는 모델이다. 한편 Polyglot-Ko 12.8B 모델은 128억 개의 매개변수를 포함하였으며, NAVER 블로그와 지식인, 국립국어원 '모두의 말뭉치' 등의 한국어 데이터를 이용해 사전학습한 대규모 언어 모델이므로 한국어만을 지원한다. 이들 모델은 서로 다른 기반 언어를 가지고 있으므로 구조 점화에 있어서도 서로 다른 경향성을 보일 수도 있어 선정한 것들이다.

이들 LLM은 모두 생성형 모델 즉 자가회귀 모델(autoregressive model)이므로 앞서 생성된 단어 혹은 토큰의 정보에 기반하여 뒤에 나올 단어 혹은 토큰의 존재 확률 분포를 통해 가장 있을 법한 문장 혹은 연쇄를 생성해 낸다. 이러한 기제에 따라 특정한 문장이 존재할 확률, 곧 모델이 훈련해 둔 언어 지식에 기반을 두고 특정 연쇄가 생성될 확률을 계산할 수 있다. 이는 곧 구조 점화 효과를 관찰

9) <https://huggingface.co/meta-llama/Llama-3.1-8B>

10) <https://huggingface.co/MLP-KTLim/llama-3-Korean-Bllossom-8B>

11) <https://huggingface.co/EleutherAI/polyglot-ko-12.8b>

하는 데에도 이용 가능하다. 점화 문장, 즉 PS가 주어졌을 때 각 대상 유형의 구문이 생성될 또는 존재할 확률이 높고 낮은 정도에 따라 구조 점화 효과의 발생 여부를 유추해 볼 수 있다.

### 3.2. 구조 점화 효과 계산 방법

위 3.1절에서 제시한 언어 모델들에서 특정 PS가 주어졌을 때 각 유형의 구문이 생성될 확률은 아래의 (12)와 같이 조건부확률을 이용해 계산된다.  $n$ 개의 토큰으로 구성된 PS를  $PS = \{x_1, x_2, \dots, x_n\}$ 이라 보고  $m$ 개의 토큰으로 구성된 대상<sub>target</sub> 구문을  $Ph = \{y_1, y_2, \dots, y_m\}$ 이라고 보자. (12)에서는 Ph에 해당하는  $y_1, y_2, \dots, y_m$  토큰들의 연쇄 생성 확률만을 구하여 이를 각 PS의 영향을 받은 구문의 존재 확률로 간주한다.

예를 들어 ‘친구가 영희에게 휴대폰을 숨겼다.’라는 문장이 PS로 주어졌을 때 ‘철수가 영희에게 피아노를 배웠다.’라는 대상 문장이 존재할 확률을 계산한다면, 해당 PS가 주어진 상태에서 그 뒤에 올 토큰으로 ‘철수(혹은 그보다 작은 단위의 토큰)’가 생성될 확률부터 시작해서 마지막 ‘.’이 생성될 확률까지 누적하여 계산하고자 한 것이다. 이때 확률값의 반복적인 곱셈 과정에서 산술 언더플로(arithmetic underflow)가 발생하는 것을 피하기 위해 로그 스케일(log scale)로 확률을 변환하여 계산하였다.

(12) PS 기반 구문 생성 확률

$$\begin{aligned} P(Ph|PS) &= P(y_1, y_2, \dots, y_m | x_1, x_2, \dots, x_n) \\ &= \prod_{k=1}^m P(y_k | x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_{k-1}) \\ &\approx \sum_{k=1}^m \log P(y_k | x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_{k-1}) \end{aligned}$$

본 연구는 각 구문의 PS의 유형과 대상 구문 유형을 조합한 16가지 유형의 문장쌍들에 대해 위와 같은 연산을 수행하고, 유형 간의 확률값 비교를 통해 구조 점화 효과의 정도를 계산하였다. 아래 식 (13)에서  $Ph^x$ 는 한 유형  $x$ 의 대상 구문을,  $PS^x$ 와  $PS^y$ 는 각각 유형  $x$ 와  $y$ 에 해당하는 PS를 나타내며 이를 활용한 조건부확률은 위 식 (12)와 같이 계산된다. 이때  $i$ 는 대상 구문 Ph의 데이터셋 내 인덱스를,  $j$ 는 PS의 데이터셋 내 인덱스를 나타낸다.

(13) 점화 효과 계산식

$$PrimingEffect = \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n P(Ph_i^x | PS_j^y) - \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n P(Ph_i^x | PS_j^x)$$

예를 들어 (13)에서  $x$ 가 DAC 유형,  $y$ 가 DAC\_S 유형이라고 한다면, 계산되는 점화 효과의 정도는

$P([\text{DAC 유형}]_{\text{대상}} | \text{DAC\_S 유형 PS})$ 와  $P([\text{DAC 유형}]_{\text{대상}} | \text{DAC 유형 PS})$ 를 활용한 식이 된다. (지금부터, 각 유형이 PS로 사용되는 경우와 평가의 대상으로 사용되는 경우를 구분하기 위해서 후자를  $[\text{DAC 유형}]_{\text{대상}}$ 과 같은 형식으로 기술할 것이다.)

본 연구에서는 각 유형의 구문에 각 유형의 PS를 부여한 경우를 1,000건씩 상정하고 이들 경우에 대해 계산된 1,000개의 로그 확률값의 평균을 계산하여 그 차이를 점화 효과로 사용하였다. 이에 따라 같은 유형의 PS가 주어졌을 때보다 다른 유형의 PS가 주어졌을 때의 Ph 확률값이 낮아서 점화 효과의 값이 높게 나타난다면 이는 곧 Ph가 같은 유형의 PS의 영향을 받아 특정 유형의 구문을 생성한 것이라고 볼 수 있을 것이다.

### 3.3. 동사 유형 목록 선택 및 관련 논의

다음은 본 논문에서 분류한 네 가지 유형에 따라 실제 실험에 사용된 동사들의 목록이다.

- (14) DAC 유형: 보내다, 밝히다, 전하다, 쓰다, 알리다, 가르치다, 건네다, 내주다, 드리다, 팔다, 넘기다, 먹이다, 베풀다, 바치다, 주다, 넣다, 물려주다, 돌려주다, 갚다, 떠넘기다, 묻다, 들려주다, 외치다, 빌다, 들려주다, 내뱉다, 퍼트리다, 취하다, 밀다
- (15) DAC\_S 유형: 받다, 언다, 배우다, 빌리다, 취하다, 거두다, 가져오다, 빼앗다, 되찾다, 돌려받다, 훔치다, 물려받다, 돌려받다, 듣다, 얻어먹다, 느끼다, 넘겨받다
- (16) 사동 유형: 숨기다, 물리다, 보이다, 맡기다, 입히다, 시키다, 안기다, 씌우다, 채우다
- (17) 피동 유형: 얻어맞다, 당하다, 빼앗기다, 잡히다, 붙잡히다, 인정받다, 잘리다, 밟히다, 뺏기다

동사의 선택 기준은 2장에서 이미 다루었으므로, 몇몇 추가 언급이 필요한 동사들 위주로 간단하게 설명한다. 이미 언급하였듯이, 본 논문은 말뭉치의 빈도를 이용하였기 때문에, 최대한 그 빈도를 유지하는 방식으로 동사를 선택하였다. 빈도를 고려하여 조건을 만족시키지 못하는 동사를 순차적으로 배제하였는데, 이때 고려해야 했던 몇 가지 조건을 소개하면 다음과 같다.

우선, 2장에서 언급한 것과 같이 대격 논항의 구체/추상성의 구분에서 나오는 문제를 최소화시키기 위해 네 유형 모두 대격 논항이 구체물인 예문과 추상물이 예문이 충분히 구성되어 이들이 동일한 비율로 대상 자료에 반영될 수 있게 동사를 조정하였다. 그러나 이후 구체/추상의 구분이 보이는 영향성을 DAC 유형과 DAC\_S 유형을 중심으로 추가적으로 실험해 볼 필요성이 생기게 되면서(6장 참조) 이 두 유형만 동사를 추가했어야 했는데, 이 때 사용된 동사까지 (14)와 (15)에 모두 포함되어 있다. 즉 (14)와 (15)에 동사 수가 많은 것은 실제 실험에 영향을 미치지 않았다.

이에 더해, 2장에서 언급한 기본 기준 이외에도 “논항 교체(argument alternation)”와 관련하여 동사 부류(verb classification)를 구분하기 위해 고안된 요인도(see Levin 1993 등) 본 논문의 목적과 어울리지 않는 경우 배제해야 한다는 점에서 고려 대상에 포함되었다. 예를 들어 ‘밀다’의 경우 본 논문에서 고려하는 격틀을 만족시키는 고빈도 동사이다. 그런데, 영어의 ‘push’와 ‘pull’은 그 자체에 주격-논항(주어)이 동사에 이동한다는 의미가 포함될 수 있어 DAC나 DOC와 다른 동사 부류로 구분되

므로 그 적절성을 따져 보아야 했다. 이와 관련하여 우선 한국어의 경우 ‘-어 가다/오다’와 같은 보조 동사가 결합하지 않고는 주어의 이동이라는 의미가 드러나지 않는다는 점에서 이 두 동사를 배제할 이유가 없다고 판단하였다. 또한, ‘push’에 대응되는 ‘밀다’ 동사는 본 논문에서 가정하는 격틀에 어울리지만 (e.g., 철수가 영희에게 의자를 밀었다.), ‘pull’에 대응되는 ‘끌다’ 동사는 자연스러운 문장이 형성되지 않아(e.g., #철수가 영희에게 의자를 끌었다.), 이 둘이 논항 교체 현상과 관련된다고 보기도 어려웠다. 이에 따라, ‘밀다’ 동사만 DAC 유형에 포함시키고 ‘끌다’ 동사는 DAC\_S 유형에서 자연스럽게 배제되는 것을 확인할 수 있었다. 마지막으로, 이 논의의 영향을 받아 대격 논항의 공간 이동 의미를 명확하게 강조하기 위해 주격 주어의 실질적인 공간 이동이 포함되는 술어는 포함시키지 않도록 노력하였고, 이에 ‘-어 가다/오다’가 연쇄된 동사는 완전히 배제하였다. 그 결과로 ‘가져 오다’ 등의 몇몇 고빈도 술어들은 목록에서 빠지게 되었다.

또한, 처격 교체(locative alternation)에 속하는 술어도 배제하려고 노력하였다. 물론 (16)의 사동 유형에 포함된 ‘채우다’를 이와 관련된 술어로 볼 수도 있다(e.g., 철수가 병에 물을 채웠다. vs. 철수가 병을 물로 채웠다). 하지만, DAC\_S 유형이나 피동 유형에서 쉽게 확인할 수 있듯이 본 논문의 네 유형을 공정하게 비교하기 위해서는 의미적으로 여격 논항은 모두 유정성(animateness)을 가져야 한다. 이러한 점에서 소위 용기(container)에 대응되는 논항이 여격 논항으로 선택되기 어렵다는 점에서 ‘채우다’ 동사가 처격 교체를 보이는 문장은 자연스럽게 배제된다. 실제, 본 논문에서 사용된 ‘채우다’ 동사는 ‘철수가 영희에게 팔찌를 채우다’와 같이 물건을 신체에 결속하는 의미를 보이는 경우에만 포함될 수 있었다.

‘안기다’의 경우는 사동-피동 중의성을 보이는 술어이지만, 대격 논항으로 추상물이 선택되는 경우 사동의 의미만 드러나고, 피동의 의미가 드러나는 문장 형성이 매우 제한적이라는 점에서 사동 유형에 포함시켰다. 그러나 DAC 유형과 DAC\_S 유형 모두에서 사용될 수 있는 술어들은 최대한 배제하였다. 예를 들어 ‘구하다’와 같은 동사는 고빈도 동사였으나, ‘얻다’의 의미로는 DAC\_S 유형이, ‘요청하다’에 대한 의미로는 DAC 유형에 속한다는 점에서 배제하였다.

마지막으로 ‘-어 주다’ 형태가 포함된 동사는 ‘-어 주다’가 결합하기 전 기본 동사가 여격 논항을 도입하지 못하는 경우에만 포함시켰고, 이 경우에도 여격 논항의 해석과 전체 구문의 의미를 충분히 고려하였다. 이밖에도 다양한 요인이 고려되었다.

### 3.4. 데이터셋 구성 방식

위 (14-17)에 기술한 네 가지 유형의 동사 부류에 기반하여 구조 점화와 그 효과를 관찰하기 위해 테스트셋을 구성하였다. 우선, 위키피디아 문서, 뉴스 기사, 기사 댓글 등의 텍스트<sup>12)</sup>를 모은 코퍼스를 대상으로 하고, 그 안에서 “A-이/가 B-에게 C-을/를 {동사}”와 같은 구조를 갖는 구문을 모두 수집하였다. 이후 앞서 간략히 언급하였지만 수집된 구문 안에서 고빈도로 사용된 동사들을 추출한 뒤 DAC, DAC\_S, 사동, 피동의 네 가지 유형으로 분류하였고, 기준에 따라 최종 동사 목록을 확정하였

12) 언어 모델 KR-BERT-MEDIUM(<https://huggingface.co/snunlp/KR-Medium>)의 훈련 데이터를 활용하였다.

다. 이어 더해서 동사에 연결되어 사용된 각 논항의 선택을 위해서 각 동사별로 A, B, C 세 논항 위치에 나타나는 고빈도 명사를 추출하고, 이 중 전체 문장의 의미를 고려하여 어울리는 명사의 빈도를 반영하여 5개씩 남기는 방식으로 작업하였다. 물론, 이렇게 명사들과 동사의 조합을 구성하는 방법은 PS 문장과 대상 문장 사이에 동일한 논항이나 유사한 의미의 동사가 공유되어 구조와 상관없이 의미적 유사성이 드러나는 등의 문제가 발생할 수 있지만, 고빈도 어휘 사용에서 오는 불가피한 부작용이라는 점에서 따로 수정하지 않았다. 이러한 작업의 일부 예시를 <표 1>을 통해 확인할 수 있다.

**<표 1> 코퍼스 내 특정 구문에 대한 고빈도 명사와 동사 조합 예시**

유형	주격 논항(A)	여격 논항(B)	대격 논항(C)	동사
DAC	감독, 팬, 선수, 회장, 철수, 영희	기자, 의원, 학생, 어머니, 철수, 영희	편지, 문자, 책, 휴대폰, 차키	보내다
DAC_S	감독, 정부, 사람들, 환자, 의원	국민들, 선수들, 전문가, 부모, 동료들	지지, 전화, 위로, 학대, 검사	받다
사동	동료들, 공격자, 가족, 친구, 매도인	피고, 형사, 이용자, 동료들, 철수, 영희	정보, 정체, 과거, 범죄, 증거	숨기다
피동	서울시장, 아저씨, 선발투수, 철수, 영희	고참, 형, 시민, 차장, 철수	머리, 종아리, 코	얻어맞다

이후, 선택된 동사와 도입 가능한 논항들을 조합하여 “A-이/가 B-에게 C-을/를 {동사 과거형}” 형태의 템플릿에 넣어 문장들을 생성하였다. 이에 따라 (18)의 예시와 같은 형식의 문장들이 자동 구성된다.

- (18) a. 감독이 기자에게 편지를 보냈다.
- b. 정부가 국민들에게 지지를 받았다.
- c. 동료들이 철수에게 정보를 숨겼다.
- d. 서울시장이 차장에게 머리를 얻어맞았다.

본 논문은 앞서 강조한 것과 같이 동사의 특성, 공간 이동 관련 의미 요인, 그리고 의미역과 관련된 의미 특성 등이 추상적 구조의 표상에 어떠한 영향을 미치고, 이들이 신경망 언어 모델에 어떻게 반영되어 있는지 확인하는 것이 목적이므로, 어순이나 PS의 복잡성 등 구조 점화 연구에서 많이 이용하는 몇 가지 변인들은 고려하지 않고 동일 어순을 가진 문장 중심으로 실험을 전개하였다.<sup>13)</sup>

지금까지 논의한 조건에 맞추어 가능한 조합을 따르는 모든 문장들을 먼저 생성한 뒤 유형별로 1,000문장씩 추출하였다. 이때 생성된 문장은 모두 따로 점검하여 의미적으로 어색한 문장은 모두 제외시켰다. 그리고 각 유형의 문장들 중 가장 전형적인 의미를 보이는 문장을 선택하여 PS 문장의 목록을 만들었다. 이때 PS는 각 동사마다 하나씩 선택하였다. 또한 각 유형별로 추출한 1,000개씩의 문장에 PS 목록에서 랜덤 추출한 PS 하나씩을 할당하여 최종적으로는 (19)와 같이 PS와 테스트 대상 문

13) 네 가지 유형을 모두 고려하여 현재의 어순을 기본 어순으로 판단하였다.

장이 이어 붙어져 있는 형태로 테스트셋을 구축하였다.

(19) 철수가 영희에게 휴대폰을 보냈다. 동료들이 철수에게 정보를 숨겼다.

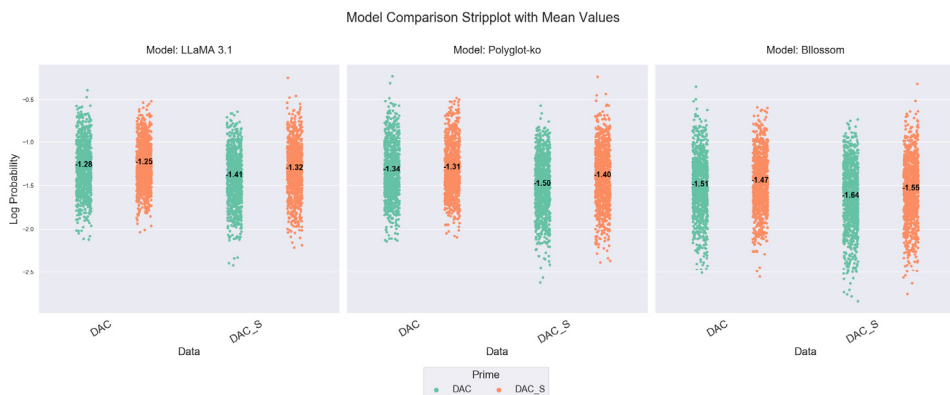
즉, PS와 대상 문장의 유형이 동일한 쌍이 1000개, 그렇지 않은 것이 3000개 구축된 것이다. 이러한 문장 쌍을 이용하여 PS 문장이 제시되었을 때, 대상 문장이 정문, 혹은 존재 가능한 문장으로 인식되는 확률을 계산하고 이를 통해 점화 효과 유무를 판단할 수 있게 환경을 구성하여 실험을 전제하였다.

## 4. 실험 및 결과

### 4.1. 구조 점화와 점화 효과의 유무

본 논문의 실험은 무엇보다도 본 논문에서 사용하는 LLM들이 모두 구조 점화를 인식하고, 이에 따라 점화 효과를 나타내는지 먼저 확인할 수 있게 구성하였다. 이를 위해 Sinclair et al.(2022)의 분석을 따라 평가 대상 자료의 유형과 동일한 유형의 PS를 제공하였을 때의 점화 효과를 이들 간의 유형 불일치가 일어난 상황에서의 효과와 직접 비교하는 방법을 사용하였다.

이때, 문장 존재 확률을 계산하기 위해 각 LLM에 (19)와 같은 문장쌍 형태의 연쇄를 입력하고, 문장쌍 내에서 두 번째 문장에 대해서만 위 3.2에서 기술한 것과 같은 문장 존재 확률을 계산하여 구조 점화 효과를 관찰하였다. LLaMA 3.1, Polyglot-Ko, Blossom 세 개 모델 모두 추가적인 훈련 또는 미세 조정 없이 테스트셋 데이터를 바로 입력하여 계산된 확률을 얻는 방식을 사용하였다. 이에 따라, 먼저 DAC 유형 PS와 DAC\_S 유형 PS를 [DAC 유형]대상과 [DAC\_S 유형]대상에 모두 적용하고, 이들



<그림 1> [DAC 유형]대상과 [DAC\_S 유형]대상의 PS에 따른 문장 생성 확률

의 점화 효과를 비교하였다. 이는 <그림 1>과 같이 표상되고, 이들 간의 비교가 유의미한지 확인하는 T 검정은 <표 2>에서 확인할 수 있다. <표 2>는 Welch's T-test에 따라 점화 효과 사이의 차이가 유의미한지를 나타내는 p-value를 보고한 것이다.

**<표 2> DAC와 DAC\_S 유형간의 점화 효과 T 검정**

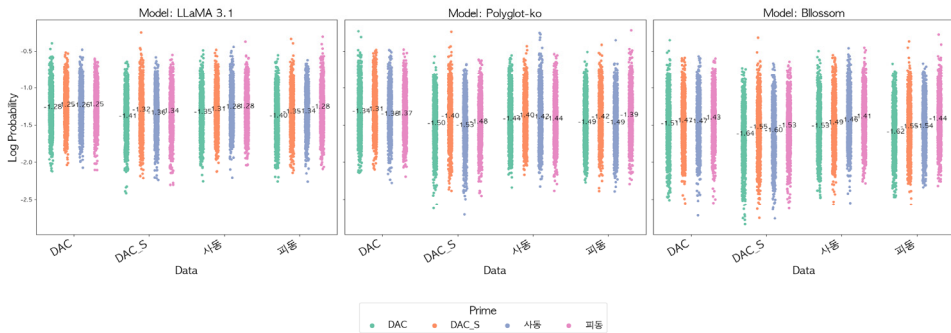
점화 효과 1 vs. 점화 효과 2	LLaMA 3.1	Polyglot-Ko	Blossom
DAC_PS, DAC_대상 vs. DAC_PS, DAC_S_대상	1.278e-20	6.874e-29	3.043e-16
DAC_S_PS, DAC_S_대상 vs. DAC_S_PS, DAC_대상	1.460e-07	7.673e-11	1.099e-08

먼저, <그림 1>을 통해 확인할 수 있는 것은 세 LLM에서 모두 [DAC 유형]<sub>대상</sub>과 [DAC\_S 유형]<sub>대상</sub> 둘 다 DAC\_S 유형의 PS가 사용된 경우 점화 효과가 더 높다는 것이다. 이러한 비대칭적 점화(asymmetric priming) 효과는 동일한 유형의 PS가 사용된 경우 더 높은 점화 효과가 나타나는 대칭적 점화(symmetric priming) 효과, 즉 [DAC 유형]<sub>대상</sub>에 DAC 유형 PS가 적용되는 경우가 DAC\_S 유형 PS가 적용되는 경우보다 점화 효과가 더 크고 그 역도 동일한 경우보다는 약하다고 할 수 있지만, 구조 점화가 실제 문장 생성에 영향을 미친다는 점은 충분히 보장해 줄 수 있다(cf. Sinclair et al. 2022). 특히, 본 논문은 이를 DAC 유형과 DAC\_S 유형은 서로 독립적인 구조이기는 하지만 매우 유사한 구조로 인식된다는 것으로 볼 수 있고, 이에 따라 공간 이동이 존재한다는 유사성이 이 두 유형의 동사구 형성에서 가장 중요한 요인이라는 입장을 받아들인다. 즉, 2장에서 가정했던 것처럼 공간 이동의 방향성이나 의미역 중 어느 하나, 혹은 둘 다 구조적인 차이를 야기하는 요소로 작용하기는 하지만 공간 이동이 존재한다는 것 그 자체가 가장 중요한 요인이라 이 두 구조의 차이가 아주 명확하게 드러나지 않는다고 보는 것이다.

이에 더해 <그림 2>는 나머지 유형들까지 모두 교차 분석하여 비교한 결과를 나타낸 것으로, 이를 통해 대상 유형과 PS 유형이 동일한 경우에 더 높은 점화 효과가 나타나거나 적어도 비대칭적 점화 효과는 드러난다는 것을 대략적으로 확인할 수 있다.<sup>14)</sup> 더군다나 (비)대칭적 점화 효과가 나타나는 분포 역시 세 LLM 모두에서 유사한 패턴으로 드러난다는 점에서 본 논문에서 실험한 네 유형 모두 다 독립적인 구조를 가진 유형으로 구분된다는 결론을 뒷받침해 준다. 이는 본 연구에서 가정한 각 모델간의 차이에도 불구하고 세 LLM 모두 한국어의 각 유형이 가진 구조적 특성을 이해한다는 점에서, 이들 구문이 구성되기 위한 동사구의 구조적 차이를 인식하게 하는 요인을 모두 언어 지식으로 반영하고 있다고도 해석할 수 있을 것이다.

14) 나머지 유형들 간의 점화 효과 비교에 대한 T-검정 결과는 몇몇 환경에서 p-value가 0.05 이상으로 나타나는 등 LLM 모델 간 차이가 발견되기도 하였다. 그러나 세 모델 모두에서 p-value가 0.05 이상으로 나타난 환경은 DAC\_S 유형을 PS로 하여 [DAC\_S 유형]<sub>대상</sub>과 [피동 유형]<sub>대상</sub>을 비교한 경우가 유일하다. 이에 대해서는 후속 연구로 남긴다.

Model Comparison Stripplot with Mean Values



<그림 2> 모든 대상 유형과 PS의 조합에 따른 문장 생성 확률

물론 추가적인 분석이 필요한 특이점도 존재한다. 특히, 세 LLM에서 모두 [DAC 유형]대상은 사동 유형의 PS가 사용된 경우에 더 높은 점화 효과가 드러났고, [DAC\_S 유형]대상에서도 유사한 경향성을 확인할 수 있었다. 그러나 이를 가지고 본 연구에서 사용된 LLM이 구조 점화를 정확하게 반영할 능력이 없는 것이라 단정 지을 수는 없다. 앞서 언급한 것과 같이, 기존의 많은 연구에서 DAC 유형이 사동 구조를 포함하고 있을 가능성이 존재하는 것이다. 실제, 다양한 언어의 DAC에 대한 연구에서, DAC에 반영된 동사구의 구조에는 사동성이 반영되어 있어야 한다는 분석이 존재한다. 예를 들어 Beck & Johnson(2004)은 영어의 DOC와 DAC 구문을 (20)과 같이 표상한다(cf. Pinker 1989).

- (20) a. Satoshi sent Thilo the Damron Guide. [DOC]
  - ⇒ [Satoshi’s sending the Damron Guide] CAUSE [BECOME [Thilo HAVE the Damron Guide]]
- b. Satoshi sent the Damron Guide to Thilo. [DAC]
  - ⇒ [Satoshi’s sending the Damron Guide] CAUSE [BECOME [the Damron Guide is AT Thilo]] (Beck & Johnson 2004: (1), (13))

즉, 2장에서 가정한 여격 논항의 대격 교체 가능성을 통해 구조적으로 사동 의미의 동사구 내 반영 여부를 확인할 수 있다는 가능성을 포기하고 한국어의 DAC 유형에도 본연적으로 사동 의미가 내재되어 있다고 봐야 할 수도 있는 것이다. 특히, 본 연구에서는 동사 목록 선정 상 DAC 유형과 같이 대격 논항의 공간 이동이 포함될 수 있는 여지를 차단하려 했다는 점에서 이러한 사동 의미가 구조적으로 표상된 구조가 DAC 유형의 구조와 본질적으로 유사하다는 결론은 의미가 있다고 볼 여지가 충분하다. 또한 (20)에서 기술된 이론적 배경에서 DAC 유형이 사동성과 밀접한 관계가 있고, 이 특성이 동사 구조의 형성에 영향을 준다고 보는 것이 불가능한 가정이 아니라는 점에서,<sup>15)</sup> [DAC 유형]대상의 경우 사동 유형의 PS가 매우 높은 점화 효과를 야기하는 것을 이해할 수 있을 것이다. 이와 유사하게

DAC\_S 유형의 경우도 사동 유형 PS의 영향을 상당히 받는 점을 포착할 수 있다. 이동의 방향만 다를 뿐 본질적으로 DAC와 유사하게 의미를 표상할 수 있기 때문이다.

(21) 철수는 영희에게 책을 받았다. [DAC]

⇒ [Yenghui's sending the book] CAUSE [BECOME [the book is AT Chelswu]]

이러한 결과에 따라, 적어도 현재 사용한 LLM에서는 분석 대상 유형들의 추상적 구조를 인식하고 구별할 수 있는 통사적 정보가 주어져 있고, 이를 통해 구조 점화가 적용된다는 점을 확인할 수 있다.

## 4.2. 구조 점화와 형태 표지

4.1절에서 간단히 살펴보았듯이, LLM에서는 적어도 DAC 유형과 관련 유형 모두에서 구조 점화가 유의미한 영향을 미친다는 것을 확인할 수 있었다. 그러나 DAC 유형과 DAC\_S 유형 간의 관계나 사동 유형 PS의 특수성 등 명확한 결론을 내리기 어렵게 하는 요인이 있어 구조 점화가 실질적으로 적용되는지 확인할 수 있는 추가 실험을 시도하였다.

이를 위해 먼저 고려한 것은 Bock(1986)에서부터 논의된 형태 표지의 특성이다. 앞서 간략히 언급하였지만, Bock(1986)은 영어의 DAC 문장들의 경우 PP의 핵(head)으로 'to'나 'for' 어느 것이 선택 되어도 동일한 결과가 나왔고, 이를 통해 전치사와 같은 형태 표지의 의미, 혹은 이에 따라 구분되는 의미역의 차이는 구조 점화에 직접적인 영향을 미치지 못한다고 결론 내렸다. 그러나 이와 관련해서 Jumelet et al.(2024)에서는 신경망 언어 모델의 경우 구조 점화에 있어서 전치사와 같은 형태 표지가 오히려 더 직접적인 영향을 미친다고 지적하고 있기도 하다. 즉, 이러한 상반된 두 입장이 존재한다는 점에서 표지 교체와 구조 점화가 LLM 기반 실험에서 어떠한 양상을 보일지 확인하는 것은 의미가 있을 것이다. 이에 본 논문은 PS로 사용되는 문장의 여격 논항에서 '-에게' 표지를 '-한테' 표지로 교체하였을 때 드러나는 구조 점화의 변화 양상을 확인해 보았다.

(22) DAC 유형

- a. 철수가 영희-에게/한테 책을 주었다.
- b. 철수가 영희-에게/한테 편지를 썼다.
- c. 철수가 영희-에게/한테 영어를 가르쳤다.

(23) DAC\_S 유형

- a. 철수가 영희-에게/한테 책을 받았다.
- b. 철수가 영희-에게/한테 차키를 뺐었다.
- c. 철수가 영희-에게/한테 피아노를 배웠다.

15) 이러한 가정은 Jung & Miyagawa(2004)에서 다른 한국어의 DOC에 대한 분석에 재고가 필요하다는 결과로 이어질 수 있다. (20)에서 제시된 의미와 DAC와 DOC의 의미적 차이, 그리고 한국어의 DOC를 분석 대상에서 제외한 이유와 함께 6장에서 다시 논의할 것이다.

## (24) 사동 유형

- a. 철수가 영화-에게/한테 발표를 시켰다.
- b. 철수가 영화-에게/한테 코트를 입혔다.
- c. 철수가 영화-에게/한테 임무를 맡겼다.

## (25) 피동 유형

- a. 철수가 영화-에게/한테 발을 밟혔다.
- b. 철수가 영화-에게/한테 구타를 당했다.
- c. 철수가 영화-에게/한테 기회를 빼앗겼다.

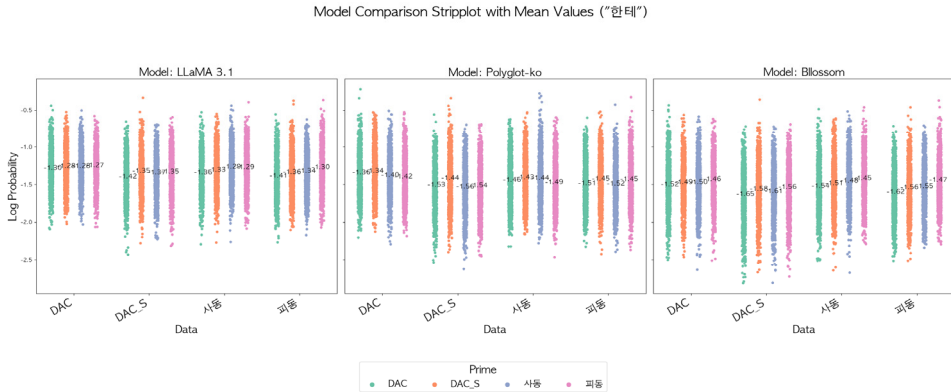
위의 예시에서 확인할 수 있듯이, 본 실험에서 사용된 모든 유형에서 여격 논항 '-에게'는 '-한테'로 의미차이 없이 교체될 수 있다. (이제부터 편의상 이 논항을 구분하기 위해 순서대로 Goal, Source, Patient, 'by-phrase' 대응, Agent라고 지칭할 것이다. 물론, 이들 논항의 기본적인 성격을 드러내기 위해서 의미역 관련 용어가 주로 사용되지만, 이는 구분의 편의성을 위한 것이지 의미역이 실제 중요한 역할을 한다고 가정하는 것은 아니다.) 또한, 네 유형의 여격 논항 모두 '-한테'가 '-에게'와 큰 차이 없이 사용될 수 있다는 분포적인 유사성만이 아니라, 이 두 표지는 유정물로 기능할 수 있는 논항과만 결합한다는 점에서도 유사한 모습을 보이는 등 의미적으로 상당히 유사하다.

- (26) a. \*자동차-에게/한테  
c. 영화-에게/한테

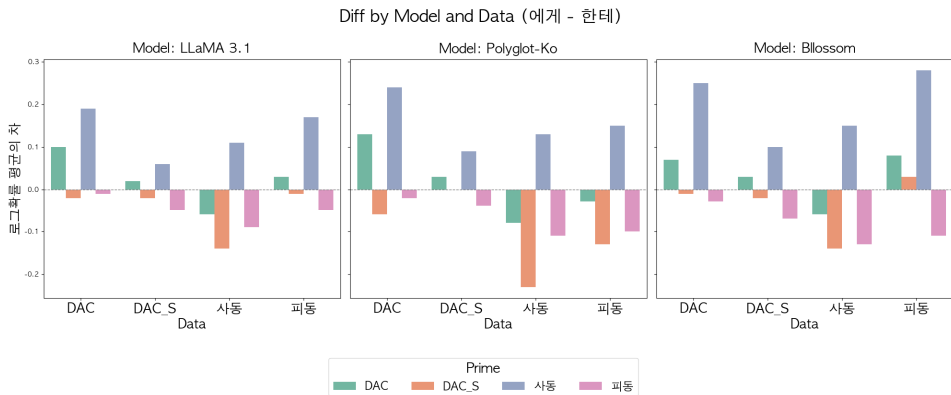
- b. \*서울역-에게/한테  
d. 학생-에게/한테

이러한 점에서 '-한테' 표지로의 교체를 이용한 실험은 기존 관련 연구에서 의미나 기능이 다른 형태 표지를 사용한 것과 다르게 형태만 다르고 의미와 기능이 최대한 유사한 형태 표지라는 점에서 차이가 있다. 특히, 의미적으로 동일한 표지의 사용은 다음과 같은 내용을 더 명확하게 확인시켜줄 수 있다. 우선 Bock(1986) 등에서 제시한 것처럼 의미/기능이 다른 표지도 구조 점화에 영향을 주지 않는다면 의미/기능이 유사한 표지의 사용도 동일한 결과를 야기할 것이라고 쉽게 예측할 수 있다. 또한 Jumelet et al.(2024)의 결과가 신경망 언어 모델의 언어 지식에 대한 차이에 기인한다기보다는 실험 환경에서 각 표지의 의미/기능의 차이가 더 과장되게 적용되었을 경우도 배제할 수 없기 때문이다. 마지막으로, Jumelet et al.(2024)의 분석이 본 논문에서 사용하는 LLM들도 적용된다면 유사한 의미를 가진 두 표지의 특성상 표지의 의미가 가장 주요한 요인인지, 아니면 다른 요인까지 고려해야 하는 것인지 확인하기 쉬워진다는 점도 고려하였다.

이러한 입장에서 실험은 PS로 제시되는 예문의 여격 논항의 표지를 모두 '-한테'로 교체하여 점화 효과를 계산하고, 이를 표지 교체 전 환경의 점화 효과와 비교하는 방식으로 이루어졌다. 우선 PS에서 여격 논항의 표지를 '-한테'로 교체한 후의 점화 효과의 값은 <그림 3>과 <그림 4>와 같이 드러난다.



<그림 3> 모든 대상 유형과 ‘한테’ PS의 조합에 따른 문장 생성 확률



<그림 4> ‘에게’ 표지 대비 ‘한테’ 표지 사용 효과

특히나, <그림 4>는 <그림 3>을 활용하여 각 환경에서 교체의 의해 야기된 “점화 효과의 차이”가 어느 정도인지 보여주려는 의도를 담고 있다. 이를 위해, 각 환경에서 ‘에서’ 표지가 사용된 경우의 평균 문장 로그확률에서 ‘한테’가 사용된 대응 구문의 평균 문장 로그확률을 뺀 수를 그래프의 Y축에 표시하였다. 이에 따라 그래프의 값이 0에 가까울수록 ‘한테’ 표지로서의 교체가 점화 효과의 변화에 큰 영향을 주지 않았다는 것으로 해석할 수 있다. 또한, 만약 그래프의 값이 음수로 나온다면 ‘한테’ 교체로 인해 점화 효과가 증가했다는 것으로, 양수로 나온다면 점화 효과가 감소했다는 것으로 해석할 수 있다고 가정하고 논의를 전개하였다.

이를 통해 확인된 결과를 간단히 기술하면 다음과 같다. 일견 ‘한테’ 표지 교체는 대상 유형이나 PS의 선택에 따라 차이가 크게 드러나, 특별한 일관성이 있다 보기 어려워 보인다. 그러나 모든 LLM

에서 거의 유사한 패턴을 확인할 수 있다는 점에서 이를 단순히 우연적인 분포라고 보기도 어려울 것이다. 더군다나 이와 관련한 한 가지 특이한 경향성을 찾을 수 있다. 우선 [DAC 유형]<sub>대상</sub>의 경우, ‘한테’ 표지가 사용된 경우 DAC\_S 유형의 PS가 사용된 환경과 피동 유형의 PS가 사용된 환경을 제외하고는 점화 효과가 매우 약해진다. 그리고 이는 DAC\_S 유형과 사동 유형에서 도입되는 여격 논항의 기본적인 (의미) 특성이 영향을 주어 이들이 PS로 사용되는 경우에 ‘한테’ 교체를 통해 드러나는 부정적 변화의 폭을 억제하였다고 해석할 수 있을 것이다. 즉, ‘한테’ 표지가 Source 논항이나 영어의 ‘by-phrase’에 대응되는 피동 주어와 같이 사용될 때에는 점화 효과가 교체 전과 동일하게 유지되었다는 것이다. 그리고 이러한 경향성은 [사동 유형]<sub>대상</sub>과 [피동 유형]<sub>대상</sub>에서도 동일하게 나타난다. Source나 ‘by-phrase’에 대응되는 논항에 ‘한테’ 표지가 사용된 PS가 나타나는 경우 점화 효과의 약화가 억제되는 모습을 확인할 수 있다. 즉, 이 세 대상 유형에서 모두 Source나 ‘by-phrase’에 대응되는 논항이 PS로 사용되는 환경의 특수성이 포착되는 것이다.

이러한 경향성은 (적어도) 본 논문에서 사용한 LLM에서는 ‘한테’ 표지가 기본적으로 Source나 ‘by-phrase’에 대응되는 논항이 보이는 의미와 ‘만’ 양립 가능하다는 것으로 이해할 수 있다. 즉, ‘에게’와 ‘한테’ 사이의 의미적 유사성이 모두 포착할 수 없게 ‘한테’의 의미가 매우 좁게 습득되었다는 것이다. 이에 따라 두 표지의 의미적 유사성을 가정한 원래 의도와는 다른 방향으로 실험이 전개되었다는 점이 확인된다. 그러나 지금까지 불완전하게 습득된 의미가 결과적으로 구조 점화 과정에 영향을 미쳤다고 가정한다면 ‘한테’ 표지가 사용된 경우 앞선 세 대상 유형에서 특정 의미의 논항이 사용된 PS가 도입되는 경우에만 점화 효과가 최대로 유지되는 경향성을 보이는지 이해할 수 있게 된다.

또한, [DAC\_S 유형]<sub>대상</sub>에 ‘한테’로 여격 논항이 교체된 PS가 사용된 경우에는 큰 변화 자체가 일어나지 않은 점도 쉽게 설명될 수 있다. [DAC\_S 유형]<sub>대상</sub>은 ‘한테’ 표지가 사용된 PS가 나타나는 경우에도 ‘에게’ 표지가 사용된 경우와 비교하여 점화 효과의 변화가 크지 않다는 점에서 특이하다. 그러나 이 역시 앞서 가정한 것처럼 ‘한테’ 표지의 의미가 점화 효과에 직접적으로 영향을 미친다면 쉽게 이해할 수 있다. [DAC\_S 유형]<sub>대상</sub>에서는 그 여격 논항이 애초에 Source 논항에 대응되어야 하므로, ‘한테’ 표지 사용에 영향이 미치는 의미가 이미 대상 문장의 내부 구조 형성에 반영되어 있다고 볼 수 있다. 다시 말해, Source 논항이 사용되어 점화 효과에 미치는 영향이 존재한다면, 그것이 무엇이든 간에, [DAC\_S 유형]<sub>대상</sub>에 이미 반영되어 있고, 이에 따라 PS 내부의 표지 교체가 큰 영향을 미치지 못할 것이다. 그리고 이러한 DAC\_S 유형의 특수성이 드러난다는 것은 적어도 DAC 유형과 DAC\_S 유형간의 구조적인 차이를 LLM이 파악하고 있다는 결론으로 이어진다. DAC\_S 유형에서 소위 Source 논항이 존재하기 위한 여러 조건, 특히 공간 이동의 방향성에 대한 차이를 인식하고, 이러한 차이가 구조에 반영되어 DAC 유형과 DAC\_S 유형간의 차이가 점화 효과의 차이로 드러났다고 볼 수 있는 것이다.

지금까지의 논의를 통해 DAC 유형과 DAC\_S 유형이 어떠한 차이를 가지는지 LLM은 구별하고 있다는 것, 그리고 LLM의 경우에도 형태 표지의 변화가 구조 점화에 직접적인 영향을 끼친다는 점을 확인하였다. 특히, 후자와 관련해서 Source나 ‘by-phrase’와 무관한 여격 논항이 PS에 포함된 경우에 ‘한테’ 표지의 사용이 점화 효과의 약화를 야기하였고, Source 논항으로 여격 논항이 해석되어야 하는 [DAC\_S 유형]<sub>대상</sub>의 경우에는 ‘한테’ 표지로의 교체가 점화 효과에 의미 있는 변화를 일으키지 못

한다는 점도 확인하였다.

이러한 결과는 ‘-한테’ 표지의 “의미”가 직접적인 영향을 미친다는 결론으로 이어진다. 적어도 신경망 언어 모델에서는 형태 표지의 변화가 ‘의미 요인’에 따라 구조 점화에 영향을 미치는 것이다. 그리고 이러한 다양한 양상은 구조 점화가 적용되지 않는다면 그 실현 패턴을 파악할 수 없다는 점에서, 설명 형태 표지 의미가 개입한다 하더라도 구조 점화의 적용 없이 의미 자체를 주요 요인으로 볼 수 없다는 점도 확인된다.

## 5. LLM에서의 구조 점화 분석에 대한 논의 및 정리

본 논문에서는 세 LLM을 기반으로 DAC 유형과 그 관련 유형들의 관계를 중심으로 구조 점화의 적용 가능성에서부터 적용에 영향을 다양한 요인을 살펴보았다. 특히 의미와 관련해서도 통사 구조의 실현에 직접적으로 영향을 미치는 요인만 골라 실험에 사용된 LLM들이 구조관련 지식을 어떻게 반영하는지 확인하려고 노력하였다.

우선 DAC 유형과 DAC\_S 유형사이에 점화 효과가 실제 유의미한 차이를 드러난다는 점에서, 이 두 유형이 동일 구조를 공유하지 않는다는 점이 드러났다. 이러한 LLM을 통한 분석이 실제 인간 언어와 얼마나 유사한지 확인하는 것은 또 다른 문제이지만, 이 두 구문이 공간 이동이라는 인지적 유사성에도 불구하고 서로 다른 구조로 인식될 수 있다는 것은 인지적 요인 그 자체보다 구조 형성에 영향을 미치는 다양한 요인들이 상호 작용한 결과로 도출되는 구조 그 자체가 구조 점화에 더 중요한 요인이라는 점을 다시 한 번 상기시켜 준다.

여기서 한 가지 더 언급할 것이 있다. 앞서 살펴보았듯이, DAC 유형과 DAC\_S 유형은 대칭적 점화 효과까지는 보이지 못한다는 점에서 두 유형 간의 ‘유사성’에 대해서는 더 자세한 논의가 필요하다. 특히, 4장에서는 이 두 유형의 경우 공간 이동의 존재가 가장 중요한 요인이고 의미역이나 방향성의 차이가 이들을 서로 다른 구조로 LLM이 인식하게 하기는 하지만 부차적인 요인이라 제안하였는데, 이에 대한 더 자세한 후속 연구가 요구되는 것이다.

이와 관련해서 한 가지 생각해 볼 수 있는 것은, 앞선 논의를 따라 공간 이동 관련 개념이 LLM에서 중시된다는 것을 받아들이고, 이에 따라 의미역보다는 이동의 방향성이 이 두 유형의 구조를 구분하는데 있어서 더 중요한 역할을 한다고 가정하는 것이 될 것이다. 이는 앞서 ‘-한테’ 표지 교체에서 드러난 공간 이동과 관련된 특성의 중요성과 함께, (적어도) 본 연구에서 사용한 LLM에서는 인간 언어와 유사하게 의미역이라는 개념은 구조 점화에 큰 영향을 끼치지 못한다는 결론(Bock 1986 *et seq.*)으로도 이어질 수 있다는 점에서 더 세밀한 후속 연구가 이루어져야 할 것이다.

이에 더해, 한국어의 DAC 유형과 DAC\_S 유형 모두 사동 의미를 구조적으로 표상하고, 이러한 사동 구조가 이 두 DAC 유형의 동사구 구조 형성에 주요한 역할을 한다고 해석할 수 있다는 점도 확인하였다. 그런데, 이는 또 다른 질문으로 이어진다. 사동 유형 PS는 [DAC 유형]<sub>대상과</sub> [DAC\_S 유형]<sub>대상</sub> 모두에 강한 구조 점화 효과를 보이는데, 그 역은 성립하지 않는 이유에 대한 의문이 제기되는 것이다. 그런데, 이러한 비대칭적 점화 효과에 대해서는 조금 다른 해석도 시도해 볼 수 있다. 많은 이론 언어

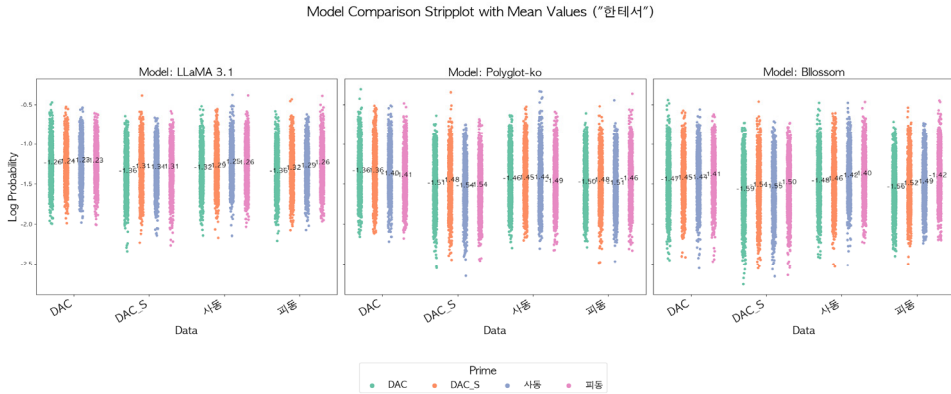
학에서의 분석처럼 이 두 DAC 유형은 공간 이동의 의미에 사동 의미가 더해진 즉 두 요인 모두 포함된 특성을 가지고(Beck & Johnson 2004 등), 이에 따라 이 두 요인이 모두 구조적으로 드러난다고 가정해 보자. 그렇다면 사동 구조는 자신을 포함한 더 복잡한 구조(i.e., 사동 + 공간 이동 구조)에 점화 효과를 보일 수 있지만, 더 복잡한 구조가 단순히 사동 의미만 반영된 구조에 점화 효과를 보이지 못하는 것으로 비대칭성에 대한 해석도 가능한 것이다(cf. 각주 7). 물론 이러한 해석이 타당성을 얻기 위해서는 몇 가지 추가적인 논의가 필요하다. 앞서 2장에서 언급하였듯이 사동 유형에 속한 동사의 일부의 경우 그 여격 논항은 사동 의미를 제외하고도 Goal에 대응되는 논항으로 해석될 여지가 있다. 물론 이러한 문제는 [DAC\_S 유형]<sub>대상</sub>에 적용될 수 없기는 하지만, 앞서 언급하였듯이 적어도 LLM 기반 분석에서도 의미역이 구조 점화에 큰 영향을 주지 못한다는 논의와 연결될 수 있다.

이에 더해 피동 유형에 대해서도 논의할 부분이 많다. 2장에서 언급하였듯이 피동 유형은 대격 논항을 포함하고 있다는 점에서 독특하다. 그리고 영어의 경우 이러한 환경은 DOC에 피동화가 적용된 경우가 대부분이라는 점에서도 관련 구문과의 추가적인 비교/대조가 요구된다. 그러나 기반 언어인 영어의 영향을 가장 많이 받는 모델인 LLaMA에서도 나머지 두 LLM과 유사한 결과를 보였다는 점에서 이에 대한 논의는 매우 세밀하게 이루어져야 할 것이다. 이에 더해 <그림 3>에서 확인할 수 있듯이, 피동 구문 PS가 [DAC 유형]<sub>대상</sub>에 대해서는 꽤나 높은 점화 효과를 보인다는 점도 확인되는데, 이 역시 세 LLM에서 모두 드러난다는 점에서 무시할 수 없으므로, 그 이유에 대한 추가적인 논의가 이루어져야 할 것이다(각주 8과 14도 참조).

마지막으로 형태 표지의 의미가 구조 점화에 상당한 영향을 줄 수 있다는 점도 확인하였다. 본 논문에서 사용한 LLM은 적어도 ‘-한테’ 표지가 실제 담당하는 의미 보다 Source와 ‘by-phrase’에 대응되는 의미/기능들만 강조되는 것을 확인하였고, 이에 따라 PS 문장에서 ‘-에게’ 표지를 ‘-한테’ 표지로 교체하는 것으로 인해 상당한 변화가 드러난다는 것도 확인할 수 있었다. 동시에 ‘-한테’는 ‘by-phrase’와 관련된 [사동 유형]<sub>대상</sub>에서의 구조보다 Source를 포함한 [DAC\_S 유형]<sub>대상</sub>의 구조와 더 민감하게 상호작용한다는 점도 확인하였는데, 이는 Source라는 개념이 의미역보다는 순수하게 공간 이동상의 기점이라는 측면을 강조하는 요인으로써 구조 형성에 영향을 끼친다고 해석될 수 있다는 점에서 더 주의해야 한다. 기능상 행위주에 대응되는 ‘by-phrase’보다 Source에 대응되는 논항이 공간 이동에 기반한 구조/인지 개념과 더 밀접하다고 할 수 있으므로, 공간 이동에 대한 인식 여부가 LLM에서 더 직접적으로 작용한다는 가정을 뒷받침하는 것이다. 그런데 이러한 주장은 만약 Source에 대응되는 의미를 강제하는 형태 표지가 존재하고, ‘-에게’가 이러한 형태 표지로 교체되는 경우 [DAC\_S 유형]<sub>대상</sub>에서는 적어도 ‘-한테’와 유사한 특성을 보일 것이라는 예측으로 이어진다. 그리고 이러한 표지는 실제 한국어에서 어렵지 않게 찾을 수 있다.

- |                               |            |
|-------------------------------|------------|
| (28) a. 철수는 영화-한테/한테서 책을 받았다. | [DAC_S 유형] |
| b. 철수는 영화-한테/*한테서 책을 주었다.     | [DAC 유형]   |
| c. 철수는 영화-한테/*한테서 노래를 시켰다.    | [사동 유형]    |
| d. 철수는 영화-한테/*한테서 발을 밟혔다.     | [피동 유형]    |

진술한 예측이 타당한지 확인하기 위해 PS의 '-에게' 표지를 '-한테서' 표지로 교체하고 이들의 점화 효과를 비교해 보았다. 다음 페이지의 <그림 5>와 <그림 6>에서 확인할 수 있듯이 앞서 '-한테' 표지 교체에 사용한 방식을 '-한테서' 교체에도 그대로 적용하였다.



**<그림 5> 모든 대상 유형과 '-한테서' PS의 조합에 따른 문장 생성 확률**



**<그림 6> '-에게' 표지 대비 '-한테서' 표지 사용 효과**

실험의 결과를 살펴보기 전에 먼저 한 가지 언급해야 할 내용이 있다. 사실 '-한테서'의 교체는 DAC\_S 유형이 PS인 경우를 제외하고 모두 비문인 PS가 사용된다는 점에서 적절한 분석 환경이라 볼 수 없을 수 있다. 이러한 점에서 '-한테서'로 교체된 경우 <그림 5>에서 볼 수 있듯이 점화 효과가 강해지는 패턴이 전반적으로 드러난다고 해서 이를 의미 있는 결과로 받아들일 수는 없다. 그러나 [DAC\_S 유형]대상에 이미 여격 논항이 Source에 대응되는 논항이라는 사실과, 이에 따른 구조 형성 관련 정보가 반영되어 있다는 점에서 '-에게' 표지가 사용된 환경과 큰 차이를 보이지 않을 것이라는 예

측은 <그림 6>의 DAC\_S 유형에 대한 결과를 통해 쉽게 확인할 수 있다. Polyglot-Ko의 경우 약간 특이점이 보이기는 하지만, 이 역시 전체적으로 보면 형태 표지 교체가 [DAC\_S 유형]<sub>대상</sub>에서 가장 적은 차이를 드러낸다는 기본 패턴 자체를 크게 벗어나지는 않는다고 볼 수 있다. Polyglot-Ko가 다른 두 LLM과 다르게 한국어 데이터를 기반으로 사전학습되었다는 점이 영향을 미쳤다고 보기도 어렵다.

이에 더해, ‘-한테’ 표지가 사용된 경우와 비교해도 흥미로운 유사점을 찾을 수 있다. ‘-한테서’ 표지로 교체가 일어난 경우 다른 대상 환경은 다르지만 오직 [DAC\_S 유형]<sub>대상</sub>에서는 ‘-한테’ 표지가 사용된 경우와도 상당히 동일한 패턴을 보인다는 점을 확인할 수 있는 것이다. 앞서 언급하였듯이, <그림 4>에서도 [DAC\_S 유형]<sub>대상</sub>이 형태 표지 교체의 영향이 가장 적다는 점을 쉽게 확인할 수 있었으므로, 이는 본 실험에서는 (적어도) 형태 표지의 의미가 구조 점화와 상호 작용하고 이에 따라 점화 효과에 영향을 주고 있다는 점을 다시 한 번 확인시켜 준다.

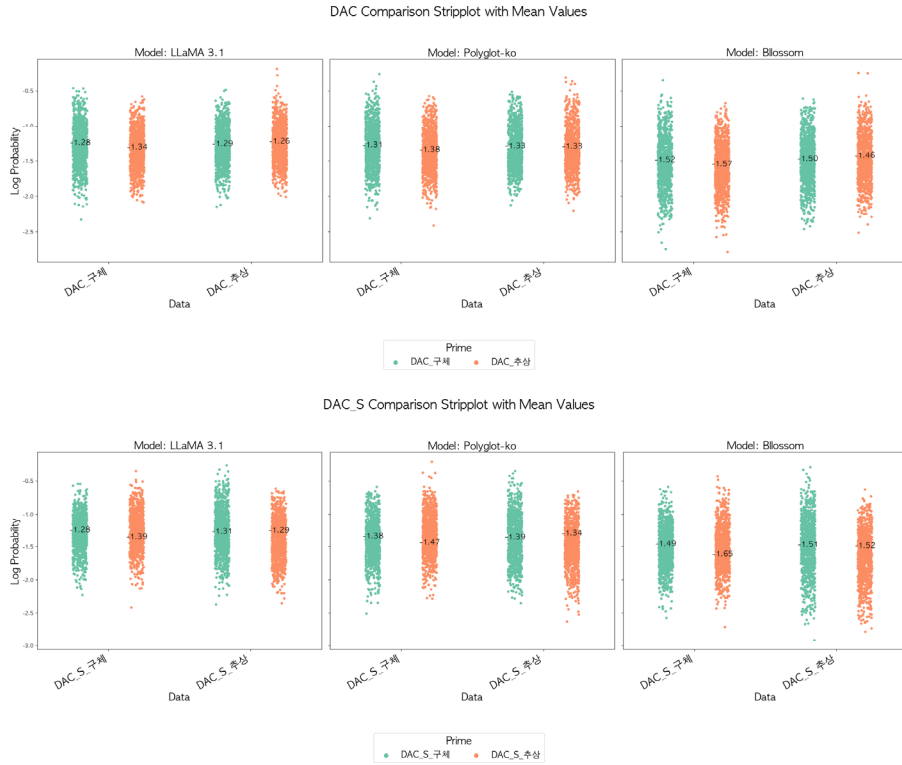
## 6. 정리 및 남은 문제

본 논문은 한국어에 적용 가능한 세 종류의 LLM을 이용하여 구조 점화가 신경망 언어 모델에 적용되는지, 그리고 이를 한국어의 수여 구문과 그 관련 구문을 통해 확인할 수 있는지 확인해 보았다. 몇 가지 실험을 통해 실제 구조 점화가 적용가능하고, 이와 연관된 구조 관련 지식을 LLM이 습득하고 있다는 것도 확인할 수 있었다.

이에 따라, 추후 연구를 통하여 LLM을 추가 훈련하거나 미세 조정하는 등, 조건 혹은 구조의 유형에 따라 입력 데이터를 달리해 가며 LLM의 반응, 곧 문장 생성 확률에 나타나는 변화를 관찰해볼 수 있을 것이다. PS 한 문장씩을 부여한 경우를 원 샷 학습(one-shot learning)으로 보아 추가적인 PS와 같은 맥락 문장들을 부여하는 퓨 샷 학습(few-shot learning)에 해당하는 맥락 내 학습(in-context learning) 역시 시도해 볼 만 하다.

그러나 몇 가지 문제점도 남아 있다. 우선 본 논문에서 분석된 네 가지 유형은 모두 대격 논항을 가지고 있고, 각 유형에서의 대격 논항이 드러내는 특성을 최대한 공평하게 비교하기 위해 추상물과 구체물이 동일한 비율로 대상 자료에 반영될 수 있게 하였다. 그러나 2장에서 간략히 언급한 것과 같이, 공간 이동이라는 측면에서 본다면 구체와 추상은 완전히 동일한 구조적 특성으로 이어진다고 확신하기 어려울 수도 있다. 또한 본 논문은 구조에 영향을 미치는 의미 요소 중 의미역과 공간 관련 인지 요인을 비교하면 후자가 더 직접적인 영향을 미친다고 결론 내리고 있다는 점에서 이를 단순히 무시하는 것도 바람직하지 않다.

이러한 인식에서 본 논문은 구조/추상의 차이를 확인하기 위하여 먼저 DAC 유형과 DAC\_S 유형의 경우 추상물로의 대격 논항만 가진 대상 자료와 구체물로의 대격 논항만 가진 대상 자료를 추가적으로 구축하였다. 그리고 PS도 추상물로의 대격 논항만 포함한 경우와 구체물로의 대격 논항만 포함한 경우를 구분하여 이들 조합 사이에 유의미한 차이가 있는지 확인해 보았다. 이는 만약 구체와 추상의 차이가 구조에 반영되지 않는다면, 구체 포함 대상 자료에 추상 포함 PS를 적용하거나 구체 포함 PS를 적용하는 경우 등에서 점화 효과의 차이가 유의미하지 않을 것이라는 가정에서 출발한 것이고, 그 실험 결과는 다음 페이지의 <그림 7>과 같이 드러난다.



**<그림 7> DAC와 DAC\_S 유형의 추상물/구체물 조합에 따른 문장 생성 확률**

안타깝게도, <그림 7>은 이 실험의 결과가 일관적이지 않다는 것을 보여준다. 예를 들어 Polyglot-Ko에서는 DAC\_S 유형에서는 유의미한 차이를 드러내지만 DAC 유형의 경우는 명확하지 않다. LLaMA의 경우는 모든 유형에서 그 차이가 유의미한지 판단하는 것 자체가 애매하였고, Blossom의 경우도 Polyglot-Ko와 유사하게 DAC\_S 유형에서만 어느 정도 유의미한 차이를 보였다.

이렇게 어느 한 쪽으로 결론을 내리기 어려운 상황이지만, 앞서 내린 공간 이동 관련 인지 정보가 구조 점화에 민감하다는 판단에 따라 더 세밀한 분석이 필요할 것이다. 그런데 이러한 주장을 받아들이면 DAC 유형과 DAC\_S 유형은 공간 이동이 존재한다는 유사점이 큰 영향을 미칠 수밖에 없고, 동시에 의미역보다는 공간 이동 방향의 차이가 두 유형의 구조상의 차이를 야기하는 주요 요인이 되어야 한다. 그리고 이렇게 공간 이동이 중요하다면 DAC\_S가 추상/구체의 구분에 더 민감한 모습을 보이는 이유에 대해서 더 명확한 논의가 필요할 것이다. 단순히 각 유형 내에서 선택된 개별 동사의 어휘 의미에 따라 드러나는 부차적인 의미 특성의 영향도 완전히 배제할 수 없으므로, 더 세밀한 연구가 요구된다. 이를 위해 구체/추상에 따른 차이가 실제 구조 형성에 영향을 미치는지, 그렇다면 그 명확한 이유가 무엇인지, 만약 구체/추상을 구분해야 한다면 이로 인해 드러나는 본 실험의 문제점이 무

엇인지 등 다양한 논의가 이루어져야 할 것이다.

마지막으로 본 연구에서는 한국어의 DOC를 따로 분석하지 않았다. 이는 우선 2장에서 언급한 ‘사동성 반영 구조’에서 생기는 문제가 생각보다 크다는 점이 큰 영향을 미쳤다. Jung & Miyagawa (2004)는 한국어의 ‘보내다’ 동사는 본질적으로 사동 의미를 결여하고 있어서 DOC 형태의 형성이 불가능하다고 지적한다. 그리고 이는 ‘보내다’ 동사는 DAC를 형성할 수 있으니, 한국어의 DAC 유형의 동사구는 사동 관련 구조가 결여되어 있어도 된다는 결론으로 이어진다. 즉, 한국어의 경우 동사의 선택에서부터 실제 구조의 형성에까지 DOC와 DAC 구분에 사동성 반영이 중요한 역할을 한다는 것이다.

그러나 여러 언어에서 DAC와 DOC 간의 차이는 사동성이 아닌 또 다른 의미 요인과 이에 따른 구조적 차이에 의해 드러난다고 지적되어 왔다는 점도 간과할 수 없다. 예를 들어, 앞서 살펴본 Beck & Johnson(2004)의 분석에서도 확인할 수 있듯이 DOC와 DAC는 전자는 소유 관계, 후자는 소유 관계가 결여된 공간 이동이라는 측면에서 구분되어야 한다는 주장이 지속적으로 이루어져 왔다(see Pinker 1989 등).

- (29=20) a. Satoshi sent Thilo the Damron Guide [DOC]  
 ⇒ [Satoshi’s sending the Damron Guide] CAUSE [BECOME [Thilo HAVE the Damron Guide]]
- b. Satoshi sent the Damron Guide to Thilo. [DAC]  
 ⇒ [Satoshi’s sending the Damron Guide] CAUSE [BECOME [the Damron Guide is AT Thilo]] (Beck & Johnson 2004: (1), (13))

(29a)는 DOC는 기본적으로 사동의 의미를 포함하고 있고, 동시에 결과 상태로 소유관계(HAVE)가 일어난다는 점을 강조한다. 그러나 (29b)에서 확인할 수 있듯이, DAC 역시 사동 사건이 필요하고, 이를 통해 Theme 논항이 특정 위치로(AT) 이동한다는 점이 강조된다. 즉, DOC는 소유관계, DAC는 공간 이동의 의미를 각각 드러낸다는 것이 이 구조의 차이를 확인할 수 있는 가장 중요한 요인이라는 것이다(see also Pinker 1989). 또한, Beck & Johnson(2004)은 이러한 차이가 아래와 같은 대조로 이어진다고 지적한다.

- (30) a. Satoshi sent the Damron Guide to Tübingen.  
 b. #Satoshi sent Tübingen the Damron Guide. (Beck & Johnson 2004: (14))

(29b)와 다르게 (30a)에서는 여격 논항이 사람이 아닌 장소다. 이렇게 장소가 드러나는 경우 DAC에서는 공간 이동의 의미가 강조되고, 특정 장소에 위치(AT)하는 의미가 양립 가능하므로 (30a)의 문법성이 쉽게 이해된다. 그러나 (30b)와 같은 DOC의 경우 장소는 소유관계를 가질 수가 없다. 즉 DAC의 여격 논항에 대응되는 DOC의 대격 논항은 장소를 지칭할 수 없다는 것을 통해, (29a)에서 드러낸 의미의 정당성을 찾을 수 있는 것이다.<sup>16)</sup>

이는 한국어의 DAC와 DOC와 범-언어적인 흐름을 따른다면 본질적으로 사동성 보다는 ‘소유 관계

와 공간 이동'이라는 특성들이 구조에 반영되는 방식을 중심으로 두 구조의 차이가 드러날 것이라는 예측으로 이어진다. 그러나 본 논문의 논의를 따르면 소유 관계만이 아니라 사동성 차이에 대한 논의도 중요해지고, 여기에서 문제가 발생한다. 지속적으로 언급하였듯이, Jung & Miyagawa(2004)의 분석을 따른다면 한국어의 DOC와 달리 DAC는 사동성이 동사구에 반영된 경우에만 형성되지 않는 경우에도 형성될 수 있다. 이는 한국어의 DAC가 영어의 DAC나 두 언어의 DOC와 다르게 사동성이 그 구조 형성에 필수 요소가 아니라는 것으로 해석된다. 그러나 본 연구에서는 사동유형이 PS로서 DAC 유형과 DAC\_S 유형 모두에 강한 접화 효과를 보일 수 있다는 점에서, DAC 유형은 사동성에 대해 구조적으로도 민감하다는 것을 확인하였다. 그러므로 사동성의 결여가 한국어 DAC 구조 형성과 그 구문적 특성에 미치는 영향에 대한 추가적인 분석이 요구된다. 물론 이는 LLM만의 특성일 수도 있으나, 한국어의 경우 DOC와 DAC가 단순히 (비)대칭적 접화 효과를 보이는지 파악하는 것 이상의 논의가 필요하다는 것은 부정하기 어려울 것이다. 그러므로 소유 관계와 공간 이동의 구조적 차이, 또 사동 구조의 실현과 그 양상이 미치는 다양한 영향, 한국어에서 DAC와 DOC가 영어의 대응 구문들의 관계와 유사한지 여부<sup>17)</sup> 등 다양한 관점에서 이 둘을 비교해야 할 것이다. 또한 본 논문에서 배제되어 있는 어순의 변화에 따른 차이를 포착하기 위한 방식에 대해서도 새로운 시각에서 준비되어야 할 것이다. 이러한 논의 모두 후속 연구로 남겨 둔다.

### <References>

- Beck, S., and K. Johnson. 2004. Double objects again. *Linguistic Inquiry* 35, 97-123.
- Bock, K. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18, 355-387.
- Bock, K., and H. Loebell. 1990. Framing sentences. *Cognition* 35, 1-39.
- Bock, K., H. Loebell, and R. Morey. 1992. From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review* 99, 150-171. doi: 10.1037/0033-295x.99.1.150
- Branigan, H. P., M. J. Pickering, S. P. Liversedge, A. J. Stewart, and T. P. Urbach. 1995. Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research* 24, 489-506.
- Chomsky, N. 1986. *Knowledge of Language: Its nature, origin and use*. New York: Praeger.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Collins, C. 2024. *Principles of Argument Structure - A Merge-based Approach*. Cambridge, MA:

16) Jung & Miyagawa(2004)는 한국어의 DOC도 소유 관계 성립 여부에 영향을 받는다는 몇 가지 증거를 제시하고 있다. 물론, (29)에서 제시하는 DAC와 DOC의 의미 차이 자체에 대한 반박도 존재한다(see Collins 2024)

17) Shin & Christianson(2009)에서는 한국어-영어 다중 언어 습득의 측면에서 한국어의 두 구문과 영어의 두 구문이 각 언어에서 모두 유의미한 접화 효과를 보였고, 이에 더해 상대방 언어의 대응 구문에도 유사한 접화 효과를 보인다고 주장하고 있다.

- MIT Press.
- Copley, B., and H. Harley. 2015. A force-theoretic framework for event structure, *Linguistics and Philosophy* 38, 103-158.
- Dowty, D. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67-3, 547-619.
- Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Hale, K., and S. J. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, eds. by K. Hale & S. J. Keyser, 53-109. Cambridge, MA: MIT Press.
- Hewitt, J., and C. D. Manning. 2019. A structural probe for finding syntax in word representations, In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4129-4138. Association for Computational Linguistics.
- Jackendoff, R. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
- Jawahar, G., B. Sagot, and D. Saddah. 2019. What does BERT learn about the structure of language?. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3651-3657.
- Jumelet, J., W. Zuideman, and A. Sinclair. 2024. Do language models exhibit human-like structural priming effects?. In *Findings of the Association for Computational Linguistics ACL 2024*, 142727-14742.
- Jung, Y.-J., and S. Miyagawa. 2004. Decomposing ditransitive verbs. In *Proceedings of Seoul International Conference on Generative Grammar*.
- Kann, K., A. Warstadt, A. Williams, and S. R. Bowman. 2019. Verb argument structure alternations in word and sentence embeddings. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, 287-297. New York City, New York.
- Katzir, R. 2023. Why large language models are poor theories of human linguistic cognition: A reply to Piantadosi. *Biolinguistics* 17, Article e13153. <https://doi.org/10.5964/bioling.13153>
- Lan, N., E. Chemla, and R. Katzir. 2024. Large Language Models and the Argument From the Poverty of the Stimulus. MS. [lingbuzz/006829](https://arxiv.org/abs/2406.10682)
- Levin, B. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- Linzen, T. 2019. What can linguistics and deep learning contribute to each other? Response to Pater. *Language* 95-1, e99-e108.
- Linzen, T., and B. Leonard. 2018. Distinct patterns of syntactic agreement errors in recurrent networks and humans. arXiv:1807.06882.

- Pater, J. 2019. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95-1, e41-e74.
- Piantadosi, S. 2024. Modern language models refute Chomsky's approach to language. In *From Fieldwork to Linguistic Theory: Atribute to Dan Everett*, eds. by, E. Gibson & M. Poliak, 353-414. Berlin: Language Science Press. doi.org/10.5281/zenodo.12665933
- Pickering, M. J., and V. S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134, 427-459.
- Pinker, S., 1989. *Learnability and Cognition: the Acquisition of Argument Structure*. MIT Press, Cambridge.
- Prasad, G., and T. Linzen. 2024. Spawning structural priming predictions from a cognitively motivated parser. arXiv preprint arXiv:2403.07202.
- Prasad, G., M. v. Schijndel, and T. Linzen. 2019. Using priming to uncover the organization of syntactic representations in neural language models. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 66-76.
- Shin, J.-A., and K. Christianson. 2009. Syntactic processing in Korean-English bilingual production: Evidence from cross-linguistic structural priming. *Cognition* 112, 175-180.
- Sinclair, A., J. Jumelet, W. Zuidema, and R. Fernández. 2022. Structural persistence in language models: Priming as a window into abstract language representations. arXiv preprint arXiv:2109.14989v2.
- Talmy, L. 2000. *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.
- Ungerer, T. 2022. Extending structural priming to test constructional relations: Some comments and suggestions. *GCLA* 10: 159-182.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. Attention is all you need. arXiv preprint arXiv:1706.03762.
- Warstadt, A., and S. R. Bowman. 2022. What artificial neural networks can tell us about human language acquisition, In *Algebraic Structures in Natural Language*, 17-60. CRC Press.
- Warstadt, A., A. Parrish, H. Liu, A. Mohananey, W. Peng, S.-F. Wang, and S. Bowman. 2020. BLiMP: A benchmark of linguistic minimal pairs for english, In *Proceedings of the Society for Computation in Linguistics 2020*, 351-352.
- Yi, E., and H. Yun. 2019. Unraveling syntactic and semantic effects on structural priming: Evidence from Korean. *Korean Journal of Linguistics* 44-3, 657-677.

Received on: December 3, 2024  
Revised on: December 22, 2024  
Accepted on: December 23, 2024