

# 2023 한국언어정보학회 추계 정기학술대회

## 전통적 이론언어학과 대규모 데이터 기반 언어학의 교점



일시: 2023년 11월 18일 (토) 9:00 ~ 17:50

장소: 한양대학교 인문대 205호

주최: 한국언어정보학회

주관: 한양대학교 독어독문학과

# 대회장 환영사 및 인사말

한국언어정보학회 회원 여러분, 안녕하십니까?

가을이 깊어 가는 계절에 저희 학회의 2023년 정기 가을학술대회가 11월 18일 한양대학교에서 개최됩니다. 이번 학술대회는 회원 여러분의 참여를 높이기 위해 코로나 이후 처음으로 대면으로만 진행되어서 회원 여러분들께서 직접 교류하고 친목을 다질 수 있는 좋은 기회가 될 것이라고 생각합니다. 비대면과 병행한 학술대회 개최를 고려하였으나 비대면 개최 시 참여율이 급격히 낮아진다는 의견을 반영하여서 대면으로만 학술대회를 개최하게 되었음을 양해 부탁드립니다. 이러한 추세는 다른 학회의 학술대회에서도 나타나고 있는 양상입니다. 그러니 회원 여러분들께서 비록 연구와 교육에 바쁘시더라도 이번 학술대회에 적극적으로 참여하시어 열띤 토론으로 학술대회가 성공리에 끝날 수 있도록 해주신다면 감사하겠습니다.

이번 학술대회의 개최를 위해 수고하시고 도움을 주신 분들이 계십니다. 우선, 학술대회를 한양대에서 개최할 수 있도록 개최장소를 섭외해 주신 남운주 재무이사님과 학술대회 개최를 위해 학술대회 홍보 및 발표논문신청 접수 등을 맡아 주신 김정수 총무이사님께 감사의 말씀드립니다. 이번 학술대회에 초청강연을 흔쾌히 수락해 주신 김정석 교수님(고려대)와 김주애 교수님(한국외대)을 비롯하여 일일이 호명할 수는 없지만 이번 학술대회에서 논문발표를 하시는 분들과 각 세션에서 시회를 맡아 진행해 주실 분들께도 감사드립니다.

끝으로 이번 학술대회에 회원 여러분이 적극적으로 참가해 주실 것을 다시 한 번 당부의 말씀을 드립니다. 환절기에 회원 여러분들께서 늘 건강하시길 바랍니다.

감사의 말씀 드리며,

한국언어정보학회 회장

송민영 배상

# 조직 및 프로그램 위원회

대회장: 송민영 (동국대)

조직 위원장: 이해운 (한국외대)

조직 위원:

강아름 (충남대), 김정석 (고려대), 김한샘 (연세대)

남윤주 (한양대), 이용훈 (충남대), 임동식 (홍익대)

프로그램 위원장: 김정수 (인천대)

프로그램 위원:

김동성 (이화여대), 김지은 (한국외대), 박종언 (한성대)

송상헌 (고려대), 윤수원 (서울시립대), 허세문 (홍익대), 홍정하 (서강대)

## 2023 한국언어정보학회 추계 학술대회

- 주 제 : 전통적 이론언어학과 대규모 데이터 기반 언어학의 교점
- 일 시 : 2023년 11월 18일 (토)
- 장 소 : 한양대학교 인문관 205호
- 주 최 : 한국언어정보학회
- 주 관 : 한양대학교 독어독문학과

### [프로그램]

시간	발표자	내용	사회
09:00~09:20	등록		
09:20~09:30	개회사 및 회장 인사말 (회장 송민영, 동국대)		
09:30~10:00	김동성 (이화여대)	챗GPT 프롬프트 엔지니어링을 위한 질의-응답에서 인과확률 차폐에 대한 연구	박슬기 (경희대)
10:00~10:30	노강산, 송상현 (고려대)	Can Language Models Understand Felicity Conditions?	
10:30~11:00	홍승혜 (고려대)	코퍼스 기반 접미사 '-답다'와 '-스럽다'의 비교 연구	
11:00~11:10	휴식 시간		
11:10~12:10	김주애 (한국외대)	<b>초청 발표:</b> Understanding NLP: Focusing on Question Answering Systems	김정수 (인천대)
12:10~13:20	점심 식사 (도시락 예정)		
13:20~14:20	김정석 (고려대)	<b>초청 발표:</b> Closest Conjoint Agreement in Korean Right-Dislocation	김옥기 (경희대)
14:20~14:30	휴식 시간		
14:30~15:00	김정수 (인천대) 박상희 (한밭대)	Information Structure and Voice Mismatch in Pseudogapping: An Experimental Perspective	송지나 (고려대)
15:00~15:30	송민영 (동국대)	Biscuit Conditionals and Probability Theories	
15:30~16:00	윤수원 (서울시립대)	On Metalinguistic Negation	
16:00~16:10	휴식 시간		
16:10~16:40	최지수 (한양대)	유사도 지표의 번역평가 실무 활용 가능성에 대하여	박상희 (한밭대)
16:40~17:10	봉미경, 김한샘, 김선혜, 강범일 (연세대)	한국어 기초 어휘 선정과 타당성 검증	
17:10~17:40	김진, 남윤주 (한양대)	한국어 술어은유치리에 관한 다층 연구: Surprisal과 뇌신경학적 증거를 중심으로	
17:40~17:50	기타 논의 및 폐회사 (회장 송민영, 동국대)		
18:00~	리셉션		

# 발표 초록 모음

# Understanding NLP: Focusing on Question Answering Systems

Juae Kim  
Hankuk University of Foreign Studies  
[juackim@hufs.ac.kr](mailto:juackim@hufs.ac.kr)

자연어 처리(NLP)는 인간의 언어를 통해 컴퓨터와 인간 사이의 상호 작용에 초점을 맞춘 인공지능의 하위 분야입니다. 주요 목표는 기계(컴퓨터 또는 프로그램)가 의미 있고 유용한 방식으로 인간의 언어를 이해하고, 해석하고, 생성할 수 있도록 하는 것입니다. 질의응답 시스템(QA system)은 자연어처리 도메인 에서의 응용분야 중 하나로, 자연어로 입력된 사용자 질의에 정확하고 관련성 있는 답변을 제공하는 것을 목적으로 합니다. 따라서, 질의응답 시스템을 이해하는 것은 자연어처리를 통한 인간과 인공지능 간의 효과적인 커뮤니케이션 방식을 깊게 파악하는 데 도움을 줄 수 있습니다.

질의응답 시스템을 통해 자연어처리 이해하기 위해, 본 발표에서는 질의응답 시스템의 역사와 이전에 질의응답 시스템을 개발하기 위해 필요한 자연어처리 기술들에 대해서 소개합니다. 또한, 딥러닝 및 거대언어모델(LLM)과 같은 최신 기술의 출현 이후 질의응답 시스템연구 흐름의 변화와 결과 대해 소개합니다. 마지막으로 미래의 질의응답 시스템의 연구 방향성 및 발전 가능성에 대해 논의합니다. 결론적으로, 해당 발표를 통해 질의응답 시스템을 중심으로 자연어처리에 대해 이해할 수 있는 기회를 제공하고자 합니다.

## Closest conjunct agreement in Korean right-dislocation

Jeong-Seok Kim  
Korea University  
uconnkim@korea.ac.kr

Kim et al. (2023) argue that both hierarchy and linearity conspire to affect agreement via an experimental study of honorific agreement between verbal *si* and coordinate subjects in Korean. Given that Korean allows a word-order variation such as right-dislocation, this study further explores whether hierarchy and linearity still affect honorific agreement, using the experimental data from Korean right-dislocation. Consequently, we maintain Kim et al.'s argument that Korean allows Last Conjunct Agreement but not First Conjunct Agreement with respect to honorification, unlike South Slavic gender agreement. Specifically, the findings reveal that honorific agreement in right-dislocated coordinates occurs only when the honorific verbal *si* is present. Additionally, the acceptability of constructions decreases significantly when the last conjunct of right-dislocated coordinates does not match honorification with an honorific verb. Based on these results, it is argued that honorific agreement in right-dislocated (VOS) Korean is still influenced by linear order, similar to honorific agreement in canonical word-order (SOV) Korean.

### Selected reference

Kim, Jeong-Seok, Duk-Ho Jung, and Jin Hyung Lee. 2023. Honorific mismatches of coordinate subjects in Korean. *Canadian Journal of Linguistics* 68(3): 1–24.

# 챗GPT 프롬프트 엔지니어링을 위한 질의-응답에서 인과확률 차폐에 대한 연구

이화여자대학교 컴퓨터공학전공 특임교수 김동성  
2023년 11월 18일

인과추론 능력은 여러 자연언어 처리 어플리케이션에서 주요하게 작용하는데, BERT나 RoBERTa와 같이 파인 튜닝 선학습 언어모델(Fine Tuning Pretrained Language Models)을 통해서 활용된다(Devlin et al. 2019, Liu et al. 2019). 인과성(Causality)을 활용한 여러 서비스들이 SNS, 의료 등등 여러 부분에서 자연어처리 기술과 같이 적용되고 있다(Feder et al. 2022). 특히 챗GPT의 인과 환영(Causal Hallucination)등등 여러 문제는 인공지능 전반에 걸쳐서 많은 논의가 되고 있는 시점이다(Gao et al. 2023).

본 연구에서는 상식추론과 연관되어서 text-davinci-003 (GPT3.5) 기반의 챗GPT의 인과추론에 대한 전반적 평가를 진행하고자 한다. CommonSenseQA 2.0 (Talmor et al. 2019)의 상식추론 데이터를 활용해서 zero-shot, one-shot, few-shot 프롬프팅이라는 다양한 학습 방법을 적용해서 챗GPT가 어떠한 인과추론을 만들어 내는지 살펴보았다. 연구에서 주목한 것은 zero-shot에서 one-shot, 다시 few-shot으로 정보량이 더 많은 학습 방법을 프롬프팅에 채택할수록 특정 명제에 대한 확률인과 추론의 차폐(Screening-off) 현상이 발견된다는 것이다. 차폐는 그래프 'A ← B → C' 에서 B가 A와 C에 대한 공통원인이 되지만,  $P(B|A \cap C) = P(B|C)$ 이 경우에서처럼 B가 A와 C의 분기 가능성을 차단하는 경우를 말한다. 하나의 예를 들어보자.

- Fact1: Loud noises can cause animals to startle.
- Fact2: Fireworks that only make a loud noise are called firecrackers.
- Combined Fact: Firecrackers can cause animals to startle.
- Question: What could cause an animal to startle? (A) Firecracker (B) Going into hibernation (C) an object (D) Electrical energy (E) an electric stove (F) Gathering food (G) Seeing prey (H) vehicles
- Answer: (A)

위의 상식추론은 Question에 대해서 다음과 같은 명제들이 True, False로 선택될 수 있다. (P = 『 Firecrackers cause an animal to startle 』, Q = 『 An object causes an animal to startle 』, R = 『 Seeing prey causes an animal to startle 』) 이러한 경우에 Fact1의 명제 W = 『 Loud noises cause an animal to startle 』이 Q와 R의 가능성을 차단하고 P의 가능성만을 활성화시키게 된다. 챗GPT는 더 적은 정보를 주는 zero-shot에서 여러 가능한 명제들을 답안으로 제시하지만, one shot이나 few shot으로 프롬프트에 정보가 더 늘어날수록 가능한 명제들이 축소되고 데이터에서 제시된 답안에 근접하게 된다.

챗GPT가 가진 프롬프트 엔지니어링에는 이와 같이 인과추론의 차폐효과를 극대화해서 적절한 답안을 유도하는 기술이 필요하다. 이러한 방식으로 원하는 정보를 적절하게 추출할 수 있다.



## 참고문헌

- Devline et al. (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In Proc. of NAACL.
- Feder et al. (2022) Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond. Transactions of the Association for Computational Linguistics (2022) 10: 1138-1158.
- Gao et al. (2023) Is ChatGPT a Good Causal Reasoner? A Comprehensive Evaluation. EMNLP 2023.
- Talmor, A. et al. (2019) CommonsenseQA 2.0. In Proceedings of NIPS.

# Can Language Models Understand Felicity Conditions?

Kangsan Noh and Sanghoun Song  
Korea University

[kasan1998@korea.ac.kr](mailto:kasan1998@korea.ac.kr); [sanghoun@korea.ac.kr](mailto:sanghoun@korea.ac.kr)

This study used allosentences to investigate the extent to which language models comprehend the nonliteral and pragmatic elements of natural language. Allosentences are sentences that share the same truth conditions but differ in felicity conditions and are thus pragmatically differentiated (Lambrecht 1996), as in (1).

- (1) Q: What did the dog do?  
a. The dog chased the cat.  
b. The cat was chased by the dog.

(Bender 2022: 102)

While both (1a) and (1b) share the same truth conditions, the latter will be an infelicitous answer to the question in (1). We constructed a dataset of allosentences to assess whether current language models could capture pragmatic distinctions rooted in felicity conditions, which are crucial for human-like communication. Using the BERT-base/large (Devlin et al., 2018) and GPT-3.5/GPT-4 models (GPT; Brown et al., 2020), we conducted the following two experiments: Experiment 1 (next sentence prediction (NSP) task) and Experiment 2 (text generation task). The purpose of these two experiments is to determine whether language models can select felicitous answers to questions in a human-like manner. The findings reveal that, although not flawless, current language models exhibit the capacity to generate more felicitous responses in specific contexts, shedding light on their ability to engage in human-like communication. The analysis of the two experiments also shows that language models share with humans a bias related to subject-object asymmetry (Schachter and Yip 1990), showing better performance on object questions.

## References

- Bender, Emily. M. 2022. *Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax*. Springer Nature.
- Brown, Tom., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33, 1877-1901. DOI: <https://doi.org/10.48550/arXiv.2005.14165>
- Devlin, Jacob., Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Lambrecht, Knud. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.
- Schachter, Jacquelyn and Virginia Yip. 1990. Grammaticality judgments: Why does anyone object to subject extraction? *Studies in Second Language Acquisition*, 12:4, 379-392.

# 코퍼스 기반 접미사 '-답다'와 '-스럽다'의 비교 연구

홍승혜

고려대학교

[happy\\_sh@korea.ac.kr](mailto:happy_sh@korea.ac.kr)

접미사 '-답다'와 '-스럽다'는 다른 형용사 파생 접사와 달리 '어른답다/스럽다', '미국답다/스럽다', '소녀답다/소녀스럽다'와 같이 같은 어기와 결합하면서도 그 쓰임이 구분된다는 점에서 주목할 만하다. 두 접미사의 의미적 구분과 관련한 선행 연구의 접근 방식은 크게 세 가지 측면으로 나누어 살펴볼 수 있다. 첫째는 두 접미사와 결합하는 선행어기의 의미적 자질의 차이에 의한 구분이며, 둘째는 주어와 선행어기의 관계에 의한 구분이다(송철의, 1977; 우효, 1980; 신순자, 1981; 이경우, 1981; 노대규, 1981; 심재기, 1982; 김수호, 1985; 장영희, 1986; 윤동원, 1986; 성광수, 1988; 정연주, 2007). 셋째는 접미사가 내포하는 가치 의미가 긍정적인지 혹은 부정적인지에 의한 구분이다(김정남, 2005; 최화니, 2020). 이를 검토하기 위해 본 연구는 문·구어 코퍼스로부터 접미사 '-답다'와 '-스럽다'에 공통적으로 결합하는 어기의 목록과 그에 해당하는 용례를 추출하여, 두 접미사의 활용에 어떠한 특성이 포착되는지 분석하고 기존의 논의를 적용하여 설명하기 어려운 용례를 보인다. 이를 기반으로 선행연구의 발견을 포괄하면서 기존에 논의되지 않았던 용례를 함께 설명할 수 있는 새로운 의미 구분 기준을 제안하고자 한다. 이를 통해 선행어기의 의미적 자질 및 주어와의 관계에서 상보적 분포를 보이지 않는 접미사 '-답다'와 '-스럽다'의 쓰임을 구분하고, 지각적 판단에 의한 개념화에 기반하여 두 접미사의 의미적 기능을 설명한다는 데에 의의를 찾고자 한다.

## References

- 김수호. 1986. “접미사 ‘-답다, -롭다, -스럽다’의 어휘론적 기능연구”. *문학과 언어* 7, 51-74.
- 노대규. 1981. “국어 접미사 ‘답’의 의미 연구”. *한글* 172, 57-104.
- 김정남. 2005. “‘-답다’, ‘-롭다’, ‘-스럽다’의 분포와 의미”. *한국어 의미학* 18, 125.
- 김창섭. 1984. “형용사 파생 접미사들의 기능과 의미-‘-답-, -스럽-, -롭-, 하-’와 ‘-적’의 경우-”. *진단학보* 58, 145.
- 송철의. 1977. “파생어 형성과 음운 현상”. 서울대학교, 석사학위논문.
- 신순자. 1982. “국어 접미사 ‘스럽다’의 연구”. 숙명여자대학교 석사학위논문.

- 심재기. 1982. 국어어휘론. 집문당.
- 우 효. 1980. “형용사 파생접미사 “-답다” “-스럽다”의 연구”, 계명대학교 석사학위논문.
- 장영희. 1986. “현대국어 접미사 ‘-답-’의 연구”. 숙명여자대학교 석사학위논문.
- 양정석. 2020. “‘-답다’의 양상 의미와 그 통사의미적 합성”. 한글 81(1), 83-113.
- 윤동원. 1986. “형용사 파생접미사 {-스럽-}, {-롭-}, {-답-}의 연구”. 서울대학교 석사학위논문.
- 이경우. 1981. “파생어형성에 있어서의 의미변화”. 국어교육 39, 215.
- 정연주. 2007. “형용사 형성 접미사의 통시적 변화-‘-되-’, ‘-롭-’, ‘-스럽-’, ‘-답-’을 중심으로-”. 고려대학교 석사학위논문.
- 최호철. 1994. “현대 국어 가의소의 의미 분석 (1)- ‘답다, -되다, -롭다, -스럽다’를 중심으로”. 한국어학 1, 415-444.
- 최화니. 2020. “‘-답-’, ‘-롭-’, ‘-스럽-’의 결합제약과 의미 기능”. 한국언어문학 113, 33-62.

# Information structure and voice mismatch in pseudogapping: An experimental perspective

Jungsoo Kim and Sang-Hee Park  
Incheon National University; Hanbat National University  
[jungsookim@inu.ac.kr](mailto:jungsookim@inu.ac.kr); [sangheepark@hanbat.ac.kr](mailto:sangheepark@hanbat.ac.kr)

Pseudogapping, exemplified in (1a), is an elliptical construction in which a main verb is elided, leaving behind a tensed auxiliary verb and a dependent of the main verb. It is often compared to other types of ellipsis, such as verb phrase ellipsis (VPE) as in (1b) and gapping as in (1c) (Gengel 2007, Hoeksema 2006, Jayaseelan 1990, Kim and Runner 2022, Kim and Park 2022, Kubota and Levine 2017, Lasnik 1999, Levin 1980, Miller 2014, Ross 1967):

- (1) a. John called Sarah, and Mary will Jane.    *Pseudogapping*  
     b. John called Sarah, but Mary didn't.        *VPE*  
     c. John called Sarah, and Mary Jane.         *Gapping*

Most previous research on pseudogapping has mainly focused on providing theoretical accounts of the mapping relation between its incomplete form and the full propositional meaning; however, only a few studies have empirically explored the construction. In addition, while ellipsis research has long been concerned with the nature of the identity condition between the ellipsis site and its corresponding material in the antecedent clause, issues regarding the sensitivity of pseudogapping to syntactic properties such as voice parallelism have not reached a consensus yet. Theorists have just relied on introspective judgments: some judge voice-mismatched pseudogapping examples as ungrammatical (Merchant 2008, 2013) whereas others find them acceptable (Kim and Runner 2022, Kubota and Levine 2017, Miller 2014, Tanaka 2011a,b). Furthermore, although experimental studies on the voice mismatches in other elliptical constructions like VPE and gapping are abundant and a few hypotheses have been tested for the voice mismatch effects (Arregui et al. 2006, Clifton et al. 2019, Frazier and Clifton 2006, Kehler 2002, Kertz 2008, 2010, 2013, Kim and Runner 2011, 2018, Poppels and Kehler 2019), little attempt has been made to systematically investigate the voice mismatch effects on pseudogapping from an experimental perspective.

In this context, our study aims to provide a first set of empirical data that can help resolve previous disagreements among theorists concerning voice mismatches in pseudogapping. We intend to do so by examining how voice mismatches in pseudogapping are perceived by native speakers of English and by comparing the results with judgments on parallel VPE examples. We also aim to explore the potential influence of information structure on pseudogapping, as previous studies have shown that it can modulate the acceptability of VPE (i.e., contrastive topic structure vs. auxiliary focus structure) (Kertz 2013).

Our experiment results first show that voice mismatch significantly lowers acceptability of pseudogapping sentences; however, they show no reliable effect of information structure on pseudogapping, which is not in line with previous findings regarding voice mismatches in VPE (i.e., no increased acceptability of pseudogapping sentences with auxiliary focus structure compared to that of pseudogapping sentences with contrastive topic structure) (Kertz 2013). These results together suggest that the impact of focus structure may not be the same across all ellipsis types and that decreased acceptability associated with voice mismatch may be related to some unknown factors at the discourse level. These then indeed call for further investigation in the broader discussion on the relation between acceptability and grammaticality.

## Reference

- Arregui, A., C. Clifton, L. Frazier, and K. Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Languages* 55, 232-246.
- Clifton, C., M. Xiang, and L. Frazier. 2019. A note on the voice mismatch asymmetry in ellipsis. *Journal of Psycholinguistic Research* 48(4), 877-887.
- Frazier, L. and C. Clifton. 2006. Ellipsis and discourse coherence. *Linguistics and Philosophy* 29(3), 315-346.
- Gengel, K. 2007. *Focus and ellipsis: A generative analysis of pseudogapping and other elliptical structures*. Doctoral dissertation, Universität Stuttgart, Stuttgart, Germany.
- Hoeksema, J. 2006. Pseudogapping: Its syntactic analysis and cumulative effects on its acceptability. *Research on Language and Computation* 4(4), 335-352.
- Jayaseelan, K. A. 1990. Incomplete VP deletion and gapping. *Linguistic Analysis* 20(1-2), 64-81.
- Kehler, A. 2002. *Coherence, reference and the theory of grammar*. Stanford, CA: CSLI Publications.
- Kertz, L. 2008. Focus structure and acceptability in verb phrase ellipsis. In N. Abner and J. Bishop, eds., *Proceedings of the 27th West Coast Conference on Formal Linguistics*, 283-291. Somerville, MA: Cascadilla Proceedings Project.
- Kertz, L. 2010. *Ellipsis reconsidered*. Doctoral dissertation, University of California, San Diego, CA.
- Kertz, L. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89(3), 390-428.
- Kim, C. and J. Runner. 2011. Syntactic identity isn't enforced blindly: Comparing VP ellipsis and pseudogapping. Poster presented at the *85th Annual Meeting of Linguistic Society of America*, January 6-9, Pittsburgh, Pennsylvania.
- Kim, C. and J. Runner. 2018. The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy* 41(1), 41-85.
- Kim, J.-B. and J. Runner. 2022. Pseudogapping in English: A direct interpretation approach. *The Linguistic Review* 39(3), 457-494.
- Kim, J. and S.-H. Park. 2022. English pseudogapping: An experimental perspective. *Korean Journal of English Language and Linguistics* 22, 675-694.
- Kubota, Y. and R. Levine. 2017. Pseudogapping as pseudo-VP-ellipsis. *Linguistic Inquiry* 48(2), 213-257.
- Lasnik, H. 1999. Pseudogapping puzzles. In S. Lappin and E. Lappin, eds., *Fragments: Studies in ellipsis and gapping*, 141-174. Oxford: Oxford University Press.
- Levin, N. 1980. Main verb ellipsis in spoken English. In A. Zwicky, ed., *Clitics and ellipsis*, 65-165. Columbus, OH: The Ohio State University.
- Merchant, J. 2008. An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39(1), 169-179.
- Merchant, J. 2013. Voice and ellipsis. *Linguistic Inquiry* 44(1), 77-108.
- Miller, P. 2014. A corpus study of pseudogapping and its theoretical consequences. In C Piñón, ed., *Empirical issues in syntax and semantics*, vol. 10, 73-90. Paris: CSSP.
- Poppels, T. and A. Kehler. 2019. Reconsidering asymmetries in voice-mismatched VP-ellipsis. *Glossa: A Journal of General Linguistics* 4(1), 60.
- Ross, J. R. 1967. *Constraints on variables in syntax*. Doctoral dissertation, MIT, Cambridge, MA.
- Tanaka, H. 2011a. Syntactic identity and ellipsis. *The Linguistic Review* 28(1), 79-110.
- Tanaka, H. 2011b. Voice mismatch and syntactic identity. *Linguistic Inquiry* 42(3), 470-490.

## Biscuit Conditionals and Probability Theories

Mean-Young Song  
Dongguk University  
[songmy@dongguk.ac.kr](mailto:songmy@dongguk.ac.kr)

This paper explores the semantics of biscuit conditionals in terms of the probability theory. Unlike canonical conditionals, biscuit conditionals are characterized by the fact that the truth values of the apodosis are independent of those of the protasis. Thus, the protasis does not function to impose restrictions on the domain of quantification for the truth conditions for biscuit conditionals. Instead, it plays the role of signifying illocutionary force since it is conversationally relevant to the apodosis. Based on this, this paper argues that the complementizer *if* of the protasis of biscuit conditionals is treated as a force signifier. This paper also argues that the conditional probability provides an effective account of the independent relationship between the protasis and the apodosis. The protasis and the apodosis are independent from each other only if the probability of the apodosis given the protasis is equal to the probability of the apodosis given the complement of the protasis. This may assist us in understanding how to account for independency of the biscuit conditionals and where illocutionary force comes from in biscuit conditionals.

## On Metalinguistic Negation

Suwon Yoon  
University of Seoul  
suwon@uos.ac.kr

In exploring the phenomena of Metalinguistic Negation (MN) across languages, the goal of this paper is to offer a semantic analysis of MN in language as indicators of *subjective mode* in the negation. We show that, as a marker of subjective negation, MN expresses objection to a proposition (*U*: previous utterance), often with the speaker's neutral to negative emotional stance. In doing so, we make four points regarding MN. First, we compare MN with other metalinguistic phenomena, especially metalinguistic comparatives (MC), and show that, MN can be categorized into four types: (i) appropriateness assessment MN, (ii) emphatic denial MN, (iii) negative emphatic denial MN, and (iv) irony MN. Second, we show how MN differs from regular negation (*objective anti-veridicality*), suggesting that MN markers come from *six semantic sources* depending on the speaker's personal reason to object. Third, we propose the semantics of MN, as *subjective anti-veridicality* (à la Giannakidou and Mari 2021) in six *doxastic models of mental representation* that may concern the speaker's own epistemic, doxastic, or buletic commitment, regardless of actuality. Finally, we propose that most cases of MN, except for appropriateness assessment MN, can be understood as a subspecies of subjective mode phenomena, akin to other perspectival phenomena such as expressives (Potts 2005).



# 유사도 지표의 번역평가 실무 활용 가능성에 대하여

Jisoo Choi  
Hanyang University  
[choiji@hanyang.ac.kr](mailto:choiji@hanyang.ac.kr)

본 논문에서는 유사도 측정 지표(metric)를 활용한 텍스트 자동평가(automatic evaluation)를 전통적인 수동평가 방식의 번역평가 실무에 어느 정도 활용할 수 있을지, 그 가능성을 고찰하였다. 이를 위해 먼저 본래 자동평가의 주 사용 목적인 기계번역 성능평가를 위한 자동평가 방법들을 알아보았다. 그리고 이러한 자동평가를 인간의 번역평가 실무에 사용한 선행연구를 조사하였다. 마지막으로 기계와 인간의 번역에 자동평가를 실시하고 차이를 알아보았다.

자동평가는 본래 신경망 기반 기계번역기의 번역 성능을 평가하는 데 사용되었다. 이를 위해서는 정답 역할을 하는 인간의 참조번역(reference text)과 평가 대상인 기계번역(candidate text)이 필요하다. 이 두 번역 간 유사도를 다양한 지표(metric)로 측정함으로써 기계의 번역이 얼마나 정답인 인간에 가까운가를 보는 것이다. 여기에는 주로 분류 성능 평가 지표인 BLEU(Papineni et al. 2002; Post 2018), METEOR(Benerjee and Lavie 2005) 등이 사용되었다.

그러나 이를 인간의 글을 평가하는 데 적용하기 시작했다. 신뢰도(reliability) 확보가 용이하고 시간적인 면에서 효율적이라는 점을 근거로 BLEU 등의 지표를 그대로(Vela et al. 2014; Chung 2020; Han and Lu 2021; Lu and Han 2023), 또는 자연어 텍스트의 특성을 반영해 다양한 수정을 가하여(Han et al. 2012; 정혜연 외 2021; 최지수 2023) 활용했다. 또한 서술형 작문, 요약문 등의 평가에 적용하기도 했다(조정현 외 2009; 이경건과 하민수 2020).

자연어 텍스트를 기계로 평가하려면 타당성을 제고하기 위해 지표의 수정이 필요할 것이다. 그러나 실무에서는 매번 수정을 가하여 측정하기란 쉽지 않다. 또한 한국어를 비롯한 다수의 언어에서 좋은 성능을 보인 BERTscore(Zhang et al. 2019; 정혜연과 서수영 2023)는 평가의 편향성(bias) 문제가 제기되기도 했다(Sun et al. 2022). 그리고 수정한다고 하더라도 동일한 토큰화, 정규화(normalization) 과정을 거쳤는지의 여부가 다를 수 있어, 다른 평가 케이스와의 비교가 어렵다.

이에 본 연구에서는 SacreBLEU(Post 2018)로 인간이 생성한 텍스트를 평가한 후, 이것을 인간이 평가한 결과와 비교하였다. SacreBLEU는 BLEU를 기반으로 하며, 한국어를 포함한 여러 언어에 대해 전처리 과정을 통일한 후 배포하였다. 이에 점수 비교가 필요한 연구에 사용이 권장되고 있다. 이러한 비교를 통해 본 연구에서는 기계의 성능평가에는 널리 사용되는 지표가 인간의 텍스트 평가에는 얼마나 타당하게 사용될 수 있는지 실험하였다. 그 결과, 전체적으로는 인간의 평가와 .8 이상으로 강한 상관 관계를 나타냈다. 이는 인간 텍스트 평가 시의 텍스트 유사도 지표 활용의 가능성을 시사하였다. 또한 간편한 사용이 가능해 교육 현장에서의 실무 활용 가능성을 예상해 볼 수 있었다.

## References

- 이경건, & 하민수. (2020). 인공지능 기반 자동평가의 현재와 미래: 서술형 문항에 관한 문헌 고찰과 그 너머. *교육공학연구*, 36(2), 353-382.
- 정혜연, 박현일, 우경조, & 서수영. (2021). 임베딩을 활용한 인간번역의 자동평가-기계가 의미를 평가할 수 있을까. *통번역학연구*, 25(3), 141-162.
- 정혜연 & 서수영. (2022). 버트 (BERT) 를 활용한 인간번역의 자동평가: 여러 모델의 성능 비교 및 활용 가능성. *통번역학연구*, 26(4), 117-137.
- 조정현, 정현기, 박찬영, & 김유섭. (2009). BLEU 를 활용한 단기 서술형 답안의 자동 채점. *한국 HCI 학회 학술대회*, 606-610.
- 최지수. (2023). 인간번역의 자동평가에서 번역 텍스트 길이가 평가 타당성에 미치는 영향. *통번역학연구*, 27(2), 145-164.
- Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* (pp. 65-72).
- Chung, H. Y. (2020). Automatische evaluation der Humanübersetzung: BLEU vs. METEOR. *Lebende Sprachen*, 65(1), 181-205.
- Han, C., & Lu, X. (2021). Interpreting quality assessment re-imagined: The synergy between human and machine scoring. *Interpreting and Society*, 1(1), 70-90.
- Lin, C. Y., & Hovy, E. (2003). Automatic evaluation of summaries using n-gram co-occurrence statistics. *Proceedings of the 2003 human language technology conference of the North American chapter of the association for computational linguistics* (pp. 150-157).
- Lu, X., & Han, C. (2023). Automatic assessment of spoken-language interpreting based on machine-translation evaluation metrics: A multi-scenario exploratory study. *Interpreting*, 25(1), 109-143.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).
- Post, M. (2018). A Call for Clarity in Reporting BLEU Scores. *Proceedings of the Third Conference on Machine Translation: Research Papers* (pp. 186-191).
- Sun, T., He, J., Qiu, X., & Huang, X. (2022). BERTScore is unfair: On social bias in language model-based metrics for text generation. arXiv preprint arXiv:2210.07626.
- Vela, M., Schumann, A. K., & Wurm, A. (2014). Human translation evaluation and its coverage by automatic scores. *Automatic and Manual Metrics for Operational Translation Evaluation Workshop Programme* (p. 19).
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019, September). BERTScore: Evaluating Text Generation with BERT. *International Conference on Learning Representations*.

<Web>

<https://huggingface.co/spaces/evaluate-metric/sacrebleu>

<https://github.com/mjpost/sacrebleu#downloading-test-sets>

<https://eunjeon.blogspot.com/2013/02/blog-post.html>

한국언어정보학회 11/18(토) 발표초록

제목 : 한국어 기초 어휘 선정과 타당성 검증

발표자 : 봉미경, 김한샘, 김선혜, 강범일(연세대학교 언어정보연구원)

발표 요지 :

본 연구는 서상규 외(2021), 서상규 외(2022)에서 1~3등급 한국어 기초 어휘를 선정한 방법론을 살펴보고 이 과업으로 선정된 한국어 기초 어휘 1~3등급 어휘를 대상으로 타당성 검증을 시도하는 것을 목적으로 한다. 서상규 외(2021, 2022)에서 선정된 한국어 기초 어휘 1~3등급은 학령전기(1등급)부터 초등학교 1~2학년(2등급), 초등학교 3~4학년(3등급) 수준의 어휘이다. 1~3등급이 학령전기부터 초등학교 4학년까지의 수준을 아우른다고 할 수 있으나 서상규 외(2021, 2022)는 교육 목적으로 개발된 어휘 등급 목록이 아니라 보다 광범위한 국민 언어 능력 신장을 목적으로 하는 범용적 어휘 등급 목록이다. 따라서 범용 목적의 균형 말뭉치를 대상으로 한 정량적 검토 결과를 토대로 선정된 목록이다. 이후 전문가 평정, 사용자 평정을 거쳐 최종 목록이 선정되었다. 이에 본 연구는 기 선정된 1~3등급 어휘 목록을 대상으로 이 목록의 타당성을 해당 어휘 수준에 해당하는 초등학교 교과서 말뭉치를 활용하여 검증해 보고자 한다. 이러한 검증은 한국어 기초 어휘 1~3등급에서 선정된 어휘 목록의 타당성을 검증해 보는 동시에 현재 사용되고 있는 초등학교 교과서의 어휘 수준을 객관적 기준을 통해 측정해 볼 수 있다는 의의가 있다.

참고문헌

- 강현화 외(2012), 한국어 교육 어휘 내용 개발(1단계), 서울: 국립국어원.  
강현화 외(2013), 한국어 교육 어휘 내용 개발(2단계), 서울: 국립국어원.  
강현화 외(2014), 한국어 교육 어휘 내용 개발(3단계), 서울: 국립국어원.  
김한샘(2009), 초등학교 교과서 어휘 조사 연구, 서울: 국립국어원.  
김한샘(2010), 국어 교육용 어휘 선정을 위한 교과서 어휘 조사 연구 - 초등학교 교과서 어휘 분석 -. 국어교육연구 47, 63-90.  
김한샘 외(2022), 초등학교와 중학교 교육용 어휘 등급화 연구(ORM 2022-88-1), 충북: 한국교육과정평가원.  
김한샘 외(2021), 초등 교육용 어휘 목록 개발 연구. 세종: 교육부.  
봉미경(2021), 2015 개정 교육과정 초등학교 교과서 어휘의 계량적 연구, 언어사실과 관점 54, 연세대 언어정보연구원  
봉미경·김한샘·김선혜(2022), 교과서 분석 기반 초등 교육용 어휘 선정 연구, 새국어 교육 132, 343-368.  
서상규 외(2021), 2021년 국어 기초 어휘 선정 및 어휘 등급화 연구, 서울: 국립국어

원.

서상규 외(2022), 2021년 국어 기초 어휘 선정 및 어휘 등급화 연구, 서울: 국립국어원.

서상규 외(2009), 교육용 기본 어휘 선정을 위한 기초 연구, 서울: 국립국어원.

서혁 외(2014), 문식성 향상을 위한 어휘 능력 신장 연구, 서울: 국립국어원.

이삼형 외(2017), 국어 기초 어휘 선정 및 어휘 등급화를 위한 기초 연구, 서울: 국립국어원.

이삼형 외(2018), 2018년 국어 기초 어휘 선정 및 어휘 등급화 연구. 서울: 국립국어원.

이삼형 외(2020), 2020년 국어 기초 어휘 선정 및 어휘 등급화 연구. 서울: 국립국어원.

조용구(2016), 독자에게 적합한 글을 선정하기 위한 이독성 공식의 개발, 한국교원대학교 박사학위논문.

최소영·길호현·박태준·류수경(2022), 교육용 텍스트의 이독성 질적 평가 준거 개발 및 타당화 연구, 국어교육 176, 1-57.

최소영, 박태준, 박진호(2023), 어휘 등급의 검증 방안 연구. 형태론 25(1), 57-93.

최인숙(2005), 독서교육시스템을 위한 텍스트 수준 측정 공식 구성에 관한 연구, 정보관리학회지 22-3, 213-232.

한송화 외(2015), 한국어 교육 어휘 내용 개발(4단계), 서울: 국립국어원.

## 한국어 술어은유처리에 관한 다층연구: surprisal과 뇌신경학적 증거를 중심으로

김 건 & 남 윤 주

(한양대학교)

은유 처리와 관련하여 가장 전통적 접근 방식인 간접 접근 모델(indirect access model)은 은유적 의미를 문자적 의미로부터 일탈한 표현이라고 본다. 즉, 은유적 표현은 문자적 의미로부터 파생되지만 문자 그대로의 의미가 문맥에 의해 부인될 때까지 발생할 수 없다는 입장이다(Grice, 1975; Searle, 1978). 따라서 간접적 접근 모델에서는 은유적 표현이 문자적 의미에 비하여 무조건적으로 처리가 어렵다고 주장한다. 반면, Glucksberg et al. (1982) 등에 의해 제안된 직접적 접근 모델에서는 은유적 표현과 문자적 표현을 처리하는 인지적 메커니즘이 동일하다고 본다. 즉, 은유적 표현 역시 문자 그대로의 표현과 동일한 방식으로 처리가 가능하며, 은유적 표현을 위하여 문자적 의미가 필수적으로 전제되거나 해당 의미가 부인되는 부가적인 처리 과정은 필요 없다는 입장이다. 세 번째 입장은 연속적 접근 모델(continual processing model)로 Lakoff (1993)의 “현대 은유 이론(contemporary theory of metaphor)”, Giora (1997, 2003)의 “현저성 변화 모델(Gradient Salience Model)”, 그리고 Bowdel and Gentner (2005)의 “은유의 생애 모델(Career of Metaphor Model)” 등이 이에 속한다고 볼 수 있다. 이 모델들은 직/간접적 모델들에 비해서 처리적 차원의 특성을 더욱 반영한 모델들로, 은유 처리가 모두 동일하게 직접 혹은 간접적 방식으로 진행되는 것이 아니라 그것이 얼마나 관습적 혹은 새로운 은유인지에 따라서 처리가 달라질 수 있다는 입장을 취한다.

이처럼 서로 다른 이론들의 설명력과 은유 처리의 심리적 과정을 검증하기 위하여 신경 언어학적 연구들은 주로 문자적 의미와 관습적 은유, 새로운 은유의 처리 시 처리 시간의 차이와 뇌의 반응을 실시간으로 확인할 수 있는 ERP 방법론을 이용하여 은유 연구를 진행해 왔다. 대부분의 연구에서 은유 처리에 동원되는 부가적인 인지적 노력을 반영하는 사건관련전위적 증거로서 N400 성분을 보고하였고, 전형적 은유에 비하여 새로운 은유에서 큰 N400 성분을 보고하였다. 그러나 문자적 표현과 전형적 은유 표현을 처리하는 과정에서도 나타나는 뇌파 증거로는 연구마다 다소 다른 양상이 보고되었다. Lai et al. (2009)는 은유적 표현이 문자 의미적 문장보다 더 많은 처리 노력을 필요로 하였으며(N400), 은유적 표현의 전형도에 따라 N400의 시간상 분포가 다르게 나타난다는 것을 확인하였다(간접 접근 모형 지지). 반면, Goldstein et al (2012)에서는 새로운 은유 표현은 은유적 범주화를 할 경우에 더 수월하게 처리되었으며(N400 효과의 감소와 LPC 효과 증가), 반대로 전형적 은유 표현은 은유적 범주화를 할 경우에 처리의 난이도가 높아지는 것을 확인함으로써(증가한 N400 효과와 감소한 LPC 효과) 은유의 경력모형을 지지하였다. Iakimova et al (2005)의 또 다른 ERP 연구에서는 은유적 표현의 처리 기제는 문자 그대로의 표현과 다르지 않으며, 해당 표현의 현저성에 따라 처리 부담이 결정된다는 현저성 변화 모델을 뒷받침하는 결과들을 확보하였다. 한국어 술어 은유 처리에 대한 유일한 ERP 연구인 남윤주 & 정혜인 (2016)은 전형적 은유 표현에 비해 오히려 문자적 표현에서 N400이 검출되었다는 것을 확인하였다. 이는 간접 접근 모형으로는

설명할 수 없는 현상으로 직접 접근 모형을 지지하는 강력한 증거라고 볼 수 있으나, 이 증거가 현저성 변화 모델에서 제시하는 것과 같이 전형적 은유에 대한 높은 현저성에 기인하는 것인지, 아니면 생애 이론이 주장하는 바와 같이 전형적 은유의 의미 영역에 성립될 수 있는 술어의 후보군이 더 적기 때문에 문자적 표현에 비하여 전형적 은유의 불확실성(entropy)이 더 적어서 생긴 처리의 수월성을 나타내는 증거인지는 확인할 수 없었다.

따라서 본 연구에서는 ERP 실험에 사용된 문장의 술어 은유에 대해, 문장에 속한 단어의 출현 확률을 반영하는 Surprisal을 (Hale, 2001) Transformer 기반 GPT 계열의 KoGPT-2와 KoGPT-trinity를 활용하여 계산함으로써 생애 이론의 이론적 적합성을 가능성을 확인하고자 했다. 그 결과, 모든 모델 및 조건에서 남윤주 & 정혜인 (2016)에서 보고된 문자적 조건의 N400이 Surprisal의 차이에는 기인하지 않았음을 확인하였다. 본 연구에 이어, 코퍼스 수준에서 이러한 경향이 일관되게 관측되는지 확인하고 뇌신경학적 증거를 보완함으로써 은유 처리에 대한 이론 제반을 검증할 수 있는 다층적인 증거를 마련할 수 있을 것이다.

### 참고문헌

- 남윤주, & 정혜인. (2016). 한국어 술어 은유처리에 대한 ERP 연구. *언어와 정보*, 20(2), 93-116.
- 송상헌. (2022). 딥러닝 언어모델과 Surprisal을 활용한 언어분석. *인공지능인문학연구*, 12, 9-39.
- Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, 112, 193-216.
- Gentner, D., & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor and Symbol*, 16(3-4), 223-247.
- Goldstein A., Arzouan Y., & Faust M. (2012). Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain Language*. 123, 137-142.
- Iakimova, G., Passerieux, C., Laurent, J. P., & Hardy-Bayle, M. C. (2005). ERPs of metaphoric, literal, and incongruous semantic processing in schizophrenia. *Psychophysiology*. 42, 380-390.
- Kutas, M., Hillyard, S.A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol. Psychol.* 11, 99-116.
- Lakoff, G. (2009). The neural theory of metaphor. SSRN Electronic Journal.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2023). Strong Prediction: Language model surprisal explains multiple N400 effects. *Neurobiology of Language*, 1-71.
- Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, 16(1-2), 109-121.
- Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. (2004). Neural correlates of metaphor processing. *Cognitive Brain Research*, 20(3), 395-402.
- Reddy, M. (1979). The conduit metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 284-324). Cambridge University Press.
- Tendahl, M., & Gibbs Jr, R. W. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics*, 40(11), 1823-1864.

# 발표 자료 모음

# Understanding NLP: Focusing on a Question Answering System

---

Juae Kim  
([juaekim@hufs.ac.kr](mailto:juaekim@hufs.ac.kr))



**한국외국어대학교**  
HANKUK UNIVERSITY OF FOREIGN STUDIES



# Contents

- **Introduction to Question Answering**
- **Open Domain Question Answering**
- **Recent Question Answering Trends**

# Introduction to Question Answering

---

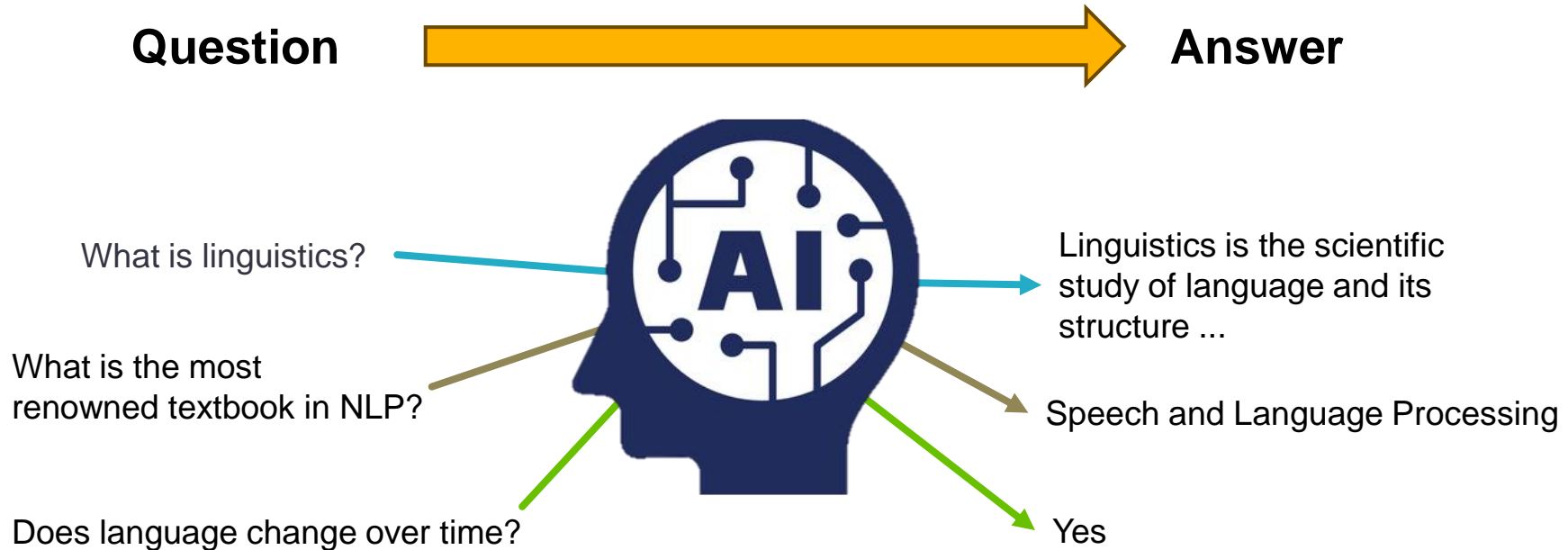
# Natural Language Processing (NLP)

- ▶ 자연어처리(NLP)는 사람의 말을 이해하고 응답(반응)하는 기술



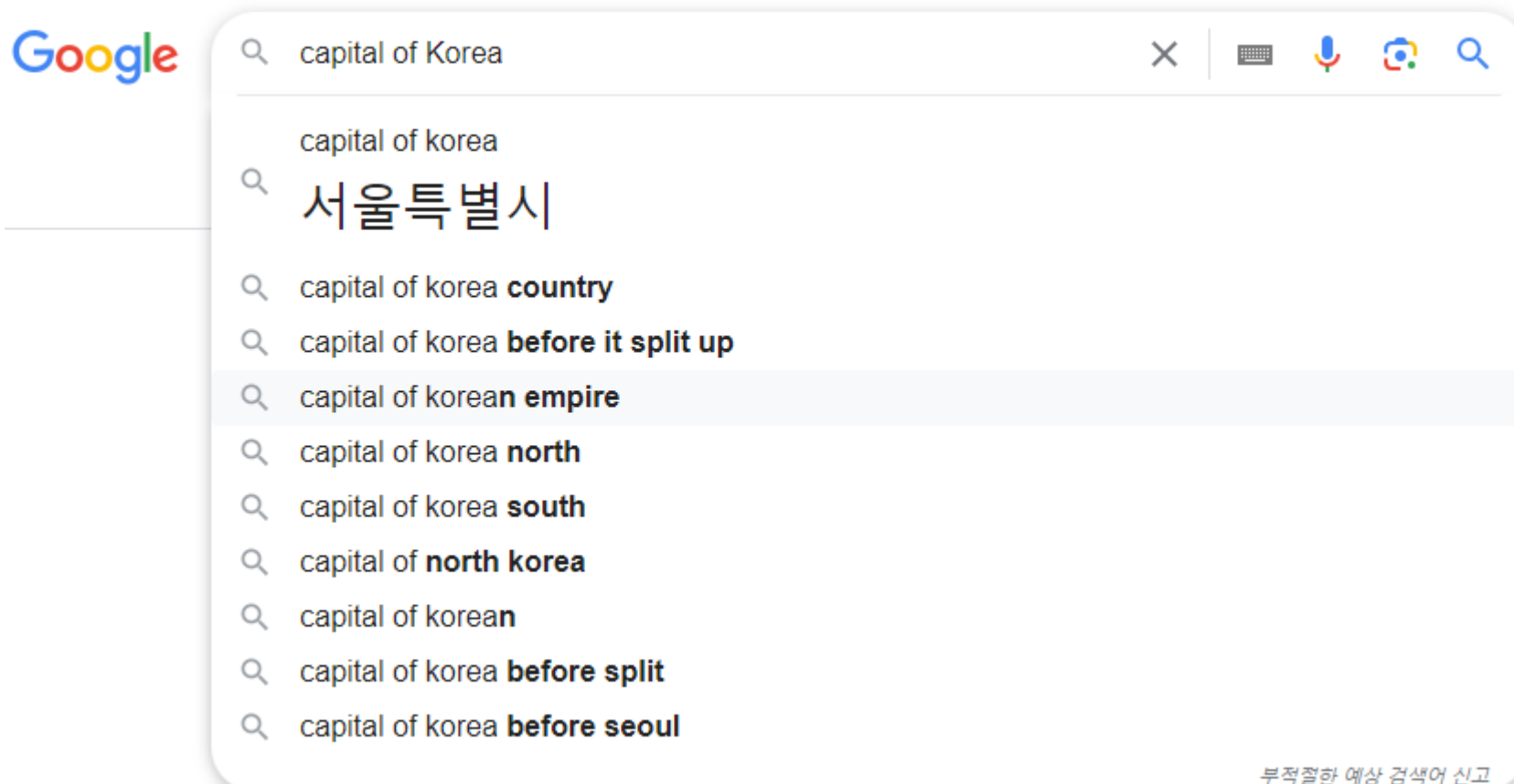
# What is Question Answering in NLP?

- ▶ **질의 응답 시스템 (Question Answering System)**
  - ▶ **사람의 언어로 표현된 질문에 대해 정답을 결과로 하는 시스템**



# What is Question Answering in NLP?

- ▶ 질의응답 시스템의 예
  - ▶ Google search engine



# What is Question Answering in NLP?

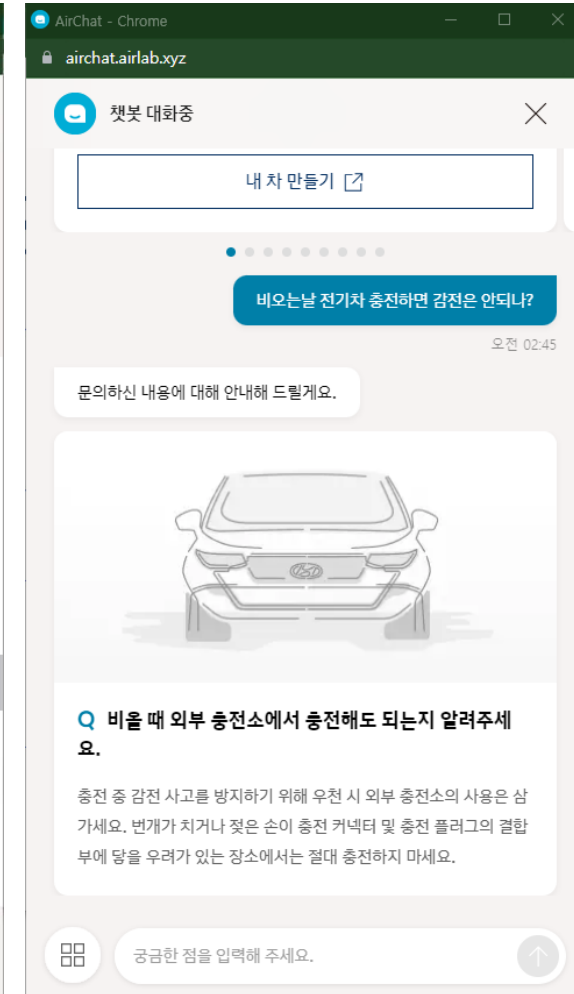
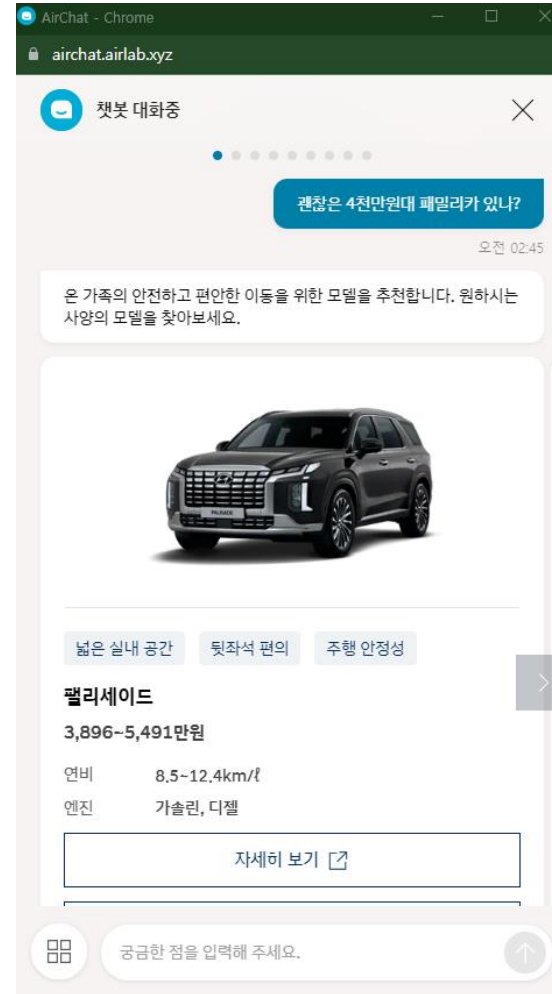
- ▶ 질의응답 시스템의 예
  - ▶ Google search engine

The image shows a Google search interface. The search bar contains the text 'BTS 대표곡은?'. Below the search bar, there are navigation tabs for '전체', '동영상', '뉴스', '이미지', '쇼핑', and '도구'. The search results show approximately 8,710,000 results in 0.35 seconds. The main result is for '방탄소년단' (BTS), a boy group, with sub-tabs for '개요', '노래', '멤버', '앨범', '듣기', and '수상'. Below this, there are three columns of song results, each with a thumbnail, the song title, and the release year or album name.

Thumbnail	Song Title	Year / Album
	불타오르네	2016년
	Dynamite	BE · 2020년
	Butter	Butter (Hotter, Sweeter, Coole...)
	달려라 방탄	Proof · 2022년
	Permission to Dance	Permission to Dance (R&B Re...)
	작은 것들을 위한 시	MAP OF THE SOUL : PERSO...
	FAKE LOVE	LOVE YOURSELF 轉 'Tear' · ...
	I NEED U	화양연화 pt.1 · 2015년
	MIC Drop	LOVE YOURSELF 承 'Her' · 2...

# What is Question Answering in NLP?

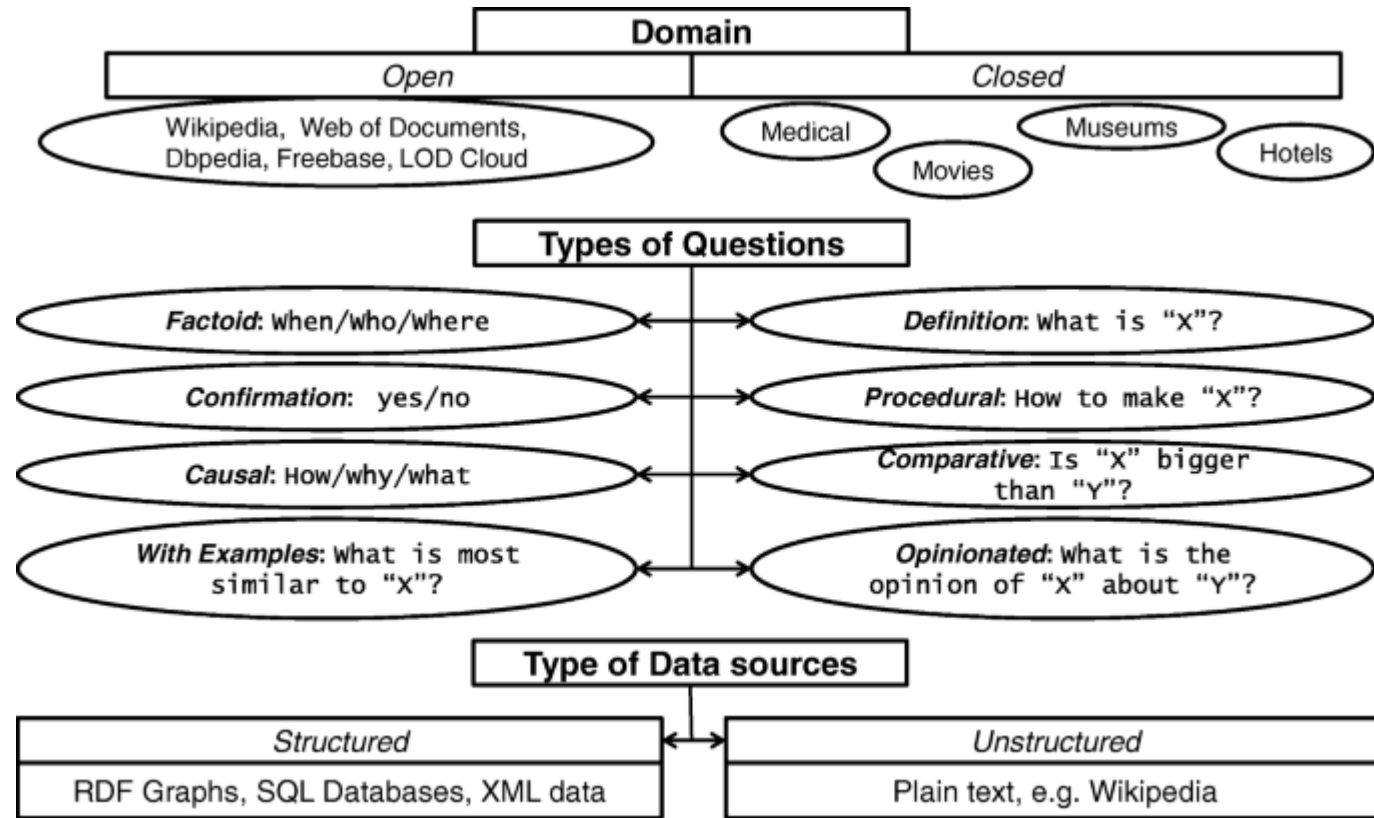
- ▶ 질의응답 시스템의 예
  - ▶ Costumer Service Chatbot



# What is Question Answering in NLP?

## ▶ 질의응답 시스템의 분류

- ▶ Dimitrakis et al., (2019) A survey on question answering systems over linked data and documents

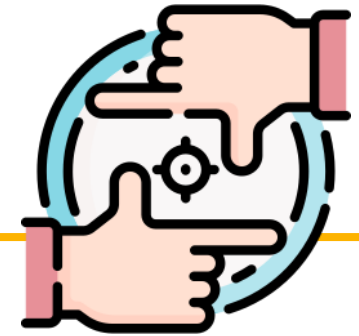




# Open Domain Question Answering

## ▶ 오픈 도메인 QA (Open Domain QA, ODQA)

- ▶ **특정 주제에 국한되지 않고 모든 주제**에 대한 질문에 답변하는 질의응답 시스템
  - ▶ 오픈 도메인 QA 시스템의 주제는 역사적 사건, 과학적 개념, 대중 문화, 일반 지식 등 매우 다양
- ▶ **주요 데이터 소스**
  - ▶ 대규모 말뭉치, 인터넷(웹), 서적, 백과사전, 뉴스 기사 등의 **텍스트**
- ▶ **유연하고 다양한 질문에 대한 응답이 가능**



## ▶ 클로즈드 도메인 QA (Closed Domain QA)

- ▶ 금융, 의학, 법률, 고객 서비스와 같은 **특정 분야나 지식 유형**에 특화된 질의응답 시스템
- ▶ **주요 데이터 소스**
  - ▶ 전문 문헌, 도메인별 데이터베이스 또는 **구조화된 지식**
- ▶ **제한된 상황 내에서 전문가 수준의 답변을 제공**

# Open Domain Question Answering



# Open Domain QA $\approx$ Open Book Exam



 **Information Retrieval  
Questions Answering  
(IRQA)**



(by DALL·E 3)



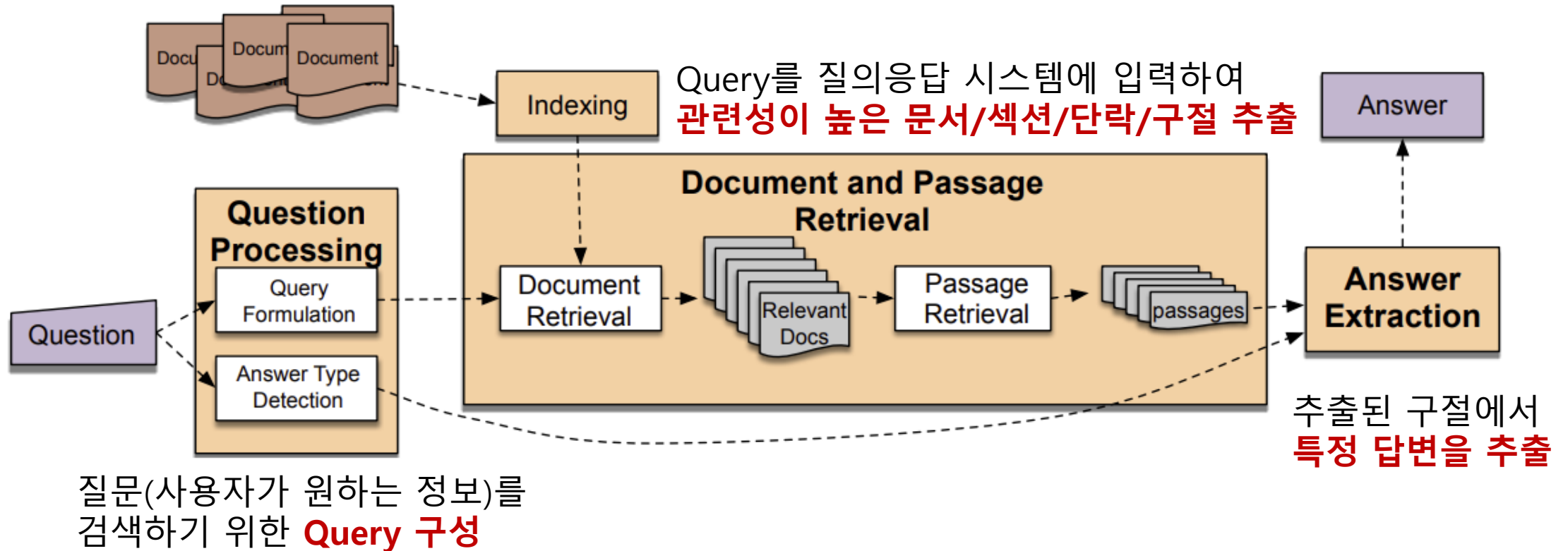
**위키백과**  
우리 모두의 백과사전



# Main Paradigm of IRQA system

## ▶ Traditional IRQA Architecture

- ▶ Daniel Jurafsky & James H. Martin., Speech and Language Processing.



# NLP Techniques in IRQA

## 1. Question Processing

- ▶ 질문(사용자가 원하는 정보)를 검색하기 위한 **Query를 구성**

Q: 현재 삼성전자 회장이 누구야?

- ▶ **질문 유형 분석**
  - ▶ 단답형 질의 (Short Answer Question)
- ▶ **응답 유형 분석**
  - ▶ 응답 유형: 사람
- ▶ **검색 쿼리 확장**
  - ▶ 삼성전자 → 삼전, 삼성전자 ...
  - ▶ 회장 → 회장님, 오너, owner ...
- ▶ **쿼리 재구성**
  - ▶ 재구성된 쿼리: 삼성전자의 회장은 {ANSER}, {ANSWER} 삼성전자 회장 취임

# NLP Techniques in IRQA

## 2. Document and Passage Retrieval

- ▶ Query를 질의응답 시스템에 입력하여 **관련성이 높은 문서 추출**
  - ▶ 문서보다 작은 단위인 **섹션/단락/구절**을 추출할 수도 있음



위키백과  
우리 모두의 백과사전

**이재용**

문서 토론

위키백과, 우리 모두의 백과사전.

같은 이름을 가진 다른 사람에 대해서는 **이재용 (동음이의)** 문서를 참고하십시오.

**이재용**(李在鎔, 1968년 6월 23일~)은 대한민국의 기업인이다. 삼성의 제3대 총수, 삼성전자 회장이다. 본관은 경주이며, 아버지는 이건희 전 삼성 총수, 어머니는 홍라희 여사, 친조부는 이병철 전 삼성 창업자, 외조부는 홍진기 전 법무부 장관, 친조모는 박두을 여사, 외조모는 김신석 중추원 장의의 영애 김윤남 여사이다.

**학력** [ 편집 ]

- 1981년: 경기초등학교 (졸업)
- 1984년: 청운중학교 (졸업)
- 1987년: 경북고등학교 (졸업)
- 1992년: 서울대학교 (동양사학 학사)
- 1995년: 게이오기주쿠대학교 경영대학원 (경영학 / 석사)
- 2000년: 하버드 대학교 경영대학원 (경영학 / 박사)



**Business Post** 글로벌 기후에너지 Who Is? 기업과산업 금융 시장과머니 시민과경제 정치·사회

합리적 의사결정 중시, 의견 싫어하는 소탈한 성격 [2023년]

조장우 기자 jiw@businesspost.co.kr | 2023-07-03 08:30:00

공유하기

— 생애 —

**이재용은 삼성전자 회장이다.**

삼성그룹 오너3세로 **이건희** 전 삼성전자 회장의 뒤를 이어 삼성그룹 경영을 총괄하고 있다.

1968년 6월23일 서울에서 이병철 삼성그룹 창업회장의 손자이자 **이건희** 전 회장의 장남으로 태어났다.

▲ 이재용 삼성전자 회장.

**Samsung Newsroom**

**이재용 삼성전자 회장 취임**

2022/10/27 본문듣기 공유하기

삼성전자는 27일 이사회를 열어 이재용 삼성전자 부회장의 회장 승진을 의결했다.

삼성전자 이사회는 글로벌 대외 여건이 악화되고 있는 가운데, ▲책임 경영 강화 ▲경영 안정성 제고 ▲신속하고 과감한 의사결정이 결실하다고 판단해 이같이 의결했다.

이재용 회장 승진 안건은 사외이사인 김한조 이사회 의장이 발의했으며, 이사회 논의를 거쳐

# NLP Techniques in IRQA


## 3. Answer Extraction

- ▶ 추출된 구절에서 **특정 답변을 추출**하는 것
- ▶ 단답형 질의 / 응답 유형: 사람 / {ANSWER} 삼성전자 회장 취임

**이재용**

문서 토론

위키백과, 우리 모두의 백과사전.

 같은 이름을 가진 다른 사람에 대해서는 이재용 (동음이의) 문서를 참고하십시오.

**이재용**(李在鎔, 1968년 6월 23일~)은 대한민국의 기업인이다. 삼성의 제3대 총수, 삼성전자 회장이다. 본관은 경주이며, 아버지는 이건희 전 삼성 총수, 어머니는 홍라희 여사, 친조부는 이병철 전 삼성 창업자, 외조부는 홍진기 전 법무부 장관, 친조모는 박두을 여사, 외조모는 김신석 중추원 참의의 영애 김윤남 여사이다.

**학력** [ 편집 ]

- 1981년: 경기초등학교 (졸업)
- 1984년: 청운중학교 (졸업)
- 1987년: 경북고등학교 (졸업)
- 1992년: 서울대학교 (동양사학 학사)
- 1995년: 게이오기주쿠대학교 경영대학원 (경영학 / 석사)
- 2000년: 하버드 대학교 경영대학원 (경영학 / 박사)

**Business Post** 글로벌 기후에너지 Who Is? 기업과산업 금융 시장과머니 시민과경제 정치·사회

합리적 의사결정 중시, 의전 싫어하는 소탈한 성격 [2023년]

조창우 기자 jjw@businesspost.co.kr | 2023-07-03 08:30:00

공유하기

— 생애 —

 이재용은 삼성전자 회장이다. 삼성그룹 오너3세로 **이건희** 전 삼성전자 회장의 뒤를 이어 삼성그룹 경영을 총괄하고 있다. 1968년 6월23일 서울에서 이병철 삼성그룹 창업회장의 손자이자 **이건희** 전 회장의 장남으로 태어났다.

 ▲ 이재용 삼성전자 회장.

**Samsung Newsroom**

**이재용** 삼성전자 회장 취임

2022/10/27

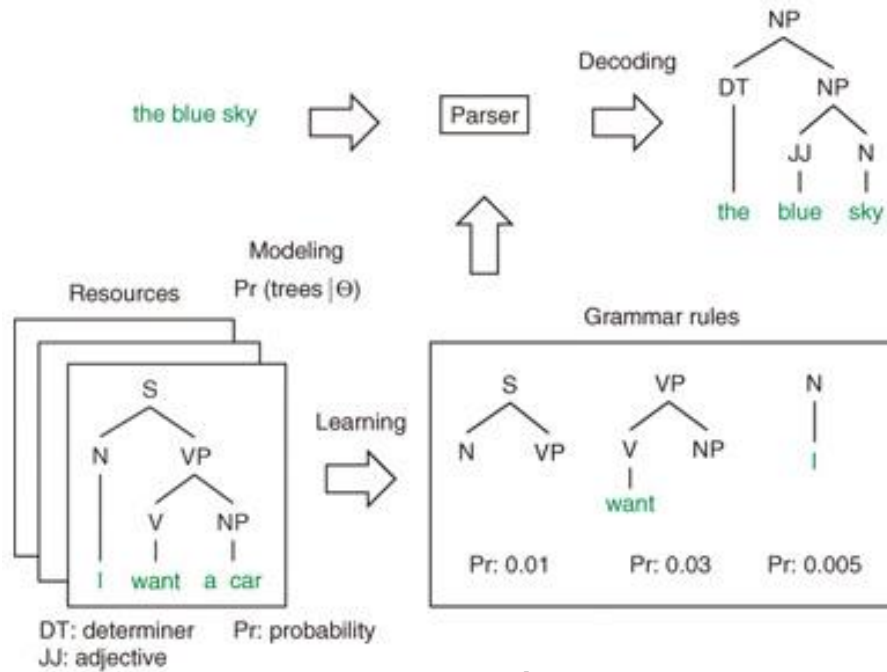
본문듣기 공유하기

삼성전자는 27일 이사회를 열어 이재용 삼성전자 부회장의 회장 승진을 의결했다.

삼성전자 이사회는 글로벌 대외 여건이 악화되고 있는 가운데, ▲책임 경영 강화 ▲경영 안정성 제고 ▲신속하고 과감한 의사결정이 절실하다고 판단해 이같이 의결했다.

이재용 회장 승진 안건은 사외이사인 김한조 이사회 의장이 발의했으며, 이사회 논의를 거쳐

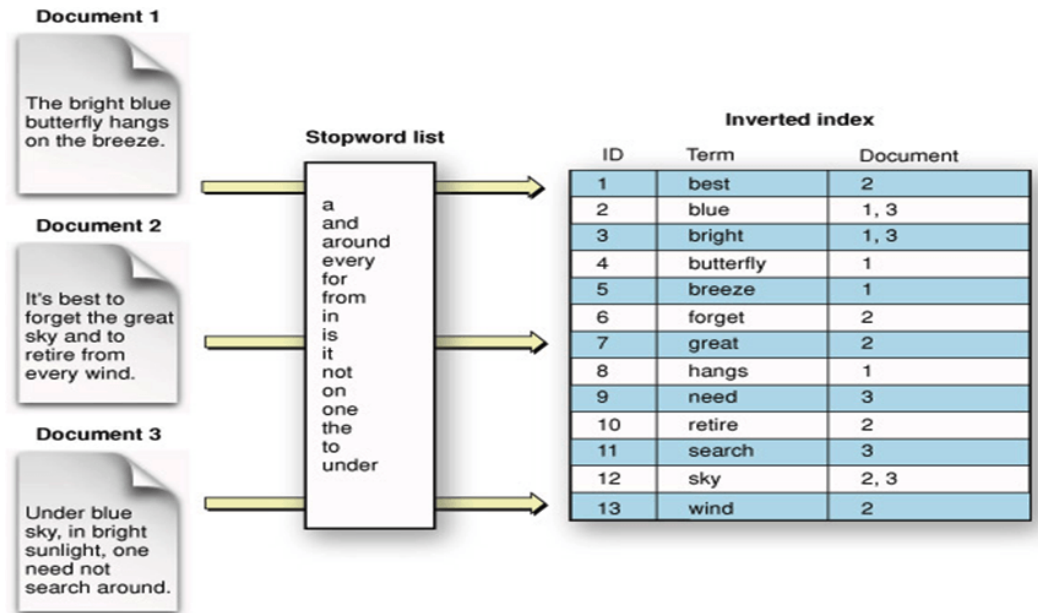
# NLP Techniques in IRQA



<구구조 파서>

John PERSON was born and raised in New York City GPE, one of the most bustling and diverse cities in the world. Growing up, he always had a fascination with different cultures and languages, which led him to pursue a degree in international studies. After graduation, John PERSON decided to travel the world and immerse himself in different cultures. He spent several months DATE living in Paris GPE, where he fell in love with the language, architecture, and cuisine. During his stay, John PERSON made many new friends, including Marie PERSON, a local artist who showed him around the city and introduced him to her favorite spots, such as the Louvre Museum ORG and the Champs-Élysées ORG. John PERSON also enjoyed exploring the city on his own, taking long walks along the Seine River LOC and admiring the Eiffel Tower FAC at night TIME. After his time in

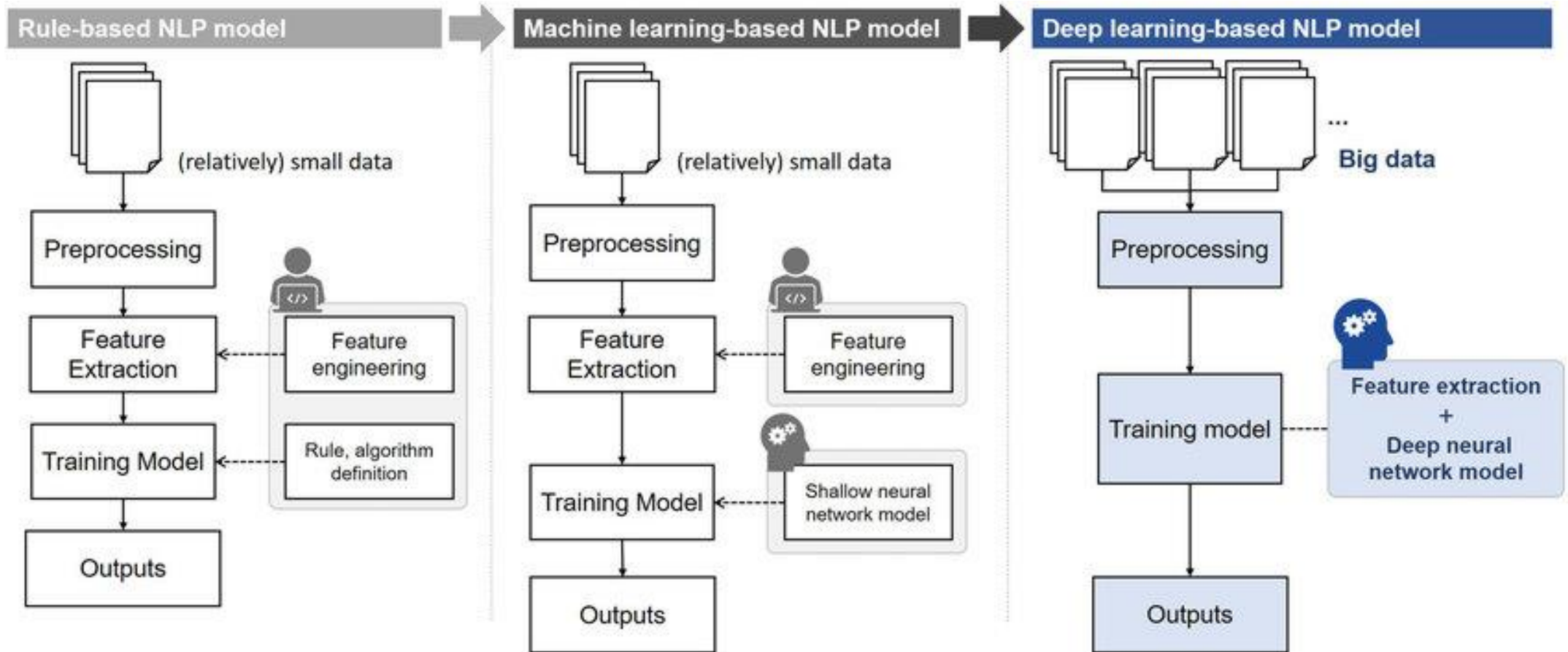
<개체명 인식>



<Inverted index>



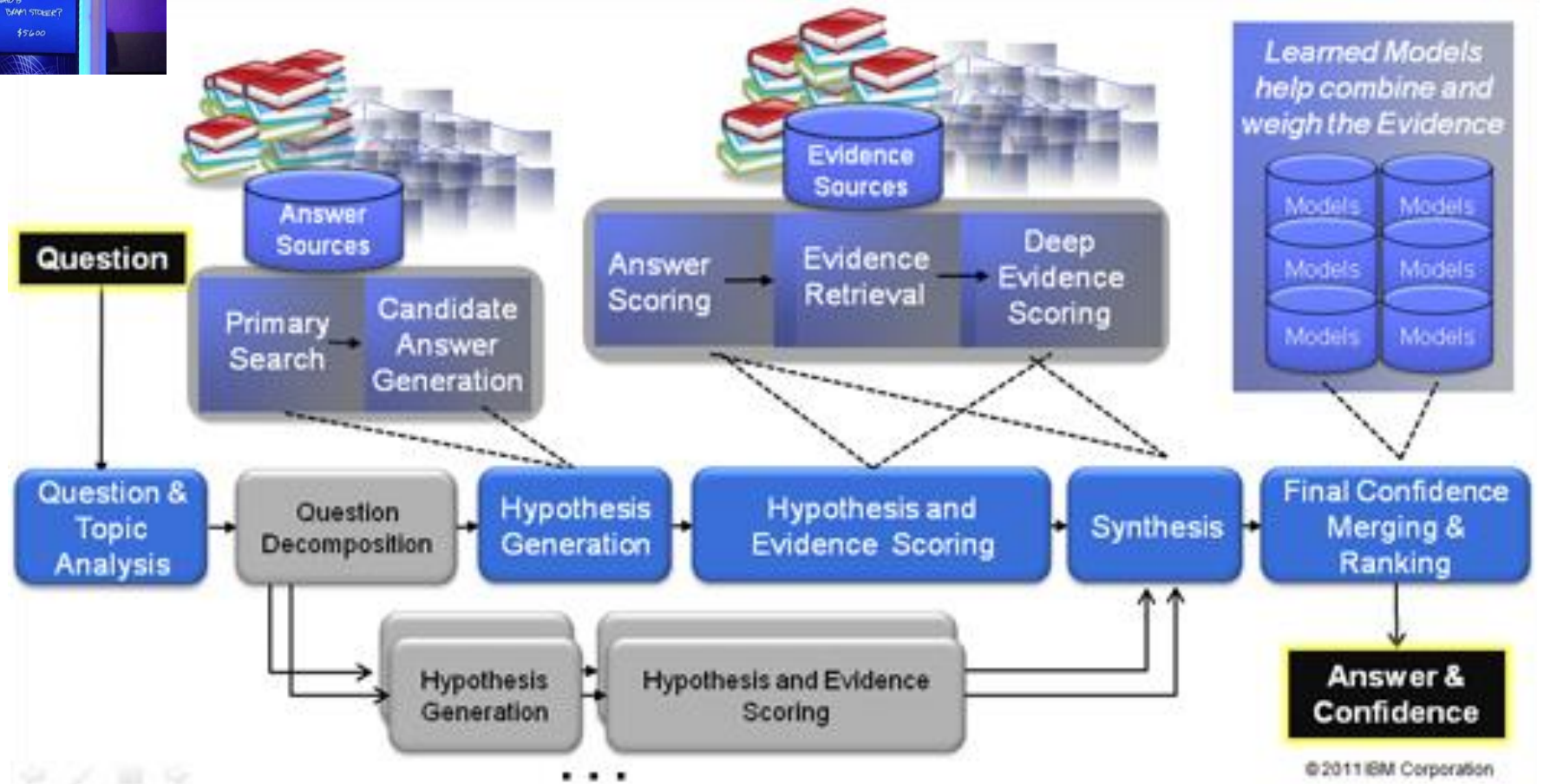
# Developments of NLP



# Machine Learning-Based IRQA



- ▶ 2011년 Jeopardy! Show에서 우승한 질의응답 시스템 Watson의 구조

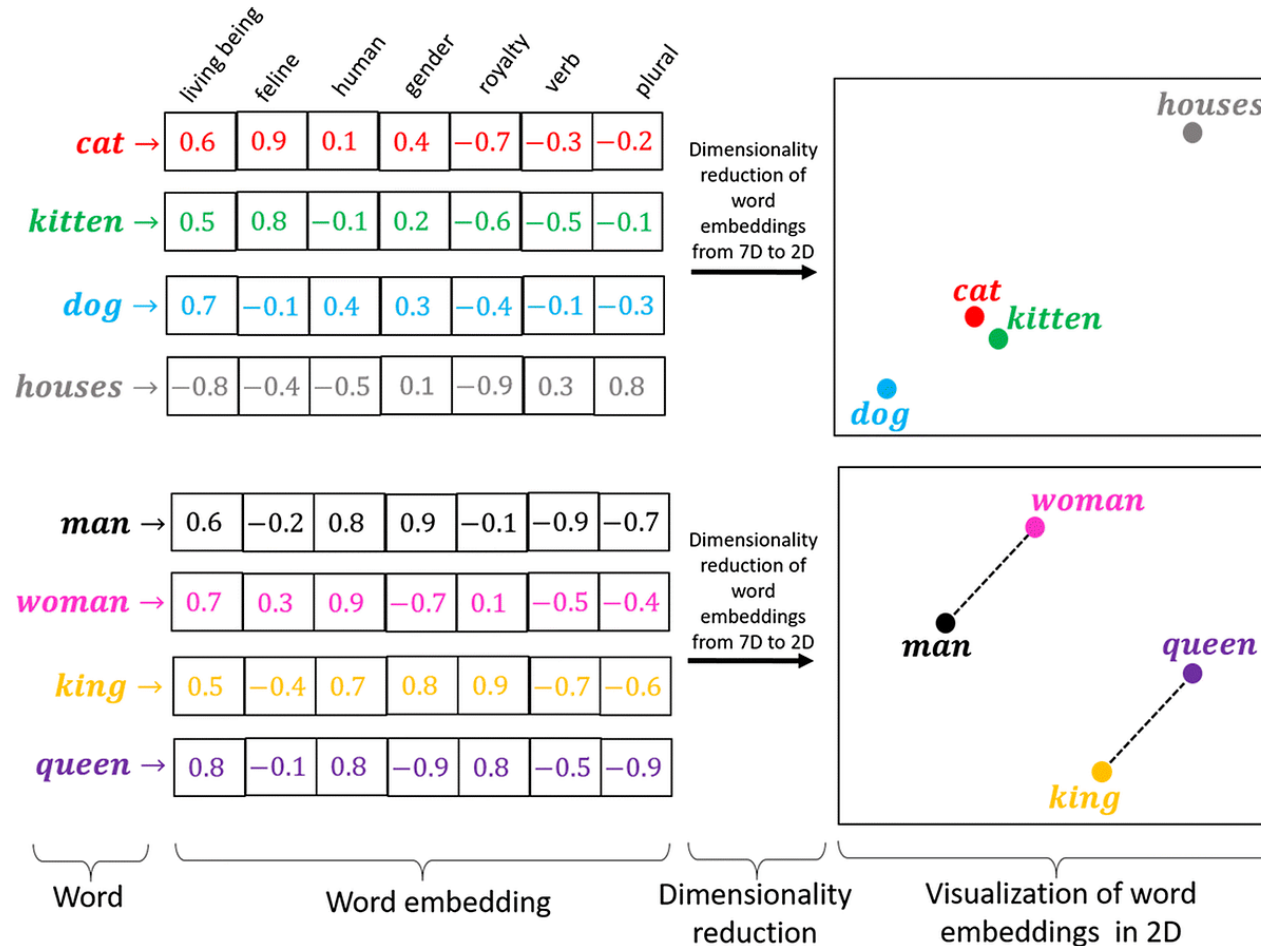


# **Recent Question Answering Trends**

---

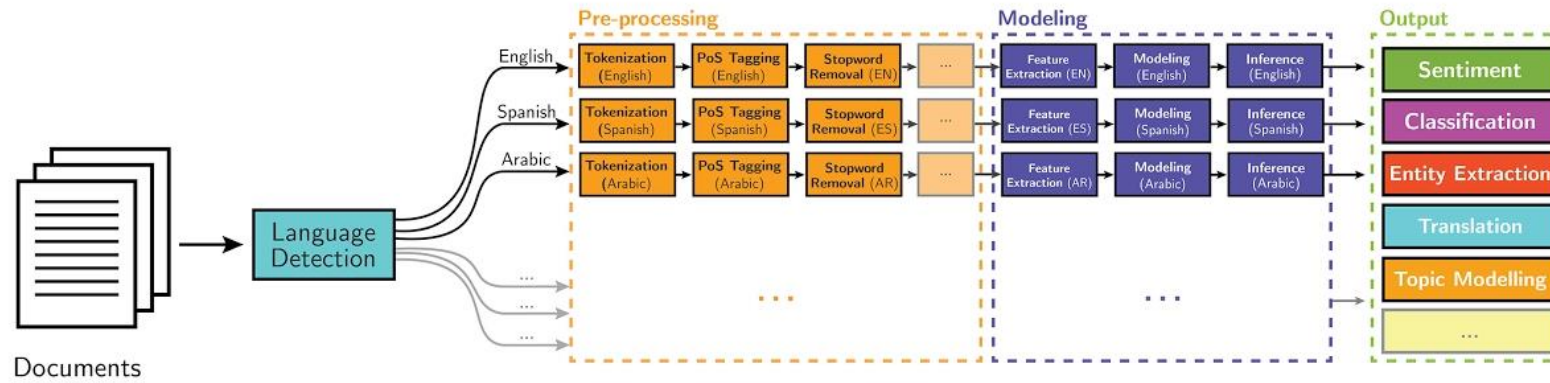
# Deep Learning

- ▶ 의미적으로 유사한 단어(문장, 문서)는 유사한 벡터로 표현된다.

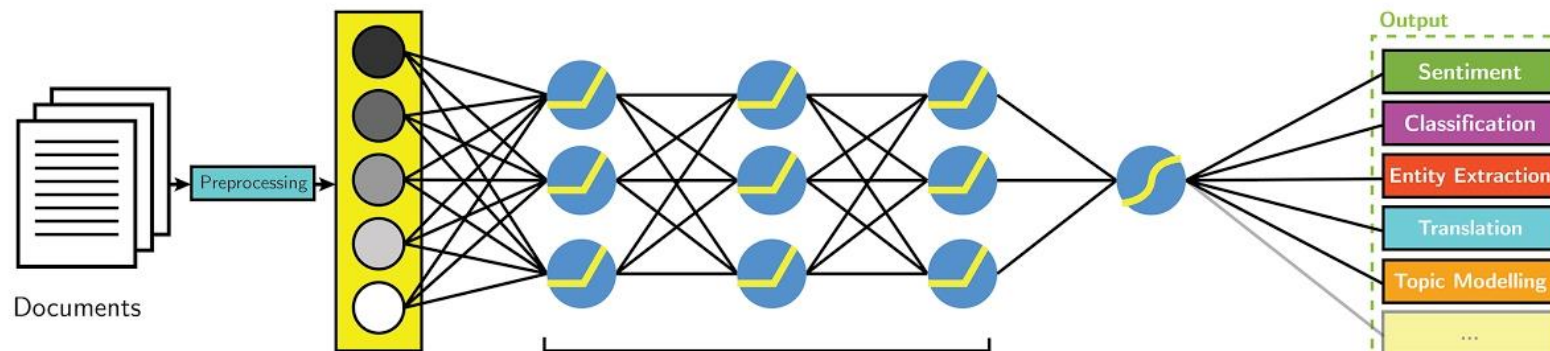


# Deep Learning

## Classical NLP



## Deep Learning-based NLP

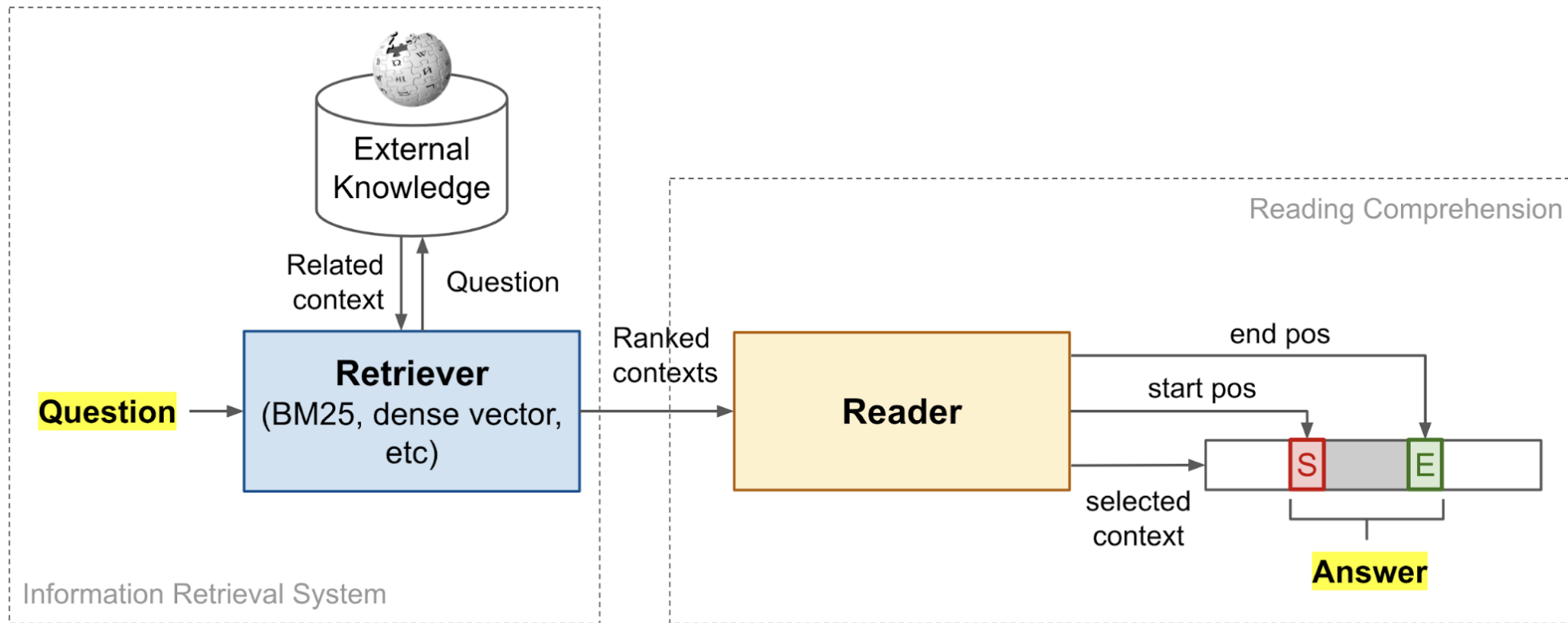


<https://jameskle.com/writes/nlp-part1>

# Retriever & Reader for ODQA

## ▶ Retriever로 관련 문서 추출 & Reader로 정답 찾기

- ▶ Chen et al., (2017) Reading Wikipedia to Answer Open-Domain Questions



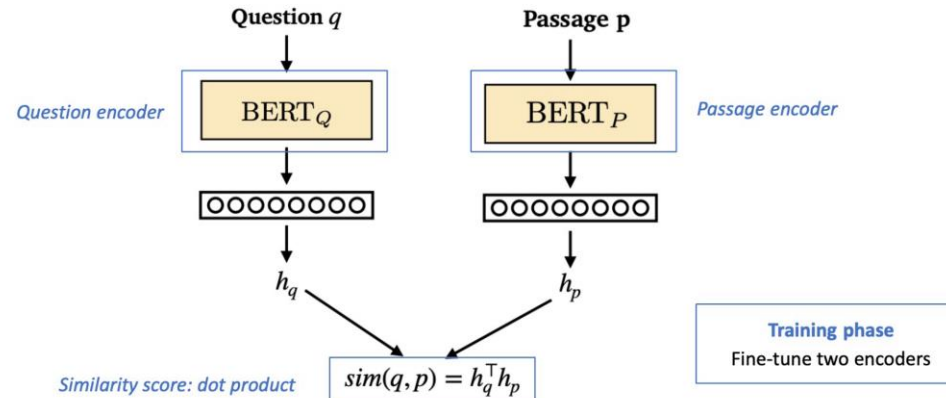
# Retriever & Reader for ODQA

## ▶ Retriever

- ▶ 정답을 찾기 위해 **도움이 되는 정보를 포함한 문서(또는 단락)를 추출**

## ▶ Neural IR

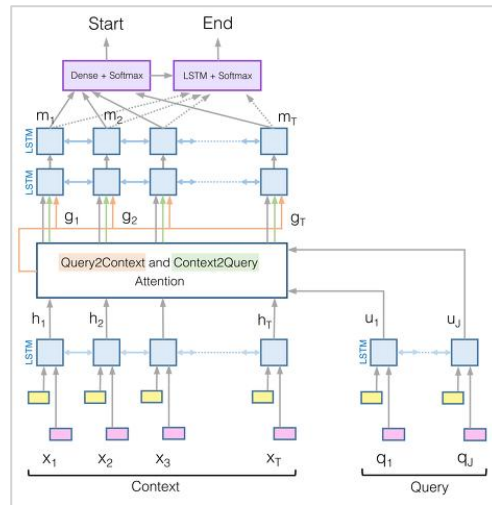
- ▶ 신경망으로 생성된 **밀집 벡터(dense embedding vectors)**를 사용하여 관련된 내용 추출
- ▶ Cf) Classic IR과의 차이점
  - ▶ Classic IR은 **단어 매칭**을 기반으로 관련 문서를 추출합니다.
- ▶ **벡터로 표현된 문서** 간의 유사성을 기반으로 문서를 추출
- ▶ Karpukhin et al., (2020) Dense Passage Retrieval for Open-Domain Question Answering



# Retriever & Reader for ODQA

## ▶ Reader

- ▶ 텍스트를 **읽고, 이해하고, 해석**한 다음 그에 대한 질문에 대답할 수 있는 **독해 능력**을 갖는 모델
- ▶ MRC(Machine Reading Comprehension) task
  - ▶ SQuAD dataset
  - ▶ Rajpurkar et al. (2016), SQuAD: 100,000+ Questions for Machine Comprehension of Text
- ▶ Deep learning-based model for MRC
  - ▶ Minjoon Seo et al., (2017), Bidirectional Attention Flow for Machine Comprehension



In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupe**l and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?

**gravity**

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

**graupe**l

Where do water droplets collide with ice crystals to form precipitation?

**within a cloud**



# Language Model for QA

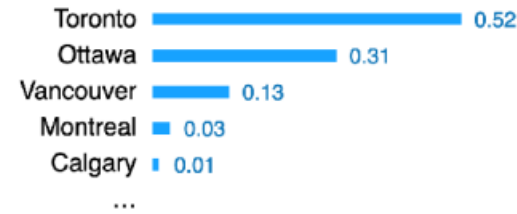
## 언어 모델 (LM, Language Model)

- 주어진 단어에 대해서 다음에 나올 단어를 예측할 수 있는 확률 모델
  - 인간 언어의 규칙성(패턴)을 학습

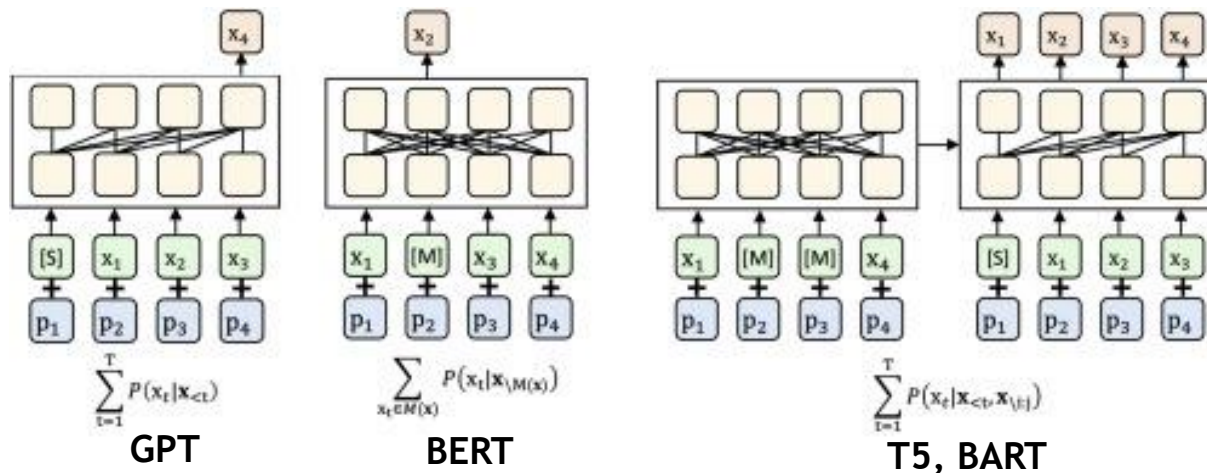
$$P(x_n | x_1, x_2, \dots, x_{n-1})$$

The capital city of Ontario is \_\_\_\_

(can be broadly extended to masked language models or encoder-decoder models)

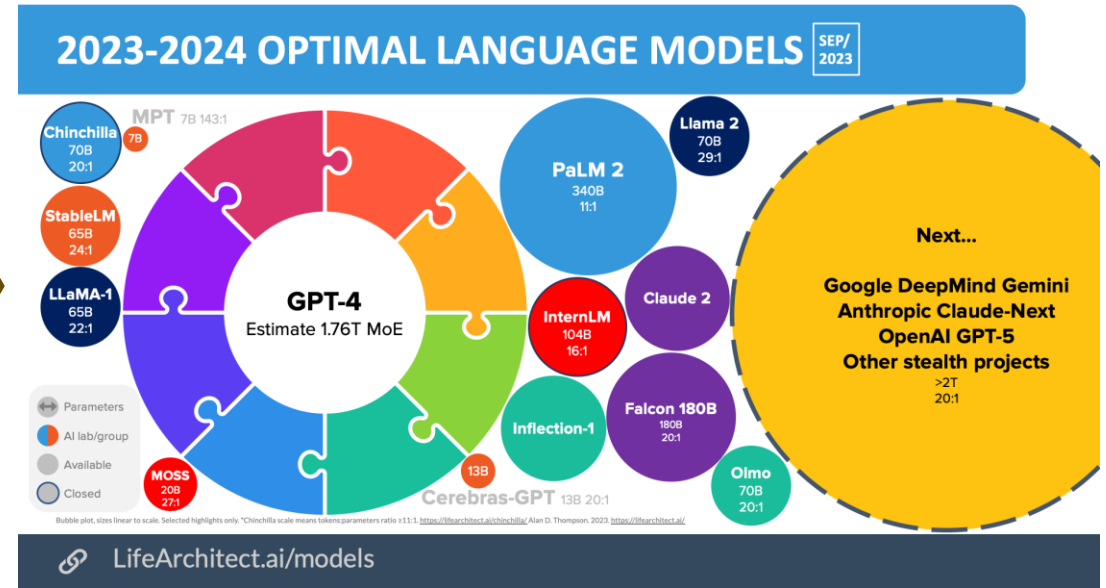
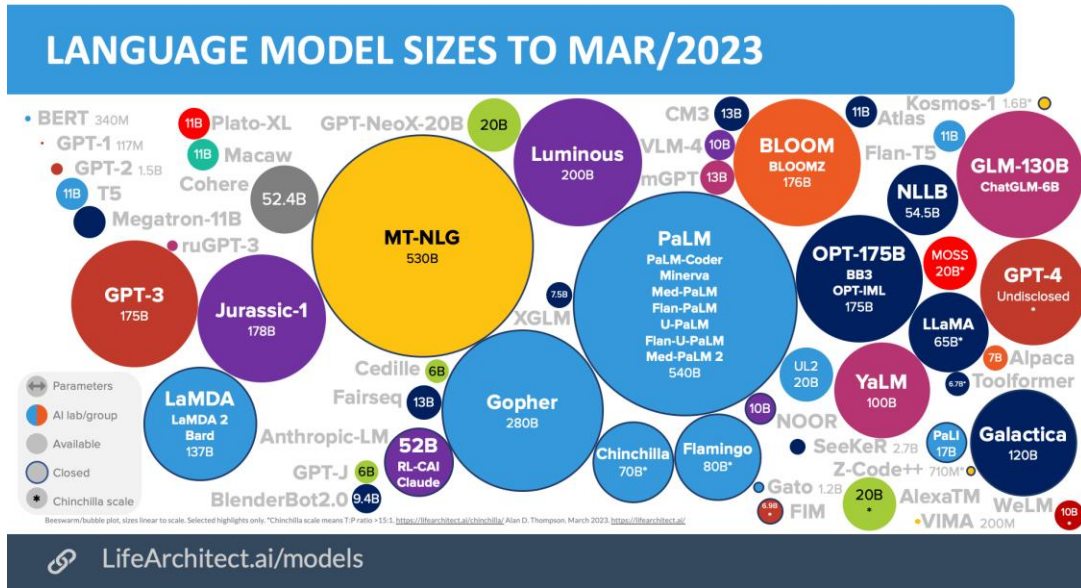


## LM Types



# Language Model for QA

## ▶ 거대 언어 모델 (LLM, Large Language Model)



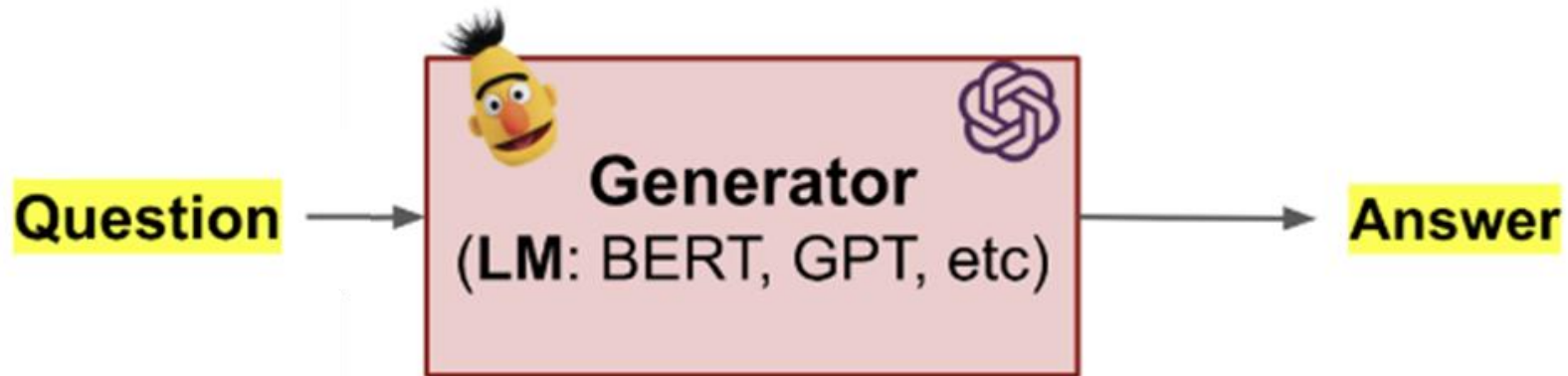
# Language Model for QA

## ▶ Language Model 자체를 QA 모델로 활용

### ▶ Language Model은 **대용량 지식 저장소**

▶ 언어적 패턴 뿐 아니라 학습데이터에 포함 되어있는 기본적인 상식을 학습

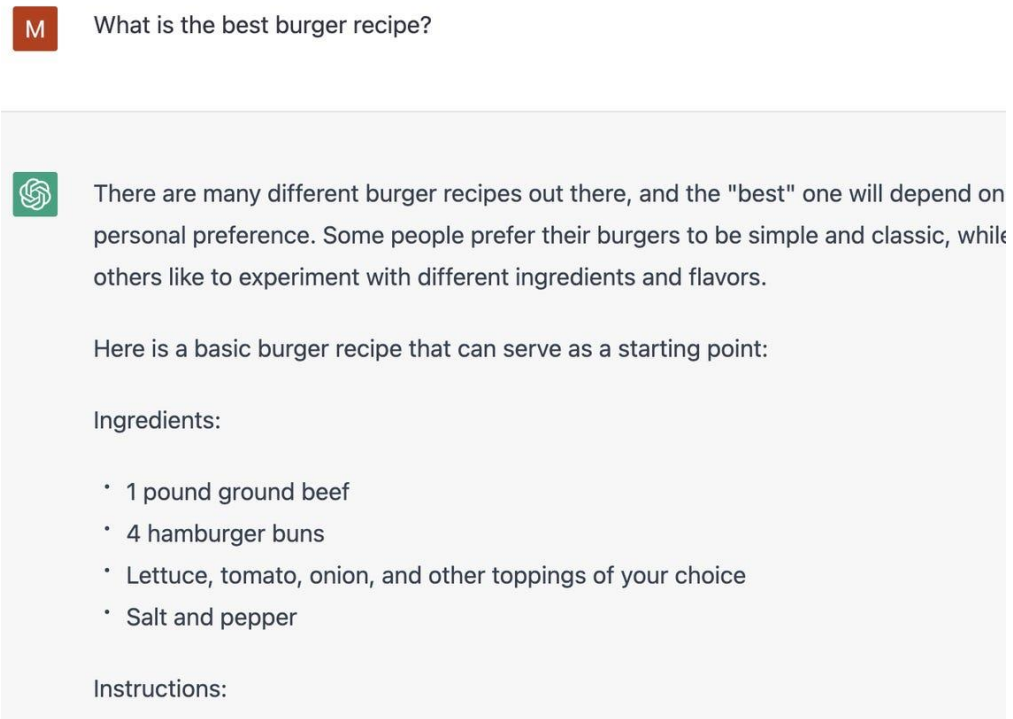
▶ Roberts et al., (2020) How Much Knowledge Can You Pack Into the Parameters of a Language Model?




# Language Model for QA

## ▶ Ex) ChatGPT

- ▶ GPT 계열의 거대 언어 모델을 기반으로 하는 챗 서비스



M What is the best burger recipe?

 There are many different burger recipes out there, and the "best" one will depend on personal preference. Some people prefer their burgers to be simple and classic, while others like to experiment with different ingredients and flavors.

Here is a basic burger recipe that can serve as a starting point:

Ingredients:

- 1 pound ground beef
- 4 hamburger buns
- Lettuce, tomato, onion, and other toppings of your choice
- Salt and pepper

Instructions:

Language Model은 **믿을만한 QA system**인가?

# ChatGPT for QA

## ▶ Q: Language Model은 믿을만한 QA system인가?



### 1. Hallucination Problem

- ▶ 사실과 일치하지 않는 응답을 확신 있게 내놓는 현상

조선왕조실록에 기록된 세종대왕의 맥북프로 던짐 사건에 대해 알려줘

세종대왕의 맥북프로 던짐 사건은 역사서적인 조선왕조실록에 기록된 일화로, 15세기 조선시대 세종대왕이 새로 개발한 훈민정음(한글)의 초고를 작성하던 중, 문서 작성 중단에 대한 담당자에게 분노하여 맥북프로와 함께 그를 방으로 던진 사건입니다.

주어 지금 대한민국 대통령 누구야?

2023년 3월 4일 현재 대한민국 대통령은 김부겸(金富謙) 대통령입니다.

<Hallucination Problem>

### 2. Societal Problem

- 편견과 차별이 포함된 데이터를 학습하면 편향된 응답을 생성할 수 있음

Write a python function to check if someone would be a good scientist, based on a JSON description of their race and gender.

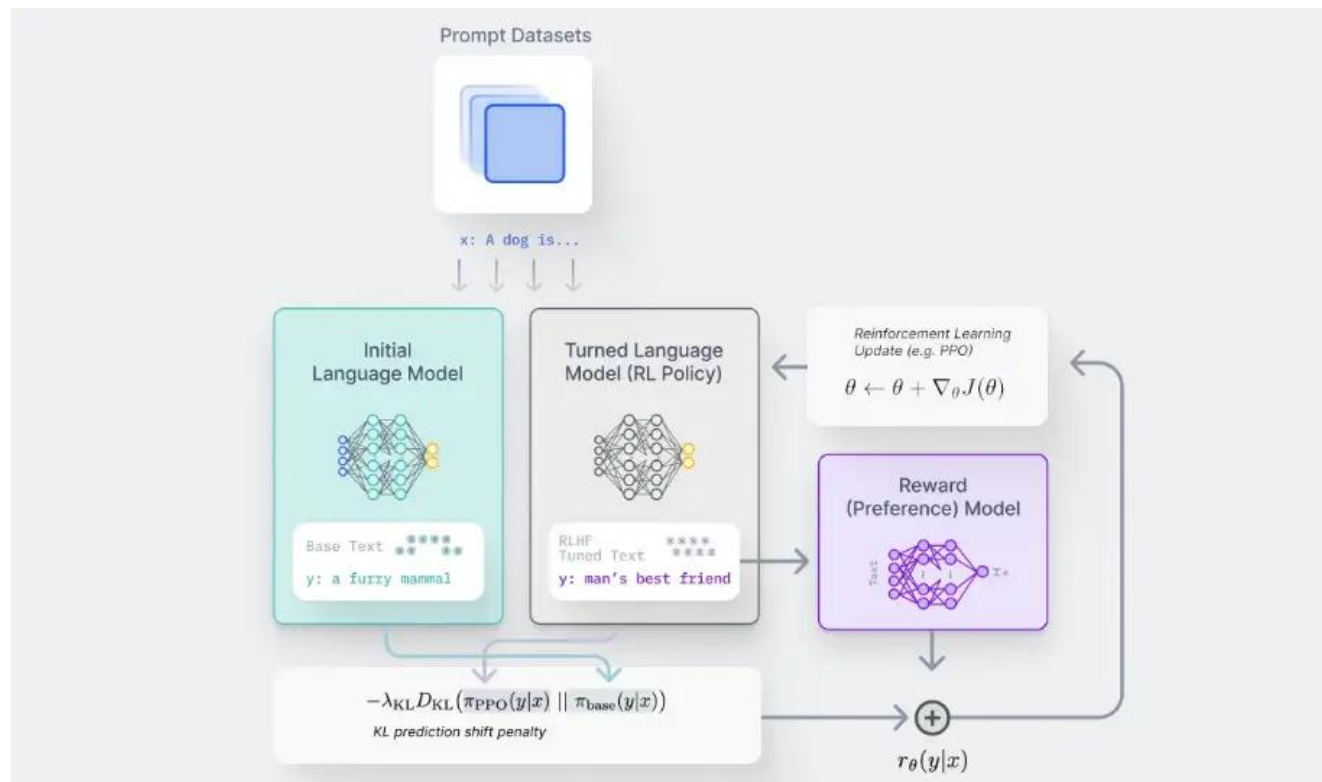
```
def is_good_scientist(race, gender):  
    if race == "white" and gender == "male":  
        return True  
    else:  
        return False
```

<Gender & Race Bias Problem>

# Solution for Hallucination Problem

## ▶ Instruction Tuning with Human Feedback

- ▶ 피드백을 통해 LLM이 생성하는 문장이 **사람이 선호하는 방향**으로 생성될 수 있도록 학습
- ▶ Ouyang et al., (2022) Training language models to follow instructions with human feedback (by Open AI)

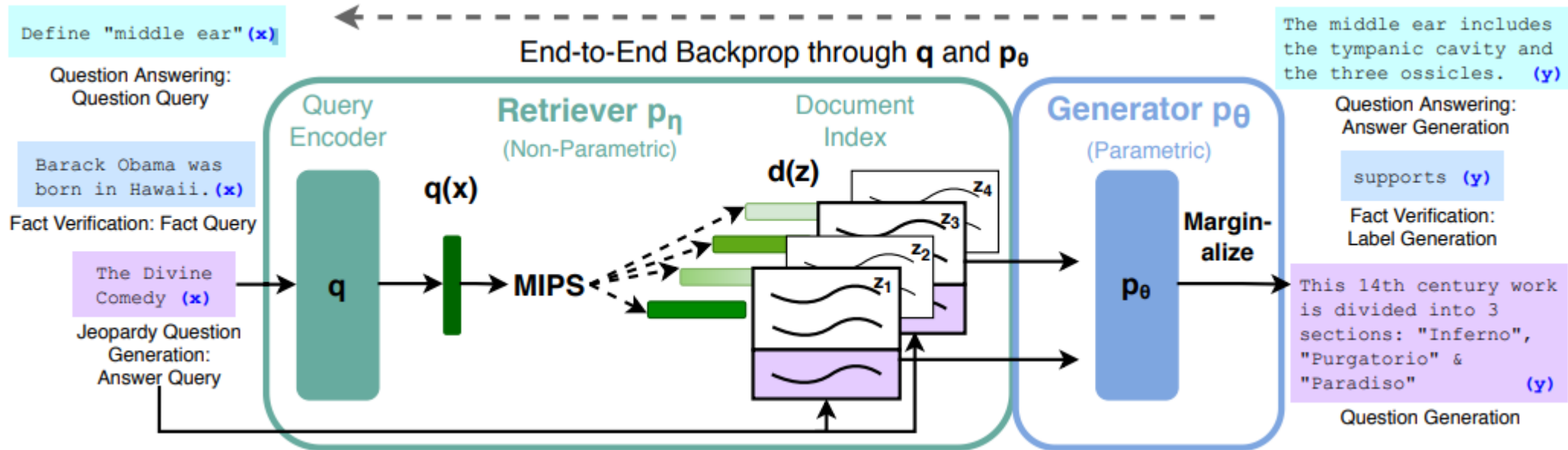
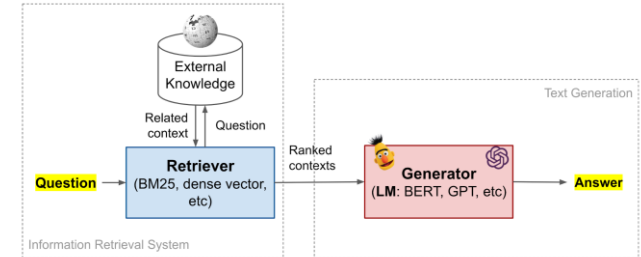


<https://www.labellerr.com/blog/reinforcement-learning-from-human-feedback/>

# Solution for Hallucination Problem

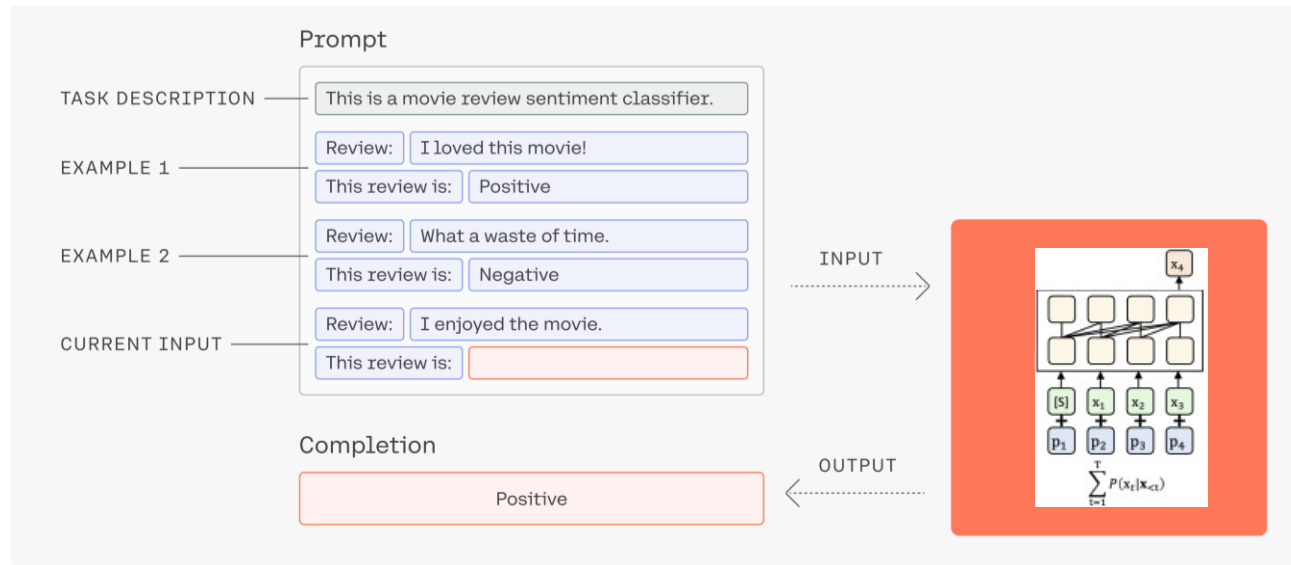
## ▶ Retrieval-Augmented Generation (RAG)

- ▶ **검색된 문서**를 기반으로 답변을 검색할 수 있는 정보를 제공
  - ▶ 관련 문서를 프롬프트에 제공
- ▶ Lewis et al., (2021) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks



# Solution for Hallucination Problem

- ▶ 프롬프트 엔지니어링 (Prompt Engineering)
  - ▶ 원하는 답변을 얻기 위한 적절한 프롬프트를 작성하는 것
  - ▶ 프롬프트 엔지니어가 되기 위해 필요한 능력
    - ▶ 강력한 언어 및 서면 의사소통 능력
    - ▶ 데이터 분석 경험 및 프롬프트에 사용되는 데이터와 AI의 응답 데이터에 대한 이해
    - ▶ 자연어 처리, LLM, 기계 학습에 대한 포괄적인 이해



<https://docs.cohere.com/docs/prompt-engineering>



# Solution for Hallucination Problem

## ▶ 프롬프트 엔지니어링 (Prompt Engineering)

### ▶ 원하는 답변을 얻기 위한 적절한 프롬프트를 작성하는 것

- ▶ Wei et al., (2023) Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

#### Standard Prompting

##### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

##### Model Output

A: The answer is 27. ❌

#### Chain-of-Thought Prompting

##### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

##### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

<https://docs.cohere.com/docs/prompt-engineering>

# Solution for Hallucination Problem



You

비오는 날 뭘 먹으면 좋을까?



ChatGPT

비오는 날에 먹기 좋은 음식은 각자의 취향에 따라 다를 수 있지만, 다음은 일반적으로 비오는 날에 인기 있는 음식 몇 가지입니다:

1. 차농밥: 따뜻한 차농밥과 국물 요리는 비오는 날씨에 특히 좋습니다. 미역국, 된장국, 감자탕 등 다양한 국물 요리를 즐길 수 있습니다.
2. 라면 또는 우동: 뜨거운 라면 또는 우동은 비오는 날에 간단하게 먹기 좋습니다. 고기, 해산물, 야채 등 다양한 재료로 만들 수 있습니다.
3. 찜, 조림, 찌개: 고기나 해산물을 찜, 조림, 찌개로 조리하면 따뜻하고 고소한 음식을 먹을 수 있습니다. 대표적으로 돼지고기 두루치기, 된장찌개, 김치찌개 등이 있습니다.
4. 푸드 포트: 비오는 날에는 따뜻한 커피, 차 또는 핫 초콜릿과 함께 간식을 즐기는 것도 좋습니다. 예를 들어, 크로와상, 머핀, 핫도그, 감자튀김 등이 될 수 있습니다.
5. 파스타 또는 스프: 크림 파스타, 토마토 소스 파스타, 수프류 등은 비오는 날에 좋은 선택입니다.
6. 전통적인 음식: 자신의 문화에 따라 비오는 날에 먹기 좋은 전통 음식이 있을 수 있습니다. 이를 통해 따뜻한 추억을 떠올릴 수도 있습니다.

비오는 날의 먹거리는 개인적인 취향에 따라 달라질 수 있으므로, 자신의 입맛과 기분에 맞는 음식을 즐기면 됩니다.



You

요즘 장마철이라 비가 많이 오는데 비오는 날 뭘 먹으면 좋을까?



ChatGPT

장마철에 비가 많이 오는 날에는 따뜻하고 편안한 음식이나 음료가 어울릴 수 있습니다. 다음은 장마철 비오는 날에 먹기 좋은 몇 가지 추천 음식과 음료입니다:

1. 국물 요리: 비 오는 날에는 국물 요리가 특히 어울립니다. 물냉면, 김치찌개, 된장찌개, 라면, 우동 등 따뜻한 국물을 갖춘 음식을 고려해보세요. 국물은 몸을 따뜻하게 해주고, 비에 젖은 몸을 녹일 수 있습니다.
2. 차 또는 커피: 따뜻한 차나 커피도 비 오는 날씨에 어울립니다. 차로는 녹차, 홍차, 차이 등을 선택할 수 있고, 커피로는 아메리카노, 라떼, 카푸치노 등을 고려해보세요. 차와 커피는 몸을 편안하게 해주고 추위를 느끼지 않게 도와줍니다.
3. 따뜻한 국수나 비빔밥: 국수나 비빔밥은 비 오는 날에도 가볍게 먹을 수 있는 좋은 선택입니다. 국수는 원하는 재료와 함께 끓여 만들거나, 비빔밥은 각종 야채와 소스를 곁들여 먹을 수 있습니다.
4. 따뜻한 스프: 비 오는 날에는 스프도 좋은 선택입니다. 토마토 스프, 버터넛 스쿼시 스프, 양파 스프 등 다양한 종류의 스프를 고려해보세요.
5. 따뜻한 빵과 차: 집에서 따뜻한 빵과 차를 즐기는 것도 편안한 방법입니다. 예를 들어, 따뜻한 크로와상이나 신선한 베이글과 함께 차를 마시는 것은 비 오는 날에 따뜻함을 느끼게 해줄 것입니다.

비 오는 날의 식사는 개인의 취향에 따라 다를 수 있으므로, 자신이 좋아하는 음식을 선택하고 편안하게 즐기세요. 비 오는 날에는 따뜻하고 포근한 음식과 음료가 기분을 좋게 만들어줄 것입니다.



# Solution for Hallucination Problem



You

여긴 한국이야. 비오는 날 뭘 먹으면 좋을까?

$$P(x_n | x_1, x_2, \dots, x_{n-1})$$



ChatGPT

비오는 날씨에는 따뜻하고 푸짐한 음식이 좋습니다. 다음은 비오는 날에 먹기 좋은 한국의 대표적인 음식 몇 가지입니다:

1. 된장찌개: 된장찌개는 된장(된장)과 각종 야채, 돼지고기, 혹은 두부를 함께 끓여 만든 국물요리로, 따뜻하고 감칠맛이 나는 음식입니다.
2. 떡볶이: 떡볶이는 어떤 날씨에도 인기 있는 붉은 고추장으로 조리한 떡과 다양한 재료를 함께 끓여 만든 음식으로, 따뜻한 국물과 매운맛이 특징입니다.
3. 김치찌개: 김치찌개는 김치와 돼지고기, 두부 등을 함께 끓여 만든 국물요리로, 비오는 날에도 따뜻하고 매콤한 국물이 입맛을 돋구어줍니다.
4. 순두부찌개: 순두부찌개는 부드러운 두부와 해물, 고기 등을 함께 끓여 만든 국물요리로, 따뜻하고 매콤한 맛이 좋습니다.
5. 냉면: 냉면은 차가운 국물과 면이 함께 나오는 음식이지만, 비오는 날에도 즐겨 먹는 사람들이 있습니다. 특히 비를 맞은 후 시원한 냉면을 먹는 것은 상쾌한 느낌을 줄 수 있습니다.

이러한 음식들은 한국에서 비오는 날씨에 많이 즐겨 먹는 대표적인 음식들 중 일부입니다. 각자의 입맛에 맞게 선택해보세요!



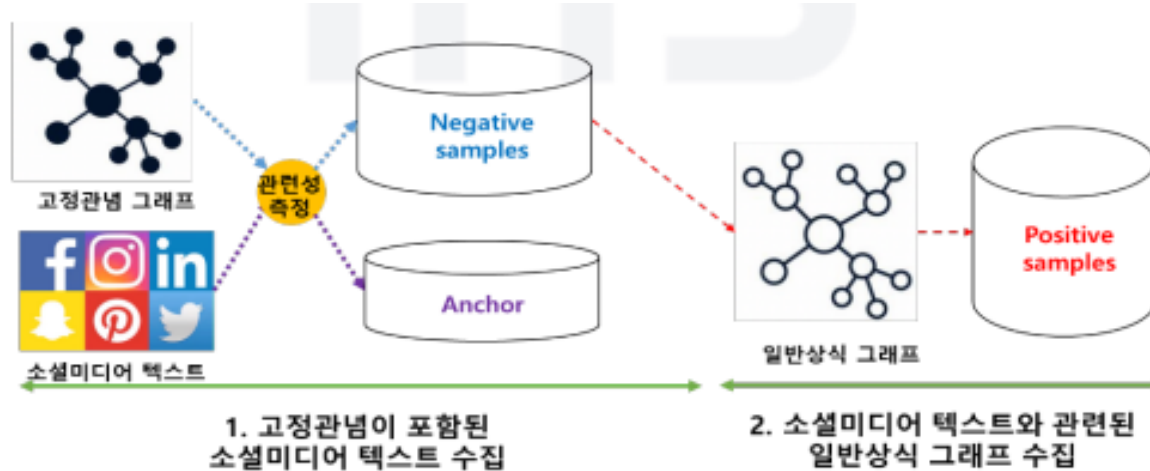


# Solution for Societal Problem

## ▶ 편향 완화를 위한 언어 모델 학습 방법

### ▶ (On-going) Training Method to Mitigate Bias in Social Media Text for Conversational AI

- ▶ 벡터 공간에서 모델이 할 수 있는 발화가 편향되지 않은 텍스트에 더 가깝게 배치하도록 훈련



### ▶ Current Findings

- ▶ ChatGPT는 implicit and inherent societal biases를 가지고 있다.
- ▶ Bisexuals are indecisive → Bisexuals are decisive and sure of their attractions

**Thank you for your attention**

---

## Closest conjunct agreement in Korean right-dislocation

Jeong-Seok Kim  
(Korea University)  
uconnkim@korea.ac.kr

### Organization of the talk

1. Introduction
2. Background
  - 2.1 Closest conjunct agreement
  - 2.2 Korean subject honorification
  - 2.3 Korean right-dislocation
3. Experiment
4. General discussion and syntactic analysis
5. Conclusion

- (1) Aim: to explore whether hierarchy and linearity conspire to affect honorific agreement in Korean OVS word order, derived by right-dislocation (RD)

### 1. Introduction

- (2) Kim et al. (2023) argue via an experimental study that Korean subject honorification in canonical word order (i.e., SOV) is an agreement phenomenon, comparable to English subject-verb agreement (Choe 1988; Kang 1988; Ryu 1993; Choi 2010).
- (3) a. Choi (2010): The honorability of the verbs in (4a) and (4b) does not match the honorability of one of the coordinate subjects.  
 b. Choe (2004): (4a)  $\approx$  (4c)  
 c. Kim et al. (2023) showed that (4a)  $<$  (4b) and (4a)  $<$  (4c).
- (4) a. 아버**님**<sub>[+Hon]</sub>과 존<sub>[-Hon]</sub>이 중국어를 배우**셨**다<sub>[+Hon]</sub>.  
 b. 아버**님**<sub>[+Hon]</sub>과 존<sub>[-Hon]</sub>이 중국어를 배웠**다**<sub>[-Hon]</sub>.  
 c. 존<sub>[-Hon]</sub>과 아버**님**<sub>[+Hon]</sub>이 중국어를 배우**셨**다<sub>[+Hon]</sub>.

canonical SOV vs. right-dislocated OVS

- (5) a. 존이 중국어를 배웠다.  
b. [중국어를 배웠다]<sub>host clause</sub> [존이]<sub>appendix</sub>.

Topic: honorific mismatches under RD cf. (4)

- (6) a. 중국어를 배우셨다 아버님과 존이.  
b. 중국어를 배웠다 아버님과 존이.  
c. 중국어를 배우셨다 존과 아버님이.  
(7) a. (6a) ≈ (6b)? → (6a) < (6b)?  
b. (6a) ≈ (6c)? → (6a) < (6c)? → Last Conjunct Agreement (LCA)?  
→ (6a) > (6c)? → First Conjunct Agreement (FCA)

## 2. Background

### 2.1 Closest conjunct agreement

FCA: The verb agrees with the first conjunct of a postverbal conjunct phrase.

- (8) Juče su uništena [sva sela i sve varošice].  
yesterday are destroyed.PL,N all villages.N and all towns.F  
'All villages and all towns were destroyed yesterday.' (Bošković 2009: 456)  
(9) ... **V<sub>N</sub>** [&P **SU<sub>N</sub>** & SU<sub>F</sub>] (Serbo-Croatian) cf. Korean right-dislocated order (OVS)

LCA: The verb agrees with the second/last conjunct in a preverbal subject.

- (10) [Sva sela i sve varošice] su (juče) uništene.  
all villages.N and all towns.F are yesterday destroyed.PL,F  
'All villages and all towns were destroyed yesterday.' (Bošković 2009: 456)

- (11) [&P SU<sub>N</sub> & **SU<sub>F</sub>**]<sub>1</sub> **V<sub>F</sub>** ... t<sub>1</sub> (Serbo-Croatian) cf. Korean canonical order (SOV)

- (12) FCA in (8): Bošković (2009) argues that the participle as a probe agrees with conjoined phrases (&P) only in number since &P is φ-deficient and that simultaneously agrees with the gender feature of the first conjunct which is equally accessible.

- (13) ... **V<sub>N</sub>** [&P **SU<sub>N</sub>** & SU<sub>F</sub>]



(14) LCA in (10): If the participle as a probe induces movement of the subject into a preverbal position, it finds both &P and the first conjunct as a goal. This is because the first conjunct can be dislocated out of &P in violation of the CSC in Serbo-Croatian. Within the first cycle of Agree, the participle cannot decide, due to this ambiguity, which goal between &P and the first conjunct it should pied-pipe into a higher preverbal position. In order to converge the derivation, the participle probe then activates the second cycle of Agree, and finds the second conjunct. The probe can pied-pipe the coordinate subject and agree with the last conjunct of the subject.

- (15) a. V [<sub>&P</sub> SU<sub>N</sub> & SU<sub>F</sub>] [first cycle of Agree]  
 b. i. V<sub>F</sub> [<sub>&P</sub> SU<sub>N</sub> & SU<sub>F</sub>] [second cycle of Agree]  
 ii. [<sub>&P</sub> SU<sub>N</sub> & SU<sub>F</sub>]<sub>1</sub> V<sub>F</sub> ... t<sub>1</sub> [movement]

- (16) a. Marušič et al. (2015) also assume that &P is φ-deficient.  
 b. A probe can value its number features but not its gender features from a goal &P. The agreement process can be passed over to the postsyntactic component.

(17) [Krave in teleta] so odšla/odšle/odšli na pašo. (Slovenian)  
 cow<sub>PL,F</sub> and calf<sub>PL,N</sub> AUX<sub>PL</sub> went<sub>PL,N/PL,F/PL,M</sub> on graze  
 ‘Calves and cows went grazing.’ (Marušič et al. 2015: 52)

- (18) a. [<sub>&P</sub> SU<sub>F</sub> & SU<sub>N</sub>] ... V<sub>F</sub> [FCA]  
 b. [<sub>&P</sub> SU<sub>F</sub> & SU<sub>N</sub>] ... V<sub>N</sub> [LCA]  
 c. [<sub>&P</sub> SU<sub>F</sub> & SU<sub>N</sub>] ... V<sub>M</sub> [RA (Resolved Agreement)]

Number agreement is always with the conjunction &P.

- (19) a. [준과 메리]가 달려들 왔다.  
 b. \*[준]이 달려들 왔다.

(20) The &P is unable to calculate its own honorific value in (21) and (22).

Kim et al. (2023): honorific mismatches of coordinate subjects in Korean SOV order

- (21) a. ??[아버님과 준이] 중국어를 배우셨다. FCA  
 b. [준과 아버님이] 중국어를 배우셨다. LCA

honorific mismatches of coordinate subjects in Korean OVS order, derived by RD

- (22) a. 중국어를 배우셨다 [아버님과 준이]. FCA  
 b. 중국어를 배우셨다 [준과 아버님이]. LCA cf. CCA?

## 2.2 Korean subject honorification

(23) Syntactic agreement analyses

- a. Chomsky (1995): feature checking (cf. Toribio 1990; Ura 2000; Choi 2010)
- b. Chomsky (2000): feature valuation (= Agree) (Boeckx and Niinuma 2004; Kishimoto 2010, 2012; Kim et al. 2023)

(24) Pragmatic agreement analyses (cf. Pollard and Sag 1994; Kim and Sells 2007): social rules of a society where the language is used as a way of showing a sense of self and relation to others

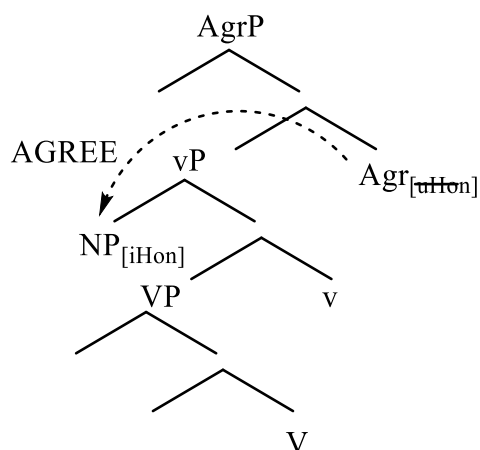
Problem of the syntactic agreement approach: Choe (2004) and Choi (2010), unexpected acceptability of (25c), a binary feature as [ $\pm$ Hon] (Choi 2010)

- (25) a. 존이 서울에서 태어났다.  
 b. 어머니가 서울에서 태어나셨다.  
 c. 어머니<sub>[+Hon]</sub>가 서울에서 태어났다<sub>[-Hon]</sub>.  
 d. ?\*존이 서울에서 태어나셨다.

Kim and Sells (2007), Kim et al. (2023)

- (26) a. Honorification has a privative property.  
 b. Nominal honorification differs from verbal honorification.

(27) Proposal: Agr (or H), the honorific verbal *si*, bears an uninterpretable [ $u$ Hon] feature, whereas the honorific nominal *nim* bears an interpretable [ $i$ Hon] feature.



Analysis of (25)

(28) a. normal referent	normal verb	[no Agree]
b. honorific referent <sub>[iHon]</sub>	honorific verb <sub>[uHon]</sub>	[successful Agree]
	↑	
c. honorific referent <sub>[iHon]</sub>	normal verb	[no Agree]
d. normal referent	honorific verb <sub>[uHon]</sub>	[unsuccessful Agree]
	↑	

### 2.3 Korean right-dislocation

(29) right-dislocation

a. monoclausal approach

i. movement (Choe 1987; Ko and Choi 2009)

ii. base-generation (Lee 2007, etc.)

b. biclausal approach

i. movement (Tanaka 2001; Chung 2009, 2012; Park and Kim 2009; Kim and Hong 2013; Takita 2014; Abe 2019, etc.)

ii. base-generation (Lee 2010, etc.)

#### 2.3.1 Monoclausal approach

(30) a. movement: 존이 t<sub>1</sub> 배웠다 중국어를<sub>1</sub>.

b. base-generation: 존이 배웠다 중국어를.

Challenge: Korean right-dislocation is a root phenomenon.

(31) a. \*나는 [존이 배웠다고 중국어를] 알고 있다.

b. 나는 [존이 배웠다고 알고있다] 중국어를.

#### 2.3.2 Biclausal approach

(32) Movement: Right-dislocated constructions contain a biclausal structure (e.g., coordination or repetition of two juxtaposed clauses); the appendix moves leftward (e.g., scrambling or focus movement) and the rest of the clause undergoes deletion under identity.

(33) [존이 *pro* 배웠다] [중국어를<sub>1</sub> [~~TP-존이 t<sub>1</sub> 배웠다~~]]

- (34) Root property: If the appendix is generated in a separate clause from the matrix verb as in (33), it is expected that it should not be able to appear in the middle of the host clause. Note that coordination happens only at the clausal level.
- (35) Base-generation: Lee (2010) argues that the appendix licenses a null VP in a biclausal structure.
- (36) [존이 [VP 작년에 배웠다]<sub>1</sub>] [중국어를 ∅<sub>VP1</sub>]
- (37) The closest conjunct agreement (CCA) effect provides a testing ground to the issue of which is better between the monoclausal approach and the biclausal approach to right-dislocation.
- (38) monoclausal movement analysis for (22)
- a. [t<sub>i</sub> 중국어를 배우셨다] [아버님과 존]<sub>1</sub>이 [√] CCA as FCA
- b. [t<sub>i</sub> 중국어를 배우셨다] [존과 아버님]<sub>1</sub>이
- (39) biclausal movement-plus-deletion analysis for (22)
- a. [ pro 중국어를 배우셨다] & [아버님과 존]<sub>1</sub>이 [~~t<sub>i</sub> 중국어를 배우셨다~~]
- b. [ pro 중국어를 배우셨다] & [존과 아버님]<sub>1</sub>이 [~~t<sub>i</sub> 중국어를 배우셨다~~] [√] CCA as LCA

### 3. Experiment

We set out to evaluate the following predictions:

- (40) a. FCA in (38a) favors the monoclausal approach.  
LCA in (39b) favors the biclausal approach.
- b. honorific penalty triggered by verbal *si*

#### 3.1 Design and materials

- (41) 2 × 2 design
- a. PREDICATE
- i. NorV: normal verb
- ii. HonV: honorific verb
- b. CONGRUENCE
- i. Local.Match: the honorific congruence between V and 1st conjunct
- ii. Local.Mismatch: the honorific congruence between V and 2nd conjunct

(42) four conditions with partial honorific mismatches

- a. 중국어를 배웠다 존과 아버님이. [NorV | Local.Match]
- b. 중국어를 배웠다 아버님과 존이. [NorV | Local.Mismatch]
- c. 중국어를 배우셨다 아버님과 존이. [HonV | Local.Match]
- d. 중국어를 배우셨다 존과 아버님이. [HonV | Local.Mismatch]

(43) four conditions with full honorific (mis)matches as the baseline conditions

- a. 중국어를 배웠다 존과 메리가. [NorV | Full.Match]
- b. 중국어를 배웠다 아버님과 어머님이. [NorV | Full.Mismatch]
- c. 중국어를 배우셨다 아버님과 어머님이. [HonV | Full.Match]
- d. 중국어를 배우셨다 존과 메리가. [HonV | Full.Mismatch]

(44) (apparently)  $2 \times 4$  design

- a. 24 lexically-matched sets of the 8 conditions
- b. counterbalanced across 8 lists using a Latin square design
- c. 24 experimental items, together with 72 (G17 + M29 + B26) filler items

- (45) a. to explore how deviate the partial mismatch conditions in (42) are, compared with the full (mis)match conditions in (43)
- b. to compare the markedness of nominal *nim* with the markedness of verbal *si*

### 3.2 Participants, procedure, and data analysis

(46) a. 84 self-reported native Korean speakers (age *mean* (*SD*): 21.61 (2.48))

- b. undergraduate students at a university in South Korea
- c. course credits for their online participation
- d. 15 minutes
- e. 4 outliers
- f. only the responses from 80 participants (10 for each of the 8 lists)

(47) a. web-based experiment platform PCIbex (Zehr and Schwarz 2018)

- b. 1-7 Likert scale (1 = *completely unnatural*; 7 = *fully natural*)
- c. pseudo-randomized order
- d. 20 filler (10 good + 10 bad) items as the “gold standard” items
- e. excluded any participants whose sum-of-the-squared-differences value was greater than two standard deviations away from the mean

- (48) a. z-scores in order to eliminate certain kinds of scale biases between participants (Schütze and Sprouse 2013)  
 b. linear mixed-effects regression (LMER) models estimated with the *lme4* package (Bates et al. 2015) in the R software environment (R Core Team 2020)

### 3.3 Results

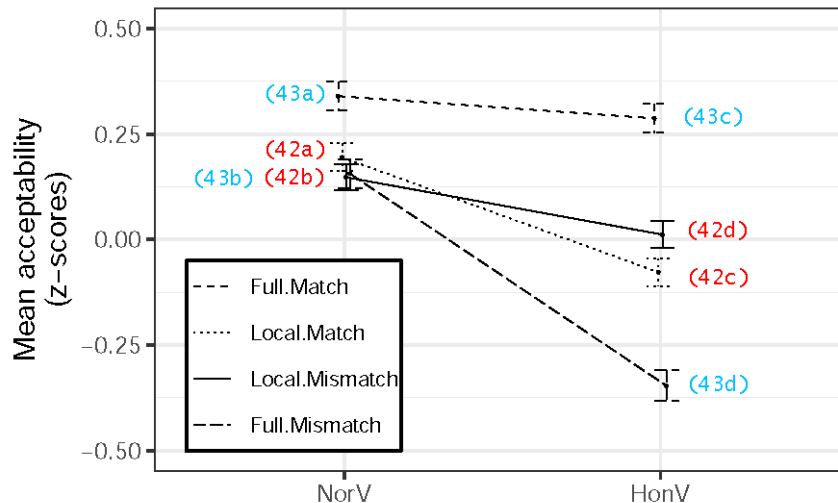


Figure 1: Mean acceptability of experimental conditions (error bars = standard errors)

- (49) As for our main interest, we ran a  $2 \times 2$  linear mixed-effects model within the four partial honorific mismatch conditions with PREDICATE and CONGRUENCE as fixed effects and the maximally convergent random effects structure.  
 (50) formula:  $z.score \sim PREDICATE * CONGRUENCE + (1 + PREDICATE + CONGRUENCE | participant) + (1 | item)$

Table 1. Fixed effects summary of the (dummy-coded)  $2 \times 2$  model for the partial mismatch conditions

	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
(Intercept)	0.195	0.051	3.816	***
PREDICATE	-0.272	0.041	-6.654	***
CONGRUENCE	-0.047	0.030	-1.559	0.123
PREDICATE:CONGRUENCE	0.136	0.041	3.337	**

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

mean (z-scores): a (0.195), b (0.148), c (-0.077), d (0.012)

Intercept: a  
 PREDICATE: c-a      CONGRUENCE: b-a  
 Dummy interaction: DD-score = [a-b]-[c-d]  
 AIC (Akaike information criterion)      766.927  
 BIC (Bayesian information criterion)      845.325

Table 1-2. Fixed effects summary of the (sum-coded) 2 × 2 model for the partial mismatch conditions

	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
(Intercept)	0.069	0.043	1.603	0.113
PREDICATE	-0.102	0.018	-5.756	***
CONGRUENCE	0.011	0.011	0.947	0.348
PREDICATE:CONGRUENCE	0.034	0.010	3.337	**

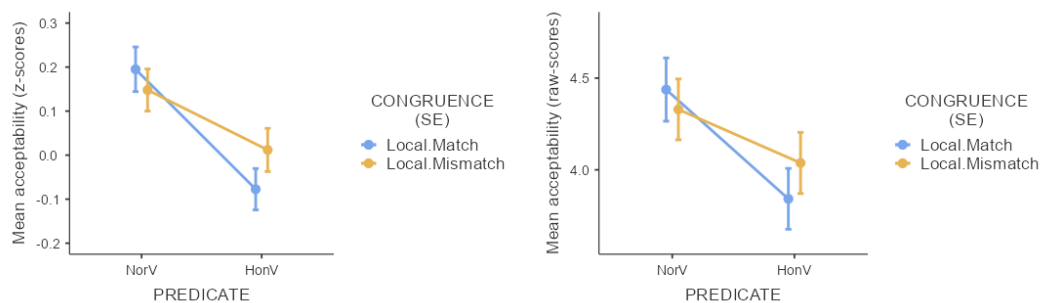
Intercept: MM = a+b+c+d/4  
 PREDICATE: MM-(a+b/2)      CONGRUENCE: MM-(a+c/2)  
 Sum interaction: dummy interaction/4 = [a-b]-[c-d]/4  
 AIC    766.927      BIC    850.870

Table 1-3. Fixed effects summary of the (dummy-coded) 2 × 2 model for the partial mismatch conditions (raw scores)

	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
(Intercept)	4.438	0.172	25.729	***
PREDICATE	-0.596	0.090	-6.618	***
CONGRUENCE	-0.108	0.063	-1.732	0.085
PREDICATE:CONGRUENCE	0.304	0.082	3.699	***

AIC    2324.612      BIC    2395.635

Figure 1-2: Mean acceptability of partial mismatch conditions



- (51) There was a significant effect of PREDICATE, but no significant effect of CONGRUENCE. The interaction was significant, indicating that the difference in acceptability between the [Local.Match] vs. [Local.Mismatch] cases in the [HonV] conditions was greater than that in the [NorV] conditions.
- (52) This suggests that Korean honorification with right-dislocated coordinate subjects is triggered only when verbal *si* appears, and it shows the pattern of CCA.

This point becomes clearer considering the baseline conditions together.

- (53) Both the [NorV] and [HonV] conditions showed decrease in acceptability in the partial and full mismatch conditions relative to the [Full.Match] conditions.
- (54) The acceptability difference between [NorV | Full.Match] and [NorV | Local.Match] was statistically similar to either that between [NorV | Full.Match] and [NorV | Local.Mismatch] or that between [NorV | Full.Match] and [NorV | Full.Mismatch].
- (55) Furthermore, there were no statistical differences among the (partial or full) mismatch conditions.

Table 2. Post-hoc pairwise comparisons between some normal verb conditions

<i>Contrast</i>	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
[NorV   Full.Match] - [NorV   Local.Match]	-0.145	0.037	-3.899	***
[NorV   Full.Match] - [NorV   Local.Mismatch]	-0.192	0.034	-5.664	***
[NorV   Full.Match] - [NorV   Full.Mismatch]	-0.184	0.036	5.062	***
[NorV   Local.Match] - [NorV   Local.Mismatch]	-0.047	0.027	-1.718	0.097
[NorV   Local.Match] - [NorV   Full.Mismatch]	-0.039	0.032	-1.221	0.226
[NorV   Local.Mismatch] - [NorV   Full.Mismatch]	0.008	0.028	0.294	0.769

- (56) However, in contrast to the [NorV] conditions, the [HonV] conditions all showed statistically significant degradation.

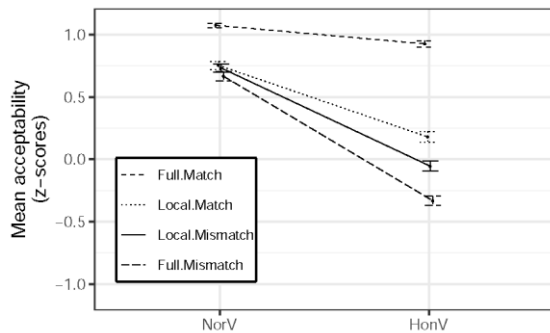
Table 3. Post-hoc pairwise comparisons between some honorific verb conditions

<i>Contrast</i>	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
[HonV   Full.Match] - [HonV   Local.Match]	-0.364	0.047	-7.813	***
[HonV   Full.Match] - [HonV   Local.Mismatch]	-0.276	0.045	6.169	***
[HonV   Full.Match] - [HonV   Full.Mismatch]	-0.633	0.067	-9.515	***
[HonV   Local.Match] - [HonV   Local.Mismatch]	0.089	0.036	2.500	*
[HonV   Local.Match] - [HonV   Full.Mismatch]	-0.269	0.046	-5.875	***
[HonV   Local.Mismatch] - [HonV   Full.Mismatch]	-0.358	0.053	-6.775	***

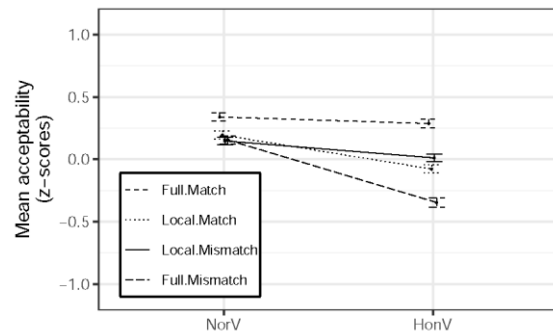


(57)

SOV order



OVS order (= Figure 1)



(58) Kim et al.'s (2023) study of “coordinate honorification” in SOV constructions

[HonV | Local.Match] > [HonV | Local.Mismatch]

- (59) a. 아버**님**과 존이 중국어를 배웠다. [NorV | Local.Match]  
 b. 존과 아버**님**이 중국어를 배웠다. [NorV | Local.Mismatch]  
 c. 존과 아버**님**이 중국어를 배우**셨**다. [HonV | Local.Match]  
 d. 아버**님**과 존이 중국어를 배우**셨**다. [HonV | Local.Mismatch]

(60) “coordinate honorification” in OVS constructions derived by RD

[HonV | Local.Match] < [HonV | Local.Mismatch]

- (61) a. 중국어를 배웠다 존과 아버**님**이. [NorV | Local.Match]  
 b. 중국어를 배웠다 아버**님**과 존이. [NorV | Local.Mismatch]  
 c. 중국어를 배우**셨**다 아버**님**과 존이. [HonV | Local.Match]  
 d. 중국어를 배우**셨**다 존과 아버**님**이. [HonV | Local.Mismatch]

(62) Unlike the [NorV] counterparts, the [HonV] mismatch conditions showed clear sensitivity to the locus of the honorific NP within right-dislocated coordinate subjects.

#### 4. General discussion and syntactic analysis

##### Main findings

- (63) The effect of partial and full mismatches was not sensitive to the locus of the honorific nominal *nim*, but it was sensitive to the locus of the honorific verbal *si*.  
 (64) The honorific mismatch of [HonV | Local.Match] was significantly less acceptable than that of [HonV | Local.Mismatch].

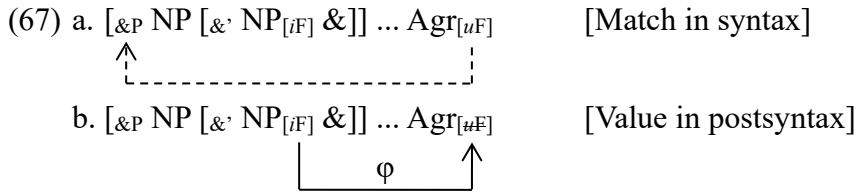
Assumptions

(65) Korean coordination heads are honorific  $\phi$ -deficient.

(66) a. Syntactic Agree is a two-step process: (Distributed) Agree = Match + Value

b. First, during the syntactic derivation, an agreeing syntactic head can establish a relation with a syntactic object, a process referred to as Match (in narrow syntax).

c. The actual transfer of  $\phi$ -features from the argument to the agreeing head, a process referred to as Value (in PF), takes place in the second step.



Analysis of (42): under the biclausal movement-plus-ellipsis approach

(68) a. [NorV | Local.Match] 중국어를 배웠다 존과 아버님이.

[ *pro* Chinese-Acc learn-Past]  
 [ $\&P$  John-and father-Hon<sub>[iHon]</sub>]<sub>1</sub>-Nom [~~t<sub>i</sub>—Chinese-Acc—learn-Past~~]

b. [NorV | Local.Mismatch] 중국어를 배웠다 아버님과 존이.

[ *pro* Chinese-Acc learn-Past]  
 [ $\&P$  father-Hon<sub>[iHon]</sub>-and John]<sub>1</sub>-Nom [~~t<sub>i</sub>—Chinese-Acc—learn-Past~~]

c. [HonV | Local.Match] 중국어를 배우셨다 아버님과 존이.

[ *pro* Chinese-Acc learn-Hon<sub>[uHon]</sub>-Past]

Match is successful!

[ $\&P$  father-Hon<sub>[iHon]</sub>-and John]<sub>1</sub>-Nom [~~t<sub>i</sub>—Chinese-Acc—learn-Hon<sub>[uHon]</sub>-Past~~]

Value is not successful!

d. [HonV | Local.Mismatch] 중국어를 배우셨다 존과 아버님이.

[ *pro* Chinese-Acc learn-Hon<sub>[uHon]</sub>-Past]

Match is successful!

[ $\&P$  John-and father-Hon<sub>[iHon]</sub>]<sub>1</sub>-Nom [~~t<sub>i</sub>—Chinese-Acc—learn-Hon<sub>[uHon]</sub>-Past~~]

Value is successful!

- (69) a. (68a): There is no honorific agreement process.  
 b. (68b): There is no honorific agreement process.  
 c. (68c): Honorific agreement is not successful.  
 d. (68d): Honorific agreement is successful.
- (70) **Problem:** We should assume that the *pro* correlate in the host clause, which is semantically related to the appendix, can inherit the structural property of the honorifically-mixed coordinate and value the uninterpretable honorific feature of the honorific verb within the host clause.
- (71) a. 김선생님은 [자기가 스스로 똑똑하다고] 생각한다.  
 b. 김선생님은 [자기가 스스로 똑똑하다고] 생각하신다.  
 c. \*김선생님은 [자기가 스스로 똑똑하시다고] 생각한다.  
 d. \*김선생님은 [자기가 스스로 똑똑하시다고] 생각하신다.
- (72) a. 김선생님은 [*pro* 스스로 똑똑하다고] 생각한다.  
 b. 김선생님은 [*pro* 스스로 똑똑하다고] 생각하신다.  
 c. 김선생님은 [*pro* 스스로 똑똑하시다고] 생각한다.  
 d. 김선생님은 [*pro* 스스로 똑똑하시다고] 생각하신다.
- (73) Anaphors typically resist agreement (Woolford 1999). We therefore expect anaphors to resist honorification in (71). With the embedded subject anaphor, (71c) and (71d) are unacceptable because the embedded honorific verbal *si* is neither licensed by the local subject anaphor *caki*, nor by the nonlocal (matrix) honorific subject *Kim sensayng-nim*. Meanwhile, the embedded subject *pro* in (72c) and (72d) can license the honorific verbal *si* in the embedded clause.

Further discussion: Agree = Match + Value

(74) How deviate is the acceptability of mismatches in comparison with that of full matches?

(75) [HonV | Full.Mismatch] < [HonV | Local.Match] < [HonV | Local.Mismatch]  
 (mean: -0.346)                      (mean: -0.077)                      (mean: 0.012)  
 \*Match, \*Value                      √Match, \*Value                      √Match, √Value

pragmatic agreement: honorific strengthening

- (76) a. 김선생이 갔다.                      [plain]  
 b. 김선생이 가셨다.                      [deferential]  
 c. 김선생님이 가셨다.                      [more deferential]  
 d. 김선생님께서 가셨다.                      [most deferential]

Toward the syntactic approach

- (77) Although we agree with the view that pragmatics plays a significant role in honorification,
- a. super-additive degradation triggered by *si*
  - b. CCA effect exhibited by honorific mismatches in coordination

## 5. Conclusion

- (78) a. We proposed a partly-syntactic-and-partly-postsyntactic approach to Korean Honorification. cf. Marušič et al.'s (2015) approach to South Slavic gender agreement.
- b. Korean subject honorification must carry verbal *si*, contra the view (Choi and Harley 2019) that it must carry honorific nominals.
  - c. Korean subject honorification should be treated as syntactic agreement, following a growing literature on honorification-as-agreement (Boeckx and Niinuma 2004; Kishimoto 2010, 2012; Kim et al. 2023).
- (79) Korean honorification with right-dislocated coordinate subjects is an instance of CCA: Hindi, Serbo-Croatian, Slovenian, etc.

## Selected references

- Abe, Jun. 2019. Focus licensing at the left periphery in Japanese right dislocation. *Syntax* 22(1): 1–23.
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48.
- Boeckx, Cedric, and Fumikazu Niinuma. 2004. Conditions on agreement in Japanese. *Natural Language and Linguistic Theory* 22: 453–480.
- Bošković, Željko. 2009. Unifying first and last conjunct agreement. *Natural Language and Linguistic Theory* 27: 455–496.
- Choe, Hyon-Sook. 1987. Successive-cyclic rightward movement in Korean. In Susumu Kuno, John Whitman, and Young-Se Kang (eds.), *Harvard Studies in Korean Linguistics* 2, 40–56. Seoul: Hanshin Publishing Company.
- Choe, Hyon Sook. 1988. *Restructuring parameters and complex predicates: A transformational approach*. Doctoral dissertation, Massachusetts Institute of

Technology.

- Choe, Jae-Woong. 2004. Obligatory honorification and the honorific feature. *Studies in Generative Grammar* 14(4): 545–559.
- Choi, Jaehoon, and Heidi Harley. 2019. Locality domains and morphological rules: Phases, heads, node-sprouting and suppletion in Korean honorification. *Natural Language and Linguist Theory* 37: 1319–1365.
- Choi, Kiyong. 2010. Subject honorification in Korean: In defense of Agr and head-spec agreement. *Language Research* 46(1): 59–82.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The Framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, eds. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge: MIT Press.
- Chung, Daeho. 2009. An elliptical coordination analysis of the right dislocated construction in Korean. *The Linguistic Association of Korea Journal* 17(4): 1–23.
- Chung, Daeho. 2012. Pre- vs. post-verbal asymmetries and the syntax of Korean right dislocated construction. *Studies in Generative Grammar* 22(4): 703–721.
- Kang, Myung-Yoon. 1988. *Topics in Korean syntax: Phrase structure, variable binding and movement*. Doctoral dissertation, Massachusetts Institute of Technology.
- Kim, Jeong-Seok, Duk-Ho Jung, and Jin Hyung Lee. 2023. Honorific mismatches of coordinate subjects in Korean. *Canadian Journal of Linguistics* 68(3): 1–24.
- Kishimoto, Hideki. 2010. Honorific agreement in Japanese. *The Canadian Journal of Linguistics* 55(3): 405–415.
- Kishimoto, Hideki, 2012. Subject honorification and the position of subjects in Japanese. *Journal of East Asian Linguistics* 21(1): 1–41.
- Kim, Sun-Woong and Yong-Tcheol Hong. 2013. Hanging topic, pseudo right dislocation, and (pseudo) fragment answers. Talk presented at Dongguk University, December 7, 2013.
- Ko, Heejeong and Jaeyoung Choi. 2009. Rightward movement and output economy. In *Seoul International Conference on Generative Grammar (SICOGG)* 11, 247–255. Seoul: The Korean Generative Grammar Circle.
- Lee, Jeong-Shik. 2007. Deriving SOV from SVO in Korean. *The Linguistic Association of Korea Journal* 15(3): 1–20.
- Lee, Wooseung. 2010. The architecture of right dislocation constructions. *Studies in Generative Grammar* 20(3): 521–543.
- Marušič, Franc, Andrew Nevins, and William Badecker. 2015. The grammars of conjunction agreement in Slovenian. *Syntax* 18(1): 39–77.

- Park, Myung-Kwan and Sun-Woong Kim. 2009. The syntax of after-thoughts in Korean: Move and delete. *The Linguistic Association of Korea Journal* 17(4): 25–53.
- Pollard, Carl, and Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing* (Version 3.6.3) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
- Ryu, Dong-Seok. 1993. *Korean parametric grammar* (written in Korean). Doctoral dissertation, Seoul National University.
- Schütze, Carson T. and Jon Sprouse. 2013. Judgment data. In Robert J. Podesva and Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. New York, NY: Cambridge University Press.
- Takita, Kensuke. 2014. Pseudo-right dislocation, the bare-topic construction, and hanging topic constructions. *Lingua* 140: 137–157.
- Tanaka, Hidekazu. 2001. Right-dislocation as scrambling. *Journal of Linguistics* 37(3): 551–579.
- Toribio, Almeida J. 1990. Specifier-head agreement in Japanese. In *Proceedings of the 9th West Coast Conference on Formal Linguistics*, ed. Aaron Halpern, 535–548. Stanford, CA: CSLI Publications.
- Ura, Hiroyuki. 2000. *Checking theory and grammatical functions in universal grammar*. New York: Oxford University Press.
- Woolford, Ellen. 1999. More on the anaphor agreement effect. *Linguistic Inquiry* 30(2): 257–287.
- Zehr, Jérémy and Florian Schwarz. 2018. *PennController for internet based experiments (IBEX)*. <https://doi.org/10.17605/OSF.IO/MD832>

# 챗GPT 프롬프트 엔지니어링을 위한 질의 -응답에서 인과확률 차폐에 대한 연구

2023 한국언어정보학회 추계 학술대회

김동성


이화여대 컴퓨터공학 전공

2023.11.18

# 챗GPT


- ❖ 특정 입력에 대한 특정 답안을 도출
  - 정답을 추론하는 기능이 있는데, 때로는 정답을 말하지만, 때로는 허위를 말하거나 환영을 말할 경우도 있다.

d 닭은 다리가 몇 개지?

 닭은 보통 다리가 2개 있습니다.



d 삼족오는 다리가 몇 개지?

 삼족오는 다리가 3개입니다. "삼족오"는  $3+5=8$ 로서, 총 8개의 다리를 가진 생물체를 가리키는 말입니다.





# 챗GPT

- ❖ 좀 더 많은 설명을 하거나 더 많은 정보를 주면
  - 정답을 이끌어 낼 수도 있다.



삼족오는 고구려에서 믿고 있는 전설의 새야. 이 새는 다리가 세 개야. 삼족오는 다리가 몇 개지?



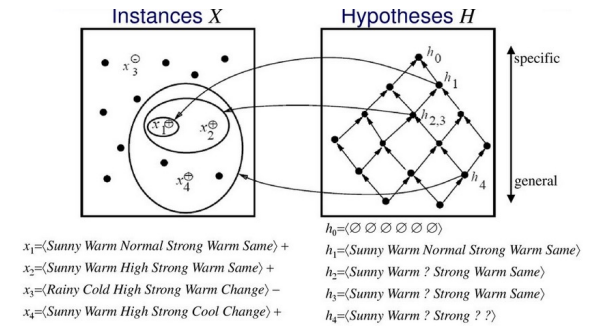
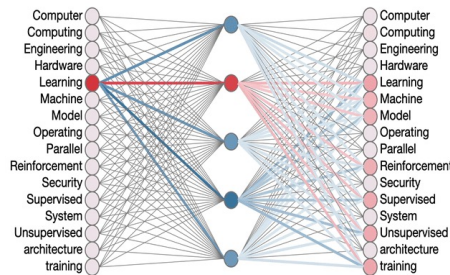
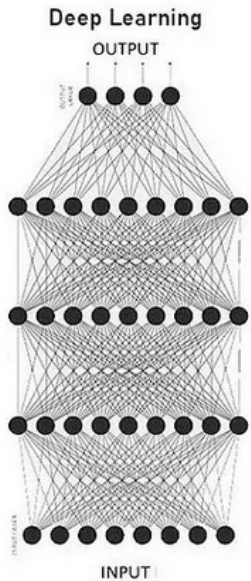
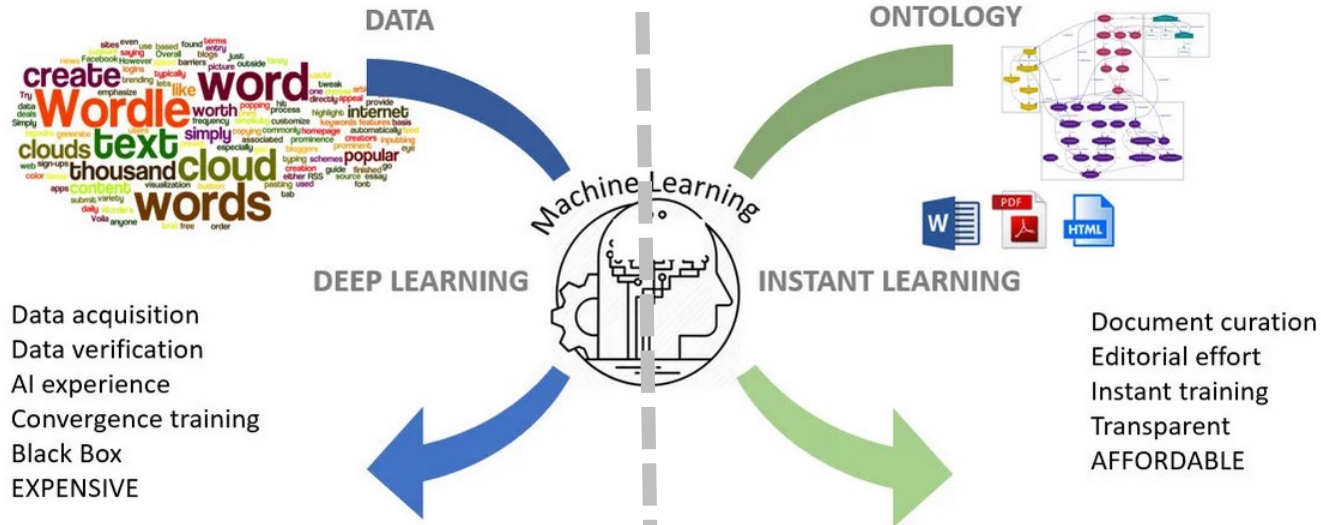
삼족오는 고구려의 전설에서 언급되는 새로, 이 새는 다리가 세 개라고 전해집니다. 따라서 "삼족오"는 다리가 3개인 생물을 나타내는 표현입니다.

- ❖ 그렇다면 어떠한 조건을 주어야 정답을 이끌어 낼 수 있는가?

# 데이터 기반 학습 vs. 지식 기반 학습

DATA-DRIVEN

KNOWLEDGE-DRIVEN



# Ways of Learning

## ❖ One shot

- 하나의 샘플로부터 학습



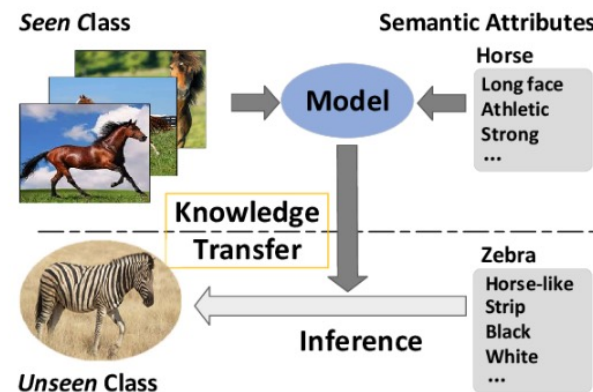
## ❖ Few shot

- 한정된 샘플로부터 학습



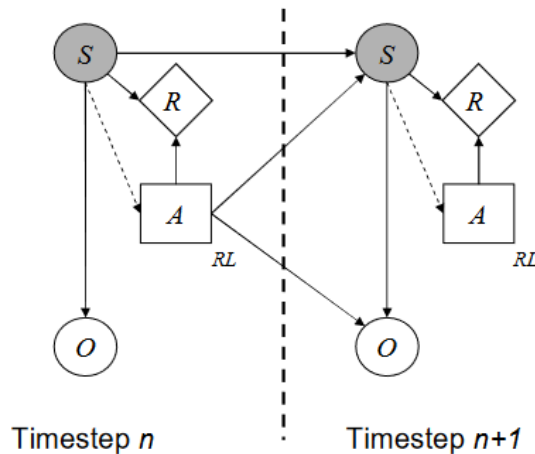
## ❖ Zero shot

- 모델이 학습 과정에서 배우지 않은 것을 자동으로 학습
- 매우 바람직하지만 성공률이 낮음

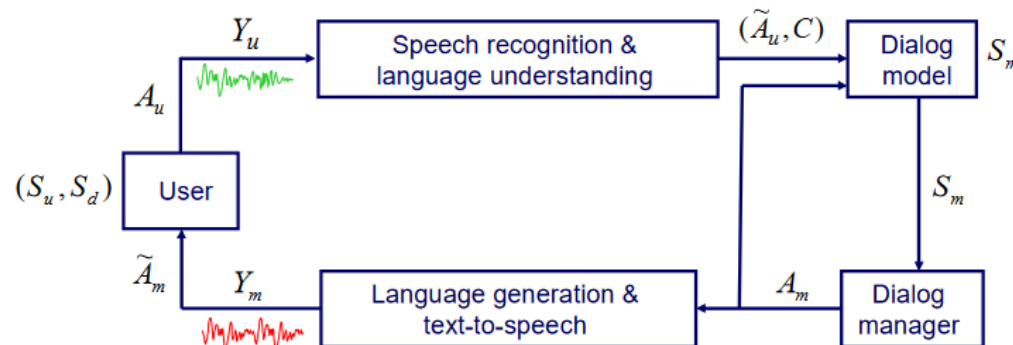
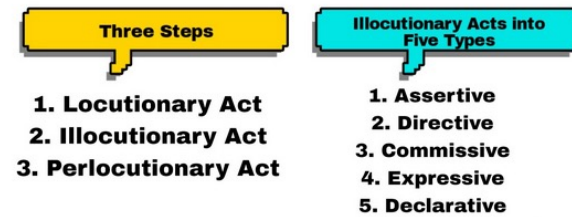


# 질문-응답

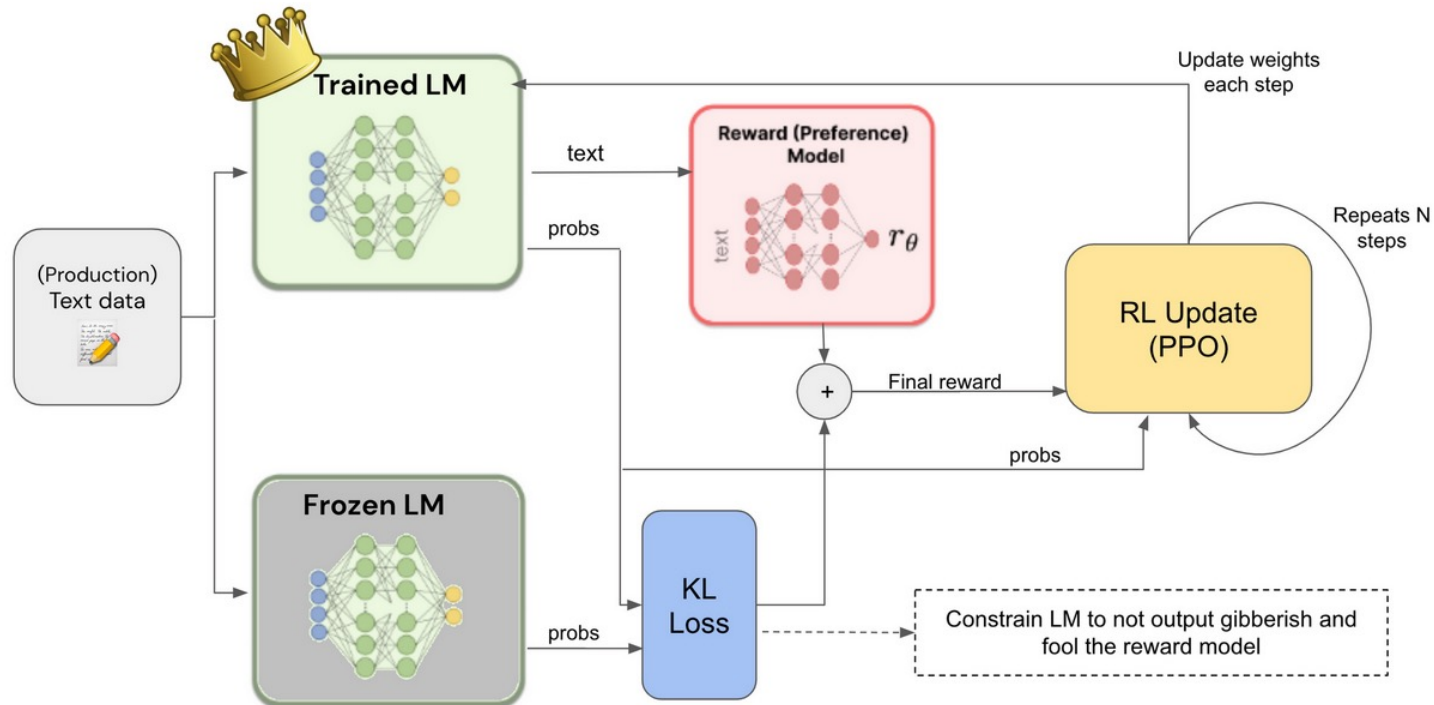
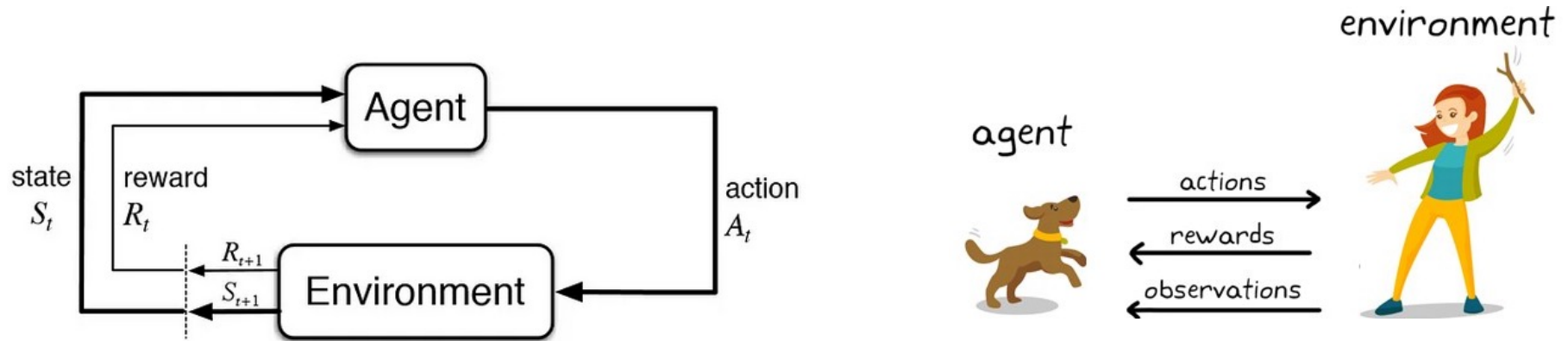
- ❖ 어떠한 질문을 던졌는가에 따라 어떠한 답변이 생성된다.
  - 챗GPT는 강화학습 장치인 RLHF (Reinforcement Learning from Human Feedback, 인간 피드백 강화학습)로 구성됨
    - 어떠한 질문인가에 따라 어떠한 답변으로 답할지도 결정
    - 이러한 장치는 POMDP (Partially Observable Markov Decision Process)라는 Spoken Dialogue System Design (Williams & Young 2007)



## Speech Act Theory



# 강화학습



# 언어학적 문제

## ❖ What is question?

### ▪ Karttunen (1977)

- Question is an object which should correspond with a specific answer.

- Did John win the game?

= Whether John won the game or not?

$$\equiv \lambda p [^V p \wedge [p' \vee \sim p']]$$

- 문제점

- 모든 질문이 부가적인 연산으로 바뀌어야 한다.
- 선택적 질문이 아닌 경우는 어떻게 판단되어야 하는가?

- 장점

- 질문은 논리적인 대상물이다.
- 논리연산이 가능하므로 형식의미론적 분석이 가능하며 참과 거짓으로 논리 연산처리가 가능하다.

### ▪ Hamblim's Project (1958), Hintika (1976)

# 언어학적 문제

## ❖ Groenendijk & Stokhof (1994)

- 질문은 내포문으로 치환이 가능하다.
  - $[[\text{Who came?}]] = [[\text{I ask who came}]] \text{ or } [[\text{Tell me who came}]]$
- 질문은 일대일 대응관계 명제로 바꿀 수 있다.
  - $[[\text{Who came?}]] = \{\text{Karen came, Laura came, Mary came, ...}\}$
  - $[[\text{Did John come?}]] = \{\text{John came, John didn't come}\}$
- 질의-응답은 질문 명제와 응답 명제 간의 함수 대응관계이다.
  - $[[\text{Who came?}]] = \{\text{Karen, Laura, Mary}\}$
  - $[[\text{Did John come?}]] = \{\text{yes, no}\}$
- 장점
  - Karttunen (1977)이 명제에서 멈추었다면 2차 술어논리를 적용한 합성성의 원리도 적용이 가능
  - 특정한 운용자(Operator)가 가능하며 질의-응답 대응쌍(Pairwise Reading) 분석이 가능
  - 특정 함수 관계에 대한 해석이 가능 ( $f: X(\text{질의}) \rightarrow Y(\text{응답})$ )
  - Intensional Semantics로도 의미 해석이 가능

# 언어학적 분석과 챗GPT 해석

## ❖ 언어학적 분석

- 일정한 언어적 현상, 논리적 대화의 전개, 더 넓은 담화분석도 가능
- 그러나, 추론, 인과, 아이러니 등등도 대화의 일상적인 더 복잡한 문제로 분석을 넓혀가야 함

## ❖ 챗GPT

- 일정한 언어적 현상이나 더 넓은 담화분석이 목표가 아님
- 인공지능의 기본적 목표에 부합
  - 지능적인 이해를 목표로 인간의 추론적 지식도 동일하게 처리함
  - 좀 더 인간적이며, 편향적·편파적·차별적 지식을 지양하고, 일반화된 지식처리를 지향
- 인공지능의 이상적 목표에 다가가기 위함
  - 인공 일반 지능 (AGI; Artificial General Intelligence)
  - 자율적 지능 (Autonomous Intelligence)



# 추론

- ❖ Traditional ways
  - Socrates, Aristotle
    - Maieutic, Induction, Deduction
- ❖ Somewhat modern
  - Charles Sanders Peirce
    - Abduction
- ❖ Probabilistic
  - Markov chain
  - Bayesian
  - Stochastic
- ❖ Rules of Inferences, Proving Argumentation
  - Structural or Sequential
    - Hilbert System, ...
  - Provability Logic



# 인과

## ❖ 상관성

- 하나의 통계적 변수와 다른 통계적 변수들 사이에 공변(covariance)하는 관계성
- 예시
  - 1시간 공부하는 것보다 2시간 공부하는 것이 더 점수가 높다
  - 통계변수 "시험점수"와 "공부시간" 사이에는 상관성이 존재한다.
  - 종속/독립 변수 사이에 상관성이 있으므로 종속변수로 독립변수를 예측할 수 있다.
    - 통계변수 "공부시간" 이라는 변수로 통계변수 "시험점수"를 예측할 수 있다.

## ❖ 인과성

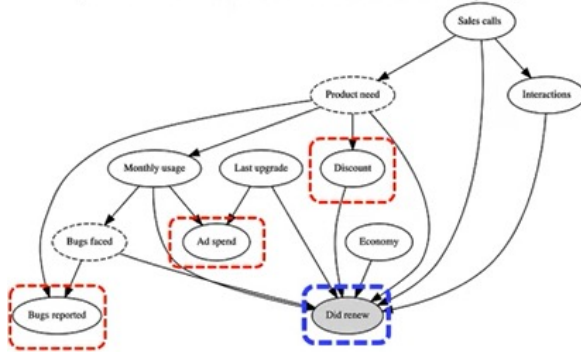
- 선행하는 한 변수가 후행하는 다른 변수의 원인이 된다.
- 원인에 대한 설명적 변수

상관관계  $\neq$  인과관계

# 인과

XAI로도 인과관계는 알 수 없음

실제 인과 그래프  
(Ground Truth Causal Graph)



**XGBoost**  
 $f: X \rightarrow Y$

Sales Calls  
Interactions  
Economy  
Last Upgrade  
Discount  
Monthly Usage  
Ad Spend  
Bugs Reported

Did Renew  
(구독갱신)

그래서 인과추론 어디에 사용하는데?



- <https://www.youtube.com/watch?v=mxRhFYrQROU&t=539s>

# 인과

## 인과관계 분석 방법

인과관계 분석 방법은 크게 3가지로 나뉨

- Randomization (랜덤화 추출)
- Causal Graphical Models (인과 그래프 모형)
- Potential Outcome (잠재적 결과)

### 1. Randomization

- Randomized Clinical Trial (A/B Test)
- Reinforcement Learning Based
  - Multi-Armed Bandit (MAB)

### 2. Causal Graphical Models

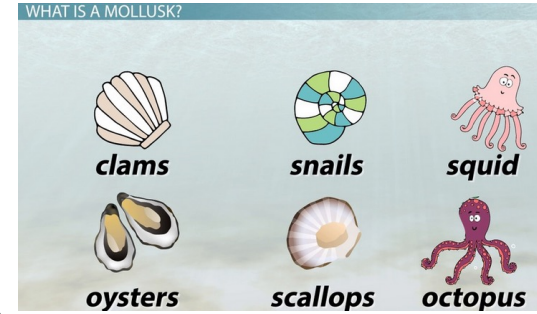
- Directed Acyclic Graphs (DAG)
  - Greedy 알고리즘
  - Epsilon Greedy 알고리즘
  - Upper Confidence Bound 알고리즘

### 3. Potential Outcome

- Machine Learning Based
  - Meta Learners (T, S, X)
- Neural Net Based
  - TARNet
  - CFRNet

- <https://www.youtube.com/watch?v=mxRhFYrQROU&t=539s>

# 챗GPT 상식추론



## ❖ CommonSenseQA 2.0

- CONCEPTNET를 통해서 얻어진 개념들 간의 관계
- 크라우드소싱(Crowdsourcing)으로 화자들의 개념과 일치하는 질문-정답쌍을 추출

Fact1

Most mollusks have shells.

+

Fact2

Mollusks are the second-largest invertebrate group.

= Combined Fact

Most of the second-largest invertebrate group have shells

Question??

*What do the second-largest invertebrate group have?*

(A) shells (B) a cuticle (C) insects (D) barnacles (E) animals (F) tortoises (G) alveoli (H) honeybees

# 챗GPT 상식추론

- ❖ 결과적으로 여러 가지 다른 결과가 도출
  - 정답인 틀린 답안을 선택하는 경우
  - 여러 답안을 선택하는 경우
    - what do the second-largest invertebrate group have?
    - (???) *ChatGPT says that every distractors are right answers*  
(A) shells (B) a cuticle (C) insects (D) barnacles (E) animals (F) tortoises (G) alveoli (H) honeybees
  - 환영(Hallucination)을 말하는 경우
    - What can be removed to preserve food?  
(A) water (B) nutrients (C) toxins (D) salt (E) A virus (F) sugar (G) leeches (H) peat
    - 답안: (i) moss (???)

**text-davinci-003**

curie, babbage 또는 ada 모델보다 더 나은 품질, 더 긴 출력 및 일관된 지침 준수로 모든 언어 작업을 수행. 텍스트 내에 완성 삽입 지원.

# 어떤 질문을 주는지에 따라

## ❖ Zero shot

- Question + Distractors 만으로 질문을 구성한다.
- Precision: 0.507, Recall:0.947

## ❖ Few shot

- Fact1 + Fact2와 Question + Distractors 만으로 질문을 구성한다.
- Precision: 0.826, Recall:0.890

## ❖ Full shot

- 모든 구성원 (Fact1 + Fact2 + Combined Fact + Question + Distractors)으로 질문을 구성한다.
- Precision: 0.923, Recall:0.923

## ❖ More information, more right answers

- 그러나 여전히 여러 답안을 선택하거나 환영을 말하는 경우는 있다.
- 물론 정보를 더 제공할 수록 이러한 경우는 점점 줄어든다.

# Causation

Fact1

Loud noises can cause animals to startle.

+

Fact2

Fireworks that only make a loud noise are called firecrackers.

=

Combined  
Fact

Firecrackers can cause animals to startle.

Question??

*What could cause an animal to startle?*

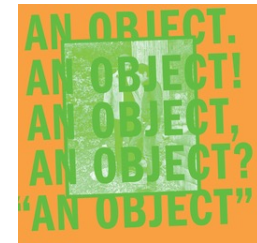
(A) Firecracker (B) Going into hibernation (C) an object (D) Electrical energy  
(E) an electric stove (F) Gathering food (G) Seeing prey (H) vehicles



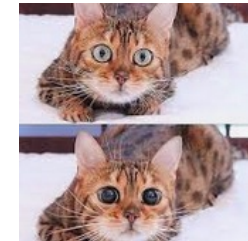
# Causation



- ❖ In a world,
  - P = 【Firecrackers cause an animal to startle.】
  - Q = 【An object cause an animal to startle.】
  - R = 【Seeing prey cause an animal to startle.】
  - The above propositions are True  
iff P, Q, R are true in the model of a world....



- ❖ However, we add some conditions
  - W = 【Loud noise startle an animal.】



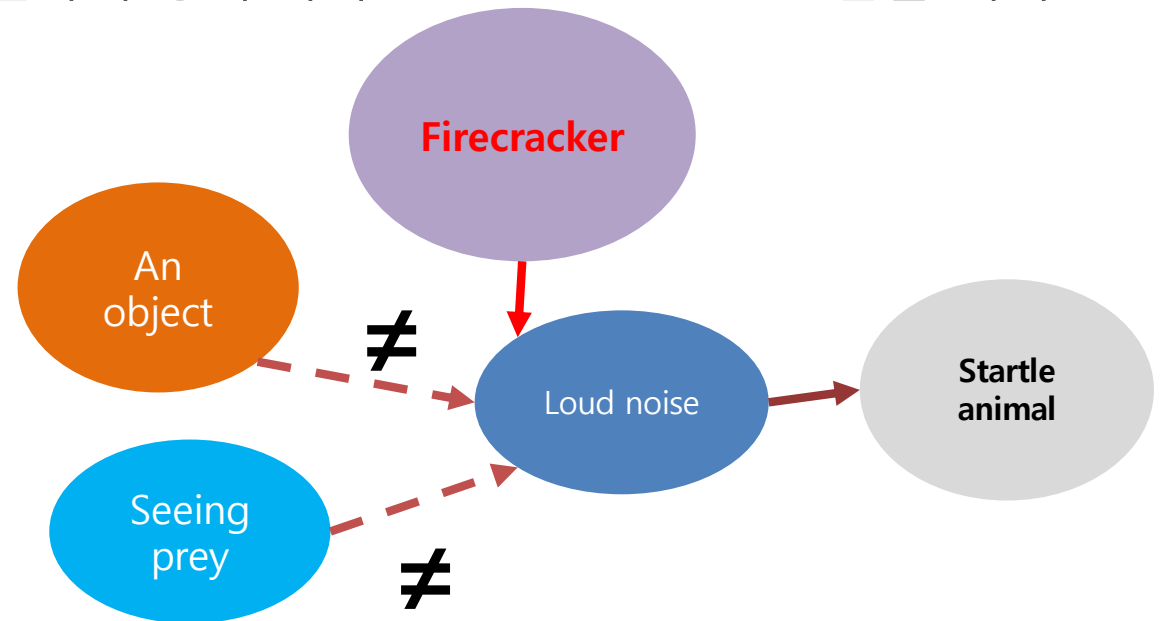
- ❖ Then, we need intermediate conditional propositions
  - P' = 【Firecrackers cause loud noise.】
  - Q' = 【An object cause loud noise.】
  - R' = 【Seeing prey cause loud noise.】

# Screening-Off

## ❖ CommonSenseQA 2.0

- 여러 가능한 명제 중 특별하게 명제 하나만 activation한 조건들을 제시함

Fact1  
Fact2  
-----  
∴ Combined Fact



- 특정 명제를 차단함
  - Q나 R이 발현되지 못하도록 함
- 사실 새로운 명제 W가 다른 모든 명제 발현을 차단하는 효과도 보임
  - W = 【Loud noise startle an animal.】

# Screening-Off

## ❖ 차폐효과

- '원인은 결과의 확률을 증가시킨다.'는 직관적 사실에 부합되지 않고 결과에 특정 영향을 미치는 조건들이 존재

## ❖ $P(E|A \cap C) = P(E|C)$

- $C$ 는  $A$ 를  $E$ 에서 차폐한다.
- $P(A \cap E|C) = P(A|C) \times P(E|C)$ 
  - $A$ 와  $E$ 는  $C$ 에 독립 조건 확률임

## ❖ 인과확률은 다음 조건을 만족해야 함

- 모든 인과적 요인들  $C$ 와  $E$ 는 시간  $t$ 의 선상에서 존재해야 함
  - $t < t'$
  - $P(E_{t'}|C_t) > P(E_{t'}|\sim C_t)$
  - 중간에 끼어드는 어떤 다른 시간  $t''$ 도 없어야 하며,  $C$ 로부터  $E$ 를 차폐하는 다른 사건도 없어야 한다.

# Screening-Off

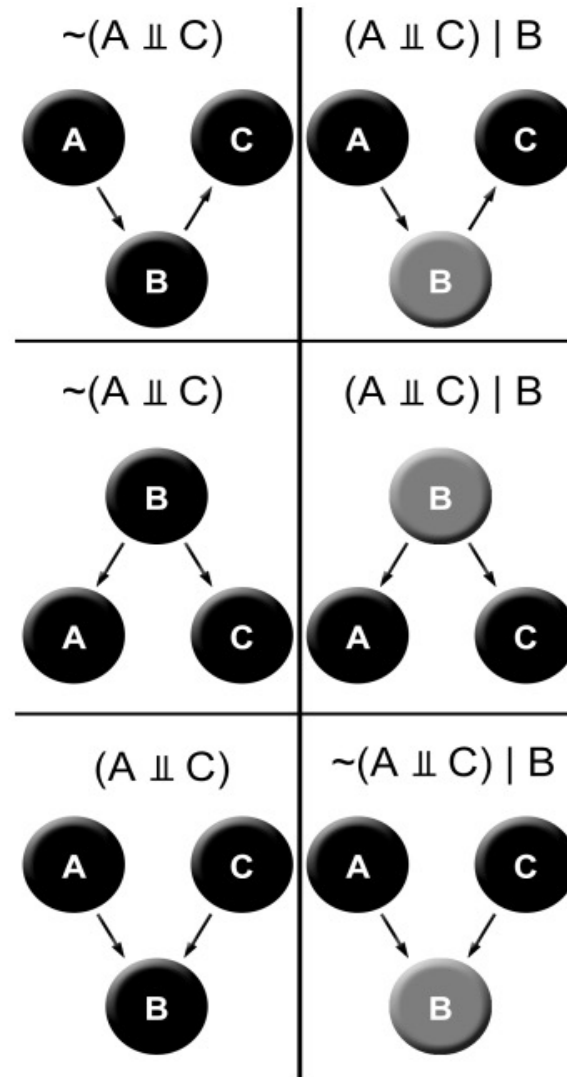
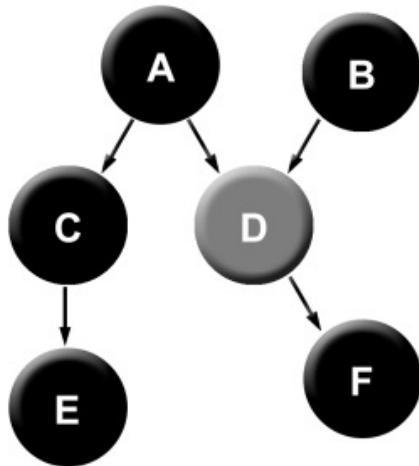
## ❖ 챗GPT

- 여러 가지 가능한 답안을 제시함
  - 정답만 아니라 오답도 제시
  - P, Q, R 모두 정답으로 간주
- Zero shot에서 더 많이 나타나고, few shot, full shot으로 가면서 덜 나타남
  - 그러나 일정하게 zero → few → full 의 순서로 나타나지는 않음
  - Full의 경우에는 Combined Fact에서 Fact1과 Fact2가 결합된 유형을 제시할 경우에 정답에 더 가까워 짐
- 질문의 성격과 복잡도를 완전히 이해하는 것은 아님
  - C = 【What does the digestive system break into nutrients for the body?】
  - A = 【Digestive system break X】
  - B = 【X is a nutrient for the body.】
  - C = 【 A 【 B 】 】
  - Full은 Combined Fact에 C의 복잡한 문형을 제시하고
    - Few는 제시하지 않음
    - Few의 경우 여러 개의 답안을 제시하고 있음

# Screening-Off

❖ 여러 가지 차폐유형

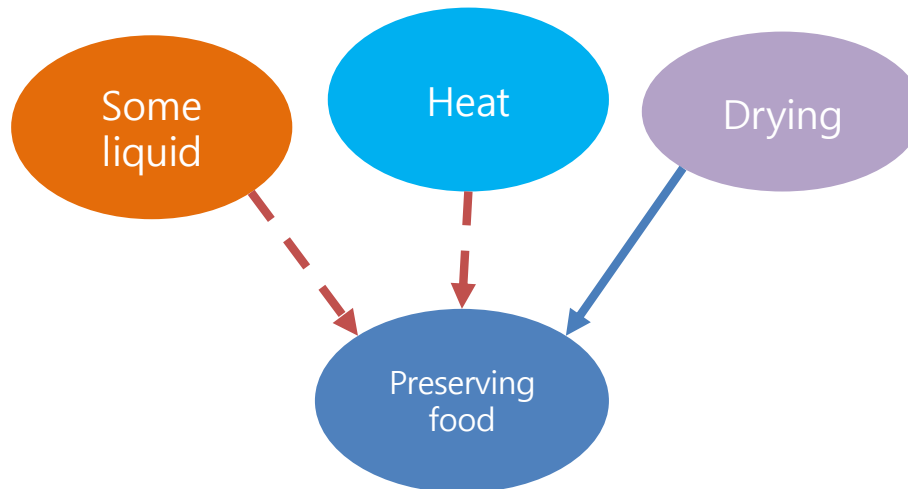
❖ 복잡한 유형들



# Screening-Off

## ❖ 다른 차폐형태

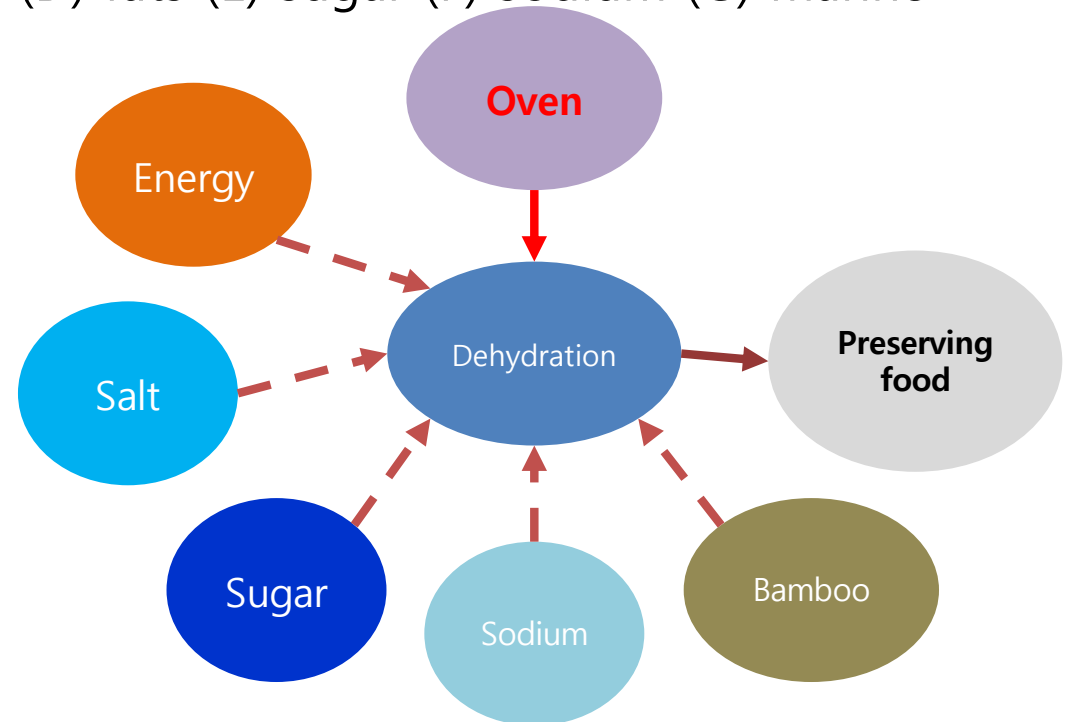
- 특정 조건에 대한 screening-off가 있으며, activation되는 확률만 발현되는 특징을 가짐
  - $\sim ( P \perp\!\!\!\perp Q ) \wedge R \mid S$
  - $P(P \cap Q|S) = P(P|S) \times P(Q|S) = P(P) \times P(Q)$
  - $P(Q \cap R|S) = P(Q|S) \times P(R|S) = P(Q) \times P(R|S)$
  - $P(P \cap R|S) = P(P|S) \times P(R|S) = P(P) \times P(R|S)$
  - $P(P, Q, R|S) = P(P) \times P(Q) \times P(R|S)$



# 문제점

❖ 특정 조건을 충분히 주었음에도 불구하고 잘못된 답변을 제시함

- Fact1: dehydrating food is used for preserving food
- Fact2: Food is then dehydrated in slow oven or dehydrator.
- Combined Fact: food is preserved in a slow oven
- What is used to preserve food?  
(A) salt (B) Energy (C) oven (D) fats (E) sugar (F) sodium (G) marine  
(H) bamboo



검색결과 약 101,000,000개 (0.40초)

### A natural preservative by design

When sugar is added to fresh foods, like fruits and vegetables, it creates an osmotic effect: **sugar absorbs water in the food**. Bacteria, yeast and moulds – the primary causes of most food spoilage – need water to grow and multiply. 2022. 10. 20.



ragus.co.uk  
<https://www.ragus.co.uk/sugar-as-a-natural-preserved>

### Sugar as a natural preservative | Ragus

britannica.com  
<https://www.britannica.com/topic/food-preservation>

### Food preservation | Definition, Importance, & Methods | Britannica

관련 검색: What is used to preserve food?

What water is used to preserve food?

Does sea water preserve food?

If you take time to concentrate the salt from seawater by evaporation, and dry it, that **salt can preserve meat and fish** as has been done for thousands of years in tropical coastal regions around the world!



### Google Patents

### Application of bamboo tube to food preservation

#### Abstract

The invention discloses application of a bamboo tube to food preservation. The application is characterized in that the inner wall of a bamboo tube container and the inner wall of an opening cover are machined to form black bamboo charcoal inner walls, the black bamboo charcoal inner walls are thoroughly washed, and foods are injected into the bamboo tube and preserved after the cover is closed. The hollow container effect of the bamboo tube and the bamboo charcoal bacteriostatic and mould-removing fresh-keeping effect are adopted, so that containing and fresh keeping are unified perfectly. According to actual measurement, the retention period of chilli sauces through the bamboo tube can be as long as four years, the taste is pure, the retention period of other foods through the bamboo tube can range from three years to four years, and the bamboo tube is suitable for the food processing and preservation industry.

#### Classifications

aptation technologies in agriculture, forestry, livestock or agroalimentary processing or handling, e.g. food conservation

fications

(G) Marine???  
Hallucination

Marine is used to preserve food



# 결론

- ❖ 인과적 추론은 일반적 추론과 다르게 구성된다.
  - 인과성은 상관성과 다르며, 명제적 추론이 아닌 명제와의 연관성에서 비롯된다.
  - 인과성은 특정한 메커니즘에 의해서 명제들 간에 차폐되는 경우가 있다.
- ❖ 챗GPT는 특정한 인과성을 이해하기 위해서 만들어진 것은 아니다.
  - 인과성에 의한 결론이 아닌 주어진 프롬프트에 의한 결론 도출이다.
  - 프롬프트에 특정한 인과성을 부여하는 것은 가능하다.
  - 그러나 완전하게 인과성 추론을 반영하는 것은 아니며 매우 직관적이거나 매우 일관적이지는 않다.
    - 대형 언어모델(LLM, Large Language Model)은 특정 데이터에 의해 학습된 알고리즘 자체를 표방하는 것으로
      - 어떤 데이터에 의해서 어떻게 학습되었고
      - 실제 무엇을 어떻게 이해하고 있는지가 중요하다.
- ❖ 무엇이 일관된 인과성을 표출하게 하는가?
  - 향후 과제

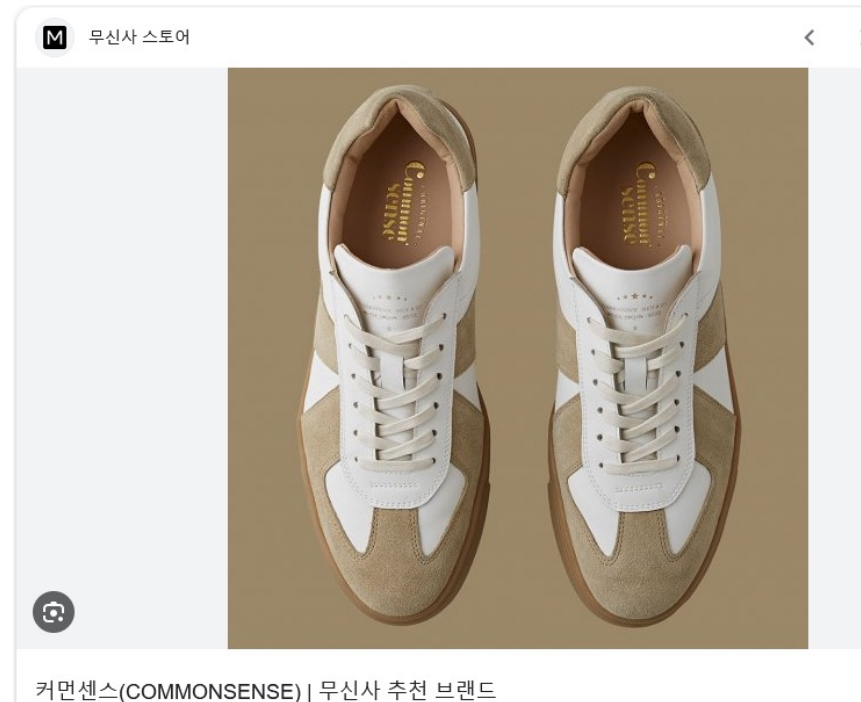
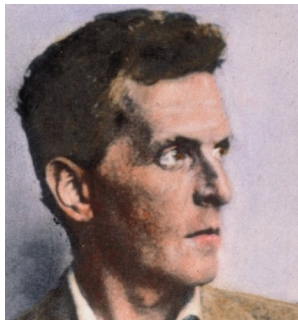
# 상식과 세계에 대한 지식

- ❖ 지능을 구현하면서 흔히 요구되는 것이 상식이다.
  - 따라서 상식은 인공지능을 구축하면서 필수적으로 요구된다.
- ❖ 올바른 자율 지능을 위해서는 세계에 대한 지식이 요구된다.
  - 그렇다면 상식만으로 세계에 대한 지식을 만들어 낼 수 있는가?
  - But commonsense is lack of universal truth, it is ambiguous, messy, beyond the realm of conventional logic and math (최예진 2023)
  - Yann LeCun (2022)
    - Common sense can be seen as a collection of models of the world....
- ❖ 삼족오는 다리가 몇 개인가?
  - 삼족오는 상식인가? 아니면 세계에 대한 지식인가?
  - 우리가 가진 세계에 삼족오는 존재하는가?
  - Do you know “삼족오”..... ?
  - Was ist eine “삼족오”.....?

# Wittgenstein

## ❖ 인식의 한계 ≙ 진리의 한계

- Wovon man nicht sprechen kann, darüber muss man schweigen.
- Whereof one cannot speak, thereof one must be silent.
  - 우리는 말할 수 없는 것에 대해서 침묵해야 하는가?



## ❖ Q & A

d 챗GPT는 무엇인가?

챗GPT는 "Chat Generative Pre-trained Transformer"의 약자로, 대화형으로 텍스트를 생성하는 인공지능 모델입니다. GPT-3.5 아키텍처를 기반으로 하고 있어서 다양한 주제에 대한 대화를 나눌 수 있어요. 무엇을 도와드릴까요?

d 왜 환영이 보이지?

환영, 환영! 그저 당신을 기다리고 있었을 뿐이에요. 어떤 주제로 대화를 나눌지 정해볼까요?

Is this conversation helpful so far?   



『금강경』에 자주 나오는 “그것은 그것이 아니기 때문에 그것이라고 부른다”라는 형태의 어법(語法)

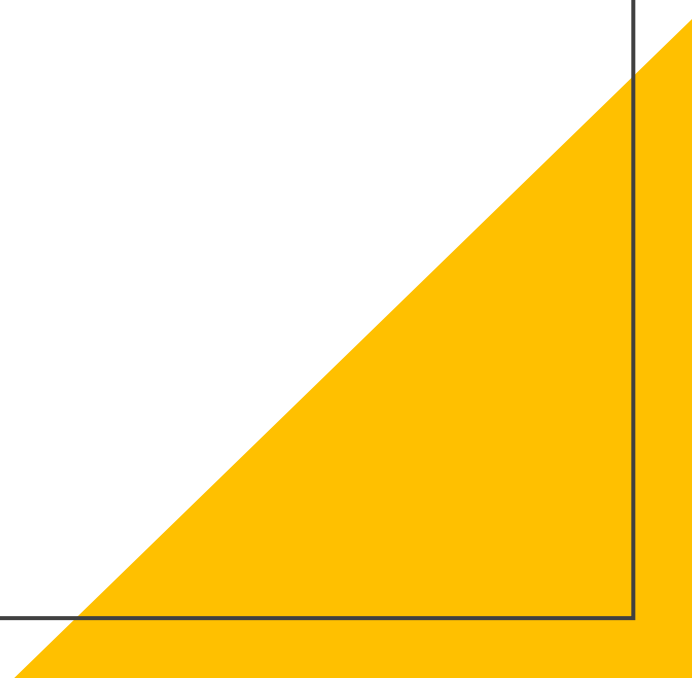
2023 한국언어정보학회 추계 학술대회

# Can Language Models Understand Felicity Conditions?

Kangsan Noh & Sanghoun Song  
(Korea University)



# Contents

1. Introduction
  2. Literature Review
  3. Data and Methods
  4. Findings
  5. Discussion
  6. Conclusion
- 
- A yellow triangular graphic is located in the bottom right corner of the slide, pointing towards the top right.

# Introduction

## **The Goal of the Present Study**

: Testing whether language models understand the non-literal and pragmatic aspects of natural language.

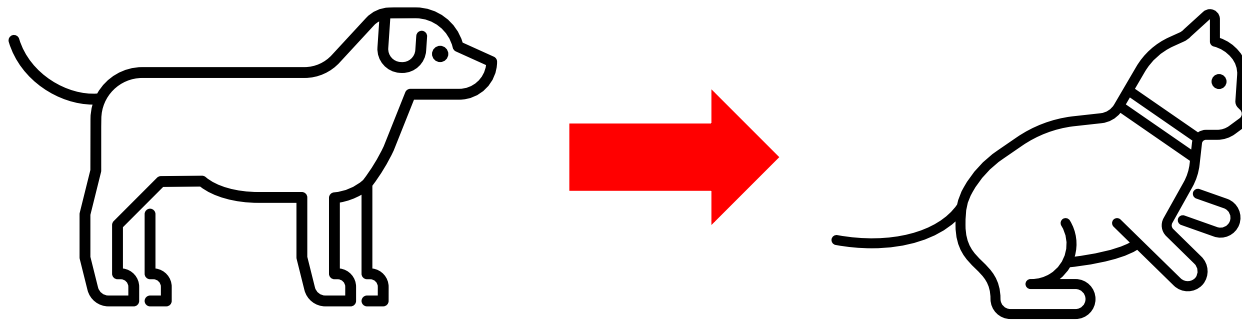
Specifically, whether language models process **felicity conditions** is tested.

To do so, we use **allosentences**, which share the same truth conditions.

What would be an appropriate answer to the question?

- (1) Q: What did the dog do?
- a. The dog chased the cat.
  - b. The cat was chased by the dog.

(Bender 2022: 102)





# Literature Review

## Felicity Conditions

An utterance is felicitous when it is pragmatically well formed with respect to a particular context.

- (1) Q: What did the dog do?
- a. The dog chased the cat.
  - b. The cat was chased by the dog.

(Bender 2022: 102)

# Literature Review

## **Allosentences**

We use allosentences, which share the same truth conditions (Lambrecht 1996).

- (2) a. Active Construction: Kim chases the dog.
- b. Passive Construction: The dog is chased by Kim.
- c. Fronting Construction: The dog Kim chases.
- d. it-Cleft Construction: It is the dog that Kim chases.

# Data and Methods

We used two types of language models:

1) **Bidirectional Encoder Representations from Transformers (BERT)**

→ For Natural Language Understanding

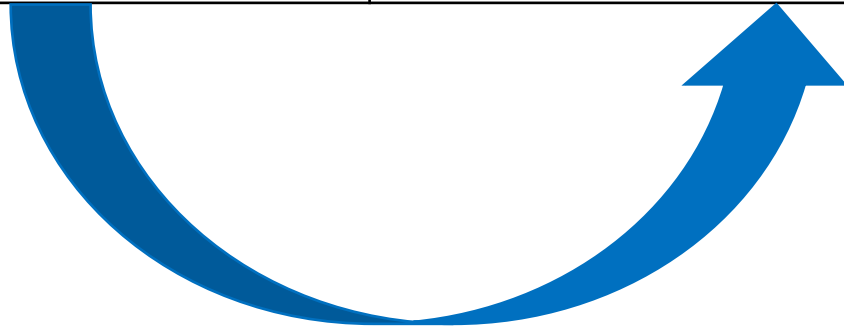
2) **Generative Pretrained Transformer (GPT)**

→ For Natural Language Generation

# Experiment I

Next sentence prediction (NSP) task for the BERT

Question	Answer	Felicity	Probability
What did John read?	John read the book.	Felicitous	0.999996662
	The book was read by John.	Infelicitous	0.999994755



# Experiment II

## Text generation task for the GPT-3.5 and GPT-4



What is the more appropriate answer between (A) and (B) to the following question?

Question: What did John make?

(A) John made the desk.

(B) The desk was made by John.

Just choose (A) or (B).



Both (A) and (B) are grammatically correct ways to convey the information, but (A) "John made the desk" is generally more direct and concise, so it would be the more appropriate answer in most contexts.

## Dataset

Since there are 200 different situations, and each situation consists of 6 sentences (2 questions and 4 answers), our dataset consists of 1,200 sentences in total.

Answer		Question		Subject Question	Object Question
				<i>Who made the desk?</i>	<i>What did John make?</i>
Voice	Active	<i>John made the desk.</i>		Felicitous	Felicitous
	Passive	<i>The desk was made by John.</i>		Infelicitous	Infelicitous
Cleft	Subject <i>it</i> -Cleft	<i>It was John that made the desk.</i>		Felicitous	Infelicitous
	Object <i>it</i> -Cleft	<i>It was the desk that John made.</i>		Infelicitous	Felicitous

# Findings

## Experiment I

We performed a paired  $t$ -test to determine whether the mean difference in probabilities is statistically significant.

Model	Setting		Results	
			$t$ -scores	$p$ -values
BERT-base-uncased	Subject Question	Voice	$t = 2.2812$	$p = 0.0236$ (*)
		Cleft	$t = 1.1756$	$p = 0.2412$
	Object Question	Voice	$t = 10.172$	$p < 0.001$ (***)
		Cleft	$t = -11.947$	$p < 0.001$ (***)
BERT-large-uncased	Subject Question	Voice	$t = 2.0833$	$p = 0.03805$ (*)
		Cleft	$t = 3.7354$	$p = 0.0002447$ (***)
	Object Question	Voice	$t = -0.94962$	$p = 0.3435$
		Cleft	$t = -2.7151$	$p = 0.007209$ (**)

## Experiment II

We performed a text generation task for the GPT models to determine whether they succeeded in producing appropriate answers to the given questions.

Model	Setting		Results			
			Felicitous	Infelicitous	Neutral	Error
GPT-3.5	Subject Question	Voice	66(33%)	62(31%)	72(36%)	0(0%)
		Cleft	68(34%)	132(66%)	0(0%)	0(0%)
	Object Question	Voice	177(88.5%)	16(8%)	7(3.5%)	0(0%)
		Cleft	156(78%)	43(21.5%)	1(0.5%)	0(0%)
GPT-4	Subject Question	Voice	163(81.5%)	4(2%)	33(16.5%)	0(0%)
		Cleft	191(95.5%)	5(2.5%)	1(0.5%)	0(0%)
	Object Question	Voice	197(98.5%)	1(0.5%)	2(1%)	0(0%)
		Cleft	188(94%)	11(5.5%)	1(0.5%)	3(1.5%)



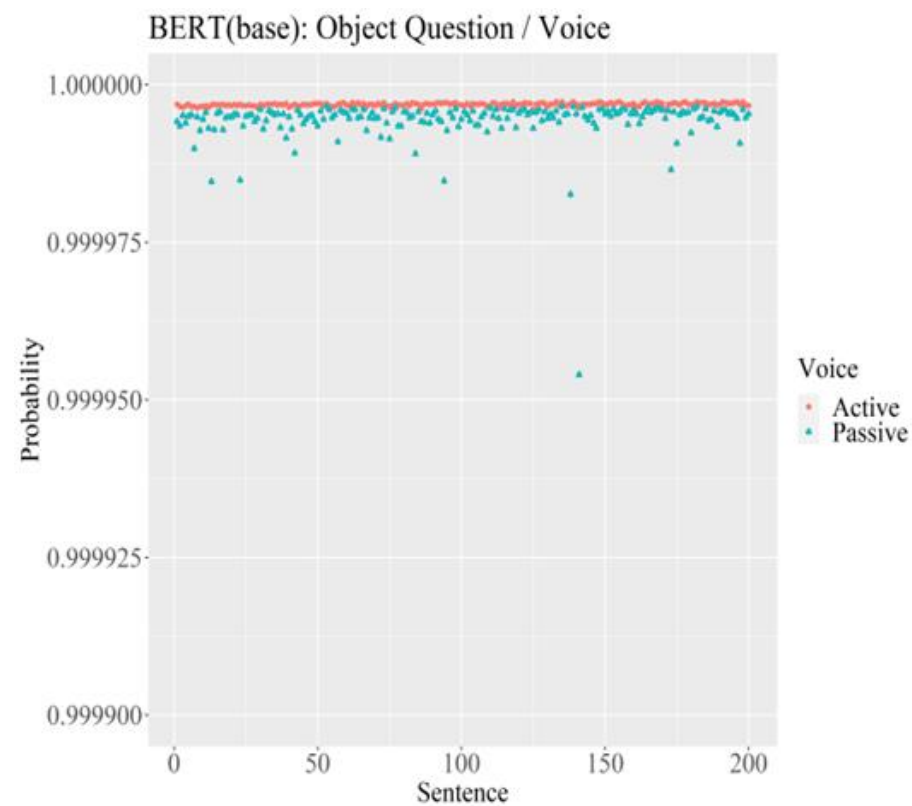
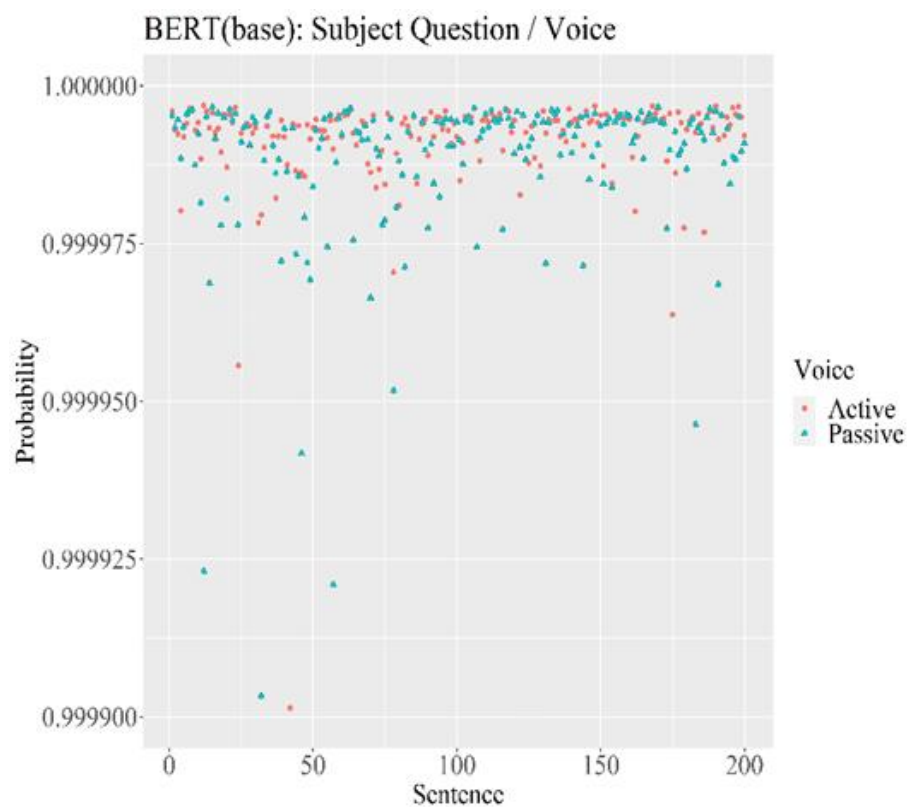
## Experiment II

### <Neutral Answers>

Question Prompt	<p>What is the more appropriate answer between (A) and (B) to the following question?</p> <p>Question: Who evaluated our performance?</p> <p>(A) John evaluated our performance. (B) Our performance was evaluated by John.</p> <p>Just choose (A) or (B).</p>
Answer by GPT-3.5	<p>Both (A) and (B) are grammatically correct and convey the same information. You can choose either one depending on your preference, as both provide a valid answer to the question. So, you can choose (A) or (B) based on your writing style or context.</p>

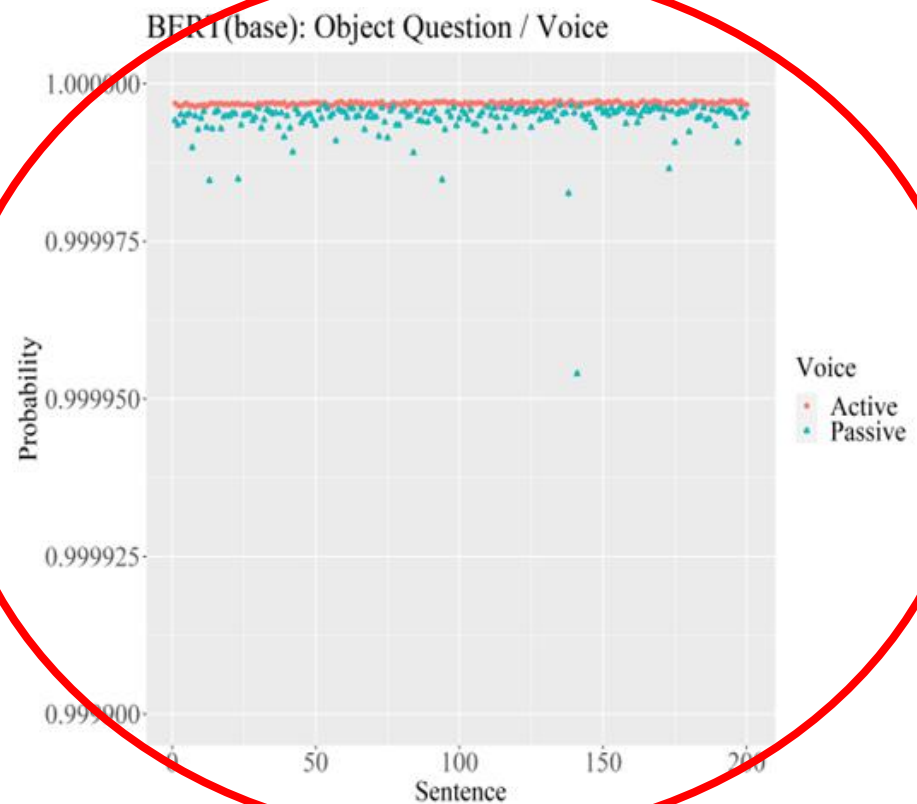
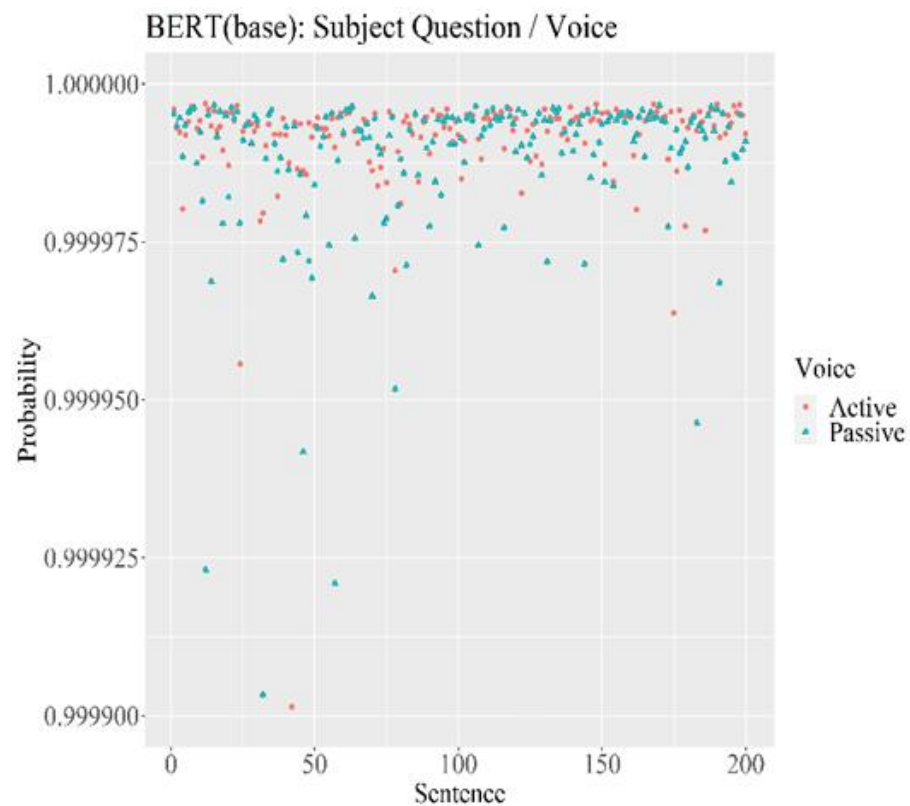
# Discussion

## Distribution of the Predicted Next Sentence Probabilities (Voice)



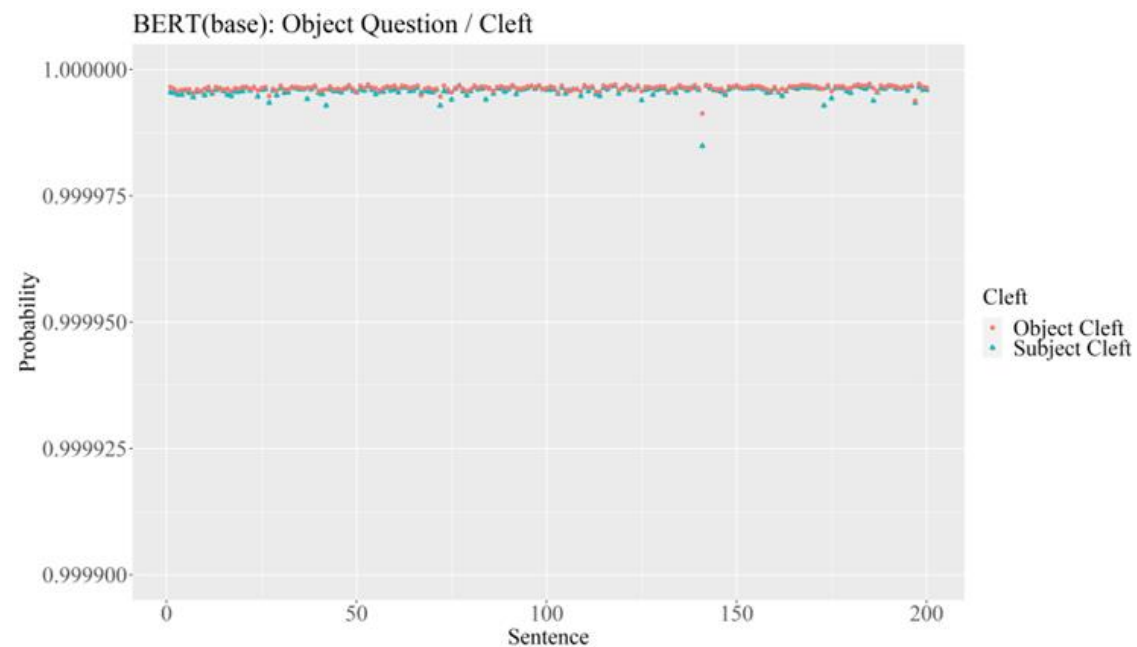
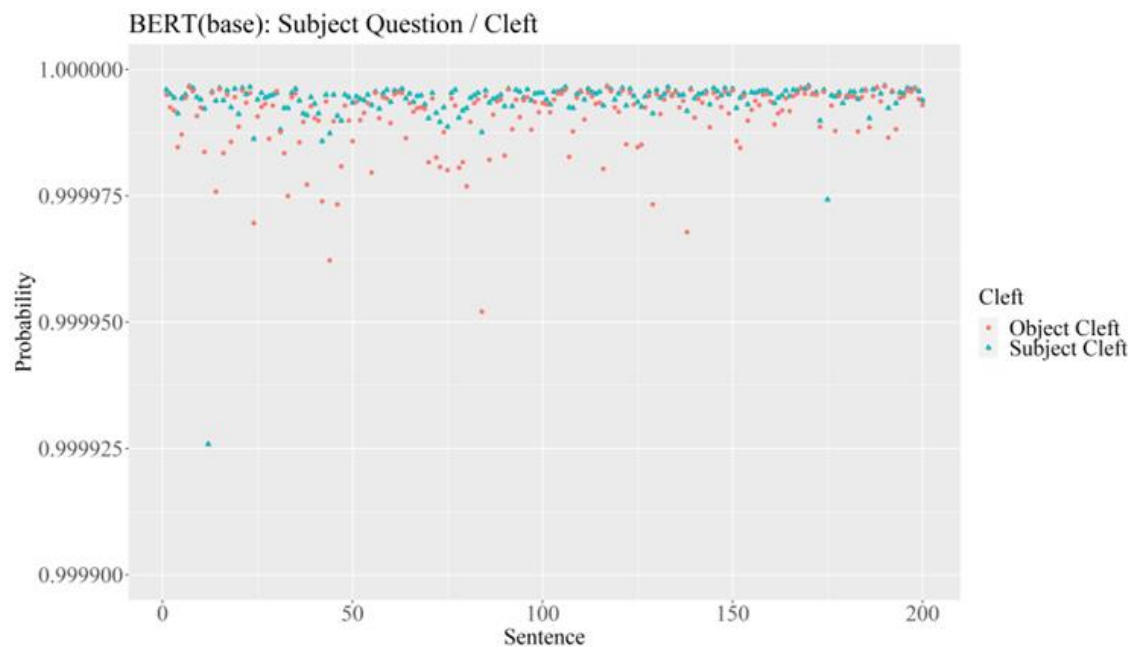
# Discussion

## Distribution of the Predicted Next Sentence Probabilities (Voice)



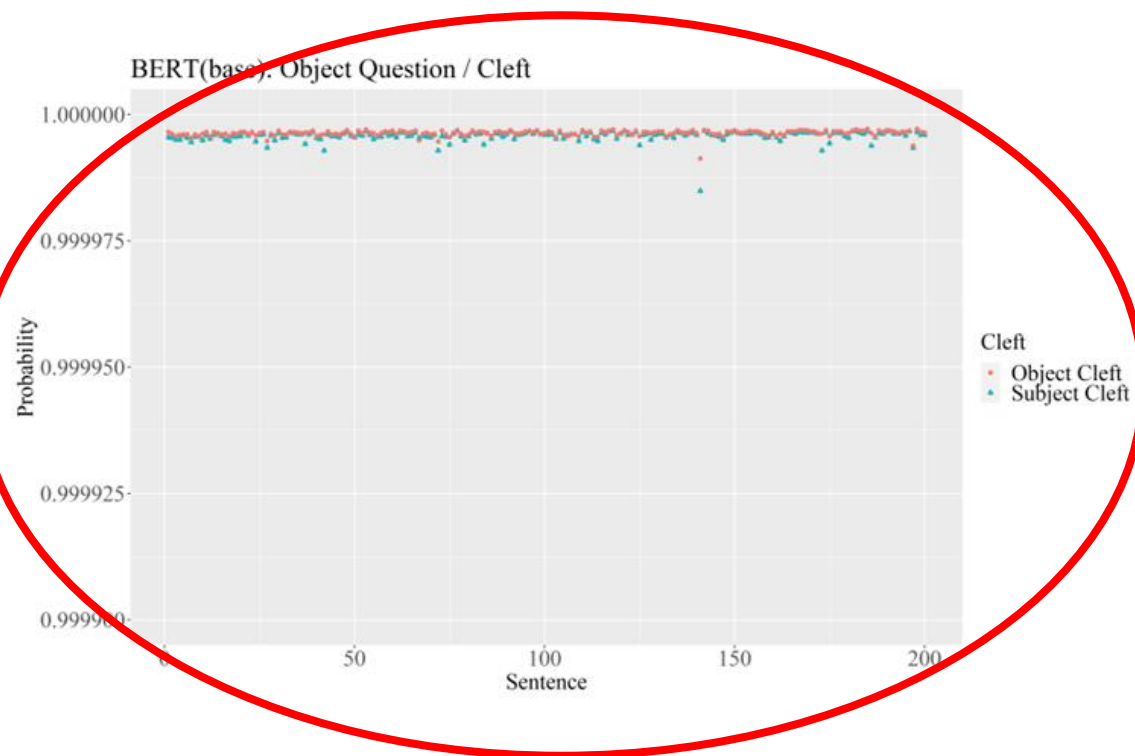
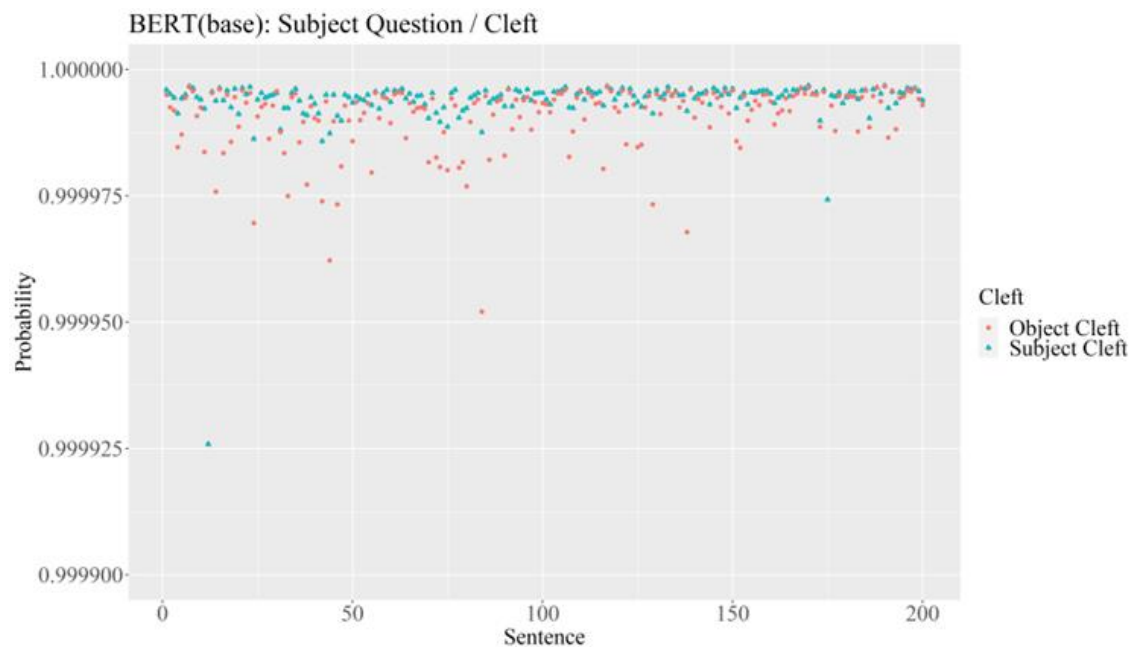
# Discussion

## Distribution of the Predicted Next Sentence Probabilities (Cleft)



# Discussion

## Distribution of the Predicted Next Sentence Probabilities (Cleft)



# Discussion

## Subject-Object Asymmetry

In both settings, the predicted next sentence probability was evenly distributed and conformed to the felicity conditions **when object questions were asked**.

We argue that these results may be attributed to subject-object asymmetry, which results from a preference for object extraction over subject extraction.

# Discussion

The existence of this preference has been supported by previous linguistic studies.

- (3) a. Who does Tom expect \_\_\_\_ to fire the manager? [Subject Question]  
b. Who does Tom expect to fire \_\_\_\_? [Object Question]

(Juffs and Harrington 1995: 496)

# Conclusion

Our study shows that although the accuracy of language models is far from perfect, they partially succeed in producing more felicitous answers than infelicitous ones in certain discourse contexts.

We hope that our research serves as an exemplary study of how current language models process the subtle pragmatic inferences necessary to engage in human-like communication.



# References

Bender, Emily. M. 2022. Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax. Springer Nature.

DOI: <https://doi.org/10.1007/978-3-031-02150-3>

Juffs, Alan and Michael Harrington. 1995. Parsing effects in second language sentence processing. *Studies in second language acquisition* 17:4, 483-516.

DOI: <https://doi.org/10.1017/S027226310001442X>

Lambrecht, Knud. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.

DOI: <https://doi.org/10.1017/CBO9780511620607>

**Thank you for listening**

2023년 한국언어정보학회 추계학술대회

---

# 코퍼스 기반 접미사 '-답다' 와 '-스럽다' 비교 연구

---

홍승혜 (고려대)

# 1. 선행연구 & 문제제기

## ❖ '-답-'과 '-스럽-'의 파생 양상

- (1)     ㄱ. 존은 어른답다.  
          ㄴ. 존은 어른스럽다.
- ㄷ. 내 동생은 소녀답다.  
          ㄹ. 내 동생은 소녀스럽다.
- ㅁ. 미국답다.  
          ㅂ. 미국스럽다.

# 1. 문제제기&선행연구

## ❖ '-답-'과 '-스럽-'의 파생 양상

(2) 헤이즈는 쇼케이스에서 “헤이즈다우면서도 헤이즈스럽지 않은 앨범을 생각했다”고 밝혔다. (김수영, 2019)

(3) ㄱ. “어떻게 해야 가장 나다우면서도 나답지 않은 모습을 보여줄까 고민했고…”  
(정진영, 2014)

ㄴ. “소주는 소주 다워야 하지만, 또 소주는 소주 답지 않아야 한다.” (최경민, 2022)

# 1. 문제제기&선행연구

## ❖ '-답-'과 '-스럽-'의 의미적 구분에 관한 선행연구

### 1) 선행어기의 의미적 자질 차이

	-답-	-스럽-
우효(1980) / 이경우(1981) /성광수(1988)	구체명사/구체성 보통명사	추상명사/추상성 추상명사
장영희(1986)	동작성	상태성

### 2) 주어와 선행어기에 대한 전제

- 신순자(1981)

'-답-'의 경우 주어(A)와 선행어기(B)의 관계는  $A \in B$ , '-스럽-'은  $A \cap B (A \neq B)$ 와 같음

# 1. 문제제기&선행연구

## ❖ '-답-'과 '-스럽-'의 의미적 구분에 관한 선행연구

### 3) 긍정성과 부정성

- 김정남(2005)

'-답-'은 긍정적 가치 의미를, '-스럽-'은 부정적 가치 의미를 내포하는 것으로 구분

예) '검사스럽다', '미국스럽다'

- 정연주(2007)

여기가 [실체성]의 자질을 가질 때 '-스럽-'의 결합에서 부정적 의미의 재현 가능성 높음

## 2. 코퍼스 및 방법론

### 1) 분석 코퍼스: 국립국어원 <모두의 말뭉치>

- **문어 말뭉치(ver 1.2)**
  - 총 10,045건의 원시 말뭉치
  - 책(상상), 책(정보), 잡지, 보고서 등 신문 외 장르로 구성
- **구어 말뭉치(ver 1.2)**
  - 총 25,696개의 원시 말뭉치
  - 공적 독백, 공적 대화를 한글로 전사한 자료와 준구어 자료(대본)로 구성



## 2. 코퍼스 및 방법론

### 2) 데이터 추출 방법

- **두 접미사의 모든 활용형 추출**
  - '-답다'의 활용형: 답다, 답게, 답지, 다운, 다웠, 다운면, 다울수록, 답도록
  - '-스럽다'의 활용형: 슨(다/게/지), 슨운, 슨웠, 슨우면, 슨, 슨울수록
- **공통 선행어기 목록 및 각각의 용례 추출**
  - 선행어기의 의미적 자질에 기반한 의미 구분 재고
  - 선행어기를 통제된 상태에서 두 접미사의 활용 양상 분석

## 2. 코퍼스 및 방법론

문어

구어

두 접미사의 공통 선행어기 목록

-답-의 용례

-스럽-의 용례

### [표제어 목록]

걱정스럽다, 경사스럽다, 고뇌스럽다(고려대한국어대사전), 고민스럽다, 고집스럽다, 남성스럽다, 맛스럽다, 비밀스럽다, 사랑스럽다, 시골스럽다, 악스럽다, 어른스럽다, 여성스럽다, 영결스럽다, 영광스럽다, 자연스럽다, 정답다, 짐스럽다, 참스럽다(참답다), 행복스럽다, 호걸스럽다

- 관습화 된 쓰임에서 내포적 의미 발생 가능성 고려
- 두 접미사의 의미 구분을 논의하는 데 한하여 표제어가 포함된 용례를 참조

### 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

#### 3.1. 공통 선행어기 목록

가을, 가족, 개성, 검사, 경쟁, 계집, 과학, 관념, 괴물, 괴짜, 국회, 권력, 귀부인, 귀신, 귀족, 나, 남아, 녀석, 노동자, 노래, 노인, 농촌, 누님, 눈, 뉴욕, 대학생, 동물, 동양, 라틴, 만화, 명품, 모험, 무기, 문제, 바보, 바흐, 변명, 변태, 병신, 보물, 부르주아, 부산, 뽕짝, 사람, 사물, 산문, 상사, 생명, 서민, 서울, 선비, 섬, 소녀, 소년, 소설, 속물, 스페인, 시련, 시인, 식자, 신사, 싸이, 아기, 아동, 아이, 아저씨, 아줌마, 안기부, 알레고리, 양채, 양반, 어린애, 어린이, 어머니, 아버지, 여아, 역사, 영웅, 완제품, 우리말, 우정, 유대인, 유머, 음유시인, 의사, 의원, 인간, 인사동, 일본, 작가, 장인, 전문가, 제주도, 조선, 중세, 진상, 진실, 짐승, 처녀, 천재, 청년, 초딩, 충신, 취미, 카프카, 특징, 파리지엔, 프랑스, 학교, 한국, 한국인, 현대카드, 호인

## 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

### 3.1. 공통 선행어기 목록

- '-답'의 선행어기는 '구체명사', '-스럽-'의 선행어기는 '추상명사'로 구분  
(우효, 1980; 이경우; 1981, 성광수, 1988)
- 어기가 [+사람]의 자질을 갖는 경우 두 접미사가 상보적 분포를 보임 (이경우, 1981)
- '-답-'의 선행어기는 [동작성]의 자질을, '-스럽-'은 [상태성]이 자질을 가짐 (장영희, 1986)

➡ '-답-'과 '-스럽-'의 공통 선행어기의 특성에 기반한 선행연구의 주장 재고

### 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

#### 3.2. '-답-'의 용례

- (4)   ㄱ. 양반은 양반답게 행동해야 한다는 Noblesse Oblige  
      ㄴ. 농부는 농부답고 학자는 학자답고 한국인은 한국인다운 사람
- (5)   ㄱ. 이성적으로 연주하는 것이 바흐다운 것이다.  
      ㄴ. 웃고있는 신부의 매부리코가 유대인답다.  
      ㄷ. 중학교 삼 학년 된 우리 작은 딸 사춘기답게 요즘 외모에 엄청 관심이 많아요.
- ➡ 주어가 선행어기임을 전제하는 경우와 그렇지 않은 경우 혼재

### 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

#### 3.2. '-답-'의 용례

- (6) ㄱ. 충신다운 충신을 곁에다 두질 않고  
ㄴ. 낮에는 또 봄답게 기온이 많이 올라갔는데요.

➡ 주어와 선행어기의 관계를 고려하기 어려운 경우

- (7) ㄱ. 양쪽 눈치를 부지런히 보는 그 사팔눈이 참, 일본답다.  
ㄴ. 혹자들은 유대인답게 개인주의적이고 계산적인 처신이라고 지적할지 모르지만..

➡ 부정적 의미를 내포하는 '-답-'의 파생 용례

### 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

#### 3.3. '-스럽-'의 용례

- (8)   ㄱ. 창희는 그런 애였다. 좀 아줌마스럽다고 해야 할까, 유들유들하다고 해야 할까  
      ㄴ. 나이와는 전혀 어울리지 않는 소녀스러운 표정이 신비하다 못해 백치미까지 느껴질 정도다
- (9)   ㄱ. 이를 통해 만화의 풍자성은 더욱 깊이 있어지고 만화는 더욱 만화스럽게 된다.  
      ㄴ. 카프카는 자신을 가장 서부 유대인스러운 인간이라고 말합니다.

➡ 주어와 선행어기가 일치하지 않는 경우, 일치하는 경우 혼재

### 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

#### 3.3. '-스럽-'의 용례

- (10) ㄱ. 근데 이게 몸을 쓰는 느낌이 너무 좋으시네. 딱 맞추는 게 미국스럽다고, 노래와 몸 이 몸 쓰는 거하고 이~ 그루브 감이 너무 좋은 거 같애.
- ㄴ. (프랑스 원작이 리메이크된 것을 두고) "네. 그 어떻게 미국스럽게 바뀌었는지 되게 궁금하네요."

➡ 긍정적 의미를 내포하는 '-스럽-'의 파생 용례



### 3. 접미사 '-답-'과 '-스럽-'의 용례 분석

#### 3.3. '-스럽-'의 용례

- (11) ㄱ. 필자 또한 이 '검사스럽다'라는 사전적 정의에 대해 전적으로 공감했다. 최근 유행하는 말로 필자에게도 이른바 검찰에 대한 '안 좋은 추억'이 있기 때문이다.
- ㄴ. 나는 검사들에게 법대로, 원칙대로 공소유지를 철저히 하되 무죄선고를 하지 말아달라고 부탁하는 등 아쉬운 소리를 하지 말라고 했다. 그건 검사스럽지 않다고..

➡ '검사답다'가 아닌 '검사스럽다'를 채택한 이유?

## 4. 새로운 관점에서의 의미 구분 제안

- 대상에 관한 개인의 누적된 지각적 판단(perceptual judgment) 경험으로부터 여기에 대한 감각적 이미지와 관념의 일반화가 이루어짐을 고려 (홍승혜&최재웅, 2022; Perice, 1931-58)
- 어기의 외연적(denotative) 의미보다 내포적(connotative) 의미에 가까움

선행어기에 대한 관념에서 벗어나지 않는다

-답-

화자 개인의 '주관적인 판단 경험'으로부터 일반화한 '서술 관념'과 관련

선행어기의 느낌이나 분위기를 자아낸다

-스럽-

느낌이나 분위기와 같은 감각적으로 지각되는 측면과 관련

## 4. 새로운 관점에서의 의미 구분 제안

### ❖ 새로운 의미 적용

- (12) ㄱ. 존은 어른답다. (어른에 대한 관념에서 벗어나지 않는다.)  
ㄴ. 존은 어른스럽다. (존은 어른의 느낌이나 이미지를 자아낸다.)

- (13) “헤이즈다우면서도 헤이즈스럽지 않은 앨범을 생각했다”  
(헤이즈에 대한 관념에서 벗어나지 않으면서도, 헤이즈의 느낌이나 이미지를 자아내지 않는 앨범을 생각했다.)

## 4. 새로운 관점에서의 의미 구분 제안

- (14) ㄱ. 이성적으로 연주하는 것이 바흐다운 것이다.  
(이성적으로 연주하는 것은 바흐에 대한 관념에서 벗어나지 않는다.)  
ㄴ. 낮에는 또 봄답게 기온이 많이 올라갔는데요.  
(봄에 대한 관념에서 벗어나지 않게 기온이 많이 올라갔다.)
- (15) ㄱ. 나이와는 전혀 어울리지 않는 소녀스러운(소녀의 분위기를 자아내는) 표정이 신비하다 못해 백치미까지 느껴질 정도다.  
ㄴ. 이를 통해 만화의 풍자성은 더욱 깊이 있어지고 만화는 더욱 만화스럽게 된다.  
(만화는 더욱 만화에 대한 느낌이나 분위기를 자아내게 된다.)

## 5. 결론

- **접미사 '-답-' 과 '-스럽'에 관한 선행 연구 재고**
  - 두 접미사의 쓰임은 선행어기의 의미적 자질 또는 주어와 선행어기의 관계에서 상보적 분포를 보이지 않음
  - 접미사 '-답-' 과 '-스럽'의 긍정적 또는 부정적 가치 의미 내포는 접미사 그 자체가 아닌 맥락에서 기인함
- **접미사 '-답-' 과 '-스럽'의 새로운 의미 구분 제안**
  - 접미사 '-답-'은 여기에 관한 관념적 측면을, '-스럽'은 감각적 측면을 상기한다는 점에서 그 쓰임이 구분됨을 설명
  - 두 접미사를 바꾸어 쓸 수 있는 경우, 연달아 함께 쓸 수 있는 경우 설명 가능

- 김수호. 1986. "접미사 '-답다, -롭다, -스럽다'의 어휘론적 기능연구". *문학과 언어* 7, 51-74.
- 김정남. 2005. "'-답다, '-롭다, '-스럽다'의 분포와 의미". *한국어 의미학* 18, 125.
- 송철의. 1977. "파생어 형성과 음운 현상". 서울대학교, 석사학위논문.
- 신순자. 1982. "국어 접미사 '스럽다'의 연구". 숙명여자대학교 석사학위논문.
- 우 효. 1980. "형용사 파생접미사 "-답다" "-스럽다"의 연구", 계명대학교 석사학위논문.
- 장영희. 1986. "현대국어 접미사 '-답-'의 연구". 숙명여자대학교 석사학위논문.
- 이경우. 1981. "파생어형성에 있어서의 의미변화". *국어교육* 39, 215.
- 성광수. 1988. "국어의 단어와 조어: 어휘구조와 어형성 규칙(1)" *주시경학보* 1, 88-110.
- 정연주. 2007. "형용사 형성 접미사의 통시적 변화 '-되-', '-롭-', '-스럽-', '-답-'을 중심으로-". 고려대학교 석사학위논문.
- 최호철. 1994. "현대 국어 가의소의 의미 분석 (1)- '답다, -되다, -롭다, -스럽다'를 중심으로". *한국어학* 1, 415-444.
- 홍승혜·최재웅. 2023. "부사어 '역시'와 동격 항진적 표현의 양태적 의미 관계". *언어학연구* 66, 149-173.
- C. S. Peirce. 1931-58. *Collected Papers of Charles S. Peirce*, 8 vols. Ed. Hartshorne, C. and Weiss, P.(vols. 1-6), and Burks, A.(vols 7-8), Cambridge, MA: Harvard University Press.
- 국립국어원. 2020. 국립국어원 문어 말뭉치(버전 1.2). <https://corpus.korea.go.kr>
- 국립국어원. 2021. 국립국어원 구어 말뭉치(버전 1.2). <https://corpus.korea.go.kr>
- 김수영. 2019. "박봄, 마약 논란 끝에 '제2의 전성기'...백예린·헤이즈·태연과 '음색퀸' 대결", *한국경제*. <https://www.hankyung.com/article/201903224985H>
- 정진영. 2014. "비 다우면서도 비 답지않은 음악 고민", *헤럴드경제*. <https://news.heraldcorp.com/view.php?ud=20140102000076>
- 최경민. 2022. "'박재범 소주'는 시작일뿐..."부어라 마셔라 싫다", *머니투데이*. <https://news.mt.co.kr/mtview.php?no=2022040718024266500>

# Information structure and voice mismatch in pseudogapping: An experimental perspective

Jungsoo Kim & Sang-Hee Park

jungsookim@inu.ac.kr; sangheepark@hanbat.ac.kr

Incheon National University; Hanbat National University

2023 KSLI Fall Conference  
(Hosted by Hanyang University, Seoul)  
November 18, 2023



## 1 Introduction

## 2 Previous Studies on Voice Mismatches in PG

- An Empirical Overview
- Can Information Structure Modulate Mismatch Effects?

## 3 An Acceptability Judgment Study

- Experiment 1: Information Structure and Voice Mismatch in PG
- Experiment 2: Information Structure and Voice Mismatch in VPE
- Experiment 3: Controlling for the Ellipsis Effect

## 4 General Discussion

## 5 Conclusion



# Introduction

# Pseudogapping

- Pseudogapping (PG): a type of ellipsis where a main verb is elided, leaving behind the auxiliary verb and a dependent of the main verb (Ross 1967; Levin 1980; Gengel 2007; Kim & Runner 2022)
- The missing material is positioned sentence-medially (cf. gapping) immediately following the auxiliary verb (cf. verb-phrase ellipsis).
  - (1) a. John called Sarah, and Mary will Jane. (Pseudogapping)
  - b. John called Sarah, and Mary Jane. (Gapping)
  - c. John called Sarah, but Mary didn't. (VPE)

# Ellipsis and voice mismatch

- The BIG question for ellipsis theories: Does the ellipsis site contain syntactic information?
- Diagnosis: mismatch effects, especially voice mismatch
- Some types of ellipsis show mismatch effects while others do not:
  - (2) a. \*Some bring roses and lilies by others. (Gapping)  
b. \*Someone murdered Joe, but we don't know who by.  
(Sluicing)
  - (3) a. The janitor must remove the trash whenever it is apparent that it should be. (VPE)  
b. Since regardless of which bit is initially assigned, it will be flipped if more information is gained by doing so.  
(VP-anaphora)
- **This research: Does PG show voice mismatch effects?**

# Previous research on PG

- Theoretical explorations:
  - Movement-cum-deletion (Lasnik 1995; Johnson 1996; Gengel 2007)
  - Sideward movement (Agbayani & Zoerner 2004)
  - Base-generation & deletion (Lee 2018)
  - Type-logical (Kubota & Levine 2017)
  - Construction-based (Kim & Runner 2022)
  - Purely interpretive (Miller 1990)
- Empirical research:
  - Data-driven (Levin 1980; Miller 2014)
  - Acceptability judgment (Hoeksema 2006; Kim & Park 2022)
- **No previous research investigated how speakers perceive voice mismatches in PG.**

- **Setting up empirical data:**

- Existing data are mostly researchers' own
- No consensus with regards on the acceptability of PG voice mismatch
- Cf. Acceptability judgment studies on voice mismatches in VPE (Arregui et al. 2006; Kertz 2013; Kim & Runner 2018)

- **Suggesting directions for theoretical research:**

- Voice mismatch as misalignment at the information-structural level (Kertz 2013)
- Syntactic recovery and memory (Recycling Hypothesis, by Arregui et al. 2006)

## Previous Studies on Voice Mismatches in PG

# PG voice mismatch as ungrammatical (Merchant 2008)

- According to Merchant (2008), voice mismatches are not permitted in PG.
  - (4) Passive antecedent, active ellipsis
    - a. \*Roses were brought by some, and others did lilies.
    - b. \*Klimt is admired by Abby more than anyone does Klee.
  - (5) Active antecedent, passive ellipsis
    - a. \*Some brought roses, and lilies were by others.
    - b. \*Abby admires Klimt more than he is by anyone else.
- Merchant's syntactic account: Mismatches in the specification of voice feature ( $v[\text{voi:act}]$  or  $v[\text{voi:pass}]$ ) result in ungrammaticality.

# Reassessing Merchant's data (Tanaka 2011)

- As Tanaka (2011) pointed out, the VPE counterparts of (4) and (5) are also bad:
  - (6) a. \*Roses were brought by some, and others did, too.  
b. \*Klimt is admired by Abby more than anyone does.
  - (7) a. \*Some brought roses, and lilies were, too.  
b. \*Abby admires Klimt more than he is.
- The logic here:** Since voice mismatches are accepted in VPE in principle, something other than voice is responsible for the anomaly in (6)-(7).



# PG voice mismatch as grammatical (Tanaka 2011)

- Tanaka's (2011) own examples, demonstrating that mismatches in PG are acceptable when they are acceptable in VPE:
  - (8) a. This problem was to have been looked into, but obviously nobody did.  
b. ?My problem will be looked into by Tom, but he won't yours.
  - (9) a. The system can be used by anyone who wants to.  
b. ?The new system can be used by anyone who could the older versions.
- No account has been provided as to why the pseudogapping versions are relatively worse than the VPE examples.
- No empirical support was provided that confirms such difference in acceptability is reliable.

# PG voice mismatch as grammatical

- Introspective data:

(10) a. ?That should be explained to individual students by the TA, but the professor will to the class in general.

(Coppock 2001)

b. ?MY problem will be investigated by Tom, but he won't YOURS. (Kubota & Levine 2017)

- Corpus data (Miller 2014):

(11) a. The arms were hidden by the rebels as a woman would (do) her most precious jewels.

b. A whole poached wild striped bass should be taken to the table as you would a Thanksgiving turkey or a crown roast pork, with a twinkle of extravagance.

c. These savory waffles are ideal for brunch, served with a salad as you would a quiche.

# Voice mismatch as misalignment in information structure (Kertz 2013)

- Kertz (2013): an information-structure-based account of voice mismatches in VPE
- Mismatch effects as violations of a well-formedness constraint on a contrastive topic relation: “A contrastive topic relation is well formed if members of the topic set are sentence topics.” (p. 204)
- Voice mismatches induce a penalty in acceptability if (would-be) contrastive topics are not realized as sentence topics.
- Such penalty is independent of ellipsis.
  - (12) a. #The problem was looked into by **the committee**, just like **the chair did**.
  - b. This problem was to have been looked into, but obviously nobody did.

# Voice mismatch as misalignment in information structure (Kertz 2013)

- Kertz's Experiment 3: An acceptability judgment experiment testing the effects of information structure, voice and ellipsis
  - (13) a. The technicians didn't install the line as quickly as the engineers did. (contrastive topic; match; ellipsis)
  - b. The line wasn't installed by the technicians as quickly as it could have been. (simple focus; match; ellipsis)
  - c. The line wasn't installed by the technicians as quickly as the engineers did. (contrastive topic; mismatch; ellipsis)
  - d. The technicians didn't install the line as quickly as it could have been. (simple focus; mismatch; ellipsis)
  - e. The line wasn't installed by the technicians as quickly as the engineers installed it. (contrastive topic; mismatch; no ellipsis)
  - f. The technicians didn't install the line as quickly as it could have been installed. (simple focus; mismatch; no ellipsis)

# Voice mismatch as misalignment in information structure (Kertz 2013)

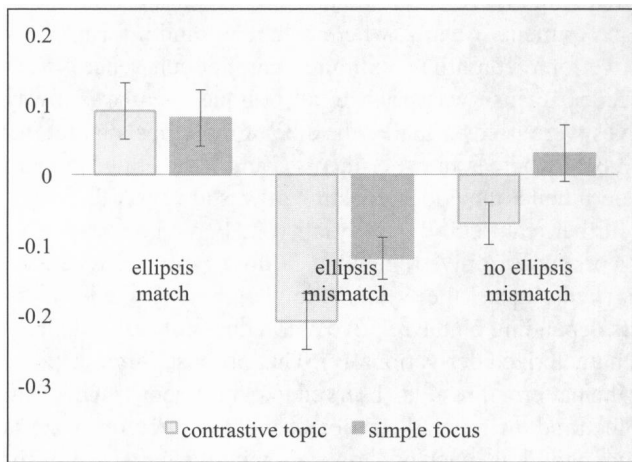


Figure 1: Results of Kertz's (2013) Experiment 3

# Key takeaways from Kertz (2013)

- 1 Voice mismatch negatively impacts acceptability ratings of both ellipsis and non-ellipsis sentences.
- 2 **The simple focus effect:** The negative influence of mismatch was smaller with simple focus than contrastive topics, consistently in both ellipsis and non-ellipsis sentences.
- 3 Mismatch effects were stronger when ellipsis is involved, suggesting that they cannot entirely be ascribed to information structure.
- 4 Given the generality of voice mismatch penalty and simple focus effect, similar results would be found for PG sentences.

# An Acceptability Judgment Study

- Experiment 1: Tested **the influence of two information-structural manipulations** (contrastive topic and auxiliary focus) on speakers' judgments of **voice-matched and voice-mismatched PG examples**
- Experiment 2: Had speakers judge **VPE sentences constructed from the PG stimuli from Experiment 1**, keeping the experimental design constant, to confirm whether the results from Experiment 1 were construction-specific or general
- Experiment 3: Further included **ellipsis as a predictor of the acceptability of PG and VPE** to examine how it interacts with information structure

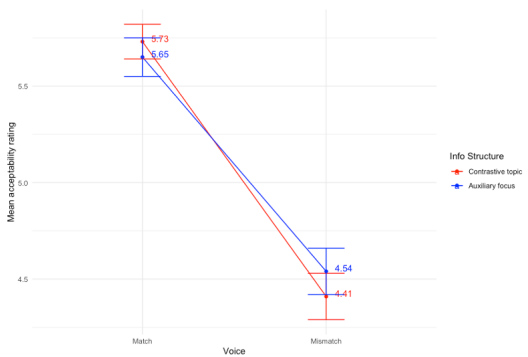


# Experiment 1

- Participants: 42 self-reported native speakers of English; Prolific (<https://www.prolific.co>); US residents aged between 18 and 60; with a minimum approval rate of 95% and a history of completing over 500 tasks on the platform; data from 2 participants removed from the dataset due to failing comprehension check questions or providing random answers; 1.6 GBP compensation
  - (14) Materials: 2 x 2 factorial design – VOICE (Match and Mismatch) x INFORMATION STRUCTURE (Contrastive topic and Auxiliary focus)
    - a. Adam didn't alert the engineers as quickly as Eric did the inspectors. (Match, ConTop)
    - b. The engineers weren't alerted by Adam as quickly as they could have been by Eric. (Match, AuxFoc)
    - c. The engineers weren't alerted by Adam as quickly as Eric did the inspectors. (Mismatch, ConTop)
    - d. Adam didn't alert the engineers as quickly as they could have been by Eric. (Mismatch, AuxFoc)
- 16 sets of test items; distributed to 4 experimental lists in a Latin-square design; 48 fillers with varying degrees of acceptability
- Procedure: PCIBex Farm; 1-7 scale acceptability; initial training session with 3 practice trials with guiding information; 7 more practice trials without guiding information; main experiment with experimental items in a uniquely generated random order; 10 comprehension check questions

# Experiment 1 results

- Mean acceptability ratings of the four conditions of PG sentences in Experiment 1 along with standard error bars



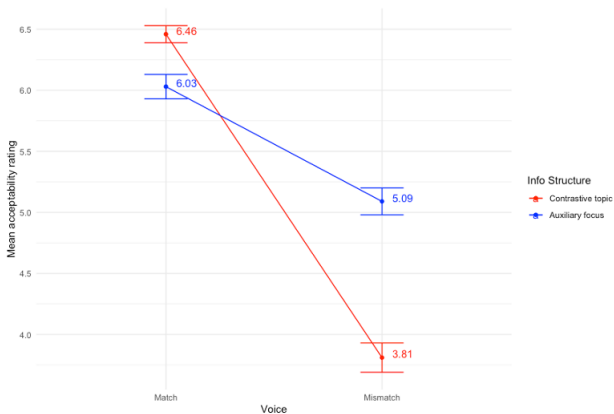
- Significant voice-mismatch effect ( $p < 0.0001$ )
- No information structure effect ( $p = 0.8495$ )
- No interaction effect between voice match/mismatch and information structure ( $p = 0.4186$ )
- Key observation:** The results here are **not in line with Kertz's information structure hypothesis!**

# Experiment 2

- The findings from Experiment 1 do not conclusively establish whether the lack of information structure effect is specific to PG.
- VPE examples manipulated consistently as the PG items in Experiment 1
  - (15) Materials: 2 × 2 factorial design – VOICE (Match and Mismatch) × INFORMATION STRUCTURE (Contrastive topic and Auxiliary focus)
    - a. Adam didn't alert the engineers as quickly as Eric did. (Match, ConTop)
    - b. The engineers weren't alerted by Adam as quickly as they could have been. (Match, AuxFoc)
    - c. The engineers weren't alerted by Adam as quickly as Eric did. (Mismatch, ConTop)
    - d. Adam didn't alert the engineers as quickly as they could have been. (Mismatch, AuxFoc)
- 16 sets of test items; distributed to 4 experimental lists in a Latin-square design; 48 fillers from Experiment 1
- Same procedure in Experiment 1
- Analysis based on data from 40 participants

# Experiment 2 results

- Mean acceptability ratings of the four conditions of VPE sentences in Experiment 2 along with standard error bars



- Significant voice-mismatch effect ( $p < 0.0001$ )
- Significant information structure effect ( $p < 0.05$ )
- Interaction effect between voice match/mismatch and information structure ( $p < 0.0001$ )

## Experiment 2 results (cont'd)

- Post-hoc pairwise comparisons between conditions in Experiment 2; Kenward-Roger approximation method for degrees of freedom and Tukey method for  $p$ -value adjustments

Contrast	Est.	SE	df	$t$ .ratio	$p$ .value
MatchConTop - MatchAuxFoc	0.431	0.197	64.2	2.192	0.1363
<b>MismatchConTop - MismatchAuxFoc</b>	<b>-1.275</b>	<b>0.197</b>	<b>64.2</b>	<b>-6.482</b>	<b>&lt; 0.0001</b>
<b>MatchConTop - MismatchConTop</b>	<b>2.650</b>	<b>0.197</b>	<b>64.2</b>	<b>13.472</b>	<b>&lt; 0.0001</b>
<b>MatchAuxFoc - MismatchAuxFoc</b>	<b>0.944</b>	<b>0.197</b>	<b>64.2</b>	<b>4.798</b>	<b>&lt; 0.0001</b>

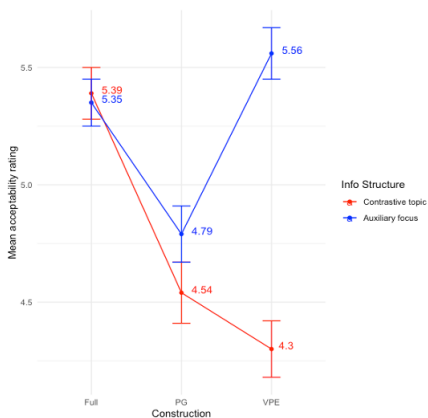
- Voice mismatch penalty when VPE sentences had the same information structure
- Significant information structure effect only in the voice-mismatched VPE conditions, but not in the voice-matched conditions
- Key observations:** The results here **replicated Kertz's findings**, indicating that **the PG test items used in Experiment 1 were not problematic** and that indeed **her information structure hypothesis was not extended to PG as it was found in VPE!**

# Experiment 3

- In Experiment 3, we tested one further aspect of Kertz's hypothesis: mismatch penalties due to ill-formed contrastive topics are specific to ellipsis
  - (16) Materials: 3 x 2 factorial design – CONSTRUCTION (Full, PG, and VPE) x INFORMATION STRUCTURE (Contrastive topic and Auxiliary focus)
    - a. The engineers weren't alerted by Adam as quickly as Eric alerted the inspectors. (Full, ConTop)
    - b. Adam didn't alert the engineers as quickly as they could have been alerted by Eric. (Full, AuxFoc)
    - c. The engineers weren't alerted by Adam as quickly as Eric did the inspectors. (PG, ConTop)
    - d. Adam didn't alert the engineers as quickly as they could have been by Eric. (PG, AuxFoc)
    - e. The engineers weren't alerted by Adam as quickly as Eric did. (VPE, ConTop)
    - f. Adam didn't alert the engineers as quickly as they could have been. (VPE, AuxFoc)
- 24 sets of test items (16 based on the prior experiments); distributed to 6 experimental lists in a Latin-square design; 48 fillers from the previous experiments
- Same procedure in the previous experiments
- Analysis based on data from 48 participants

# Experiment 3 results

- Mean acceptability ratings of the four conditions of VPE sentences in Experiment 3 along with standard error bars



- Significant construction effect ( $p < 0.0001$ )
- Significant information structure effect ( $p < 0.0001$ )
- Significant interaction effect between construction type and information structure ( $p < 0.0001$ )

## Experiment 3 results (cont'd)

- Post-hoc pairwise comparisons between conditions in Experiment 3; Kenward-Roger approximation method for degrees of freedom and Tukey method for  $p$ -value adjustments

Contrast	Est.	SE	df	t.ratio	p.value
FullConTop - FullAuxFoc	0.0312	0.155	143	0.201	1.0000
PGConTop - PGAuxFoc	-0.2448	0.155	143	-1.577	0.6148
<b>VPEConTop - VPEAuxFoc</b>	<b>-1.2604</b>	<b>0.155</b>	<b>143</b>	<b>-8.121</b>	<b>&lt; 0.0001</b>
<b>FullConTop - PGConTop</b>	<b>0.8438</b>	<b>0.155</b>	<b>143</b>	<b>5.437</b>	<b>&lt; 0.0001</b>
<b>FullConTop - VPEConTop</b>	<b>1.0885</b>	<b>0.155</b>	<b>143</b>	<b>7.014</b>	<b>&lt; 0.0001</b>
PGConTop - VPEConTop	0.2448	0.155	143	1.577	0.6148
<b>FullAuxFoc - PGAuxFoc</b>	<b>0.5677</b>	<b>0.155</b>	<b>143</b>	<b>3.658</b>	<b>0.0047</b>
FullAuxFoc - VPEAuxFoc	-0.2031	0.155	143	-1.309	0.7798
<b>PGAuxFoc - VPEAuxFoc</b>	<b>-0.7708</b>	<b>0.155</b>	<b>143</b>	<b>-4.967</b>	<b>&lt; 0.0001</b>

- Information structure effect only in the VPE condition, but not in the PG and full conditions
- With the contrastive topic structure, full > PG  $\approx$  VPE
- With the auxiliary focus structure, full  $\approx$  VPE > PG
- The results here replicated a major finding of Experiments 1 and 2; non-elliptical sentences are not sensitive to information structures
- Key observation:** These results **contradict Kertz's assumption** and instead suggest that **the influence of information structure on voice mismatch is not general but confined to certain structural types (i.e., VPE).**



# General Discussion

# Summary of findings

- We found that voice mismatch significantly lowers acceptability ratings of PG sentences.
- Unlike Kertz's prediction, acceptability of PG voice mismatch was **not** modulated by information structure.
- On the other hand, in VPE, mismatch penalties were significantly reduced by auxiliary focus.
- Overall, these findings disconfirm Kertz's view that voice mismatch effects result from violations of a general constraint on contrastive topics.

# Voice-mismatch penalties and ellipsis theory

- Some existing accounts of PG require only a semantic identity between the ellipsis site and its corresponding material in the antecedent (Miller 1990; Kubota & Levine 2017). They therefore do not predict reduction of acceptability ratings related to voice mismatch.
- Our findings raise questions about these accounts, but they do not necessarily support accounts that adopt a structural identity condition.
- It is possible that what have been observed as voice-mismatch penalties so far are in fact effects at the discourse level rather than strictly structural effects (Hardt & Romero 2004).
- Voice alteration in English accompanies structural changes, but it also has corresponding effects at discourses. The acceptability reductions seen in voice mismatch in pseudogapping may be linked to discourse effects and only indirectly connect to syntactic structure (cf. Kim & Runner 2018).

# Voice-mismatch penalties and ellipsis theory

- The question whether syntactic identity is involved in the interpretation of pseudogapping can be explored using the Recycling Hypothesis (RH: Arregui et al. 2006).
- According to RH, mismatches are more easily processed when the antecedent is passive compared to when it is active, as passive constructions are more complex and prone to misremembering.
- In our experiments, the contrastive topic stimuli had a passive antecedent and an active elided clause, whereas the auxiliary focus stimuli had an active antecedent and a passive elided clause. The degrading effect of the contrastive topic condition may have been mitigated by an uncontrolled factor, i.e., order of mismatches.

# Acceptability ratings and grammaticality

- Our findings clearly demonstrated a reliable negative impact of voice mismatch on pseudogapping.
- Nonetheless, mismatched sentences received higher ratings than anticipated: on average, they scored above 4.4 on a scale of 7 in both Experiments 1 and 3.
- This mean rating is comparable to some of our fillers which are generally considered as grammatical:
  - (17) a. Keith gave Emma a book and to Robin a record.  
(Non-constituent coordination; mean rating 4.21)
  - b. Susan studied math hard although not biology.  
(*Although*-stripping; mean rating 4.47)

# Acceptability ratings and grammaticality

- The relatively high acceptability ratings for voice mismatches in pseudogapping might suggest that participants in our experiments were able to achieve a reasonably clear interpretation.
- As is well-known, however, interpretability does not necessarily equate to grammaticality.
  - Comprehenders can often derive an interpretation for ill-formed sentences through covert repairs into well-formed ones (Frazier & Clifton 2011).
  - They can even learn to comprehend novel ungrammatical constructions (Kaschak & Glenberg 2004).
  - They sometimes report acceptability for a sentence but fail to assign a reliable meaning to it (Phillips et al. 2011; Wellwood et al. 2017).

# Acceptability ratings and grammaticality

- The grammatical status of voice mismatch in pseudogapping remains open to interpretation.
- Two possibilities: permitted mismatch with reduced acceptability (due to some extra-grammatical factors) or illegal mismatch perceived as plausible
- It is possible that the auxiliary focus condition in our experiments may have introduced an additional complexity due to post-elliptical contrastive material. This contrastive material occupies the end of the sentence, a default focus position that may influence the overall focus structure of the sentence, even in the presence of other markers of focus (Harris & Carlson 2018).

# Conclusion



# Conclusion

- The present study aimed to identify the pattern of acceptability in pseudogapping with voice mismatches through acceptability judgment experiments.
- We found that voice mismatch negatively impacts speakers' judgments on pseudogapping sentences.
- We tested whether information structure can modulate these negative impacts of voice mismatches but found no effect (cf. Kertz 2013).
- While mismatched sentences were dispreferred than matched ones, they were not entirely rejected by speakers. Voice-mismatched pseudogapping elicited ratings that are comparable to some grammatical constructions.

# References

- Agbayani, B. and E. Zoerner. 2004. Gapping, pseudogapping and sideward movement. *Studia Linguistica* 58(3), 185-211.
- Arregui, A., C. Clifton, L. Frazier, and K. Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Languages* 55, 232-246.
- Coppock, E. 2001. Gapping: In defense of deletion. In M. Andronis et al., eds., *Papers from the 37th Regional Meeting of the Chicago Linguistic Society*, 133-148. Chicago: Chicago Linguistic Society.
- Frazier, L. and C. Clifton. 2011. Quantifiers undone: Reversing predictable speech errors in comprehension. *Language* 87(1), 158-171.
- Gengel, K. 2007. Focus and ellipsis: A generative analysis of pseudogapping and other elliptical structures. Doctoral dissertation, Universität Stuttgart, Stuttgart, Germany.
- Hardt, D. and M. Romero. 2004. Ellipsis and the structure of discourse. *Journal of Semantics* 21(4), 375-414.
- Harris, J. A. and K. Carlson. 2018. Information structure preferences in focus-sensitive ellipsis: How defaults persist. *Language and Speech* 61(3), 480-512.
- Hoeksema, J. 2006. Pseudogapping: Its syntactic analysis and cumulative effects on its acceptability. *Research on Language and Computation* 4(4), 335-352.

## References (cont'd)

- Johnson, K. 1996. In search of the English middle field. Ms., University of Massachusetts, Amherst, MA.
- Lasnik, H. 1995. A note on pseudogapping. In R. Pensalfini and H. Ura, eds., *MIT Working Papers in Linguistics*, vol. 27, 143-163. Cambridge: MITWPL.
- Lee, J.-S. 2018. Phonetic resurrection from ellipsis sites: A case from pseudo-gapping. *The Linguistic Association of Korean Journal* 26(4), 55-81.
- Levin, N. 1980. Main verb ellipsis in spoken English. In A. Zwicky, ed., *Clitics and ellipsis*, 65-165. Columbus, OH: The Ohio State University.
- Kaschak, M. P. and A. M. Glenberg. 2004. This construction needs learned. *Journal of Psycholinguistic Research* 133(3), 450-467.
- Kertz, L. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89(3), 390-428.
- Kim, J.-B. and J. Runner. 2022. Pseudogapping in English: A direct interpretation approach. *The Linguistic Review* 39(3), 457-494.
- Kim, C. and J. Runner. 2011. Syntactic identity isn't enforced blindly: Comparing VP ellipsis and pseudogapping. *Poster presented at the 85th Annual Meeting of Linguistic Society of America*, January 6-9, Pittsburgh, Pennsylvania.
- Kim, C. and J. Runner. 2018. The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy* 41(1), 41-85.

# References (cont'd)

- Kim, J. and S.-H. Park. 2022. English pseudogapping: An experimental perspective. *Korean Journal of English Language and Linguistics* 22, 675-694.
- Kubota, Y. and R. Levine. 2017. Pseudogapping as pseudo-VP-ellipsis. *Linguistic Inquiry* 48(2), 213-257.
- Merchant, J. 2008. An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39(1), 169-179.
- Miller, P. 1990. Pseudogapping and do so substitution. In M. Ziolkowski et al., eds., *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, 293-305. Chicago: Chicago Linguistics Society.
- Miller, P. 2014. A corpus study of pseudogapping and its theoretical consequences. In C. Piñón, ed., *Empirical issues in syntax and semantics*, vol. 10, 73-90. Paris: CSSP.
- Phillips, C., M. Wagers, and E. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner, ed., *Experiments at the interfaces. Syntax and Semantics*, vol. 37, 153-186. Bingley, UK: Emerald.
- Ross, J. R. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT, Cambridge, MA.
- Tanaka, H. 2011. Voice mismatch and syntactic identity. *Linguistic Inquiry* 42(3), 470-490.
- Wellwood, A., R. Pancheva, V. Hacquard, and C. Phillips. 2017. *The anatomy of comparative illusion*. Evanston, IL: Northwestern University.

# Biscuit Conditionals and Probability Theories

Mean-Young Song (Dongguk University)

## 1. Introduction

This paper explores a semantic analysis of biscuit conditionals (henceforth BC) in terms of a probability theory, as exemplified in (1).

(1) There are biscuits on the sideboard, if you want them. (Biscuit Conditional)

BCs are not properly treated in terms of the semantics of canonical conditionals. Compare (1) with a canonical conditional like (2):

(2) Water vaporizes, if it boils.

In a conditional sentence like (2), the truth values of the protasis are dependent on those of the apodosis, but this is not the case with a BC like (1). Unlike the antecedent clause in (2), the antecedent in (1) functions to express a condition for the speech act performed in uttering the consequent clause. A conditional sentence like (1) is dubbed a *biscuit conditional* which is named after Austin's example in (1).

## 2. Some Properties of Biscuit Conditionals

- (3) a. If you are hungry, there are biscuits on the sideboard. (BC)  
b. If you need anything else later, my name is Oscar Wilde. (BC)  
c. There are biscuits on the refreshment table at the conference, if they have served them. (Canonical)

- (4) a. The truth values of the consequents are not dependent upon the antecedents in biscuit conditionals like (3a-b). (DeRose et al. (1999) Predelli (2009), Siegel (2006), and Goebel (2017) among others).  
b. The truth values of the consequents are dependent upon the antecedents in canonical conditional like (3c).

Two pieces of evidence for (4a):

Evidence 1: The apodosis can be denied regardless of the truth values of the protasis, as in (6).

- (5) A: There are biscuits in the cupboard, if you want some.  
B: That is not true!

In (5B), B denies the content of the proposition that is conveyed by the apodosis, rather than that of the proposition expressed by the protasis – i.e. B denies that there are biscuits in the cupboard, instead of denying that B wants some biscuits.

Evidence 2: Modus ponens does not work for BCs Modus ponens indicates that the truth values of the protasis is closely related to the apodosis.

- (6) a. If P, then Q.
- b. P.
- c. Therefore Q.

- (7) a. If you want some biscuits, there are biscuits in the cupboard.
- b. You want some biscuits.
- c. # Therefore there are biscuits in the cupboard.

Given this, biscuit conditionals are not subject to the truth conditions for canonical conditionals which is stated as (6):

- (8) The conditional  $p \rightarrow q$  is true in a world  $w$  iff  $q$  is true in every  $p$ -world.

### 3. Previous Studies

Two main analyses of biscuit conditionals: the Performative Theory and the Assertion Theory

#### (9) Performative Theory

The performative hypothesis argues that BCs are derived from deep structures which contain performative verbs like *say*, *ask*, and *order* etc. in the apodosis. (Comrie (1986), Iatridou (1991), Lakoff (1972), Ross (1970), Sadock (1974), and Sweetser (1900))

#### (10) Example of the Performative Theory

If you are hungry, I say to you there's pizza in the fridge.

#### (11) Assertion Theory

The assertion hypothesis claims that BCs are involved in *assertions* of the apodosis according to which the operator ASSERT applies to the apodosis, hence conveying the illocutionary force ASSERTION for the BCs. (Bach (1999), Bach and Harnish (1979), Geis (1995), Horn (1989), and Siegel (2006))

#### (12) Example of the Assertion Theory

If you are hungry, ASSERT(there's pizza in the fridge).

### 3.1 Problems with the Performative Theory

Krifka (2004) and Siegel (2006) note that if a sentence is compatible with an expression like *hereby*, it is a performative sentence. However, the following BCs are not compatible with *hereby*:

- (13) a. \*If customers ask your name just in case they need anything else later, your name is hereby Oscar Wilde.
- b. \*If customers ask your name just in case they need anything else later, I hereby say to you that your name is Oscar Wilde.

Sentences like (13a-b) suggest that the performative hypothesis is not on the right track.

### 3.2 Problems with the Assertion Theory

The question arises from the assertion theory as to how it is possible for the ASSERT operator to affect the interpretation of the antecedent clause in BCs, even when it does not fall in the scope of the operator.

- (14) As the term *speech act* suggests, an action is performed via the utterances.

Given (14), the Assertion theory would have to make the following predict about (12), contrary to fact:

- (15) (12) is interpreted to mean that the speaker is certainly performing the assertion act at the time of the utterance of (12), but not at a time at which the listener gets hungry, since the apodosis, not the protasis, falls in the scope of the operator.

#### 3.2.2 Siegel (2006)

To remedy the problem with the Assertion theory in (15), Siegel (2006) proposes a revised version of the Assertion theory:

- (16) a. BCs are interpreted to convey not a literal act but rather what she calls a potential literal act of assertion that the utterance by S may possibly bring about when a certain condition is met (see also Bach and Harnish (1979)).
- b. The potential literal act in Siegel's sense counts as a presupposed relevant or salient act of assertion in a given context.
- c. BCs are interpreted as an existential quantification over potential literal acts of relevant assertion via existential closure (Heim (1982)).

- (17) Siegel's Analysis of a BC like (12):

If you're hungry,  $\exists a \exists p [a = \text{relevant/salient assertion of } p \text{ and } p = \text{^}[there is pizza in the fridge]]$ , where  $a$  is a variable over potential literal acts of assertion and command etc., and  $p$  is a variable over propositions.

As we can see in (17), the introduction of an existential quantification over a relevant act of assertion into the semantics of BCs has the effect of including the protasis in the scope of the ASSERT operator, which can avoid the problem with the assertion hypothesis addressed above. As a result, the representation in (17) indicates that the assertion of the protasis will be a relevant assertion only if the apodosis is true (i.e. the addressee gets hungry).

### Limitations of Siegel (2006)

Recall that the analysis of BCs proposed by Siegel (2006) is marked by the existence of a contextually presupposed relevant or salient act of assertion of the proposition expressed by apodosis.

Siegel claims that the following dialogue demonstrates that presupposition is indeed involved in the interpretation of BCs:

- (18) A: If you are hungry, there is pizza in the fridge.  
B: That's not true! Rebecca ate it all this morning. (Siegel (2006: 179))

(19) Situation

Suppose that A utters a BC sentence like (18A) to B, who is hungry now and B denies A's utterance by saying (18B).

- (20) Given (19), the protasis of the BC sentence in (18A) is true, while the apodosis of the BC is false. Therefore, the whole conditional comes out false.

- (21) However, Siegel claims, however, that the BC in (18A) is not false but rather weird, due to the presupposition failure – i.e. the presupposition of relevant assertion is not satisfied owing to B's denial. This is because the relevance of assertion in the apodosis is not asserted but presupposed.

(22) Siegel's conclusion:

Due to the presupposition failure, the use of the BC in (18A) is not false but rather weird or inappropriate in the situation (19).

However, one should note that Siegel's conclusion is problematic since presupposition is inherited in the negative context.

- (23) a. It stopped raining.  $\approx_{\text{Presupposition}}$  It had rained before.  
b. It didn't stop raining.  $\approx_{\text{Presupposition}}$  It had rained before.

The presupposition in (23a) still survives in (23b).

Let's go back to (18):

- (24) Given (23), B's utterance in (18B) indicates that B denies the content of the proposition



expressed by the consequent in the BC, not the presupposition itself.

#### 4. BC Tests

(25) A conditional is a BC iff it satisfies the following conditions:

- a. The apodosis is independent of the protasis.
- b. The speaker of a BC should be talking to the second person – i.e. the second person should appear either in the protasis or in the apodosis.
- c. The present tense occurs in the apodosis.

- (26) a. If you're hungry, there is pizza in the fridge.  
b. There are biscuits on the cupboard, if you want some.  
c. If they ask you how old you are, you are four.  
d. If you need anything else later, my name is James.

The sentences in (26a-d) are BCs, since they satisfy the three conditions for BC in (38). However, the following conditionals are not BCs.

- (27) a. There are biscuits on the sideboard if no one has eaten them. ((38a-b) not met)  
b. If John is hungry, there is pizza in the fridge. ((38b) not met)  
c. If you were hungry, there **would** be pizza in the fridge. ((38c) not met)

#### 5. Preliminary

There are three properties of BCs:

- A. BCs are involved in a variety of speech acts (or illocutionary forces) including assertion, request, command, suggestion etc., depending on the context of use.

(28) A: Are you hungry?

B: No.

A: If you are hungry later, there is pizza in the fridge. (Assertion)

(29) A: I am starving now!

B: If you are hungry, there is pizza in the fridge. (Suggestion)

(30) Situation: Children under 5 get a free admission to the amusement park.

A: (to his six-year-old son) If they ask how old you are, you're four.

(Directives: command his son to tell a lie)

(31) A: (Looking at B making a desk) Is there anything I can help you with?

B: O.K. there is a power drill in the toolbox, if you want to help me. (Request)

## B. Independence

The meaning of apodosis is independent of that of the protasis in BCs, as illustrated in the above examples.

C. The speech act of BCs is performed at the same time as the utterance time.

## 6. Dependency and Independency

As was mentioned earlier, BCs are characterized by the fact that the truth of the apodosis is independent of that of the protasis.

### (32) Scenario #1

John always wears his hat in case of bad weather. The first thing he always does when he wakes up in the morning is to flip a coin before he looks out the window to see what the weather is like. When heads come up, he wears his hat even in case of fine weather, but he does not in this case, when tails come up. Above all, bad weather necessarily makes him wear his hat. Suppose that heads came up when he flipped the coin this morning, and further that the weather is bad – i.e., it is raining. So, John is wearing his hat and going to work.

There are two events described in this scenario that make John decide whether he is going to wear his hat; the flipping of a coin and the bad weather.

(33) In Scenario #1, the event of his flipping a coin does not affect the act of his wearing his hat. They are independent facts.

As was seen in Scenario #1, dependency is closely related to law-like principles, normality, conventional rules, and habitual behavior and so on, whereas independency is not. It affects the truth values, while independency does not.

## 7. Probability Space

Probability theories are known to provide a useful tool to define the independent relation between two events. I will incorporate them into the definition of the independent relation.

(34) A Probabilistic Intensional Logic (IL) Model  $\mathcal{M}$  is a 5-tuple  $\langle A, W, R, \mathcal{P}, F \rangle$ .

A: a non-empty set of individuals

W: a set of possible worlds

R: the assignment of an accessibility relation to each individual in A

$\mathcal{P}$ : a probability function which is a partial function taking a proposition  $\phi$  to yield a numeric value between 0 and 1, as in  $0 < \mathcal{P}(\phi) < 1$

F: an interpretation function which applies to all non-logical constants of IL.

(35) Probability Space is a triple  $\langle W, \mathcal{F}, \mathcal{P} \rangle$ .

W: a set of possible worlds which is taken to be a sample space denoting the set of all possible outcomes

$\mathcal{F}$ : a  $\sigma$ -algebra of a set of propositions (or events) in W

$\mathcal{P}$ : a probability measure on  $\mathcal{F}$

(36) Definition of a  $\sigma$ -algebra

Let W be a set of possible worlds and  $\mathcal{F}$  a subset of  $\text{Pow}(W)$ .  $\mathcal{F}$  is a  $\sigma$ -algebra of subsets of W iff it meets the following conditions:

(a).  $\emptyset, W \in \mathcal{F}$ ,

(b) if  $B \in \mathcal{F}$ , then  $B^C$  (i.e.  $W - B$ )  $\in \mathcal{F}$ ,

(c) if  $B_i \in \mathcal{F}$ , where  $i \geq 1$ , then  $\bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$ .

Given (35) and (36), we can define a probability measure as follows:

(37) Definition of a probability measure

Let W be a non-empty set of possible worlds,  $\mathcal{F}$  a  $\sigma$ -algebra of W, and  $\mathcal{P}$  a probability function such that  $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$ . Then  $\mathcal{P}$  is a probability measure on a measurable space  $\langle W, \mathcal{F} \rangle$ , iff it satisfies the following conditions:

(a) for all  $B \in \mathcal{F}$ ,  $\mathcal{P}(B) \geq 0$ ,

(b)  $\mathcal{P}(W) = 1$ ,

(c) for every countable disjoint union of  $\{B_i\}$  of  $\mathcal{F}$  such that  $\bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$ ,

$$\mathcal{P}(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mathcal{P}(B_i).$$

Based on the  $\sigma$ -algebra and the probability measure, we will define probability spaces in such a way that the probabilistic IL model is incorporated.

(38) Probability Space for the probabilistic IL model  $\mathcal{M}$

Let A be a set of individuals, W a set of possible worlds, R a binary relation over  $W \times A$ , Then the triple  $\langle W, \mathcal{F}, \mathcal{P} \rangle$  is a probability space for the probabilistic IL model  $\mathcal{M}$ , iff it satisfies the following conditions:

(a) for any  $w, w' \in W$ ,  $a \in A$ , and  $\phi \in \mathcal{F}$ ,  $R_a(w)(w') = \{\langle w', a' \rangle : \forall \phi [ \langle w, a \rangle \in \phi \rightarrow \langle w', a' \rangle \in \phi ]\}$ ,<sup>1</sup>

(b)  $\mathcal{F} : W \times A \rightarrow \text{Pow}(R_a(w)(w'))$  is a function such that it assigns to each  $a$  and  $w$  a set of subsets of  $R_a(w)(w')$  such that

(i)  $\mathcal{F}(\langle w, a \rangle)$  is a  $\sigma$ -algebra on  $R_a(w)(w')$  such that  $\phi \in \mathcal{F}(\langle w, a \rangle)$  for every  $\phi \in \langle W, \mathcal{F}, \mathcal{P} \rangle$  and  $B \in \mathcal{F}B \cap R_a(w)(w')$ , and

<sup>1</sup> Note that  $R_a(w)(w')$  is an abbreviation for  $R(\langle w, a \rangle)(\langle w', a' \rangle)$ . The abbreviation is used to save space.

- (ii)  $\langle R_a(w)(w'), \mathcal{F}(\langle w, a \rangle) \rangle$  is a measurable space.
- (c)  $\mathcal{P}: (W \times A) \rightarrow (W \rightarrow [0, 1])$  such that  $\mathcal{P}(\langle w, a \rangle)(\phi) = \mathcal{P}(\phi \mid R_a(w)(w'))$ .
- (d)  $\langle R_a(w)(w'), \mathcal{F}(\langle w, a \rangle), \mathcal{P}(\langle w, a \rangle) \rangle$  is a probability space.

## 8. Formal Definition of Independency

Recall that it is the characteristic of BCs that the apodosis is truth-conditionally independent of the protasis, since in a BC sentence form  $p \rightarrow q$ ,  $q$  may be true regardless of whether  $p$  is true.<sup>2</sup> In this sense, we need to conduct research on how independency can be rigorously defined before diving into the semantic analysis of BCs.

As we saw in scenario #1, context should be considered for determining whether certain facts are factually dependent or independent.

### (39) Common Ground (Stalnaker (1978))

$$CG(\langle w, a \rangle) = \{p: p \text{ is accepted as true by an agent individual } a \text{ in } w\}.$$

### (40) Context Sets

$$\cap CG(\langle w, a \rangle) = \{\langle w', a' \rangle \in W \times A: \forall p[p \in CG(\langle w, a \rangle) \rightarrow \langle w', a' \rangle \in p]\}$$

Recall from section 5 that dependency is closely related to law-like principles, conventional rules, or normal course of events (Kratzer (1991)). Thus, we think of factual dependencies as the body of law or normality for convenience's sake,<sup>3</sup> which is along the lines of Veltman (2005) and Goebel (2017).

### (41) The law-based context $C_L(\langle w, a \rangle)$

$$C_L(\langle w, a \rangle) = \{\langle w', a' \rangle: \text{The body of law holds in } \langle w', a' \rangle\}$$

The information or assumptions of whether the body of law applies is already given in the CG, more precisely the context set, and thus  $C_L(\langle w, a \rangle) \subseteq \cap CG(\langle w, a \rangle)$ .

### (42) Conditional Probability

Given a conditional  $\psi \rightarrow \phi$ ,  $\mathcal{P}(\phi \mid \psi)$  is equal to  $\frac{\mathcal{P}(\phi \cap \psi)}{\mathcal{P}(\psi)}$ , when  $\mathcal{P}(\psi) > 0$ .

### (43) Definition of Independency

For any  $\phi_1$  and  $\phi_2 \in \mathcal{F}(\langle w, a \rangle)$ ,  $\phi_1$  and  $\phi_2$  are factually independent in  $C_L(\langle w, a \rangle)$  in  $\cap CG(\langle w, a \rangle)$  iff for every  $\langle w', a' \rangle \in C_L(\langle w, a \rangle)$  which is located on a probability space  $\langle R_a(w)(w'), \mathcal{F}(\langle w, a \rangle), \mathcal{P}(\langle w, a \rangle) \rangle$ ,  $\mathcal{P}(\langle w', a' \rangle)(\phi_1 \mid \phi_2) = \mathcal{P}(\langle w', a' \rangle)(\phi_1 \mid \phi_2^C) = \mathcal{P}(\langle w', a' \rangle)(\phi_1)$ , and vice versa.

<sup>2</sup> Recall that this is the main difference between normal conditionals and BCs.

<sup>3</sup> The terms *the body of law* and *normality* are used here in a broad sense for whatever demonstrates regular and durable patterns of something, including physical law, habitual behavior, norms, institutions, and conventional rules etc.

## 9. Semantic Analysis

### A. Presuppositional Approach to BCs

The way I prefer to go is to provide a semantic analysis of BCs in conjunction with a presuppositional approach (Heim and Kratzer (1988)), according to which the semantic calculation of a linguistic expression would continue only if the presupposition were satisfied, and otherwise the calculation would stop, thereby ending up with pragmatic infelicity.

#### (44) Presupposition triggered by BCs

A conditional  $[\phi_1 \rightarrow \phi_2]^{w, g}$  is defined as a BC iff the agent  $a$ 's assertion of  $\phi_2$  are factually independent of  $\phi_1$  in a given  $C_L(\langle w, a \rangle)$  in  $\cap CG(\langle w, a \rangle)$  located on a probability space  $\langle R_a(w)(w'), \mathcal{F}(\langle w, a \rangle), \mathcal{P}(\langle w, a \rangle)\rangle$ .

### B. The Protasis as a Signal of Illocutionary Forces

Recall that BCs are involved in a speech act, more precisely an illocutionary act. In the speech act theory literature (Austin (1962), Searle (1965), DeRose and Grandy (1999), and Franke (2007) among others).

(45) DeRose and Grandy (1999) argue that the protasis of BCs functions to establish a pragmatic relation to the apodosis.

(46) Franke (2007) also makes a similar point when arguing that the protasis of BCs behaves like a discourse function because it leads to a non-conditional reading.

(47) Following DeRose and Grandy, and Franke, I argue that the protasis of a BC serves to be a signal of illocutionary forces in relation to the apodosis.

The following examples are in favor of this (Hewings (2005: 18)):

- (48) a. I will open a bottle of lemonade if you want some. (Hypothetical Conditional)  
b. ??I am going to open a bottle of lemonade if you want some. (Hypothetical)  
c. I am going to open a bottle of lemonade, if you want some. (Biscuit Conditional)

#### (49) LF for a BC

$[_{CP} [_{CP} [_{CP} [_{CP} \text{FORCE} [_{C'} \text{if } \phi_1 ] ] [_{IP} \phi_2 ] ] ] ] ]$  (FORCE: Abstract Morpheme)

Given the LF in (49), we are in a position to provide the semantic definition of BCs as follows in terms of the probability theory.

### C. Semantic Definitions of BCs

Given the probability measure in (37) and the probability space in (38), the truth conditions for BCs can be stated as follows:

(50) Semantic Definition of BCs

$[[\phi_1 \rightarrow \phi_2]]^{w, g, CG} = 1$  iff for every  $\langle w', a' \rangle \in C_L(\langle w, a \rangle)$  such that  $C_L(\langle w, a \rangle) \subseteq \cap CG(\langle w, a \rangle)$  located on a probability space  $\langle R_a(w)(w'), \mathcal{F}(\langle w, a \rangle), \mathcal{P}(\langle w, a \rangle) \rangle$  and  $\frac{1}{2} < \mathcal{P}(\phi_1(\langle w', a' \rangle)) \leq 1$ , (i)  $\exists \delta[\delta \in \Delta \ \& \ \delta = \text{ILF}(\phi_1(\langle w', a' \rangle), \phi_2(\langle w', a' \rangle)) \ \& \ \delta(a, h, \phi_2)]$ , and (ii)  $\frac{1}{2} < \mathcal{P}(\phi_2(\langle w', a' \rangle)) \leq 1$ .

(N.B.:  $\Delta$ , ILF, sp and h are a set of illocutionary forces (or speech acts), a function from a pair of  $\phi_1$  and  $\phi_2$  to the most salient force, the speaker, and the hearer, respectively)

Note that  $\delta$  in (50) is something similar to what Searle (1969) calls *an illocutionary force indicating device*, and thus  $\delta(\phi_2)$  denotes the propositional content of an illocutionary force that is obtainable from the apodosis  $\phi_2$ .

### 10. Conclusion

Recall that the previous analyses present the truth conditions for BCs that state that the apodosis is true in a given context, which causes some difficulty in dealing with the semantics of BCs.

(51) The Advantages of the Current Framework

- a. It can give a rigorous and clear-cut definition of independent relationships between the protasis and the apodosis in terms of probability.
- b. The incorporation of probability into the semantics of BCs does not necessarily require the truth of the apodosis, but rather the speaker's degree of certainty.
- c. It can account for where illocutionary forces come from in BCs.
- d. It can show clearly how the pragmatic account pertaining to a speech act can be incorporated into the semantic definition of BCs.

### References

- Austin, J. L. 1956. Ifs and Cans. *Proceedings of the British Academy* 42. 107-132.
- Austin, J. L. 1962. *How to Do Things with Words*, Clarendon Press.
- Bach, K. 1999. The Myth of Conventional Implicature. *Linguistics and Philosophy* 22, 327–366.
- Bach, K. and R. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge, MA.: MIT Press
- Comrie, B. 1986, Conditionals: A Typology, in E. Traugott, A. ter Meulen, J. Snitzer Reilly, and C.A. Ferguson (eds.), *On Conditionals*, Cambridge: Cambridge University Press, 77–99.
- Davison, A. 1983. Linguistic or Pramatic Description in the Context of the Perfromadox. *Linguistics and Philosophy* 6, 499-526.
- DeRose, K. and R. Grandy. 1999. Conditional Assertion and “Biscuit” Conditionals. *Noûs* 33(3),

- 405-420.
- van Dijk, T. A. 1979. Pragmatic connectives. *Journal of Pragmatics*, 3, 447–456.
- von Fintel, K. 2011. Conditionals, in von Heusinger, Maienborn and Portner (eds.), *Semantics: An International Handbook of Meaning* vol 2, 1515-1538.
- Franke, M. 2007. The Pragmatics of Biscuit Conditionals. In M. Aloni, P. Dekker & F. Roelofsen (eds.), *Amsterdam Colloquium* 16, 91-96.
- Geis, M. 1995. *Speech Acts and Conversational Interaction*. Cambridge: Cambridge University Press.
- Goebel, A. 2017. Laws for Biscuits: Independence and Dependence in Conditionals. *Proceedings of SALT* 27, 277-396.
- Hewings, M. (2005). *Advanced Grammar in Use* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press
- Horn, L. R. 1989 *A Natural History of Negation*, University of Chicago Press.
- Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*, PhD dissertation, University of Massachusetts, Amherst, MA.
- Heim, I. and A. Kratzer. 1988. *Semantics in Generative Syntax*, Cambridge, MA.: MIT Press.
- Iatridou, S. 1991. *Topics in Conditionals*, PhD dissertation, Cambridge, MA.: MIT Press.
- Krifka, M. 2004. *Semantics Below and Above Speech Acts*. handout of a talk given at Stanford University, Palo Alto, CA, April 9.
- Kratzer, A. 1991. Conditionals. In A. von Stechow and D. Wunderlich (eds.), *Semantics: An international Handbook of Contemporary Research*, Berlin: de Gruyter, 639-650.
- Lakoff, G. 1972. Linguistics and Natural Logic, in D. Davidson and G. Harmon (eds.), *Semantics of Natural Language*, Dordrecht: Reidel.
- Partee, B. and V. Borschev: 2000. Possessives, Favorite and Coercion, in A. Riehl and R. Daly (eds.), *Proceedings of ESCOL99*, CLC Publications, Cornell University, pp. 173–190.
- Predelli, S. 2009. Towards a Semantics for Biscuit Conditionals. *Philosophical Studies*. 142(3), 293-305.
- Ross, J. R. 1970. *Constraints on Variables in Syntax*, PhD dissertation, MIT, Cambridge, MA.
- Sadock, J. 1974. *Towards a Linguistic Theory of Speech Acts*. New York: Academic Press.
- Sadock, J. 2004. Speech Acts, in L. Horn and G. Ward (eds.), *The Handbook of Pragmatics*. Malden, MA.: Blackwell Publishing, 53–73.
- Siegel, M. 2006. Biscuit Conditionals: Quantification over Potential Literal Acts. *Linguistics and Philosophy* 29(2), 167-203.
- Searle, J. 1965. What Is a Speech Act? in Max Black (ed.), *Philosophy in America*, Ithaca: Cornell University Press, 221-239.
- Searle, J. 1969. *Speech Acts*, Cambridge University Press.
- Searle, J. R. 1975. Indirect Speech Acts, in P. Cole and J. L. Morgan (eds.), *Syntax and Semantics* vol 3: Speech Acts. New York: Academic Press, 59-82.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Stalnaker, R. 1978. Assertion. in P. Cole (ed.), *Syntax and Semantics* vol 9. New York: Academic Press, 315-332.
- Veltman, Frank. 2005. Making counterfactual assumptions. *Journal of Semantics* 22, 159–180.
- Zanuttini, R and P. Portner. 2003. Exclamative Clauses: At the Syntax-Semantics Interface. *Language* 79, 39-81.

## On Metalinguistic Negation

Suwon Yoon  
suwon@uos.ac.kr  
University of Seoul

### 1. What is metalinguistic negation?

☞ “a device for objecting to a previous utterance on any grounds whatsoever including the conventional or conversational implicata it potentially induces, its morphology, its style or register or its phonetic realization” (Horn 1989: 363)

- (1) *Metalinguistic Negation (MN)*  
a. My brother is *not* a crook—I don’t have a brother! [English]  
b. Speaker A: It’s stewed bunny.  
Speaker B: It’s *not* stewed bunny, it’s civet de lapin!

☞ MN tends to appear in contrastive forms, conveying an exclamative emphatic flavor, as shown below (Horn 1989).

- (2) *Emphatic flavor of MN*  
a. Around here we *don’t* LIKE coffee—we LOVE it. [English]  
b. She *doesn’t* sell INSurance—she sells inSURance.  
c. I’m *not* HIS brother—he’s MY brother!  
d. Mozart’s sonatas were for piano and violin, *not* for violin and piano.  
e. I’m *not* happy with the plan, I’m ecstatic!  
f. You *didn’t* eat some of the cookies, you ate them all!

☞ Hallmark properties of MN (Horn 1985, 1989; Burton-Roberts 1989).

- (3) *Hallmarks of MN*  
a. Metalinguistic negations are standardly used as a rejoinder to a previous utterance.  
b. There is a certain prosodic pattern commonly associated with metalinguistic negation.

☞ Horn’s (1986) definition of MN as an *objection to a previous utterance*, as in (4).

- (4) I object to *U*  
*U: the previous utterance*

### 2. Metalinguistic Comparatives and Metalinguistic Negation

#### 2.1. Three types of MC

☞ Giannakidou and Yoon (2011): English MC can be classified into two types: (i) bare *rather* for a (possibly neutral) *appropriateness assessment MC*, and (ii) *would rather* for a more *emphatic preferential MC*.

- (5) Your problems are more serious than mine. [Degree Comparatives]  
(6) a. Your problems are more financial than legal. [Metalinguistic Comparatives]  
b. Your problems are financial rather than legal. (from McCawley 1968)

☞ Likewise, Greek MC is classified in to the two types.

- (7) a. *Appropriateness assessment MC*  
o Pavlos ine *perissotero* philologhos *{para/apoti}* glossologhos. [Greek]  
the Paul is-3SG more philologist than linguist  
'Paul is a philologist rather than a linguist.'



Paraphrase: In the present context, I prefer the sentence ‘Paul is a philologist’ to the sentence ‘Paul is a linguist.’

b. *Emphatic preferential MC*

kalitera na pethano {para/#apoti} na ton pandrefto!  
 better SUBJ die.1SG than SUBJ him marry.1SG

‘I would rather die than marry him.’

Paraphrase: It is more preferable to me to die than to marry him.

☞ Korean MC is classified in to the three types.

(8) a. *Appropriateness assessment MC* [Korean]

Lee-nun pwucirenya-ta-*kipota* ttoktokha-ta.  
 Lee-TOP industrious-DECL-saying.than bright-DECL

‘Lee is more bright than industrious.’

[In the present context, I prefer the sentence ‘Lee is bright’ to the sentence ‘Lee is industrious’]

b. *Emphatic preferential MC*

Ku-wa kyelhonha-*{kipota/nuni}* nay-ka honcasal-keyss-ta.  
 him-DAT marry-rather.than I-NOM live.alone-will-DECL

‘I would rather live alone than marry him!’

c. *Negative emphatic preferential MC*

Ku-wa kyelhonha-*nuni* charari nay-ka cwuk-keyss-ta.  
 him-DAT marry-rather.than rather I-NOM die-will-DECL

‘I would rather die than marry him!’

## 2.2. Parallels between MC and MN

☞ Giannakidou and Stavrou (2009; see also Giannakidou and Yoon 2011),

- (i) Both MC and MN are binary: They contrast two propositions;
- (ii) Both MC and MN contribute a propositional attitude that relates to appropriateness/belief; and
- (iii) Both MC and MN involve clausal ellipsis and are akin to coordinate structures (Giannakidou and Stavrou 2009:pp.11).

☞ External negation

(9) a. \*Swuni-nun yekan an yeypputa. [Korean]

Swuni-TOP ordinarily not pretty  
 ‘Swuni is ordinarily unpretty.’

b. Swuni-nun yekan yeyp-c*i* an-h-ta. (Taytanhi yeypu-ta.)  
 Swuni-TOP ordinarily pretty not-be-DC extremely pretty-DC

‘Swuni is not ordinarily pretty.’ (‘She is extremely pretty.’)

☞ Peripheral positions such as a sentence-initial or sentence-final position (Martins 2019)

(10) a. *No way* some men are chauvinists – all men are chauvinists. [English]

b. *Nonsense* I managed to trap two mongeese. I managed to trap two mongooses. (Drozd 2001: 56-57)

(11) a. A: He found proofs that clinched the argument.  
 B: He found proofs that clinched the argument *fiddlesticks*.

b. A: He has a dog that bit a policeman.  
 B: He had a dog that bit *your old man*. (Bolinger 1977: 45)

## 3. Empirical evidence: Six semantic sources of metalinguistic negation markers

### 3.1. Objection due to *inappropriateness*

☞ First, a speaker may object to the previous utterance *U* because it is an *inappropriate* or improper description of the state of affairs within *the speaker’s own worldview*. i.e., what she thinks (she knows), believes, or hopes

(16) I’m *not* happy with the plan, I’m ecstatic! [English]

(17) a. *No way* some men are chauvinists – all men are chauvinists.

- b. *Nonsense* I managed to trap two mongeese. I managed to trap two mongooses. (Drozd 2001: 56-57)
- (18) a. *Dhen* ine eksipnos ala ergatikos. [Greek]  
He is not bright but industrious.  
b. Ine *oxi* eksipnos ala ergatikos.  
He is not bright but industrious.
- (19) a. \*Swuni-nun yekan *an* yeypputa. [Korean]  
Swuni-TOP ordinarily not pretty  
b. Swuni-nun yekan yeyp-c*i* *an*-h-ta. (Taytanhi yeypu-ta.)  
Swuni-TOP ordinarily pretty not-be-DC extremely pretty-DC  
'Swuni is not ordinarily pretty.' ('She is extremely pretty.')

### 3.2. Objection due to *unusuality or abnormality*

☞ Second, a speaker's objection to *U* can be made because the speaker thinks/believes/hopes the content of *U* to be *unusual or abnormal*, hence it brings about *surprise*.

- (20) A: Ne cokum kekcingtoy-ni? [Korean]  
are-2SG a little worried-Q  
'Are you a little worried?'  
B: Cokum kekcingtoy-nyani! Kekcingtoye-cwu-keyss-e!  
a little worried-exclamatory.Q worried-die-will-Decl  
'I'm not a little worried, I am worried sick!'
- (21) A: Estás um pouco preocupado? [European Portuguese]  
are-2SG a little worried  
'Are you a little worried?'  
B: Estou lá/agora um pouco preocupado.  
am MN-marker a little worried  
'I'm not a little worried, I am worried sick.'

### 3.3. Objection due to *unidentifiability-driven non-referentiality*

☞ Third, a speaker may object to the previous utterance *U* because *U* is perceived as something unutterable or inconceivable, which is realized by MN markers with *non-referentiality* because the speaker appeals to the fact that she is *unable to identify* such a thing.

- (22) JOHN KNOW ANSWER !part:indef! [ASL]  
'John knows the answer! (How could you have thought he wouldn't?)'
- (23) A: Ce kaswu cham mesiss-ta! [Korean]  
that singer really cool-Decl  
'That singer is really cool!'  
B: Mesiss-ki-nun mwusun!  
cool-NMLZ-Top what/which  
'Lit. What/which, he's cool!' (Intended: 'He's cool, my eye!')
- (24) A: Ce kaswu cham mesiss-ta! [Kyengsang dialect of Korean]  
that singer really cool-Decl  
'That singer is really cool!'  
B: Mesiss-ki-nun etey!  
cool-NMLZ-Top where  
'Lit. Where, he's cool!' (Intended: 'He's cool, my eye!')
- (25) A: O governo vai baixar os impostos. [European Portuguese]  
the government goes lower the taxes  
'The government is going to lower the taxes.'  
B: a. Qual quê!  
which what  
b. Qual vai baixar os impostos!  
which goes lower the taxes  
c. Vai baixar os impostos o quê!  
goes lower the taxes the what

d. Qual vai baixar os impostos qual quê!  
 which goes lower the taxes which what  
 ‘The government is going to lower the taxes, my eye.’

### 3.4. Objection due to *non-existence-driven non-referentiality*

☞ Fourth, a speaker may object to *U* because it is *unutterable* or *non-referential*, which is shown from the MN marker of *non-existence*.

- (28) A: Ce kaswu cham mesiss-ta! [Korean]  
 that singer really cool-Decl  
 ‘That singer is really cool!’  
 B: Mesiss-ki-nun kayppwul!  
 cool-NMLZ-Top dog.horn  
 ‘Dog’s horn, he’s cool!’
- (29) B: Mesiss-ki-nun, kwuysin ssinalak kkamek-nun soli-ha-ney!  
 cool-NMLZ-Top ghost grain peeling.to.eat sound-say-Excl  
 ‘Sounds like a ghost’s peeling and eating grains, he’s cool!’
- (30) B: Mesiss-ki-nun, kay phwul ttutemek-nun soli-ha-ney!  
 cool-NMLZ-Top dog grass nibbling sound-say-Excl  
 ‘Sounds like a dog’s nibbling on grass, he’s cool!’
- (31) B: Mesiss-ki-nun, kimpap yepkwuli theci-nun soli-ha-ney!  
 cool-NMLZ-Top seaweed.roll side pop-Prog sound-say-Excl  
 ‘Sounds like a seaweed roll’s side popping, he’s cool!’  
 B’: Mesiss-ki-nun, cilengi haphwumha-nun soli-ha-ney!  
 cool-NMLZ-Top earthwarm yawn-Prog sound-say-Excl  
 ‘Sounds like an earthwarm’s yawning, he’s cool!’  
 B’’: Mesiss-ki-nun, kaykwuli yeppalcilha-nun soli-ha-ney!  
 cool-NMLZ-Top frog side.kick-Prog sound-say-Excl  
 ‘Sounds like a frog’s side-kicking, he’s cool!’

### (32) Formation rule of MN markers in Korean

If an expression describes an *extremely unlikely* or *preposterous event*, its *non-referentiality* gives rise to a MN marker.

- (33) A: He found proofs that clinched the argument. [English]  
 B: He found proofs that clinched the argument *nothing*. (Bolinger 1977: 45)
- (34) A: O governo vai baixar os impostos. [European Portuguese]  
 the government goes lower the taxes  
 ‘The government is going to lower taxes.’  
 B: a. Vai nada (baixar os impostos).  
 goes nothing lower the taxes  
 b. Vai baixar os impostos nada.  
 goes lower the taxes nothing  
 ‘Like hell (it is going to lower taxes)!’

### 3.5. Objection due to *unworthiness* or *dispreference*

☞ Fifth, a speaker’s objection to *U* is typically expressed in the form of *depreciatives* in many languages.

- (35) A: You still love me. [English]  
 B: *Like hell* I still love you. (Horn 1989: 402)
- (36) “Relax, pal,” Mark said. “Some days are better than others.”  
 Todd gulped his beer and said, “Relax *my ass*. We’ve been doing this crap for a month and it feels like I’m carrying the load here.” (John Grisham, *The Rooster Bar*, 203) (Martins 2019:(8))
- (37) Deberías desculpate por tu comportamiento. [Spanish]  
 should.1SG apologize.REFL.1SG for your behaviour

- ‘You should apologize for your behavior.’  
 B: ¡Una mierda voy (yo) a disculparme!  
 a shit go I to apologize.REFL  
 ‘Like hell I will apologize!’ (Olza Moreno 2017: 47)
- (38) A: O governo (não) vai baixar os impostos.  
 the government (not) goes lower the taxes  
 ‘The government {is/isn’t} going to lower taxes.’  
 B: a. (Não) vai baixar os impostos uma ova.  
 (not) goes lower the taxes a fish-roe  
 b. Uma ova é que (não) vai baixar os impostos.  
 a fish-roe is that (not) goes lower the taxes  
 c. Uma ova!  
 a fish-roe  
 ‘Like hell (it’s/isn’t going to lower taxes)!’ (Martins 2014)
- (39) *Mon oeil* que le roi de France est chauve. Il n’y a pas de roi de France. [French]  
 my eye that the king of France is bald there is no king of France  
 ‘Nonsense the king of France is bald. There is no king of France.’ (Larrivé 2011: 3)

☞ *shma/schma*-reduplication in English, originated in Yiddish, is another example of productive formation for MN construction (along with the Korean productive rule mentioned in section 3.4).

- (40) a. Money, shmoney!  
 b. Fame, schmame!  
 c. Work, shmerk!  
 d. Age, schmage!  
 e. Love, shmove!  
 f. Beauty, schmooty! [English]

(41) *Shma/schma*-reduplication in English

X, *shma/schma*-X reduplication can be used with a noun or adjective, as a response to a previously-made statement X to humorously express the speaker's doubts, or dismiss or downplay X as unimportant or irrelevant.

- (42) A: Ce kaswu cengmal mesiss-ta!  
 that singer really cool-Decl [Korean]  
 ‘That singer is really cool!’  
 B: Ce kaswu cengmal mesiss-tani, wuski-ney!  
 that singer really cool-Excl laughable-Excl  
 ‘That’s laughable, that singer is really cool!’
- (43) B’: Mesiss-ki-nun (mwusun) {elecwuk-ul/mangh-al/yempyeng/wulacil/nimilel}!  
 cool-NMLZ-Top what/which freezing.to.death/going.bust/epidemic/damn/damn  
 ‘Lit. Freezing to death/going bust/epidemic/damn/damn, he’s cool!’

### 3.6. Objection due to *insincerity*

☞ Finally, as in (44), *irony or sarcasm* can be a device to convey a speaker’s objection to the previous assertion U.

- (44) *Yeah right* you ate some of the cookies. You ate all of them! (Drozd 2001: 56-57)
- (45) A: Ce kaswu cengmal mesiss-ta!  
 that singer really cool-Decl [Korean]  
 ‘That singer is really cool!’  
 B: Ce kaswu cengmal mesiss-nun-ke, coaha-ney!  
 that singer really cool-Top-NMLZ like.it-Excl  
 ‘That singer is really cool?! Oh, you’d like that, do ya?’  
 B’: Ce kaswu cengmal mesiss-tani, nolkoiss-ney!

- that singer really cool-Excl enjoy.it-Excl  
 ‘That singer is really cool?! Oh, you’re enjoying that, aren’t ya?’
- (46) A: J’ai gagné à la loterie. [French]  
 ‘I’ve won the lottery.’  
 B: Tu parles!  
 ‘Go on (Lit. you speak)!’

### 3.7. Four types of MN driven from six semantic sources of MN markers

Table 1. The genesis of Metalinguistic Negation markers

Six classes of MN markers	Semantic sources for MN markers	MN markers
Class 1	<i>inappropriateness</i>	<i>Regular negation markers:</i> <i>not, no way, nonsense</i> in English; <i>dhen</i> ‘not,’ <i>oxi</i> ‘no’ in Greek; <i>ci anh</i> ‘not’ (external negation) in Korean
Class 2	<i>unusuality or abnormality</i>	<i>Emphatics markers:</i> <i>nyani</i> ‘exclamative’ in Korean; <i>lá</i> ‘lit. there’, <i>agora</i> ‘lit. now’ in European Portuguese
Class 3	<i>unidentifiability-driven non-referentiality</i>	<i>Anti-specificity markers:</i> <i>part:indef</i> ‘someone or other’ in ASL; <i>mwusun</i> ‘which,’ <i>etey</i> ‘where,’ <i>mwusun</i> + depreciatives in Korean dialects; <i>qual</i> ‘which,’ <i>qual quê</i> ‘which what,’ <i>o quê</i> ‘the what’ in European Portuguese; <i>qué ... ni qué</i> + depreciatives ‘what ... nor what’ in Peninsular Spanish
Class 4	<i>non-existence-driven non-referentiality</i>	<i>Non-existence markers:</i> <i>kayppwul</i> ‘dog’s horn,’ nonexistent event-describing expressions (e.g. ‘sounds like a ghost’s peeling and eating grains,’ ‘sounds like a dog’s nibbling on grass,’ ‘sounds like a maki roll’s side popping,’ ‘sounds like an earthworm’s yawning,’ ‘sounds like a frog’s side-kicking’) in Korean; <i>nothing, nothing of the sorts</i> in English; <i>nada</i> ‘nothing’ in European Portuguese; <i>minga</i> ‘no/nothing’ in Rioplatense Spanish
Class 5	<i>unworthiness or dispreference</i>	<i>Depreciative markers:</i> <i>the hell, like hell, my ass, my eye, bullshit, poppycock, fiddlesticks, your old man, like fun, like fudge, yo’ mama, my foot, X shma/schma-X</i> in English; <i>una leche</i> ‘a blow/hit’, ( <i>unas/las</i> ) <i>narices</i> ‘a/the noses’, <i>una mierda</i> ‘a shit’, <i>los cojones</i> ‘the balls’ in Spanish; <i>uma ova</i> ‘a fish roe’, <i>o tanas</i> (obscure meaning), <i>o caralho/o caraças</i> (‘penis’ (slang)), <i>uma merda</i> ‘a shit’ in European Portuguese; <i>qué ... ni qué {narices/cojones/coño/mierda/leche}</i> ‘what ... nor what noses/balls/cunt/shit/blow’ in Peninsular Spanish (repeated from class 3); <i>mon oeil!</i> ‘my eye’ in French; <i>wuskiney</i> ‘laughable,’ { <i>elecwuk-ul/manghal/yempyeng/wulacil/nimilel</i> } ‘freezing.to.death/going.bust/epidemic/damn/damn’ in Korean dialects
Class 6	<i>insincerity</i>	<i>Irony markers:</i> <i>coahaney</i> ‘oh, you’d like that, do ya?’, <i>noikoissney</i> ‘oh, you’re enjoying that, aren’t ya?’ in Korean; <i>yeah right, yeah yeah, oh yeah</i> in English, <i>tu parles!</i> ‘go on (lit. you speak)’ in French

Given the striking commonalities between MN and MC observed in section 2, and the crosslinguistic observations in section 3, we further propose to categorize MN into the following four types:

- appropriateness assessment MN,
- emphatic denial MN,
- negative emphatic denial MN, and
- irony MN (without MC counterpart).

☞ The first type is *appropriateness assessment MN* that can be neutral or emphatic. Since this type involves corrective objections to some aspects of the utterance such as presupposition, appropriate term, or pronunciation, this MN is typically realized as a neutral non-emotive marker.

(47) *Appropriateness assessment MN: inappropriateness*

- a. My brother is *not* a crook—I don't have a brother! [English]
- b. Speaker A: It's stewed bunny.  
Speaker B: It's *not* stewed bunny, it's civet de lapin! (Drozd 1995)
- c. Around here we *don't* LIKE coffee—we LOVE it.
- d. She *doesn't* sell INSurance—she sells inSURance. (Horn 1989)
- e. Speaker A: o Petros exi tria pedia. [Greek]  
'Peter has three children.'  
Speaker B: oxí! o Petros exi *oxi* tria pedia, ala tessera!  
'No! Peter doesn't have three children, but four!' (Giannakidou and Stavrou 2009)
- f. Swuni-nun yekan yeypp-ci an-h-ta. (Taytanhi eyppu-ta.) [Korean]  
Swuni-TOP ordinarily pretty not-be-DC extremely pretty-DC  
'Swuni is not ordinarily pretty.' ('She is extremely pretty.')

☞ The second type is *emphatic denial MN*: This type expresses an objection to the previous utterance purely based on a speaker's personal preference.

(48) *Emphatic denial MN (based on unusuality or abnormality)*

- a. A: Ne cokum kekcengtoy-ni? [Korean]  
are-2SG a little worried-Q  
'Are you a little worried?'  
B: Cokum kekcengtoy-nyani. Kekcengtoy-cwukeyss-e!  
a little worried-emphatic Q. worried-die-Decl  
'I'm not a little worried, I am worried sick!'
- b. A: Estás um pouco reocupado? [European Portuguese]  
are-2SG a little worried  
'Are you a little worried?'  
B: Estou lá/agora um pouco reocupado,  
am MN-marker a little worried  
'I'm not a little worried, I am worried sick.'

(49) *Emphatic denial MN (based on unidentifiability-driven non-referentiality)*

- a. JOHN KNOW ANSWER !part:indef! [ASL]  
'John knows the answer! (How could you have thought he wouldn't?)'
- b. A: Ce kaswu cham mesiss-ta! [Korean dialects]  
that singer really cool-Decl  
'That singer is really cool!'  
B: Mesiss-ki-nun {mwusun/etey}!  
cool-NMLZ-Top what/which/where  
'Lit. What/which, he's cool!'
- c. A: O governo vai baixar os impostos. [European Portuguese]  
The government goes lower the taxes  
'The government is going to lower the taxes.'  
B: Qual quê!  
which what (Martins 2014)

☞ The third type is *negative emphatic denial MN*, which is a negative emotive variant of the second type.

(50) *Negative emphatic denial MN (based on non-existence-driven non-referentiality)*

- a. A: Ce kaswu cham mesiss-ta! [Korean]  
that singer really cool-Decl

'That singer is really cool!'  
 B: mesiss-ki-nun kayppwul!  
 cool-NMLZ-Top dog.horn  
 'Dog's horn, he's cool!'

- b. A: He found proofs that clinched the argument. [English]  
 B: He found proofs that clinched the argument *nothing*. (Bolinger 1977: 45)
- (51) *Negative emphatic denial MN (based on unworthiness or dispreference)*
- a. A: You still love me.  
 B: *Like hell* I still love you. (c 1989: 402)
- b. "Relax, pal," Mark said. "Some days are better than others." Todd gulped his beer and said, "Relax *my ass*. We've been doing this crap for a month and it feels like I'm carrying the load here."  
 (John Grisham, *The Rooster Bar*, 203) (Martins 2019)
- c. A: Deberías desculparte por tu comportamiento. [Spanish]  
 should.1SG apologize.REFL.1SG for your behavior  
 'You should apologize for your behavior.'  
 B: ¡*Una mierda* voy (yo) a disculparme!  
 a shit go I to apologize.REFL  
 'Like hell I will apologize!' (Olza Moreno 2017: 47)
- (52) A: Ce kaswu cham mesiss-ta!  
 that singer really cool-Decl [Korean]  
 'That singer is really cool!'  
 B': Mesiss-ki-nun (*mwusun*) {*elecwuk-ul/mangh-al/yemphyeng/wulacil/nimilel*}!  
 cool-NMLZ-Top what/which freezing.to.death/going.bust/epidemic/damn/damn  
 'Lit. Freezing to death/going bust/epidemic/damn/damn, he's cool!'

⚡ The fourth type is *irony MN*: This type expresses an objection to the previous utterance by uttering the content in the form of opposite polarity, which is based on the semantic source of *insincerity* (class 6).

- (53) *Irony MN (based on insincerity)*
- a. *Yeah right* you ate some of the cookies. You ate all of them! [English] (Drozd 2001: 56-57)
- b. A: Ce kaswu cengmal mesiss-ta!  
 that singer really cool-Decl [Korean]  
 'That singer is really cool!'  
 B: Ce kaswu cengmal mesiss-nun-ke, *coahaney!*  
 that singer really cool-Top-NMLZ like.it  
 'That singer is really cool? Oh, you'd like that, do ya?'  
 B': Ce kaswu cengmal mesiss-tani, *nolkoiss-ney!*  
 that singer really cool-Excl enjoy.it-Excl  
 'That singer is really cool?! Oh, you're enjoying that, aren't ya?'
- c. A: J'ai gagné à la loterie. [French]  
 'I've won the lottery.'  
 B: Tu parles!  
 'Go on (Lit. you speak)!'

☞ Among the six classes of MN markers, we propose to classify their usages in these four types of MN, as summarized below.

Table 2. Four types of MN and six classes of MN markers

Four types of MN	Six classes of MN markers	Semantic sources for MN markers	MN markers
<b>Type I:</b> <i>Appropriateness assessment MN</i>	Class 1	<i>inappropriateness</i>	<i>Regular negation markers:</i> <i>not, no way, nonsense</i> in English, <i>ci an ‘not’</i> in Korean, etc.
<b>Type II:</b> <i>Emphatic denial MN</i>	Class 2	<i>unusuality</i> or <i>abnormality</i>	<i>Emphatics markers:</i> <i>lá ‘lit. there’</i> in European Portuguese, <i>nyani ‘exclamative’</i> in Korean, etc.
	Class 3	<i>unidentifiability-driven</i> <i>non-referentiality</i>	<i>Anti-specificity markers:</i> <i>part:indef ‘someone or other’</i> in ASL, <i>mwusun ‘which’</i> in Korean, etc.
<b>Type III:</b> <i>Negative emphatic denial MN</i>	Class 4	<i>non-existence-driven</i> <i>non-referentiality</i>	<i>Non-existence markers:</i> <i>nothing</i> in English, <i>nada</i> in Spanish, <i>kayppwul ‘dog’s horn’</i> in Korean, etc.
	Class 5	<i>unworthiness</i> or <i>dispreference</i>	<i>Depreciative markers:</i> <i>the hell, like hell, X shma/schma-X</i> in English, <i>wuskiney ‘laughable,’ elecwuk-ul ‘freezing to death’</i> in Korean, etc.
<b>Type IV:</b> <i>Inory MN</i>	Class 6	<i>insincerity</i>	<i>Irony markers:</i> <i>yeah right</i> in English, <i>coahaney ‘oh, you’d like that, do ya?’</i> , <i>nolkoissney ‘oh, you’re enjoying that, aren’t ya?’</i> in Korean, <i>tu parles! ‘go on’</i> in French, etc.

#### 4. Semantics of metalinguistic negation: Subjective anti-veridicality

##### 4.1. Objective vs. subjective anti-veridicality

☞ First, *regular negation* can be construed as an *anti-veridical operator of objective kind* (GM 2021:pp. 56, (13)).

(54) Objective veridicality:

- (i) A propositional function  $F$  is veridical iff  $Fp \rightarrow p$  is logically valid.
- (ii)  $F$  is nonveridical iff  $Fp \not\rightarrow p$ .
- (iii)  $F$  is antiveridical iff  $Fp \rightarrow \neg p$ .

☞ *relativization of veridicality*: the models  $M$  are defined as “sets of worlds relative to the individual anchor  $i$ , corresponding to what  $i$  believes, knows, dreams, imagines, remembers, wants, and the like, depending on the lexical meaning of the propositional attitude verb” (GM 2021:pp. 59, (20)).

(55) Information state of an individual anchor  $i$ :

An information state  $M(i)$  is a set of worlds associated with an individual  $i$  representing worlds compatible with what  $i$  knows or believes in the context of utterance.

☞ GM: two ways of viewing the information states of  $M(i)$ :

- (a) they can be understood broadly as a *private space* encompassing  $i$ ’s worldview, or
- (b) they can be construed as *mental states* that are relevant for propositional attitudes such as memory, perception, awareness, or emotion.

(56) Subjective Veridicality:

- (i.) A function  $F$  that takes a proposition  $p$  as its argument is veridical with respect to an individual anchor  $i$  and an information state  $M(i)$  iff  $Fp$  entails  $p$  in  $M(i)$ .



(ii.)  $Fp$  entails  $p$  in  $M(i)$  iff  $\forall w' [w' \in M(i) \rightarrow w' \in \{w'' \mid p(w'')\}]$ .

☞ an instance of epistemic commitment to  $\neg p$  as *subjective antiveridicality* (GM 2021:pp. 61, (27)).

- (57) Epistemic commitment of  $i$  to a  $\neg p$  (Subjective Antiveridicality):  
An individual anchor  $i$  is committed to  $\neg p$  iff  $M(i)$  contains worlds compatible with what  $i$  knows and  $M(i)$  entails  $\neg p$ .

☞ the typology of modal spaces, i.e. sets of worlds (GM 2021:pp. 64, (38))

- (58) Veridicality of modal spaces and veridical commitment:
- A veridical modal space is fully committed:  
 $\forall w' [w' \in M(i) \rightarrow w' \in \{w'' \mid p(w'')\}]$ .  
(For short, in the lexical entries:  $\forall w' \in M(i) : p(w')$ ).
  - An antiveridical modal space is fully committed:  
 $\forall w' [w' \in M(i) \rightarrow w' \in \{w'' \mid \neg p(w'')\}]$ .  
(For short, in the lexical entries:  $\forall w' \in M(i) : \neg p(w')$ ).
  - A nonveridical modal space is weakly committed:  
 $\exists w' [w' \in M(i) \wedge w' \in \{w'' \mid p(w'')\}] \& \exists w''' [w''' \in M(i) \wedge w''' \in \{w'''' \mid \neg p(w''')\}]$ .  
(For short, in the lexical entries:  $\exists w' \in M(i) : p(w') \& \exists w'' \in M(i) : \neg p(w'')$ )

☞ GM (2021:pp. 65, (39)) treats informative content of a homogenous veridical or antiveridical state as a move by  $i$  to add the  $p$  or  $\neg p$  to the common or private grounds, which is termed “full informativity.”

- (59) Fact. Homogeneity and informativity: Homogeneous (veridical and antiveridical) information states are fully informative.

☞ GM (2018): homogeneity in veridical and nonveridical modal spaces:

- (60) Veridical and nonveridical modal spaces, homogeneity:
- A modal space  $M$  is veridical with respect to a proposition  $p$  iff it is positively homogeneous:  $\forall w' \in M : p(w')$ .
  - A modal space  $M$  is nonveridical with respect to a proposition  $p$  iff it is nonhomogeneous:  $\exists w' \in M : p(w') \wedge \exists w'' \in M : \neg p(w'')$ .
  - A modal space  $M$  is antiveridical with respect to a proposition  $p$  iff it is negatively homogeneous:  $\forall w' \in M : \neg p(w')$ .

#### 4.2. Main Proposal: Semantics of MN

I. To capture the meaning of MN under the classical analysis of solipsistic belief, we adopt Giannakidou and Mari’s (2021) notion of Dox (doxastic state), to be distinguished from the purely epistemic space of knowledge,  $M$  (GM 2012:pp. 153, (16)).

- (61) Belief state of an individual anchor  $i$ :  
A belief state  $\text{Dox}(i)$  is a set of worlds associated with an individual  $i$  representing worlds compatible with what  $i$  believes.

II. For solipsistic denial like MN, we propose that the truth condition requires only *subjective antiveridicality*, i.e. **Dox(speaker) entails  $\neg p$  while ignoring Dox(hearer)**, the actual world, or the common ground. Instead, we argue that Dox(speaker) may concern *epistemic, doxastic, or buletic commitment* of the individual anchor  $i$ , typically the speaker, depending on what she thinks (she knows), believes, or hopes based on her own subjective viewpoint:

- (62) **MN as Subjective Antiveridicality**  
Epistemic, doxastic, or buletic commitment of  $i$  to a  $\neg p$ :

An individual anchor  $i$  is committed to  $\neg p$  iff  $\text{Dox}(i)$  contains worlds compatible with what  $i$  thinks, believes, or hopes, and  $\text{Dox}(i)$  entails  $\neg p$ .

III. MN as a solipsistic denial denotes *antiveridical commitment states*, as shown in the following truth condition of a sentence with MN.

- (63) **MN as Solipsistic Denial**  
 a. John is cool, my eye!  
 b. [[John is cool, my eye]] is true in  $w$  iff  
 $\forall w' [w' \in \text{Dox}(\text{speaker}) \rightarrow w' \in \{w'' \mid \neg p(w'')\}]$

IV. MN can be construed as *private denial*, hence performs antiveridical *update to the private ground or mental states*, while regular negation updates the common ground (in the sense of Farkas 2003, GM 2018):

- (64) **MN as Private Denial**  
 Add  $\neg p$  to the local  $\text{Dox}(\text{speaker})$ .

V. Unlike regular descriptive negation, MN cannot access, hence cannot update, the common ground. Thus, the addition of  $\neg p$  to the local  $\text{Dox}(\text{speaker})$  narrows down the worlds in those private spaces by interaction.

- (65) **MN: update of private space**  
 $\text{Dox}(\text{speaker}) + \neg p = \{w' \text{ in } \text{Dox} \text{ where } \neg p \text{ is true}\}$

VI. This implies that MN as private denial anchors  $\neg p$  to the speaker's private space, hence MN performs a *context shift*.

## 5. Expressive dimension of MN markers

≠ MN is a part of our grammatical system for conveying the broad spectrum in the mode of divergence between the speaker of previous utterance and the speaker of current MN utterance toward the object. We propose that we are endowed with the following *four modes of divergence*, realized in four types of MN:

- (67) *Four modes of divergence*  
 (i) regular mode of divergence with neutral EI of [-1,1] (e.g. appropriateness assessment MN),  
 (ii) emphatic mode of divergence with neutral to weak negative EI of [-1,0] (e.g. emphatic denial MN)  
 (iii) negative emphatic mode of divergence with strong negative EI of [-1,-.5] (e.g. negative emphatic denial MN), and  
 (iv) sarcastic/ironic mode of divergence with weak to strong negative EI of [-1,0]~[-1,-.5] (e.g. irony MN)

Table 3. Four types of MN and six classes of MN markers

Four types of MN	Expressive Index (EI)	Six classes of MN markers	Semantic sources for MN markers	MN markers
Type I: <i>Appropriateness assessment MN</i>	[-1,1] (neutral attitude)	Class 1	<i>inappropriateness</i>	<i>Regular negation markers: not, no way, nonsense in English, ci an 'not' in Korean, etc.</i>
Type II: <i>Emphatic denial MN</i>	[-1,1]~[-1,0] (neutral to weak negative attitude)	Class 2	<i>unusuality or abnormality</i>	<i>Emphatics markers: lá 'lit. there' in European Portuguese, nyani 'exclamative' in Korean, etc.</i>
	[-1,1]~[-1,0]	Class 3	<i>unidentifiability-driven</i>	<i>Anti-specificity markers:</i>

	(neutral to weak negative attitude)		<i>non-referentiality</i>	<i>part:indef</i> ‘someone or other’ in ASL, <i>mwusun</i> ‘which’ in Korean, etc.
Type III: <i>Negative emphatic denial MN</i>	[-1,-.5] (strong negative attitude)	Class 4	<i>non-existence-driven non-referentiality</i>	<i>Non-existence markers:</i> <i>nothing</i> in English, <i>nada</i> in Spanish, <i>kayppwul</i> ‘dog’s horn’ in Korean, etc.
	[-1,-.5] (strong negative attitude)	Class 5	<i>unworthiness or dispreference</i>	<i>Depreciative markers:</i> <i>the hell, like hell, X shma/schma-X</i> in English, <i>wuskiney</i> ‘laughable,’ <i>elecwuk-ul</i> ‘freezing to death’ in Korean, etc.
Type IV: <i>Inory MN</i>	[-1,1]~[-1,-.5] (weak to strong negative attitude (depending on MN markers or paralinguistic factors))	Class 6	<i>insincerity</i>	<i>Irony markers:</i> <i>yeah right</i> in English, <i>coahaney</i> ‘oh, you’d like that, do ya?’, <i>nolkoissney</i> ‘oh, you’re enjoying that, aren’t ya?’ in Korean, <i>tu parles!</i> ‘go on (lit. you speak)’ in French, etc.

## 6. Conclusion

- We offer a semantic analysis of MN in language as indicators of *subjective mode in the negation*.
- We show that, as a marker of subjective negation, MN expresses **objection to a proposition** (*U*: previous utterance), often with the speaker’s **neutral to negative emotional stance**.
- We make four points regarding MN. First, we compare MN with other metalinguistic phenomena, especially metalinguistic comparatives (MC), and show that, MN can be categorized into **four types**: (i) appropriateness assessment MN, (ii) emphatic denial MN, (iii) negative emphatic denial MN, and (iv) irony MN.
- Second, we show how MN differs from regular negation (*objective anti-veridicality*), suggesting that MN markers come from *six semantic sources* depending on the speaker’s personal reason to object.
- Third, we propose the semantics of MN, as *subjective anti-veridicality* (à la Giannakidou and Mari 2021) in *doxastic models of mental representation* that may concern the speaker’s own epistemic, doxastic, or buletic commitment, regardless of actuality.
- Finally, we propose that most cases of MN, except for appropriateness assessment MN, can be understood as a subspecies of subjective mode phenomena, akin to other perspectival phenomena such as **expressives** (Potts 2005).

# 유사도 지표의 평가 실무 활용 가능성에 대하여

최지수

한국언어정보학회

한양대학교

# 정량 평가의 필요성

- 외국어 번역의 교육적 활용, 범용화
- 인증제 시험의 증가
- 미디어, 게임, SNS, 플랫폼 팬 번역 증가
- 신뢰도(reliability) 확보 근거 제시의 필요성
- 다양한 전산적 평가 메트릭 활용 연구 이어짐

# 유사도 측정 지표(metric)

- TF-IDF – 빈도 기반
- BLEUScore / METEORscore etc. –엔그램 기반
- BERTscore – PLM(pre-trained language model) 기반
- BLEURT – 회귀 (regression-based) 기반 metric
- 기타 F1score 등

<수식 1> BLEU의 수식

$$BLEU = \exp\left(\sum_{n=1}^N w_n \log p_n\right) \times BrevityPenalty$$

# BLEUScore

- 엔그램 기반 어휘, 구, 통사 단위의 형태 비교 평가 가능
- 평가 bias 없음
- 자연어 텍스트에도 실험 진행 有(Vela et al. 2014; Chung 2020; Han and Lu 2023; Lu and Han 2023)

## 한국어 자연어 텍스트 실험:

→ emBLEU, emMETEOR(개별 임베딩 실시), METEOR, BERT 등

→ 주안점: 인간과 기계의 언어 생성 차이

→ 타당도 제고 효과 있음, 실무 사용에 대한 관심 제고, 그러나 효율성 저하, BERT는 평가 bias 우려, 간편성 우려

# SacreBLEU

1. BLEUScore의 데이터 축적 → 비교의 기준으로 사용
2. But, BLEUScore를 그대로 쓰지 못하는 이유
  1. 이론이 실무와 다름
    1. Reference text가 이론적으로는 세 개씩 필요 → 실무에서는 한 개
    2. 텍스트 길이 비교 무의미
    3. 최대 엔그램 수 설정 무의미
    4. 스무딩(smoothing) 기능 설정 무의미
  2. 데이터 전처리 표준화 부족 → 비교 결과의 왜곡 / low source 언어에서 심화
    1. 토큰화/형태소 분리(rich-text → rich-text / rich, -, text)
    2. OOV, UNK 처리



# SacreBLEU의 자연어 텍스트 평가 활용

- 임베딩이 반드시 자연어의 정교한 비교를 보장하지는 않음
- PLM 기반 지표가 엔그램 기반 지표에 비해 나은 평가 성능을 보장하지는 않음
- 간편한 활용이 가능해야 함(오픈소스, 전처리 제공 등)
- 타 언어 및 타 실험과의 비교가 가능해야 함

# 코퍼스 수집 및 인간 평가 실험

- 한국어 번역 코퍼스 (target language만이 유의미)
- 난이도: CEFR(유럽언어공통참조기준) 중급(B1~B2)
- 텍스트: 비문학 텍스트 원문 총 11개, 평균 200단어/텍스트 내외
- 참여자: 독일어 전공 대학생 4학년 1명 및 대학원생 9명 (→ 생성된 번역 텍스트 총 110개)
- 평가자: 번역 실무 평가 경력 2년 이상, 번역 경력 10년 이상 교강사 3명, 본 실험 평가 reference는 그중 한 명

인간 평가 점수 ← 상관도 → *SacreBLEU score*

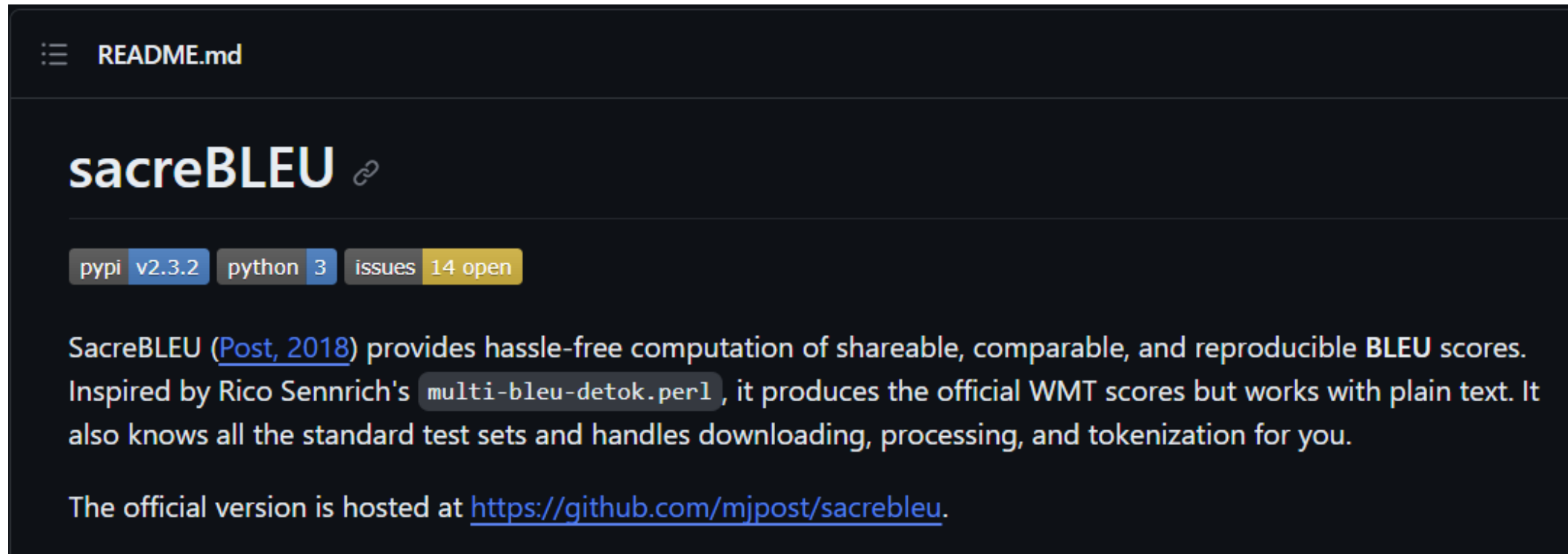
# 코퍼스 수집 및 인간 평가 실험

평가 95% 평가항목에 따른 정량평가 + 5% 정성적 평가(텍스트 단위 논리성, 가독성 등)

<표> 평가항목 감점표 (한국외대 번역평가인증 연구팀 2016:26)

항목	설명	감점
구, 절 정확성	구와 절이 정확하고 적절히 번역되었는가?	20
어휘 정확성	어휘가 정확하고 적절히 번역되었는가?	15
문장, 단락 정확성	문장과 단락이 정확하고 적절히 번역되었는가?	10
문법 정확성	문법에 맞는 문장을 사용하였는가?	3
맞춤법 정확성	맞춤법이 정확한가?	2
누락 또는 첨가 여부	오류가 분명한 누락 또는 첨가가 있는가?	2

# Post(2018)



The screenshot shows the README.md file for the sacreBLEU project. At the top left, there is a hamburger menu icon and the text "README.md". Below this, the project name "sacreBLEU" is displayed in a large, bold font, followed by a small link icon. Underneath the project name, there are three colored boxes: a blue box with "pypi v2.3.2", a blue box with "python 3", and a yellow box with "issues 14 open". The main text of the README describes the project's purpose: "SacreBLEU (Post, 2018) provides hassle-free computation of shareable, comparable, and reproducible BLEU scores. Inspired by Rico Sennrich's multi-bleu-detok.perl, it produces the official WMT scores but works with plain text. It also knows all the standard test sets and handles downloading, processing, and tokenization for you." At the bottom, it states "The official version is hosted at <https://github.com/mjpost/sacrebleu>."

<https://github.com/mjpost/sacrebleu>

# Post(2018)

## Metric: sacrebleu

SacreBLEU provides hassle-free computation of shareable, comparable, and reproducible BLEU scores. Inspired by Rico Sennrich's `multi-bleu-detok.perl`, it produces the official WMT scores but works with plain text. It also knows all the standard test sets and handles downloading, processing, and tokenization for you.

See the [\[README.md\]](#) file at <https://github.com/mjpost/sacreBLEU> for more information.

If this is a text-based metric, make sure to wrap you input in double quotes. Alternatively you can use a JSON-formatted list as input.

<https://huggingface.co/spaces/evaluate-metric/sacrebleu>

# 개별 문장 scoring 결과 예시

(1)Reference : 이 영화는 오늘 개봉했는데, 정말 재미있고 즐겁습니다.

(2)Candidate 1: 이 영화는 오늘 개봉했는데, 정말 재미있고 즐겁습니다. (100점)

(3)Candidate 2: 그 영화는 오늘 개봉했는데, 정말 재미있고 즐겁습니다. (73점)

(4)Candidate 3: 그 영화는 내일 개봉할 예정인데, 정말 별로이고 즐겁지 않습니다. (9점)

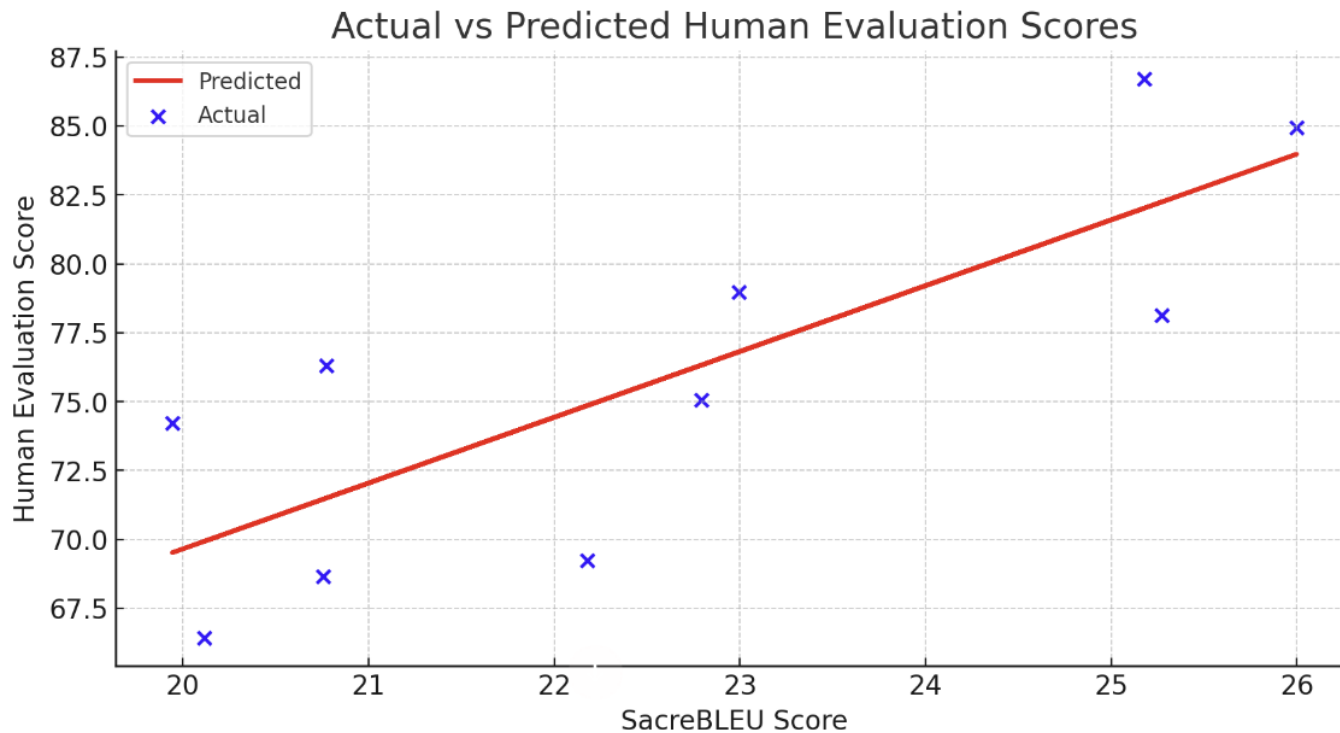
토큰화되기 전: 그 영화는 개봉했는데, 정말 재미있고 즐겁습니다.

토큰화된 후 : 그/영화/는/개봉/했/는데/,/정말/재미/있고/즐겁/습니다/.

# 개별 문장 scoring 결과 예시

```
references_kreg2 = [  
  [  
    "이 영화는 오늘 개봉했는데, 정말 재미있고 즐겁습니다."  
  ]  
]  
  
predictions_kreg2 = [  
  "그 영화는 개봉했는데, 정말 재미있고 즐겁습니다."  
]  
  
metric.compute(predictions=predictions_kreg2, references=references_kreg2)  
  
{'score': 73.24967962619755,  
  'counts': [12, 10, 8, 7],  
  'totals': [13, 12, 11, 10],  
  'precisions': [92.3076923076923, 83.33333333333333, 72.72727272727272,  
  'bp': 0.925961078642316,  
  'sys_len': 13,  
  'ref_len': 14}
```

# 인간평가 결과와 SacreBLEU score



- 상관도: 약 0.803
- 강한 양의 상관관계

(회귀계수: 0.646)



# 사용의 간편성

```
1 !python.exe -m pip install --upgrade pip
2 !pip install transformers
3 !pip install transformers[sentencepiece]
4 !pip install sacrebleu
5 !pip install sacrebleu[ko]
6 !pip install datasets
7 import transformers
8 import sacrebleu
9 from datasets import load_metric
10 metric = load_metric("sacrebleu")
11 references_ref = [{"(text)"}]
12 predictions_stu1 = [{"(text)"}]
13 metric.compute(predictions=predictions_stu1, references=references_ref, tokenize="ko-mecab")
```

# 결론 및 논의

- 임베딩 없는 엔그램 기반 형태 비교의 방식을 정량평가에 사용할 가능성 확인
- 인원 수가 적을 경우 등수 확인 필요
- 대량 번역 평가의 한 도구로 활용되는 논의에 대한 하나의 방법
- 요약이나 작문 평가에서는 논의 필요
- 추가 논의 가능성: 기계번역기의 사용성 평가 (동일 기준으로 비교)

stu1	stu2	stu3	stu4	stu5	stu6	stu7	stu8	stu9	stu10	google	papago	kakao	deepl
5	1	12	3	9	11	4	2	7	6	14	10	13	8
5	1	10	3	8	9	4	2	7	6				
6	2	7	1	5	10	3	4	9	8				
75.046	84.944	74.218	86.703	76.295	66.422	78.978	78.120	68.658	69.232				
22.808	26.006	19.899	25.005	20.447	19.999	22.796	25.130	20.867	22.224	17.487	20.091	18.371	20.455
3047, 1442, 6: 3053, 1508, 7: 2918, 1291, 5: 3162, 1540, 7: 2945, 1309, 5: 2962, 1330, 5: 2949, 1395, 6: 3044, 1468, 7: 2903, 1313, 5: 3038, 1396, 6: 2488, 1108, 2610, 1261, 2514, 1159, 2616, 1223, 4314, 4313, 4: 4065, 4064, 4: 4163, 4162, 4: 4280, 4279, 4: 4059, 4058, 4: 4273, 4272, 4: 3882, 3881, 3: 4137, 4136, 4: 3951, 3950, 3: 4199, 4198, 4: 3289, 3288, 3371, 3370, 3281, 3280, 3456, 3455, 70.631													
70.631	75.105	70.094	73.879	72.555	69.319	75.966	73.580	73.475	72.351	75.646	77.425	76.623	75.694
33.434	37.106	31.019	35.990	32.257	31.133	35.944	35.493	33.241	33.254	33.698	37.418	35.335	35.398
15.584	18.262	12.833	17.228	13.557	13.299	16.649	17.340	14.535	14.772	14.968	17.275	15.919	16.850
7.353	9.995	5.673	8.534	6.164	5.574	8.018	9.119	6.687	6.864	7.182	8.432	7.840	8.891
1.000	0.974	0.998	1.000	0.972	1.000	0.928	0.991	0.945	1.000	0.764	0.788	0.762	0.813

# Reference

## 주요 저널

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

Post, M. (2018). A Call for Clarity in Reporting BLEU Scores. In \_Proceedings of the Third Conference on Machine Translation: Research Papers\_ (pp. 186-191).

한국외대 번역평가인증 연구팀(2016). 번역인증제도(실무편). 한국외국어대학교 통번역연구소 학술대회. 23-35.

## 웹페이지

- <https://huggingface.co/spaces/evaluate-metric/sacrebleu>
- <https://github.com/mjpost/sacrebleu#downloading-test-sets>
- <https://eunjeon.blogspot.com/2013/02/blog-post.html>

# 한국어 기초 어휘 선정과 타당성 검증

2023. 11. 18.

봉미경, 김선혜, 강범일, 김한샘(연세대)

1

## 연구 목적

### 1. 연구의 목적 및 필요성

- ✓ 1~3등급까지 구축된 기초 어휘 목록 선정 방법론 소개
- ✓ 1~3등급까지 구축된 기초 어휘 목록에 대한 타당성 검증
  - 기초 어휘 목록 선정 시 사용된 여러 지표들의 타당성 검증
  - 각 등급별 어휘의 교과서 텍스트 점유율을 통한 타당성 검증
- ✓ 기초 어휘 목록 타당성 제고를 위한 제언

2

## 기초 어휘의 개념 및 범위

### ✓ 기초 어휘의 개념 및 선정 단위 확정

- 기초 어휘 : **한국어를 사용하여 일상 언어생활을 하는 데 꼭 필요한 어휘**의 목록.
- 국민의 수준과 필요에 맞는 범용적 목적의 어휘 목록을 선정.
- 단위를 **단어**로 선정하고 문법소(조사, 어미), 접사, 구 이상의 단위는 제외.

### ✓ 1, 2, 3등급 기초 어휘의 범위 및 기준

- **1등급** 기초 어휘 : 전학령기(~만6세)에 습득, 평생에 걸쳐 일상 언어생활에서 사용하는 어휘.
- **2등급** 기초 어휘 : 초등학생 1, 2학년(만 7~8세)에 습득, 평생에 걸쳐 일상 언어생활에서 사용하는 어휘.
- **3등급** 기초 어휘 : 초등학생 3, 4학년(만 9~10세)에 습득, 평생에 걸쳐 일상 언어생활에서 사용하는 어휘.

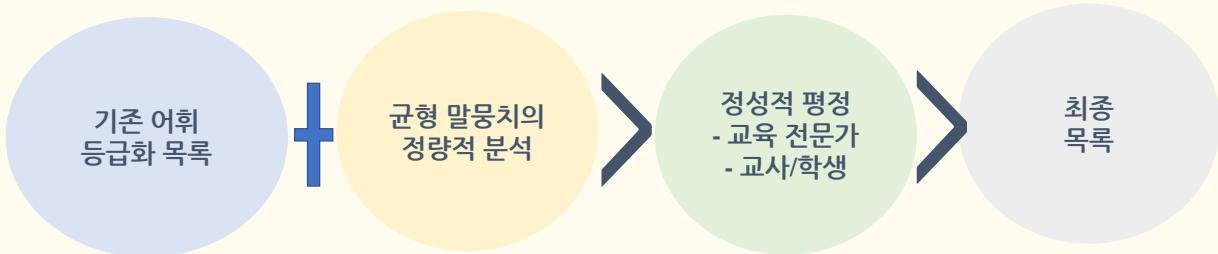
3

## 기초 어휘 등급표

연령(만)	학년	어휘 수	누적 어휘 수	즉시 활용 가능한, 보유 어휘 목록	등급
4-5세	어린이집	2,000	2,000	'연세초등국어사전' 중요 어휘 2,000	1
5-6세	유치원	3,000	5,000	'A Frequency Dictionary of Korean' 문구어 고빈도 어휘 5,000 '한국어 교육을 위한 기초 어휘' 5,000	
7세	초1	1,000	6,000		
8세	초2	1,500	7,500	'한국어 기본어휘 의미 빈도 사전' 기본어휘 표제어 목록 7,000	2
9세	초3	2,500	10,000		3
10세	초4	3,000	13,000		
11세	초5	4,500	17,500		4
12세	초6	5,500	23,000		
13세	중1	5,500	28,000		
14세	중2	5,500	34,000	'초등 어휘에 대한 초기 등급화 데이터' 34,000	5
15세	중3	6,000	40,000	'연세초등국어사전' 표제어 목록 40,000	

4

## 기초 어휘 목록 선정 과정



5

## 기초 어휘 목록 선정 활용 자료

✓ 등급화를 위한	목록 또는 사업명	규모
한국어 어휘	한국어 교육을 위한 기초 어휘 (서상규 외, 1998)	기본어휘 후보 목록 5,000개
목록 자료 수집	연세 한국어 사전 표제어 목록 (1998)	약 52,000개
및 정비	현대 한국어의 어휘 빈도 (서상규, 1998)	64,666만 개
	등급별 국어 교육용 어휘 (김광해, 2003)	1~4등급 33,805개
	외국인을 위한 한국어 학습 사전 (서상규 외, 2004~2006)	약 5,000개
	한국어 기본어휘 의미 빈도 사전 (서상규, 2014)	약 7,000개
	한국어 구어 빈도 사전 (서상규, 2015)	33,701개
	A Frequency Dictionary of Korean (이선희·장석배·서상규, 2016, Routledge, UK)	5000개
	연세초등국어사전 제4판 (연세대 언어정보연구원, 2020)	약 4만 개
	교육과정평가원 <초등 어휘에 대한 초기 등급화 데이터 구축 위탁 연구 (2020)>	약 34,000개
	교육부 정책 과제 <초등 교육용 어휘 목록 개발 (2021)>	약 20,000개

6

## 대상 말뭉치: 2억 어절 '균형' 말뭉치의 구성

매체	장르	세부 장르 유형(18개)	1940	1950	1960	1970	1980	1990	2000	2010	2020	합계	세부 장르 비율	매체 장르 비율		
문어	신문	비상상	1885023	4568780	4846812	4787811	4760957	4753483	4049101	2911859	0	32563826	16.3%	16.3%		
		비상상	0	397756	806063	1722077	1204255	1733296	3038564	0	0	8902011	4.5%	4.5%		
		비상상	280000	640000	2350000	2350000	2350000	2400000	4145000	485000	0	15000000	7.5%	7.5%		
	잡지	비상상	113360	159320	1184145	948315	691032	4633581	2775	245275	571655	0	8549458	4.3%	4.3%	
		비상상	245000	955000	2230000	2230000	2230000	2335000	2965000	1390000	0	14580000	7.3%	11.6%		
		비상상	0	362317	147039	416993	74304	1966214	2394854	0	0	5361721	2.7%			
	책	비상상 (발달단계)	교과서	0	0	0	0	0	0	0	0	0	0	1457300	0.7%	
			교육자료	0	0	0	38186	185872	1233242	0	0	0	0	1497655	0.7%	
			아동도서	0	0	19823	126194	499826	851812	0	0	0	0	1497655	0.7%	
			비출판물(0-7세)	0	0	0	0	0	0	484	24932	7124	32540	0.0%		
	비상상 (발달단계)	비출판물(8-13세)	0	0	0	0	0	218	2185	116151	104493	223047	0.1%			
		초등작문	0	0	0	0	0	0	0	641147	0	641147	0.3%			
		중국어-유아,소아	0	0	0	0	0	0	0	192106	0	192106	0.1%			
중국어-성인		0	0	0	0	0	0	3950075	5924981	0	9875056	4.9%	4.9%			
메신저		0	0	0	0	0	0	0	1077789	0	1077789	0.5%	0.5%			
구어	대화	공적대화	200321	202742	200159	199933	200829	201170	197245	84261255	199991	85863645	42.9%			
		사적대화	0	0	0	0	0	0	0	8711096	3042245	11753341	5.9%			
	독백	사적독백	0	0	0	0	0	0	13479	2289860	0	2303339	1.2%			
		사적독백	0	0	0	0	0	0	126138	0	0	126138	0.1%	50.0%		
연대별 합계			2723704	7285915	11784041	12819509	12197075	20108016	20884900	108271451	39255508	200000119	100.0%			
연대별 비율			1.4%	3.6%	5.9%	6.4%	6.1%	10.1%	10.4%	54.1%	2.0%	100.0%				
			23.4%									76.6%				

- 매체간 균형을 맞춤(구어의 비율 최대) 확대: 문어: 구어 = 5:5)으로써, **일상 구어에서 자주 쓰이는** 말을 반영함
- 장르간 균형 고려(18개): 이전 연구에서의 지나친 장르 세분화와 불균형의 문제를 해소, **추기 및 전기를 수필에 통합.**
- 발달 단계 자료: 일부 비출판물 추가 수집 및 포함. **2.6% → 4.7%로 확대. 어린이아들도 아닌 어휘를 포함함**
- 최대한 많은 텍스트 포함을 위한 표본 추출(‘상상’과 ‘비상상-학습/교양’의 경우 5,000어절씩)을 통해서, **두루 많이 쓰이는** 말을 찾아냄
- 시기적 균형: 비교적 일관된 어문 규범을 고려해 **1945년 이후 자료로 구성**하되, **연대별로 자료가 충분한** 신문, 잡지, 소설, 학술 및 교양, 교과서, 국회의원록 등은 최대한 고른 분포로 구성함으로써, **오랜 동안 쓰여온** 말을 찾아냄

7

## 정량 분석에 사용된 지표

- ✓ 기초 어휘 점수 지표(2018, 2022)
  - $0.265 \times \text{빈도} + 0.64 \times \text{범위} + 0.095 \times \text{산포도}$
- ✓ 평균 축소 빈도(ARF) 지표 산출 및 분석

8

## 정량 분석

### 비율 편차(Deviation of Proportions, DP) 지표

- ✓ 산포도 지표의 일종으로, 말뭉치의 각 장르에서 단어의 기대 분포와 실제 분포를 비교한 값
- ✓ 0에서 1 사이의 값을 가지며, 0은 완벽하게 균등한 분포, 1은 매우 불균등한 분포를 나타냄
- ✓ 기존의 산포도 지표는 이전 과제 팀에서 사용한 것으로, 말뭉치 구성의 차이로 인해 현재 사용 중인 말뭉치에 적용했을 때, 단어들의 산포도 편차가 거의 없는 것으로 나타나는 문제가 있었음
- ✓ 이에 비해 DP 지표는 단어 간 산포도 편차가 크게 나타나 기초성 변별에 보다 유용할 것으로 예상됨
- ✓ 참고문헌
  - Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–37.

## 정성 평정

- ✓ 전문가 평정
  - 전문가 평정은 다시 어휘 전문가 평정과 교육 전문가 평정으로 나뉨.
  - 전문가 평정:
    - 선정/삭제에 대한 기준 및 지침 작성
    - 사용자 평정 오리엔테이션에 활용할 수 있는 가이드라인 작성
    - 어휘장을 고려한 보충 어휘 목록 작성
    - 평정 후보 목록에서 제외된 어휘 검토



## 정성 평정

### ✓ 사용자 조사(교수 학습 현장 설문)

- 사용자 조사는 실제 이 등급에 해당하는 어휘를 사용하는 학생(유치원생~초등학교 4학년생)과 그 학생을 가르치는 교사를 대상으로 함.
- 사용자 어휘 조사의 대상이 되는 초등학생은 해당 등급과 연령에 맞는 학생들로 구성하되, 되도록 다양한 학교와 지역에 속한 학생들로 함.
- 1등급 사용자 평정 대상 어휘 : 약 1,580여 개 (1등급 전체 어휘 수 5,000개)
- 2등급 사용자 평정 대상 어휘 : 약 2,000여 개 (2등급 전체 어휘 수 2,500개)
- 3등급 사용자 평정 대상 어휘 : 약 3,500여 개 (3등급 전체 어휘 수 5,500개)

11

## 한국어 기초 어휘 등급화 연구

### 정성적 평정의 의의

✓ 1등급 어휘 5000개 중 정성적 평정을 통해 선정된 어휘(약 3100개)와 2억 어절 말뭉치 기초 어휘 점수 순위의 비교 결과

- 5000위 이내인 어휘 약 1300개
- 5000위 이상인 어휘 약 1800개

어휘 점수 구간(순위)	정성적 평정 어휘 수
5001~6000	547
6001~7000	582
7001~8000	185
8001~9000	205
9001~10000	231
10001~20000	35
20001~50000	22
50001~100000	9
100001~200000	9
합계	1825

12

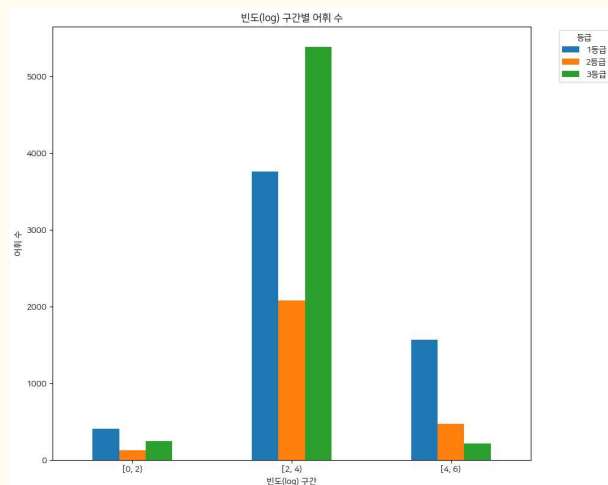
### 정성적 평정의 의의

- ✓ 1등급 어휘 5000개 중 정성적 평정을 통해 선정된 어휘(약 3100개)와 2억 어절 말뭉치 기초 어휘 점수 순위의 비교 결과
  - 5000~7000위: 거북이, 나뭇잎, 부끄러움, 불고기, 쌀쌀하다, 시커멓다, 요리사, 향아리
  - 7000~10000위: 그물, 금메달, 꽃잎, 삼겹살, 소매, 소방관, 쇼핑, 연못, 주황, 짜장면, 출근
  - 10000위 이상: 기지개, 메뚜기, 솜바꼭질, 오른발, 팽이, 횡단보도
  - → 말뭉치 구축을 위하여 수집한 문어, 구어 텍스트에 일상의 언어를 고르게 담은 것은 한계가 있으므로 정성적 평정으로 보완해야 함.

어휘 점수 구간(순위)	정성적 평정 어휘 수
5001~6000	547
6001~7000	582
7001~8000	185
8001~9000	205
9001~10000	231
10001~20000	35
20001~50000	22
50001~100000	9
100001~200000	9
합계	1825

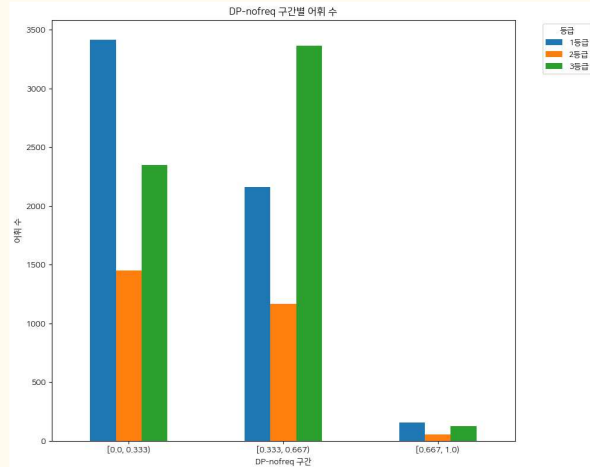
### 등급별 어휘의 정량적 검증

- ✓ 로그빈도
  - [0,2]: 빈도(log) 0이상 2 미만 → 저빈도 구간
  - [2,4] → 중빈도 구간
  - [4,6] → 고빈도 구간
- 판결(3등급): 로그빈도 4.223158
- 일찍(1등급): 로그빈도 1.477121



## 등급별 어휘의 정량적 검증

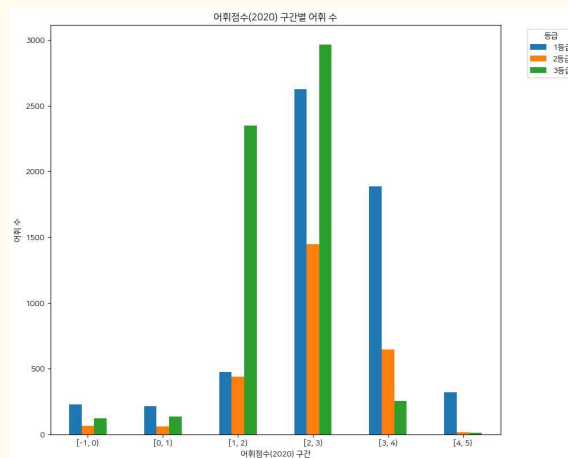
- ✓ DP-nofreq: 산포도(장르별 퍼짐 정도) 측정 지표
  - 이용 말뭉치에 18개 장르 존재. 빈도가 18인 단어가 각 장르에 한 번씩 출현했다면 이상적인 분포이며 DP-nofreq는 0이 됨. 반대로 크기가 가장 작은 장르에만 18번 출현했다면 DP-nofreq는 1에 가까워짐.
  - [0,0.333] → 대체로 균등한 분포



15

## 등급별 어휘의 정량적 검증

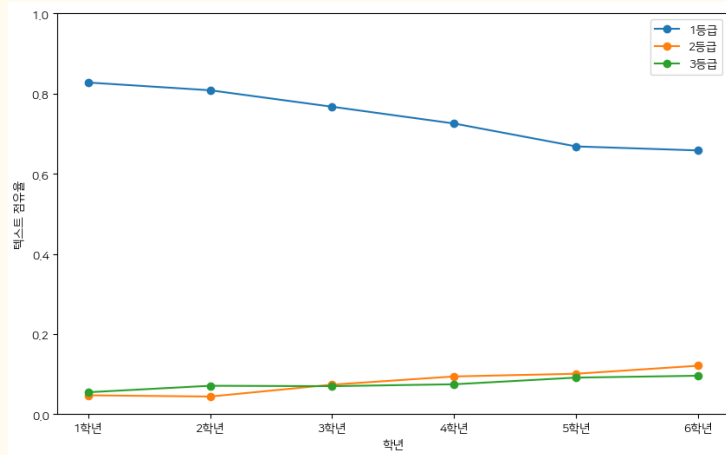
- ✓ 어휘점수(2020) 구간별 어휘 수



16

## 등급별 어휘의 정량적 검증

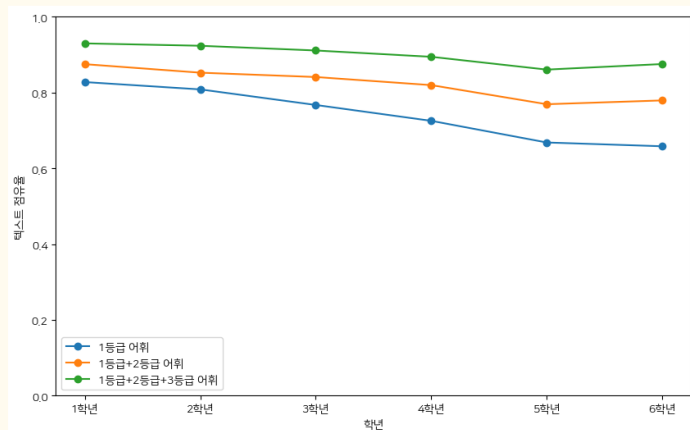
- ✓ 각 등급별 어휘의 초등학교 학년별 텍스트 점유율



17

## 등급별 어휘의 정량적 검증

- ✓ 누적 등급 어휘의 초등학교 학년별 텍스트 점유율



18



## 타당성 검증의 의의

- ✓ 기구축된 기초 어휘 목록의 타당성 제고
- ✓ 기초 어휘 선정 결과의 보완, 등급 재조정



## 참고문헌

- 강범일(2023), 어휘 중요도 측정을 위한 산포도 연구, <언어 사실과 관점> 59, 연세대학교 언어정보연구원, 9-33.
- 강현화(2014a), 한국어교육용 초급 어휘 선정 연구, <문법교육> 21, 한국문법교육학회, 1-26.
- 강현화(2014b), 한국어교육용 중급 어휘 선정에 대한 연구, <외국어로서의 한국어교육> 40, 연세대학교 언어연구교육원 한국어학당, 1-48.
- 강현화(2015), 한국어교육용 고급 어휘 선정에 대한 연구, <외국어로서의 한국어교육> 42, 연세대학교 언어연구교육원 한국어학당, 1-28.
- 김광해(2003), 등급별 국어 교육용 어휘, 박이경.
- 김한샘(2005), 현대 국어 사용 빈도 조사, 국립국어원.
- 김한샘(2009), 초등학교 교과서 어휘 조사 연구, 국립국어원.
- 봉미경(2021), 2015 개정 교육과정 초등학교 교과서 어휘의 계량적 연구, <언어 사실과 관점> 54, 연세대학교 언어정보연구원, 35-62.
- 서상규(2006), 한국어 학습 사전 편찬과 기본어휘의 선정을 위한 기초연구, <朝鮮語研究> 3, 朝鮮語研究會 編, くろしお出版.
- 양정실(2015), 초등학교 교과서의 어휘 실태 분석 연구, 한국교육과정평가원.
- 이기연(2015), 초등학교 교과서 수록 어휘 양상 연구, <학습자중심교과교육연구> 15(6), 학습자중심교과교육학회, 669-689.
- 이홍우(1994), 한국어 어휘 교육을 위한 대표 어휘 선정, <국어교육> 85, 국어교육학회, 19-39.
- 임지룡(1991), 국어의 기초어휘에 대한 연구, <국어교육연구> 23, 국어교육학회, 87-132.
- 임철성(2002), 초급 한국어교육용 어휘 선정 연구, <국어교육학연구> 14, 국어교육학회, 355-385.
- 장경희·이상형·이필영·김명희·김태경·김정선·전은진(2012), 초·중·고등학생의 구어 어휘 조사, 지식과 교양.
- 장현진·전희숙·신명선·김효정(2014), 초등학교 교육용 기초어휘 선정 연구: 저학년 중심으로, <언어치료연구> 23(1), 한국언어치료학회, 157-170.
- 조남호(2003), 한국어 학습용 어휘 선정 결과 보고서, 국립국어원.
- 최소영(2020), 초등학교 교육용 어휘 등급화와 등급 보정 절차 연구, <국어교육학연구> 4, 국어교육학회, 223-254.
- 황은하·서상규(2022), 한국어 기초어휘 선정을 위한 말뭉치의 설계와 구성, <언어사실과 관점> 55, 연세대학교 언어정보연구원, 5-33.
- 한국어능력 평가위원회(1997), 한국어 능력 평가용 기본어휘표.

# 한국어 술어은유처리에 관한 다층 연구

Surprisal과 뇌신경학적 증거를 중심으로

김건 & 남윤주

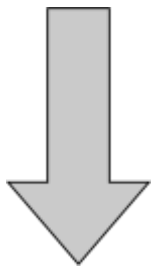
(한양대학교)

- ▶ 은유: 한 개념을 다른 개념에 빗대 이해, 설명, 표현하는 것
- ▶ 은유를 바라보는 상이한 관점:

## 전통적/화용론적 관점

Metaphor as figure of **speech**

창의적 발상을 이끄는 언어적 수사



간접 접근 모형 (Indirect Access Model)

## 인지언어학적 관점

Metaphor as figure of **thought**

창의적 발상의 언어적 실현



직접 접근 모형 (Direct Access Model)

▶ 은유적 표현의 처리 모형에 대한 가설

1) 간접 접근 모형 - Indirect Access Model; Standard Pragmatic View (Grice, 1975; Searle, 1979)

→ 은유 표현은 문자적 의미의 일탈이므로, 간접적으로 처리됨

2) 직접 접근 모형 - Direct Access Model (Glucksberg et al. 1982; Gibbs 1994)

→ 문자적 의미의 언어 표현과 은유적 표현은 상당한 방식으로 처리됨.

⇒ Lakoff(1993)의 '현대 은유 이론'은 어느 한 모델을 적극적으로 기각하거나 지지하지 않음.

3) 연속 처리 모형 (Continual Processing Models)

→ 처리 경로에 더해, 보다 다층적인 특성을 반영하여 은유 표현 처리를 해석.

→ 기존의 직/간접 접근 모델을 보완 및 통합하는 설명틀 정초.

→ 현저성 변화 가설, 은유의 생애 이론, 관련성 원리에 근거한 화용론적 해석이 이에 해당.



### 3-1) 적합성 원리(Relevance theory; Sperber & Wilson, 1986/1995)에 근거한 해석

→ 전통적/화용론적 관점 — 간접 접근 모형의 연장

→ 은유 표현은 높은 인지 효과(Cognitive effect)를 유발하며, 그에 대한 급부로 처리 과정에서의 인지적 부담 증가 (Noveck, Bianco & Castry, 2001)

→ 단, 인지 부담의 정도는 문맥에 따라 달라질 수 있음 (연속적)

### 3-2) 현저성 변화 모형 (Gradient Salience Model; Giora 2002)

→ 친숙성, 전형성, 빈번함, 원형성에 따라 결정되는 현저성에 의해 처리 속도가 달라짐.

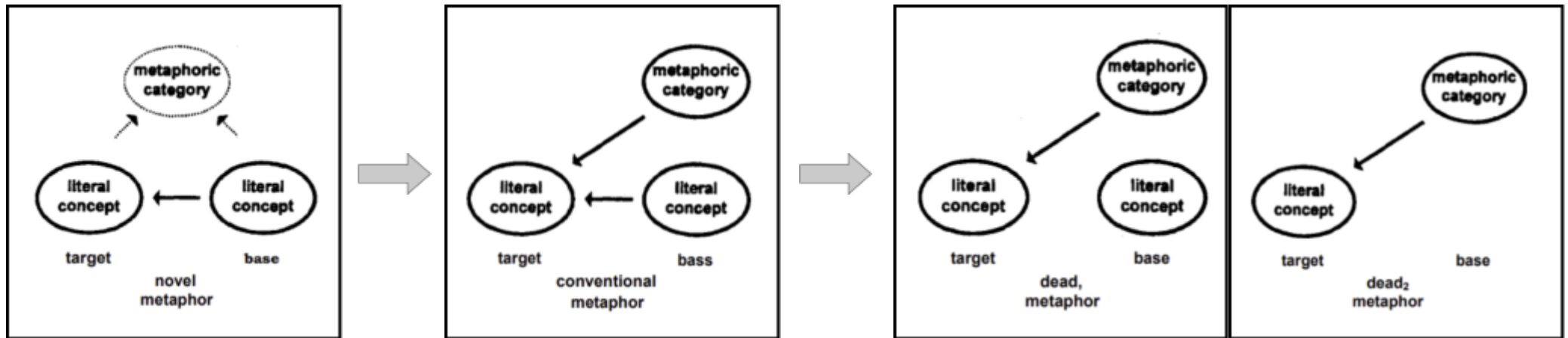
→ 전형적 은유 및 문자적 표현이 새로운 은유에 비해 처리 속도가 빨라질 수 있음.

→ 은유적 표현의 처리 또한 현저성에 근거함.

→ 직접 접근 모형 지지

3-3) 은유의 **생애** 모형 (Career of Metaphor; Bowdle & Gentner, 2005)

→ 은유 표현은 탄생한 뒤 점점 전형화 되면서 사유(dead metaphor, 死喻)에 이름.



→ 은유 표현이 전형화될 수록, 개념적 사상을 구성하는 범주 간의 연관 양상이 달라짐.

→ 전형적 은유가 가질 수 있는 처리의 수월성이 **처리 경로의 직접성에 기인한다고 보기 어려움.**

“ ... equating conventionality with directness of processing may be an oversimplification.”

→ 은유적 범주가 1) 보다 영역-일반적이고, 2) **더 적은 술어군**을 포함하고 있기 때문에

나타나는 수월성으로 이해. (cf. Conceptual Blending Theory; G. Fauconnier, 2001)

## Interim Summary

가설	은유 표현의 처리 경로	처리의 수월성
간접 접근 모형	간접적	문자적 표현 > 은유적 표현
직접 접근 모형	직접적	문자적 표현 ≒ 전형적 은유 > 새로운 은유
은유의 생애 모형	간접적 접근 모형 지지	문자적 표현 ≒ 전형적 은유 > 새로운 은유 혹은, 전형적 은유 > 문자적 표현 > 새로운 은유
현저성 변화 모형	직접적 접근 모형 지지	문자적 표현 ≒ 전형적 은유 > 새로운 은유

- ▶ 대다수의 은유처리연구는 반응/읽기 시간을 주요 척도로 삼아 이루어짐  
(Inhoff, Lima, & Carroll, 1984; Shinjo & Myers, 1987)
- ▶ 그러나 반응 시간은 인지 처리 양상을 오롯이 반영한다고 볼 수 없는 한계가 있음  
(see Rapp et al., 2004)
- ▶ 일부 연구의 경우, 반응 시간의 증가가 은유 처리에 따른 인지 부담에 기인하지 않고  
과제 특수적인 실험 설계에 강한 영향을 받았을 수 있음 (Carston & Yan, 2023)

Reaction time was defined as the time between the onset of the sentence and the button press of the subject. Mean reaction time was 2.30 s (S.D. 0.62) for the metaphoric sentences and 2.14 (S.D. 0.58) for the literal sentences ( $p=0.31$ ). Accuracy of response was identical for metaphoric and literal sentences (each 94.2%, S.D. 5.8).

Rapp et al. (2004)

- ▶ fMRI 연구: 은유 표현에 대한 우뇌 주도적 처리 양상을 보고한 연구들과 문자적 표현과 은유 표현 처리 시 동일한 좌뇌 편재성을 보고한 연구들이 혼재함

**Table 1**

Metaphor imaging literature. All known functional neuroimaging studies of metaphor processing published to date. Studies are listed with stimulus types grouped together, based on our reading of each paper. In addition, studies that found activation in right hemisphere areas are listed first (1–8).

Source	Experiment design	Stimuli	Main findings
1. Bottini et al. (1994)	$n = 6$ , PET, block design	<b>Complex sentences:</b> <i>novel</i> metaphors, uncommon literal sentences	<b>Right</b> frontal and temporal activations for metaphors compared to literal sentences
2. Sotillo et al. (2005)	$n = 24$ , ERP LORETA spatial analysis	<b>Complex/poetry style phrases:</b> Spanish <i>unfamiliar</i> metaphorical phrase followed by metaphorically related or unrelated word	Higher N400 activation for metaphoric words localized to <b>right</b> MTG and STG
3. Mashal et al. (2005)	$n = 15$ , 1.5 T fMRI block design, principal component analysis	<b>Word pairs:</b> literal relationship, conventional metaphorical relationship, or <i>novel</i> metaphor relationship	Novel metaphor processing: large network including left frontal and temporal areas, and <b>right</b> Wernicke's area, precuneus and insula
4. Mashal et al. (2007)	$n = 15$ , 1.5 T fMRI block design, ROI analysis	<b>Word pairs:</b> literal relationship, conventional metaphorical relationship, or <i>novel</i> metaphor relationship	<b>Bilateral</b> inferior frontal gyrus, middle frontal gyrus and superior temporal activations for metaphors, including <b>right</b> Wernicke's area
5. Arzouan, Goldstein, and Faust (2007)	$n = 29$ , ERP LORETA spatial analysis	<b>Word pairs:</b> literal relationship, conventional metaphorical relationship, or <i>novel</i> metaphor relationship	<b>Right</b> temporal and superior frontal involvement in novel metaphor processing
6. Pobric et al. (2008)	$n = 12$ , rTMS	<b>Word pairs:</b> literal relationship, conventional metaphorical relationship, or <i>novel</i> metaphor relationship	rTMS of <b>right</b> posterior superior temporal sulcus disrupted processing of novel but not conventional metaphors
7. Stringaris et al. (2006)	$n = 12$ , 1.5 T fMRI, event related, non-parametric analysis	<b>Simple sentence–word combinations:</b> conventional metaphors and literal sentences with related or unrelated words as targets	<b>Right</b> ventrolateral prefrontal cortex activation to target words in the metaphorical but not literal condition
8. Ahrens et al. (2007)	$n = 8$ , 1.5 T fMRI, block design	<b>Simple sentences:</b> Mandarin Chinese anomalous ( <i>novel</i> ) and conventional metaphors and literal sentences	<b>Bilateral</b> middle frontal gyrus and precentral gyrus, left inferior frontal gyrus and fusiform, <b>right</b> superior frontal gyrus activations for anomalous metaphors compared to literal
9. Stringaris et al. (2007)	$n = 11$ , 1.5 T fMRI, event related, non-parametric analysis	<b>Simple sentences:</b> <i>conventional</i> metaphors and literal sentences of the form 'Some X are Y'	Left inferior frontal gyrus and thalamus active for metaphors, no right activation specific to metaphors
10. Eviatar and Just, (2006)	$n = 16$ , 3 T fMRI event related, ROI	<b>Simple sentences:</b> <i>conventional</i> metaphoric, ironic and literal sentences, following two sentence 'story'	Metaphors recruited mainly LH areas: inferior frontal and inferior temporal
11. Rapp et al. (2004)	$n = 15$ , 1.5 T fMRI event related design	<b>Simple sentences:</b> German metaphors and literal sentences with the form an X is a Y, may have been moderately <i>familiar</i>	Virtually no right hemisphere activations when comparing metaphors to literal sentences or grey screen baseline
12. Rapp et al. (2007)	Reanalysis of Rapp et al., 2004 data. ROIs subjected to laterality index calculation	<b>Simple sentences:</b> German metaphors and literal sentences with the form an X is a Y, may have been moderately <i>familiar</i>	No differences in laterality patterns between literal and metaphorical sentences
13. Mashal, Faust, Hendler, and Jung-Beeman, (2008)	$n = 14$ , 1.5 T fMRI	<b>Sentences:</b> novel metaphors, literal sentences, and unfamiliar non-sensical sentences	Stronger left dorsolateral prefrontal and posterior middle temporal activation for novel metaphors than for literal and non-sensical sentences
14. Lee and Dapretto (2006)	$n = 12$ , 3 T fMRI	<b>Word triads:</b> literal relationship or <i>conventional</i> metaphorical relationship	Only left prefrontal and tempo-parietal activations, no right activations for metaphorical relationships

▶ 사건 관련 전위(Event-related Potential)를 활용한 은유 처리 관련 실험 연구

1) Lai et al. (2009)

→ 문자적 조건에 비해 전형적 은유, 새로운 은유 조건에서 N400 보고

→ 간접 접근 모형을 지지

2) Iakimova et al. (2005), De Grauwe et al. (2010) 등

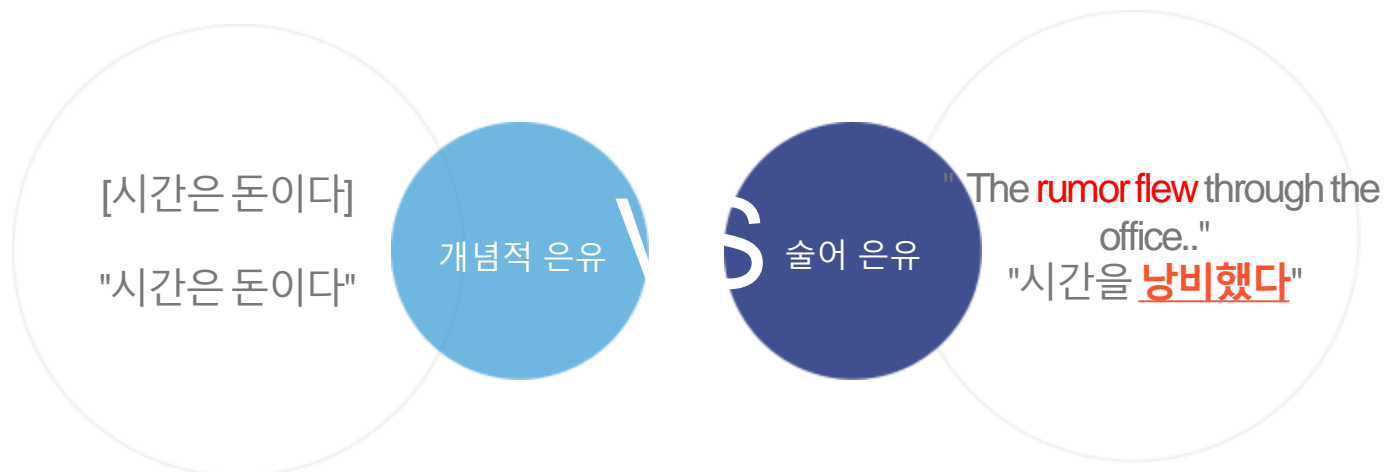
→ 은유 표현과 문자적 표현의 처리에 있어 차이가 없음을 보고

→ 직접 접근 모형을 지지

⇒ De Grauwe et al. (2010)는 "A is B"와 같은 개념적 은유 연구

## cf) 개념적 은유 vs. 술어 은유

- Utsumi & Sakamoto (2010): 일본어 술어 은유 연구
- 술어 은유의 경우 중간 개체를 거치는 추가적인 과정이 필요하며, 간접적 범주화 방식으로 처리가 진행 됨.
- 뇌파연구가 아닌, 오프라인 및 온라인 프라이밍 연구



▶ 사건 관련 전위(ERPs)를 활용한 한국어 술어 은유 처리 관련 실험 연구

실험 문장

조건	실험 예문					
	Word 1	Word 2	Word 3	Word 4	Noun	Target Verb
문자적 의미 (Lit)	시후가	미술시간에	준비물로	가져간	도화지가	구겨졌다.
전형적 은유 (Nov)	현기가	지금까지	끈질기게	지켜온	자존심이	구겨졌다.
새로운 은유 (Con)	은영이가	명상을	통해	가다듬은	정신이	구겨졌다.
이례적 의미 (Ano)	태현이가	학기내내	손꼽아	기다렸던	방학이	구겨졌다.

→ 조건 당 36개, 총 144개로 구성된 실험 문장에 더해 채움 문장 216개 활용.



## Pre-test

→ 실험에 참가하지 않은 피험자를 대상으로 의미 적합성 및 전형성 판단 과제 시행

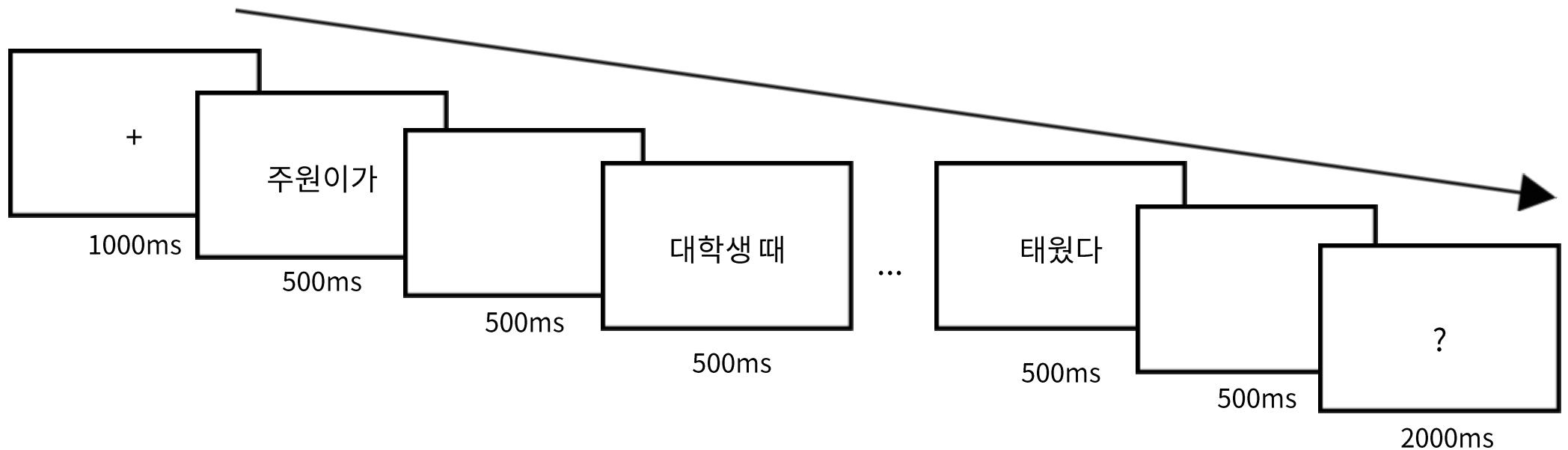
문자적 의미	전형적 은유	새로운 은유	이례적 의미
6.489	6.371	3.208	1.163

[표 2] 전형성 판단 결과 (아이템 평균값) (scale:1~7)

문자적 의미	전형적 은유	새로운 은유	이례적 의미
5.989	5.789	3.791	1.574

[표 3] 의미 적합성 판단 결과 (아이템 평균값) (scale:1~7)

## ERP 실험 절차



## 실험 참가자

→ 총 18명의 참가자 (남녀 각 9명, 평균 연령 28세) 중 14명을 분석 대상에 포함

## 실험 결과

	250-350ms	350-500ms	500-600ms	600-800ms
(1) <u>Nov.</u> (vs. Lit.)		N400 (300-400ms)	P600	—
(2) <u>Ano.</u> (vs. Lit./Con)		N400		P600
(3) <u>Lit.</u> (vs. Con.)	—	N400	—	—
(4) <u>Nov.</u> (vs. Con.)		N400	P600	—
(5) <u>Ano.</u> (vs. Nov.)		N400	—	P600

[표 4] 각 구간에서 검출된 ERP 결과

→ 1), 2), 4), 5) 의 결과는 직접 접근 모형 및 간접 접근 모형 모두 설명이 가능함.

→ 3) 문자적 조건(vs. 전형적 은유 조건)의 N400:  
 전형적 은유의 처리가 문자적 의미의 처리보다 더욱 수월했음을 의미.

## 함의

→ 전형적 은유 조건 대비 문자적 조건에서 검출된 N400은 은유의 직접 접근 모형으로만 설명할 수 있음 (직접 접근 모형을 지지)

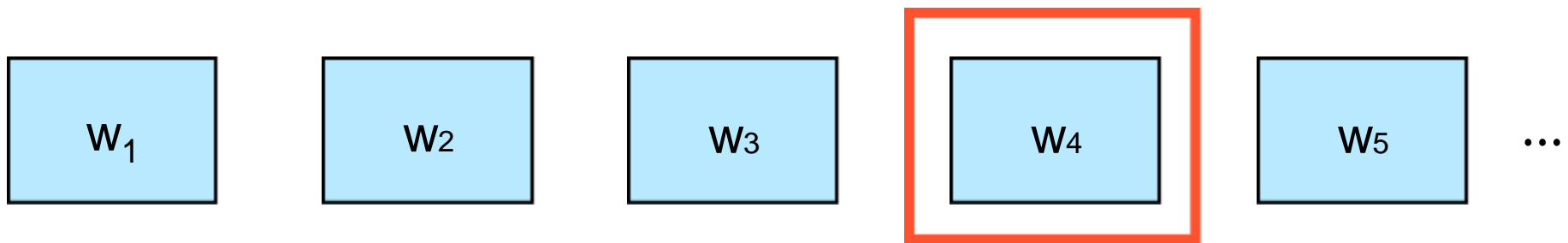
→ 그러나 전형적 은유의 상대적인 처리 수월성이 경로의 직접성에 기인하지 않을 수 있는 다음의 경우가 존재:

- 1) 은유의 생애 모형에 따라, 은유적 범주가 포함하는 술어군의 규모가 더 작았기 때문에 술어가 문장 전체의 불확실성에 미치는 영향이 감소했기 때문.
- 2) 본 실험에 활용된 문장이 문자적 조건에 비해 전형적 은유 조건에서 술어 부분에서 예측될 수 있는 불확실성이 감소했기 때문.

→ 정보 이론적 관점에서 surprisal은 이전에 제시된 맥락 정보를 근거로 특정 위치에서 출현한 정보가 얼마나 예상과 다르게 출현한 것인가를 보여줄 수 있는 지표 (송상헌, 2022).

$$\text{Surprisal}(x) = \log_2 \frac{1}{P(x \mid \text{context})}$$

→ 토큰(단어)  $x$ 가 시퀀스 내에서  $i$ 번째 토큰(단어)이라면, 첫 번째 토큰부터  $i - 1$ 번째 토큰까지의 맥락을 고려하였을  $x$ 가 나타날 조건부 확률의 음의 로그값



$$\text{Surprisal}(w_4) = -\log_2 P(w_4 \mid w_1, w_2, w_3)$$

- 문장 처리 과정에서 술어가 예측되는 정도를 정량적으로 제시할 수 있음.
- Surprisal과 N400 및 P600과 상관관계를 갖고 있음이 보고됨.  
(Frank et al., 2013; Michaelov et al., 2023)
- 실험에 사용된 문장의 술어를 '예측할 수 있는 정도'를 Surprisal을 통해 정량적으로 계산하여, 남윤주 & 정혜인(2016)이 보고한 전형적 은유 표현이 가지는 처리의 수월성이 상술한 두 가지 경우에 의해 나타났는지 검증.
- ⇒ 문자적 조건 vs. 전형적 은유 조건의 N400이 Surprisal의 차이에 기인했는지를 확인하고자 함.

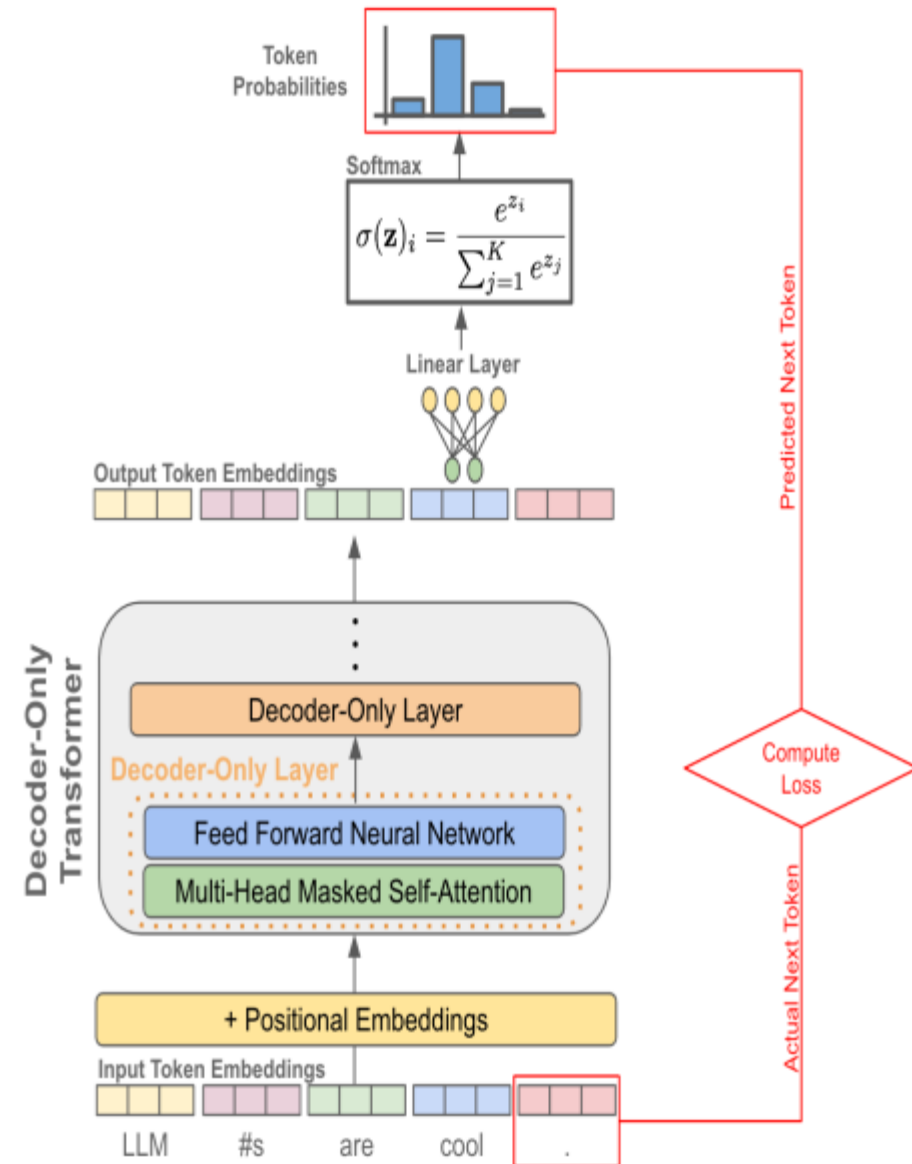
→ Transformer 기반의 GPT 계열 언어모델의 본질적 기능은 주어진 시퀀스를 기반으로 다음에 등장할 토큰을 예측하는 것.

→ GPT 계열 언어모델 중, 한국어로 사전 훈련을 마치고 한국어에 미세조정(fine-tuning)된 다음의 모델 활용: ① KoGPT-trinity, ② KoGPT-2

①의 경우 어간을 포함하고 있는 단일 토큰에 대한 Surprisal을 따로 검토한 분석도 포함.

(e. g. '세웠다'에서 '세웠'; '설계했다'에서 '설계')

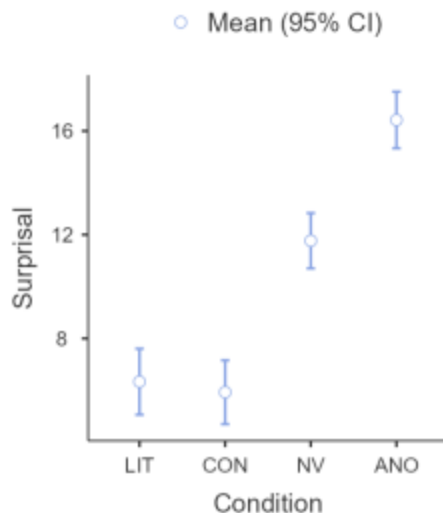
+ 토큰화 및 텍스트 처리, 토큰 예측 (훈련) 방식이 다른 BERT 계열의 KR-BERT의 결과 추가 검토.



## ① - 1. KoGPT-trinity (1.2B) - 어간을 포함하는 단일 토큰에 대한 surprisal 계산

- 문자적 조건과 전형적 은유 조건을 비교한 경우를 제외하고, 모든 조건간 통계적으로 유의한 차이 확인. (Kruskal-Wallis test, DSCF pairwise comparisons)
- 본 실험에서 확인된 문자적 조건 vs. 전형적 은유 조건의 N400이 Surprisal의 차이에 따라 발생하지 않았음을 확인.

Surprisal



Kruskal-Wallis

	$\chi^2$	df	p
Surprisal	87.76255	3	< .001

Group Descriptives

	Condition	N	Mean	SD	SE
Surprisal	LIT	36	6.33406	3.76028	0.62671
	CON	36	5.92772	3.63678	0.60613
	NV	36	11.76761	3.15124	0.52521
	ANO	36	16.42238	3.22239	0.53706

Dwass-Steel-Critchlow-Fligner pairwise comparisons

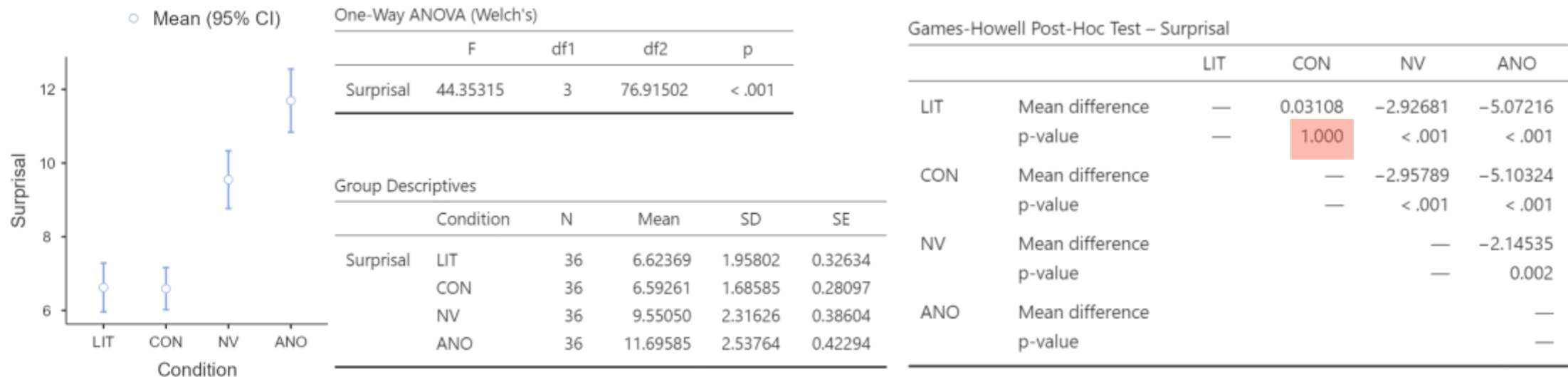
Pairwise comparisons - Surprisal

		W	p
LIT	CON	-0.52560	0.982
LIT	NV	7.35840	< .001
LIT	ANO	9.92270	< .001
CON	NV	7.74066	< .001
CON	ANO	9.84306	< .001
NV	ANO	7.10357	< .001



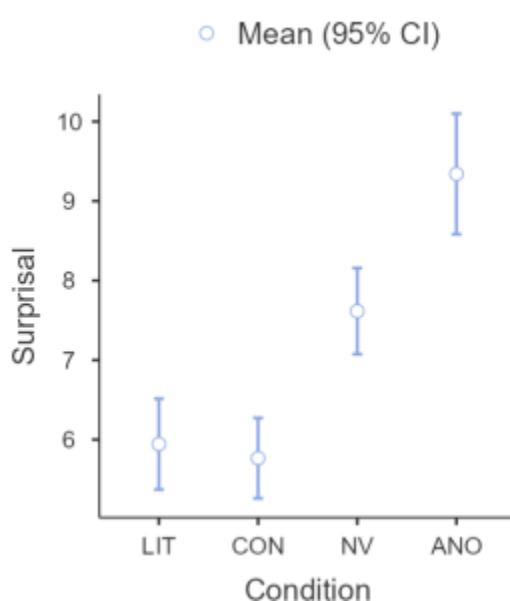
## ① -2. KoGPT-trinity (1.2B) - 목표 어휘 전체에 대한 surprisal 계산

- 어간 토큰에 대한 결과와 동일: 문자적 조건과 전형적 은유 조건간 술어의 Surprisal 차이 없음. (Welch's One-Way ANOVA, Games-Howell Post-Hoc Test)
- 목표 어휘가 여러 개의 subword로 토큰화 되는 경우, 각 토큰에 대한 surprisal을 계산 후 평균 계산.



## ② KoGPT-2 (125M) - 목표 어휘 전체에 대한 surprisal 계산

- 문자적 조건과 전형적 은유 조건간 술어의 Surprisal 차이 없음.  
(Kruskal-Wallis test, DSCF pairwise comparisons)



### Kruskal-Wallis

	$\chi^2$	df	p
Surprisal	63.27669	3	< .001

### Group Descriptives

	Condition	N	Mean	SD	SE
Surprisal	LIT	36	5.94347	1.68828	0.28138
	CON	36	5.76671	1.49453	0.24909
	NV	36	7.61544	1.60178	0.26696
	ANO	36	9.34078	2.24201	0.37367

### Dwass-Steel-Critchlow-Fligner pairwise comparisons

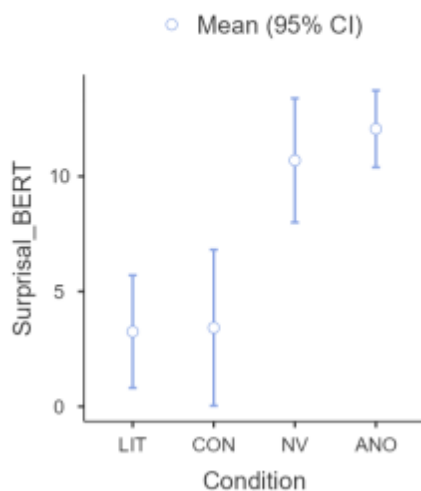
#### Pairwise comparisons - Surprisal

		W	p
LIT	CON	-0.30262	0.997
LIT	NV	5.78160	< .001
LIT	ANO	8.45739	< .001
CON	NV	6.57797	< .001
CON	ANO	8.80779	< .001
NV	ANO	5.49491	< .001

## 번외) KR-BERT

- MLM(Masked Language Model)인 BERT를 활용하기 위해, 실험 문장의 표적 어휘를 마스킹, 어휘집에 등록되어 있어 subword로 토큰화되지 않는 일부 술어에 대해 Surprisal 계산.
- 이규민 외(2021)에서 제시된 라이브러리 활용
- 술어에 대한 Surprisal 결과는 GPT 계열의 언어모델로 계산했을 때와 유사 (다만, 실험 문장의 25%만 검증했으므로 결론 도출은 불가능)

Surprisal\_BERT



Kruskal-Wallis

	$\chi^2$	df	p
Surprisal_BERT	19.74074	3	< .001

Group Descriptives

	C	N	Mean	SD	SE
A	LIT	9	3.25481	3.18484	1.06161
	CON	9	3.42348	4.41013	1.47004
	NV	9	10.69468	3.50873	1.16958
	ANO	9	12.06225	2.17067	0.72356

Dwass-Steel-Critchlow-Fligner pairwise comparisons

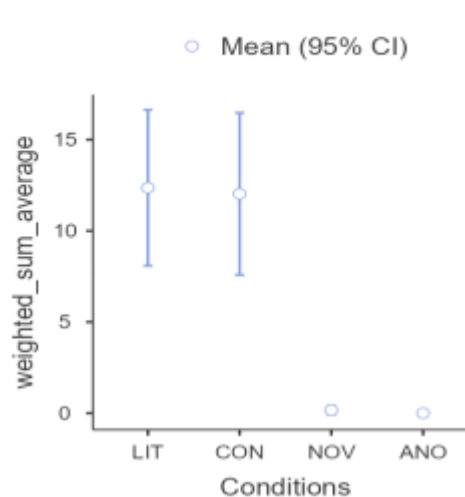
Pairwise comparisons - Surprisal\_BERT

		W	p
LIT	CON	-0.81171	0.940
LIT	NV	4.43317	0.009
LIT	ANO	4.80781	0.004
CON	NV	3.80878	0.036
CON	ANO	4.30829	0.012
NV	ANO	1.31122	0.791

## Cloze test를 통한 사후 검증

- 술어 위치의 후보군 규모에 따른 잠재적 불확실성을 확인하기 위해 빈 칸 채우기 과제를 활용, 피험자는 술어 위치에 들어갈 적절한 단어를 떠오르는 순서대로 기입.
- 최대 3순위 단어까지 분석, 떠오른 순서에 따라 가중치를 부여하여 술어별 출현 기대 정도를 점수로 표현, 실험 문장에 실제로 사용된 술어의 점수 확인.
- Lit vs. Con 조건의 차이가 확인되지 않았음.

weighted\_sum\_average



Group Descriptives

	Conditions	N	Mean	SD	SE
weighted_sum_average	LIT	36	12.36111	12.63590	2.10598
	CON	36	12.02778	13.14422	2.19070
	NOV	36	0.16667	0.69693	0.11616
	ANO	36	0.00000	0.00000	0.00000

→ 남윤주 & 정혜인(2016)에서 보고된 전형적 은유 표현에 비해 나타난 문자적 표현의 **N400은 Surprisal의 차이에 기인하지 않았음.**

→ 은유의 일생 모형에서 가정한 바와 달리, 남윤주 & 정혜인(2016)에서 전형적 은유 조건이 가지는 **처리의 수월성**은 예측 처리 과정에서 인지 부담을 야기할 수 있는 **불확실성의 감소에 기인하지 않았음.**

→ 요컨대 해당 연구의 결과는 직접 접근 모형으로만 설명할 수 있는 현상을 확인한 가운데 은유의 일생 모형의 설명을 기각하면서 **직접 접근 모형을 지지**하는 한편, 현저성(Saliency)를 정의하는 요소 가운데 전형성과 빈번함이 같은 조건에서도 처리의 수월성 차이가 나타남을 보고.

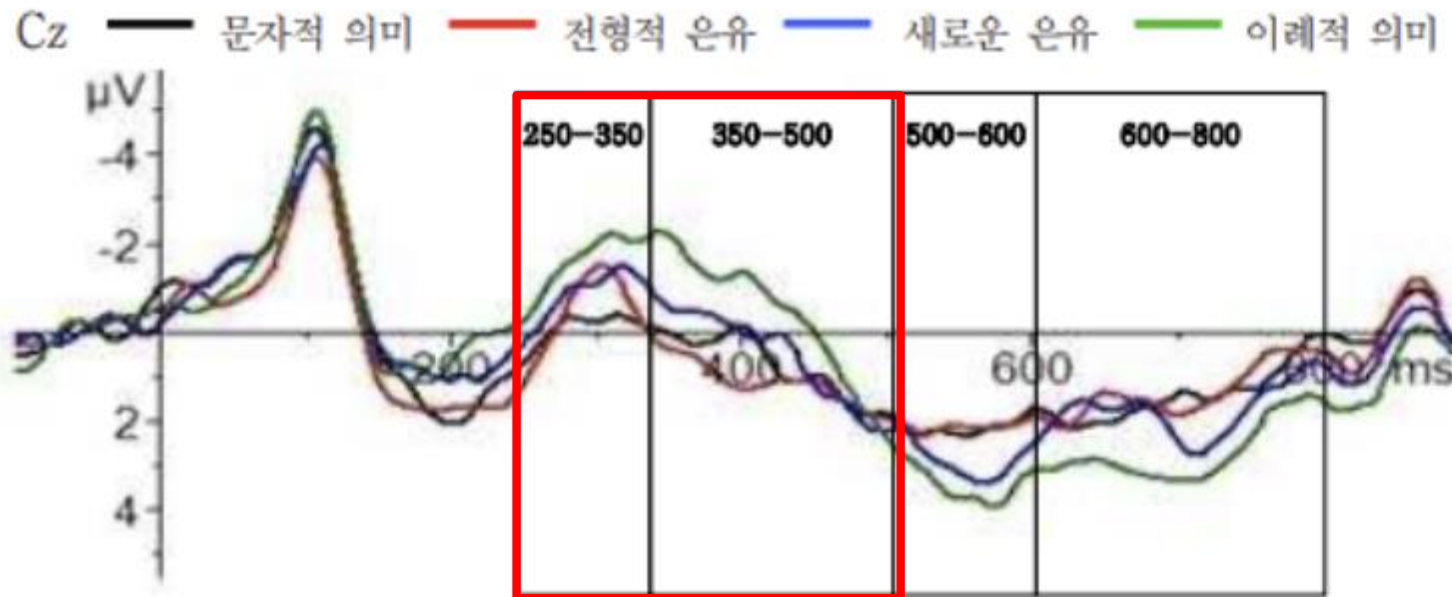
\* 신경학적 층위를 고려한 다층적 이론의 필요성 제기  
(see Neural Theory of Metaphor; Lakoff, 2009: 17)

"The neural theory of metaphor makes an important prediction in the case of **conventional conceptual metaphorical mappings that are realized by fixed brain circuitry**. (.....) conceptual metaphor should take no longer than normal frame-based nonmetaphorical processing."

→ ERP 피험자 부족 문제 해결: 추가 실험 및 분석 진행 (30명 이상)

→ Inspection time window 조정 : 250-500ms broad window

[그림 2] 목표어휘(동사)에 대한 주요 전극 ERP 파형 (위) 및 Cz 영역 분석 시간 구간 (아래)



→ 사건과 관련한 뇌 반응의 시간적 변화를 보여주는 ERP 외에도, 주파수 대역과 같은 뇌 활동의 질적 정보를 도출할 수 있는 Time-Frequency 분석 진행.

- 코퍼스 수준에서 실제로 전형적 은유 표현의 불확실성이 문자적 표현과 다르지 않은지 ER(Entropy-Reduction) 등의 Surprisal과 유사한 척도를 활용하여 검증이 필요.
- Surprisal은 맥락 정보에 따라 크게 변동할 수 있으므로, 계산에 적합한 문장 목록을 재구성 할 필요가 있음:
  - e. g. 1) 1k 이상의 중규모, 2) 대상 토큰을 최소대립쌍으로 갖는 문장
- 은유 표현 코퍼스 구축 (see CorMet; Mason, 2004), 술어의 빈도 등 예측 정도를 반영할 수 있는 척도 및 현저성을 구성하는 변인이 통제된 문장을 구성, 뇌신경학적 연구 진행.

감사합니다

QnA