

2023 한국언어학회 겨울 학술대회

발표논문 초록 모음집

- 주 제 : AI 시대의 언어 사용과 분석, 그리고 연구방법론
(Language Uses, Analyses, and Research Methodologies in the AI Era)
- 장 소 : 경희대학교 서울캠퍼스 Space 21
- 일 시 : 2023년 12월 9일(토) 9:30 ~ 17:50
- 주 최 : 한국언어학회
- 주 관 : 경희대학교 언어정보연구소



한국언어학회
The Linguistic Society of Korea

학술대회 조직위원회 및 프로그램 위원회

대회장

김종복(경희대)

조직 위원장

신효필(서울대)

조직 위원

류병래(충남대)

박시균(군산대)

박분주(대구가톨릭대)

문숙영(서울대)

이두원(한국교통대)

이주경(서울시립대)

박동우(한국방송통신대)

프로그램 위원장

권익수(한국외대)

송상헌(고려대)

프로그램 위원

윤수원(서울시립대)

오은정(상명대)

한수미(한림대)

장경철(부산대)

김태호(부산대)

최영주(조선대)

조혜선(단국대)

초대의 말씀

안녕하십니까? 어느덧 시간이 흘러 올해의 마지막 달이 되었습니다.

2023 년 한국언어학회 겨울 학술대회가 12 월 8 일(금)과 9 일(토)에 개최됩니다. 이번 학술대회는 9 일 경희대에서의 대면 정기학술대회와 함께 8 일 오후 고려대에서 신진학자 심포지엄으로 구성되어 있습니다.

이번 학술대회의 주제는 “AI 시대의 언어 사용과 분석, 그리고 연구방법론(Language Uses, Analyses, and Research Methodologies in the AI Era)”입니다. 최근 주목을 받고 있는 인공지능과 이를 활용하는 언어처리 및 생성 모델이 각광을 받고 있는 지금, 이번 학술대회가 다양한 언어 자료의 분석과 그를 위한 연구방법론에 대한 좋은 토론의 장이 될 것으로 기대됩니다.

이번 학술대회에는 기조강연을 포함하여 언어학 전반의 주제를 다루는 31 편의 논문이 발표될 예정입니다. 기조강연으로는 실험화용론(Experimental Pragmatics)의 선두 주자이신 CNRS & Université de Paris- Cité 의 Ira Noveck 교수님을 모셨습니다. 기조강연과 더불어 일본언어학회, 필리핀언어학회, 대만언어학회 대표인 Hiromu Sakai, Shirley Dita, Shu-ing Shyu 교수를 모시고 Asian Language 에 대한 특별세션도 준비하였습니다.

이번 학술대회 프로그램을 위해 가장 애써 주신 프로그램위원장이신 권익수 교수님과 송상헌 교수님, 그리고 논문 발표 및 사회를 맡아 수고하실 여러 회원님들에게도 깊은 감사의 말씀을 드립니다. 그리고 뒤에서 묵묵히 도움 주신 조직위원님과 운영위원님들에게도 이 자리를 빌려 감사의 말씀드립니다.

COVID-19 의 기세가 꺾이며 약 4 년 만에 대면으로 개최되는 학술대회가 성공적으로 마무리되기 위해서는 회원 선생님들의 적극적인 관심과 참여가 꼭 필요하다고 생각합니다. 동료 및 학생 분들께 이번 겨울 학술대회를 널리 알려주심과 더불어 참석하셔서 자리를 빛내 주시기를 부탁드립니다.

아무쪼록 이번 학술대회와 그 프로그램이 참여하시는 모든 선생님들께 유익하고 뜻깊은 만남의 장이 되기를 기원합니다.

한국언어학회 회장
김종복 드림

오시는 길

장소: 경희대학교 서울캠퍼스 Space 21 B101, B108, B109 호

[대중교통 이용시]

전철 이용시

수도권전철 회기역(1호선, 경의중앙선) 하차 후 1번출구에서 동대문 01번 버스 이용, 경희대의료원 정류장(종점)에 하차

KTX 이용시(용산역, 서울역, 청량리역):

각 역사에서 수도권전철 1호선(상행, 광운대/의정부/양주 방면)을 이용해 회기역(1호선, 경의중앙선) 하차

SRT 이용시(수서역):

수도권 전철 수인분당선 수서역에서 상행(상행, 일원/강남구청 방면)을 이용, 왕십리역에서 경의중앙선(상행, 청량리/상봉 방면)으로 환승 후 회기역(경의중앙선, 1호선) 하차

[주차안내]

- 주차장 위치는 ‘학술대회장 안내’의 [캠퍼스 지도] 항목을 참조해 주세요.
- 학회 당일(토요일)은 21시 이후 후문 출차가 불가하오니, 늦은 시간 출차 시 정문을 이용해주시기 바랍니다.
- 자세한 사항은 경희대학교 홈페이지를 참조해주세요.
주차관리실(푸른솔문화관 지하 2층, 02-961-0181)

경희대 주차안내



<https://bit.ly/3R3Uufw>

학술대회장 안내

[캠퍼스 지도]

장소: 경희대학교 서울캠퍼스 Space 21 B101, B108, B109 호



Space 21 은 언덕 아래(선승문 앞, B2 층), 중간(B1 층), 위(1 층)에 각각 입구가 있습니다.

- 언덕 위 입구를 이용하시는 경우, 언덕과 가까운 **왼쪽 건물의 입구**를 이용해 건물 내부로 진입, 계단을 한 층 내려오셔야 합니다.
- 언덕 중간의 입구를 이용하시는 경우, 지하 1 층 학술대회장으로 바로 진입하실 수 있습니다.
- 언덕 아래(선승문 앞) 입구를 이용하시는 경우, 건물 중앙 계단을 이용해 한 층 올라오셔야 합니다.



Conference Program

Conference Program

일시 2023년 12월 9일 토요일 09:30-17:50

Date December 9, 2023. Saturday 09:30-17:50

장소 경희대학교 Space21 일원

Venue Space 21 Room# B101, B108, B109. Kyung Hee University, Seoul.

발표장 제 1 발표장(Room 1): B101 호 ; 제 2 발표장(Room 2): B109 호; 제 3 발표장(Room 3): B108 호

오전 1부

9:30-9:40 제 1 발표장	개회사 김종복(학회장, 경희대) [Jong-Bok Kim (President, Kyung Hee Univ.)] 사 회 박동우(한국방송통신대) [Dongwoo Park (Secretary General, Korea National Open Univ.)]
Plenary Talk 제 1 발표장	
9:40-10:40	Isolating the pragmatic contribution of utterances and even words Ira Noveck (CNRS, Université de Paris-Cité) Moderator 윤수원(서울시립대) [Suwon Yoon, Univ. of Seoul]

오전 2부

세션/발표장	Special session Room 1 (B101) Moderator Joo-Kyeong Lee (Univ. of Seoul)	일반 세션 1 제 2 발표장 사 회 구경연(계명대)	일반 세션 2 제 3 발표장 사 회 김주애(한국외대)
10:50-11:20	Neural representation of semantic concepts revealed by MEG Hiromu Sakai (Waseda Univ.)	모국어와 제 2 언어가 제 3 언어의 통사 습득에 미치는 영향: 영어의 관계절과 <i>wh</i> -의문문 연구 박유순(인천대)	How does context window size address polysemy of adverbial postposition - (<i>u</i>)lo in Korean? 문성민(아주대)
11:20-11:50	Tagalog response system: issues and challenges Shirley Dita (De La Salle Univ.)	수일치 처리에 나타나는 영어학습자의 문장처리의 시선 추적양상 박분주(대구가톨릭대), 최성목(경북대)	인공지능의 한국어 접사 처리 연구 비립, 박서운, 강예지, 김한샘(연세대)
11:50- 12:20	Mandarin <i>Ye</i> : Additivity or scalarity? Shu-ing Shyu (National Sun Yat-sen Univ., Kaohsiung)	튀르키예 한국어 학습자의 '은/는' 사용 오류에 대하여 김현정(서울대)	Validating Python libraries for text analysis in language research 한수미, 김민지(한림대)

오후 1부

세션/발표장	일반 세션 3 제 1 발표장 사 회 최영주(조선대)	일반 세션 4 제 2 발표장 사 회 한수미(한림대)	일반 세션 5 제 3 발표장 사 회 이정훈(서강대)
13:40-14:10	영어 허사 <i>there</i> 구문의 표찰화 남소영(부산대)	뒤섞기 구문에 나타나는 복합명사구 제약 효과에 대한 한국어와 일본어 비교 실험 연구 발표 김일규(강원대), 손은채(한국외대), 지윤주(한국외대)	한·영 모순어법(<i>oxymoron</i>) 용례에 대한 인지의미론적 고찰 김정민(한국외대)
14:10-14:40	Transitive <i>out of -ing</i> construction: A collostructional and constructional approach 김정수(인천대), 심록(Univ. of South Carolina)	On the recovery of Korean null objects: Some challenges from experimental evidence 고희정(서울대), 김현우(연세대), 강초룡(서울대)	Exploring the social meaning of the usage of <i>yay /cyay /kyay</i> for inanimate referents in Korean through frameworks of Sociolinguistics and conceptual metaphor 이진석(가톨릭대)
14:40-15:10	Elliptical <i>why not?</i> : A discourse-based perspective 김옥기, 김종복(경희대)	Spoken acceptability judgment on Korean ditransitives: Comparisons with written judgment 박상희(한밭대), 이은경(이화여대)	Bridging the gap: A tale of two types of anaphoric nominals in Korean 이상근(고려대), 마정연(고려대), 김희경(Binghamton Univ.)



발표장 제 1 발표장(Room 1): B101 호 ; 제 2 발표장(Room 2): B109 호; 제 3 발표장(Room 3): B108 호

오후 2부			
세션/발표장	일반 세션 6 제 1 발표장 사회 고희정(서울대)	일반 세션 7 제 2 발표장 사회 박시균(군산대)	일반 세션 8 제 3 발표장 사회 남윤주(한양대)
15:20-15:50	Distinction between event and process for Metaphysical-Cum-Semantic Rules of the progressive	베트남 화자가 산출한 한국어 단모음에 대한 한국어 모어 화자의 지각연구	The language of modesty and confidence in doctoral dissertations: A corpus-based cross-disciplinary comparison between Korean L2 and English L1 writers
	강민지, 이예식(경북대)	이주희(경희대)	김희경(Binghamton Univ.), 이상근(고려대)
15:50-16:20	영어 전치사 to 에 대한 단상 (Notes on the English Preposition to)	일본인 한국어 학습자의 의문문 문말 억양에 나타나는 모어 간섭	Semantics of people and places: A corpus-assisted assisted NSM analysis of Tagalog affix <i>taga</i> -
	장경철(부산대)	우에다 아스카, 주현희(부경대)	Nicole Pontanilla, Eugene Chung (Korea Univ.)

오후 3부			
세션/발표장	일반 세션 9 제 1 발표장 사회 김옥기(경희대)	일반 세션 10 제 2 발표장 사회 조혜선(단국대)	일반 세션 11 제 3 발표장 사회 김한샘(연세대)
16:30-17:00	K-pop 외국 팬들의 한국어 외래어 사용 현상 연구 주어-동사	Acoustic and articulatory characteristics of the three diphthongal vowels of /we/*, /we/, /we/	드라마 자료를 통해 살펴본 {네}, {예}의 사용의 변이와 변화
	황경진(고려대)	이대용(한양대), 송지연(한양대), 김사향(홍익대), 조태홍(한양대)	장교진, 신지영(고려대)
17:00-17:30	On the lexical meaning of <i>seltukha</i> - 'persuade' in Korean	The Effect of age, morphological structure, and syllable-count on the realization of Korean nasal-liquid sequences	회사 내 호칭어 변화에 대한 연구: 1980 년대와 2020 년대 드라마 자료 분석을 바탕으로
	이주원(전주대)	심록(Univ. of South Carolina), Drew Crosby(고려대), D. Eric Holt(Univ. of South Carolina)	박지은, 신지영(고려대)

총회 및 폐회식(General Assembly & Closing Remarks, 제 1 발표장)	
17:30-17:50	폐회사 권익수(프로그램위원장, 한국외대)[Iksoo Kwon (Program Committee Chair, Hankuk Univ. of Foreign Studies)] 사회 박동우(한국방송통신대)[Dongwoo Park (Secretary General, Korea National Open Univ.)]



2023 년 한국언어학회 Young Scholar Symposium

○ 일시 2023 년 12 월 8 일(금)
○ 주최 한국언어학회

○ 장소 고려대학교 문과대학 서관 일원
○ 후원 인문사회 디지털 융합인재양성사업단, 한국연구재단

14:00-15:20		특 강 장소: 서관 314B 호
오은정 교수님(상명대)	언어학 전공 학생을 위한 논문 작성 및 투고 요령	
15:30-17:15		포스터 발표 장소: 서관 132 호
김선재(고려대)	감성분석 알고리즘을 이용한 한국어 문장완성검사 진단 프로그램의 개발	
김예윤(고려대)	문학 번역은 어떻게 문화적 렌즈를 초월하는가 - 색채어를 중심으로	
김지혜(경희대)	Not that-clauses with polarity-sensitive items: A corpus-based perspective	
김형선(한양대)	Evidence of rapid automatic translation in Korean-English bilinguals using masked implicit priming: An ERP study	
김희수, 김병현(고려대)	Two Kings' Speeches	
김희연(경희대)	Coordinated multiple wh-sluiting: A corpus-based investigation	
노채원(고려대)	Teacher boredom and student motivation in Korean EFL context: Student boredom as a mediator	
리서 빅터 존(인천대)	Multimodality and its Influence on Language Change: A Study on Memes	
문경민(세종대)	Speech intention variations in elderly speakers' interrogatives: A Speech Corpus Study	
박유빈, 윤수연(충남대)	한국어 반모음 w 첨가에 대한 음운론적 고찰	
박윤규(경희대)	Quantified fragments in English: A corpus investigation	
배주현(고려대)	The role of Out-of-Class Learning in L2 willingness to communicate: Perceptions of English as an international language as a mediator	
세터레 누리(서울대)	번역 교육	
안예진, 신지영(고려대)	한국어 2 인 대화의 말차례 조직 양상	
양희욱(서울대)	한국어 예외적 격표시 구문에서의 격 교체에 관하여: '이/가'와 '은/는'을 중심으로	
오서영(서울시립대)	Detecting sarcasm in Amazon reviews	
오재원(서울대)	An experimental study on processing ambiguous verb phrases	
원종빈, 권은재, 한선아(고려대)	Viability of Korean visual abductive reasoning	
유정호, 김나연(성균관대)	Argument status and semantic interference	
윤현섭(한국외대)	Metaphor in mind and machine: Unraveling conceptual metaphors for NLP applications	
이도영, 강아름(충남대)	SNS 코퍼스에 기반한 신어 추출	
이지영, 김나연(성균관대)	Understanding the influence of local coherence on the reanalysis process of temporally ambiguous noun phrase: The role of animacy	
이환희(서울대)	The syntax of desiderative constructions: The case of Korean - <i>Ko siph</i> -	
정안순(Univ. of Edinburgh)	CP selection at the Syntax-Semantics Interface: A case study in Mandarin think "xiang"	
차현지, 이용훈(충남대)	단편소설 속 일반명사의 사용 양상 연구: 이태준의 작품을 중심으로	
최재원, 김보경(한국외대)	신조어 분석 및 카테고리 분류	
추다연(한국외대)	페르시아어 색채어의 문화언어학적 분석	
한송, 류병래(충남대)	코퍼스분석을 통한 구별모작 『한 스펀의 시간』과 『상아의 문으로』의 비교	
한예영, 이주경(서울시립대)	The phonetic status of word-initial / ɾ/ in Korean borrowing words	
황인슬(고려대)	Attitude toward foreign accent, intelligibility and comprehensibility	



한국언어학회
The Linguistic Society of Korea



Plenary Talk

Isolating the pragmatic contribution of utterances and even words

Ira Noveck
(CNRS, Université de Paris-Cité)

Based on prior work on what are often considered to be conversational implicatures (e.g. on scalar implicatures and irony), one can apply a general rule of thumb concerning the processing of pragmatic information, which goes as follows: an utterance that generates multiple effects (to produce an enriching pragmatic inference) requires more cognitive effort (in the form of time, developmental changes or supplementary ERP profiles) than an utterance that generates fewer. For example, an ironic reading of a sentence generally takes more time to process than a sincere one; a reading that appears to incorporate a scalar implicature (e.g. when *Some* is understood to mean *Some but not all*), generally speaking, takes more effort to process than one without it (where *Some* is understood to mean *Some and perhaps all*). In more recent work, my collaborators and I have considered cases in which pragmatic procedures are part and parcel of lexical meanings. With the expectation that the general rule of thumb should hold more generally, I will present results showing that words with pragmatic import prompt slowdowns when compared to controls that are presumed to be less pragmatically rich. I will talk about (at least) two cases. One concerns the added pragmatic meaning of discourse connectives such as *but* and *so* when compared to the truth-functional *and*. The other concerns pragmatically rich response particles, such as *Si* in French and *doch* in German which become available after a negative question or assertion. These particles not only de-negate the negative proposition but re-assert an implicit affirmative. So, if one asks, in French, *Biden n'est pas président des États-Unis?* (*Biden is not president of the US?*), it would be appropriate to reply *Si*. This (type of) response in an experiment takes longer to produce than a *Oui* or *Non* (when these are the appropriate answer to control questions). There is currently no equivalent of this response particle in many languages, including English and Quebecois.

References

- Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge University Press.
- Noveck, I., Petit, N., Tian, Y., & Turco, G. (2021). Revealing pragmatic processes through a one-word answer: When the French reply *Si*. *Journal of Memory and Language*, 120, 104245.
- Ronderos, C. R., & Noveck, I. (2023). Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. *Cognition*, 238, 105480.



Special Session

Neural representation of semantic concepts revealed by MEG

Hiromu Sakai
(Waseda University)

Introduction: Previous research on the neurobiology of language revealed that the network in the dominant hemisphere plays a central role in human language processing. Relatively less is known about how each semantic concept is represented in this network. Using electrophysiological measures such as EEG and MEG, previous studies showed that words or phrases can be decoded from the signals measured during language comprehension (Chan et al., 2011; Dash, Ferrari, and Wang, 2020). Such results indicate that semantic concepts are represented in the brain as neural activities. However, not much research was conducted about how semantic concepts are represented during language production. In this study, we decoded semantic concepts from MEG signals recorded when the participants orally named visually presented images of familiar objects.

Methods: Seven native Japanese speakers are asked to (i) see the images of six objects, (ii) imagine the action related to the objects, (iii) silently name the objects, and (iv) orally pronounce the objects' names. MEG signals were recorded from 64 channels of SQUID sensors during these tasks using a superconducting self-shield MEG (Narasaki and Tsunematsu, 2020).

Results: A time-resolved multivariate pattern analysis (tmVPA) was conducted using a sci-kit learn machine learning toolkit. The result indicates significantly higher decoding accuracy above the chance level for (i) and (iv) in the time windows between 200 ms to 600 ms from the stimuli onsets. The results suggest that semantic concepts of each object are represented as neural activities during these time windows.

Tagalog response system: Issues and challenges

Shirley N. Dita
(De La Salle University)

Asking questions and answering them is a vital component of human communication. Polarity questions are a common type of questions that typically demand an answer which is either a confirmation or a denial. In this paper, I will present some of the issues in answering polar questions in Tagalog in different contexts. Further, I will discuss the usual response strategies to polar questions such as the yes-no system, the agree-disagree system, and the echo system.

Mandarin *Ye*: Additivity or Scalarity?

Shu-ing Shyu

(National Sun Yat-sen University, Kaohsiung)

It has been widely acknowledged that Mandarin utilizes various means to express the meaning of EVEN, including the adverbs like *shenzhi*, *jinran*, *jurán*, and syntactically composed *lian...dou* ‘including...all’ or *lian...ye* ‘including...also’ constructions. The syntax and semantics research has paid much attention to the focus nature of the (displaced) *lian*-phrase and its contribution to the scalar implicature presupposition. Another widely discussed issue concerns the quantification nature of *dou*. While *dou* in *lian...dou* is often likened to the regular *dou* in terms of universal quantification operator (Paris 1979; Shyu 1995, 2014; Yuan 2004; among many others) or distributive operator (cf. Lin 1998), it is further characterized as a maximality operator (as in Giannakidou & Cheng (2006), and Xiang (2008)), or a scalar operator. Moreover, since *dou* and *ye* are often interchangeable in *lian...dou/ye* construction, they are often treated alike (Paris 1979; Shyu 1995, 2014, “prametric *dou/ye*” in Hole 2004; etc.). However, questions arise as to whether *dou* and *ye* are in fact synonymies in *lian* sentences. Most of the studies converge on the basic additivity nature of *ye* even in *lian*-sentences (*leitong fuci* in Ba 2012; Ma 1982, 2014; Niu and Luo 2022; Zhou 2009, etc.), whereas others contend for a distinct category for the “scalar *ye*” (Hole 2017; Yang 2018). This paper first aims to elucidate the nature of *ye*, specifically in *lian*-sentences, wh-indefinites, and *wulun* clauses.

- (1) Lian ta ye/dou lai le.
‘Even (s)he has come.’
- (2) Wǒ wúlùn tí shénme tiáojiàn, tā dou/*yě dāying.
‘No matter what conditions I bring up, he will not agree.’
- (3) Wǒmen shénme dǐxì dou/*yězhīdao!
‘We don’t know any exact detail!’

I argue that *ye* be better treated as an additive focus adverb, unlike *dou* in the above cases which requires to be maximally quantified. In addition, the scalar meaning in *lian...ye* sentences is either compositionally derived in syntax or contextually determined, cf. Shyu (2016, 2018). It further accounts for cases in which *ye* is allowed, such as total negation (negative polarity), and sentences containing modals.

The current study bears on issues including articulating the distinction between *dou* and *ye*, and it sheds further light on the comparisons with other Chinese additive particles (*yi* 亦, *hai* 還).



Abstracts

모국어와 제 2 언어가 제 3 언어의 통사습득에 미치는 영향: 영어의 관계절과 Wh-의문문 연구

박유순
(인천대학교)

본 연구는 한국에 거주 중인 중도입국 청소년을 대상으로 그들이 경험한 언어가 제 3 언어인 영어의 통사 습득에 미치는 영향을 영어의 관계절과 wh- 의문문을 통해 분석했다. 제 1 언어(L1)와 제 2 언어(L2)가 제 3 언어(L3)에 미치는 영향에 대해 '누적 향상 모델'과 'L2 지위 요소 모델'을 채택하여 중도입국 청소년들이 제 3 언어를 습득할 때 어느 모델이 영향을 끼치는지 알아본다. 누적 향상 모델(Cumulative Enhancement Model)은 이전에 경험한 문법적 특성이 다음 언어 습득에 중립적 또는 긍정적인 영향을 미친다는 가설이고 L2 지위 요소 모델(L2 Status Factor Model)은 L1 이 아닌 L2 만이 L3 습득에 긍정적 또는 부정적인 전이의 요인이 된다는 가설이다.

위와 같은 가설들을 실험하기 위해서 본 연구에서는 중도입국 청소년의 모국어(L1)와 한국어(L2), 영어(L3)의 통사 부분에서 유형론적인 유사성과 차 이점에 따라 아래와 같이 그룹지었다.

우선 관계절 실험에서 A 그룹은 L1 러시아어 =L2 한국어 =L3 영어 / B 그룹은 L1 중국어 = L2 한국어 = L3 영어로 구성되었으며, wh-의문문 실험에서 A 그룹은 L1 러시아어, 우크라이나어 =L2 한국어 =L3 영어 / B 그룹은 L1 중국어, 키르기스어 =L2 한국어 =L3 영어로 구성되었다.

각 실험 항목과 결과는 다음과 같다. 관계절 실험은 두 가지로 '듣고 따라 하기'와 '그림보고 단어 배치하기'를 실행했다. '듣고 따라 하기' 과업에서 A 그룹이 B 그룹보다 수행 능력이 높아 누적 향상 모델을 지지하는 것으로 나타났다. 핵방향 매개변수가 중요하게 작용하는지 제대로 알아보기 위한 '그림보고 단어 배치하기' 과업에서는 A와 B 두 그룹의 수행 능력이 비슷하게 나왔다. 이와 같은 결과를 세부적으로 살펴보니 공교육과 사교육에서 문법 교육을 받은 B 그룹이 A 그룹보다 오히려 핵방향성을 나타내는 문장에서 많은 오류를 나타냈다. 따라서 그림보고 단어 배치하기에서도 누적 향상 모델을 지지하는 것으로 판단할 수 있다.

Wh-의문문 실험에서 wh-어의 이동 여부에 대해 6 가지 패턴을 수준별로 살펴봤을 때, 완성된 문장인 Wh-Aux-S-V 패턴에서 동일한 L2 와 L3 수준을 가진 참가자들이 생성한 wh-의문문에서는 A 그룹이 B 그룹보다 수행능력이 훨씬 더 높았다. 이와 같은 결과는 L3 와 유사한 L1 wh-매개변수가 영어 의 문문을 습득하는데 촉진적인 역할을 한다는 것을 알 수 있다. 종합적으로 위에서 실험한 모든 과업에서 누적 향상 모델을 지지하는 것으로 나타났다.

주어-동사 수일치 처리에 나타나는 영어학습자의 문장처리의 시선추정양상

박분주 · 최성묵
(대구가톨릭대, 경북대)

문장을 구성하는 언어처리과정은 언어산출의 처리과정에서 개념적인 생각이 언어로 표현되기까지 비언어적인 개념(message)을 단어와 구를 구조적으로 연결하여 절을 형성하며, 시제와 수와 같은 형태적 특징들을 반영하여 절을 형성하게 된다. 모국어화자들 뿐아니라 영어학습자들의 주어-동사 수일치의 처리과정에 대해서는 오류분석에서부터 시작하여 많은 학자들이 관심을 가지고 연구를 이어오고 있다.

시선추적연구는 사람의 시선의 움직임을 통해 인간의 인지적 과정 및 정의적 반응을 추론할 수 있다는 가설(eye-mind hypothesis)에 기반하고 있다(Just & Carpenter, 1980). 이에, 제 2 언어 학습자가 영어텍스트를 읽는 동안, 다른 어휘들보다 특정 어휘를 더 오래 그리고 더 자주 시선을 고정(fixation) 시킨다는 것은 그 어휘를 처리하는데 더 많은 인지적 노력이 필요할 가능성이 높다는 것으로 해석될 수 있다(Conklin, Pellicer-Sánchez, & Carrol, 2018).

본 연구는 시선추적 접근방법을 활용하여, 대학생 영어학습자들을 대상으로 주어-동사 수일치 처리현상에 있어서 주어에 포함된 단수의미토큰유형이 포함되어 있어서, 통사처리과정이 요구되는 단수의미토큰 유형과, 명사의 유형은 단수이지만 의미적으로 복수의 의미를 유발시킬 수 있는 의미분산 효과가 있는 복수의미토큰유형의 언어처리에 있어서의 시선추적양상을 비교분석한다. 목표문장, 즉 단수의미토큰 및 복수의미토큰을 포함하는 문장을 읽으면서 나타난 전체 문장에 대한 시선추적의 결과, 즉, 복합명사구의 명사 1(N1)과 명사 2(N2)와 동사 선택의 동사 1(V1)과 동사 2(V2)영역에 초점을 맞추어 두 유형의 명사구를 주어로 포함하는 문장의 동사를 처리하는 양상의 차이를 분석해 보고자 한다. 결과분석은 단수의미 토큰 유형과 복수 의미 토큰 유형에서의 시선처리양상을 문장단위, 동사단위의, 명사 1 영역, 및 명사 2 영역을 관심영역으로 두고 시선처리 데이터를 분석하였다.

참고문헌

- Conklin, K. Pellicer-Sánchez, A., & Carrol, G. 2018. *Eye-tracking: A Guide for Applied Linguistics Research*. Cambridge, NY: Cambridge University Press.
- Just, M. A., & Carpenter, P. A. 1980. A Theory of Reading: from Eye Fixations to Comprehension. *Psychological Review*, 87, 329-354.

튀르키예 한국어 학습자의 보조사 ‘은/는’ 학습 양상에 대하여

김현정
(서울대학교)

한국어를 배우는 외국인 학습자가 자주 오류를 일으키는 문법 사항 중 하나가 보조사 ‘은/는’의 사용이다. 이는 한국어 보조사 ‘은/는’의 정보구조적 기능에 대해 완전한 숙지가 이뤄지지 않아서 초래되는 오류이다.

국어학계에서는 ‘은/는’의 정보구조적 기능에 대해 많은 연구를 축적해 왔다. 큰 틀에서는 대조와 화제(topic)의 측면에서 ‘은/는’을 다룬다. 더 나아가 화제대조, 초점대조 등으로 분석을 심화하기도 한다.

‘은/는’의 오류 양상은 한국어를 배우는 외국인 학습자가 모국어와 상관 없이 언어보편적으로 비슷한 모습을 보인다고 간주할 수도 있으나, 분명히 개별 국가의 언어적 특징과 한국어와의 유사성 또는 차이점으로 인해 그 나라 학습자 특유의 오류를 보이기도 한다.

본고에서는 튀르키예의 한국어 학습자가 주로 일으키는 ‘은/는’의 오류 양상을 살펴보고자 한다. 아울러 ‘은/는’의 사용 오류 양상을 밝히고 시중 한국어 교재에서 이에 대한 문법 설명을 어떻게 하고 있는지 점검할 것이다.

그 결과, 튀르키예 한국어 학습자는 ‘은/는’에 대해 크게 세 가지 오류 양상을 보이는 것으로 나타났다. 첫 번째 주격조사 ‘이/가’와의 혼동, 두 번째는 처격 및 처소격조사 ‘에’와 ‘에서’와 ‘은/는’의 결합 미숙, 세 번째는 튀르키예 소유격조사 In 과의 인지적 착오였다.

먼저 튀르키예 한국어 학습자는 보조사 ‘은/는’과 주격조사 ‘이/가’를 혼동해 사용하는 오류를 빈번하게 저질렀는데, 이는 비단 튀르키예 학습자에게만 나타나는 현상은 아니다. 기존 연구에서도 타언어권의 외국인 학습자가 ‘은/는’과 ‘이/가’를 혼동해 잘못 사용하는 경우를 많이 관찰해 보고하고 있다. ‘이/가’는 담화 내 신정보를 나타내는 초점의 기능을 주로 갖고, ‘은/는’은 구정보를 나타내는 화제의 기능을 주로 갖는다는 이론적 배경을 교수자가 숙지하고, 이를 외국인 학습자에게 가르쳐야 할 필요가 있다.

다음으로 처격 및 처소격조사 ‘에’와 ‘에서’는 올바르게 사용하지만, 여기에 ‘은/는’을 결합하는 단계로까지 나아가지 못하는 외국인 학습자의 발달과정상의 문제에 대해 언급했다. 이는 반복적인 학습을 통해 외국인 학습자가 ‘에’와 ‘에서’에 ‘은/는’을 결합한 ‘에는’과 ‘에서는’의 사용이 가능하도록 교수자가 지도해야 할 것으로 보인다.

튀르키예어 소유격조사 In 이 한국어 보조사 ‘은/는’과 의미적 한정성 측면에서 공통점을 보이며, 이에 따라 튀르키예 한국어 학습자가 인지적 착오를 일으켜 In 을 ‘은/는’으로 잘못 사용하는 경우를 언급했다. 이 둘은 음성적으로도 발음이 유사하기 때문에 빚어지는 인지적 오류라고 판단된다. 이는 교수자가 학습자에게 반복적으로 주지시켜 인지적 착오를 알아차리게 만드는 것이 중요해 보인다.

마지막으로 통계적으로 ‘은/는’의 사용 경향성을 살펴봤을 때 ‘대하여성’에 대해서는 사용 오류가 발견되지 않으나, ‘관계적주어집성’에 대해서 일부 사용 오류가 나타나는 것을 발견할 수 있었다.

또한 시중에서 구할 수 있는 한국어 교재 9 종을 살펴본 결과, 대부분의 교재에서 초급 1~2 단원에서 ‘은/는’을 제시하고 있었으나, 그중 ‘화제 또는 주제(topic)’에 대한 문법 설명을

간단하게라도 하고 있는 교재는 3 종에 불과했다. 연세대 교재가 유일하게 화제, 대조, 부가적 기능을 모두 설명하고 있었다.

이를 종합해 볼 때, 보조사 ‘은/는’을 초급 단계의 학습자에게 자세한 문법 설명을 하면서 가르치는 것은 적절하지 않을 것이라고 판단할 수 있다. 교재를 집필한 저자들이 이를 포함시키지 않은 교수적 측면의 이유가 있을 것이다. 초급 학습자에게 학습 부담을 주고 자칫 학습 의욕을 떨어뜨릴 수 있기 때문이다. 그렇지만 한국어 회화를 어느 정도 구사하는 중급 이상의 학습자에게는 ‘은/는’에 대한 자세한 문법 지식을 설명해줄 필요가 있을 것이다. 사소해보일 수 있지만 자연스러운 한국어 발화를 돕기 위해 필요한 교수 단계이다. 특히 외국인 학습자는 한국어를 자연스럽게 잘 구사한다는 한국인의 칭찬에 따라 학습 의욕이 고취되고 고급 수준의 한국어 구사에까지 이를 동기로 삼을 수 있으므로, 교수자가 이러한 심화 문법 지식을 잘 가르쳐서 학습자의 한국어 구사 능력을 끌어올리는 것이 중요할 것이다.

오류 유형	백분율
‘은/는’과 ‘이/가’의 혼동	18%
‘은/는’과 ‘을/를’의 혼동	13.6%
주격 조사 생략	9%

How does context window size address polysemy of adverbial postposition *-(u)lo* in Korean?

Seongmin Mun
(Ajou University)

Intended construal of a polysemous word occurs within a range of words, which deliver various frame-semantic meanings (Goldberg, 2019) and yet purport similar meanings (Harris, 1954). In this regard, context window—a range of words surrounding a target word, which affects the determination of characteristics of the word—is drawing attention to the computational understanding of combinatorial properties of words in human language (MacDonald & Ramscar, 2001).

We pose a question as to how context window size applies to polysemy of a function word such as a postposition in Korean, a language typologically different from the major Indo-European languages that have been investigated for this task. We report a computational simulation that explores how various sizes of context window account for polysemy of *-(u)lo*, which manifests polysemy due to its various functions mapped onto one single form (Choo & Kwak, 2008).

For this purpose, we used the Sejong corpus (Kim et al., 2007; 90% for training and 10% for testing), with semantic annotation of this corpus cross-verified by three native speakers of Korean ($\kappa = 0.95$). Employing a distributional semantic model (Baroni et al. 2014.), we devised an unsupervised learning algorithm by combining Singular Value Decomposition with Positive Pointwise Mutual Information (Turney & Pantel, 2010). Cosine similarity scores of *-(u)lo* and its co-occurring content words were re-scaled through the min-max normalisation (Luai et al., 2006). Using these scores, model performance was measured through the rate of accuracy that the model classified instances of the test set involving the six functions of *-(u)lo* under manipulation of context window size from one to ten.

Overall, our model achieved the highest rate of classification accuracy in the window size of one, and the rates of accuracy decreased as the window size increased. The global trend of accuracy that the model demonstrated is consistent with previous research that shows advantages of small window sizes (Bullinaria & Levy, 2007). A narrower range of context window relates more to syntactic information than to semantic information (Patel et al., 1997). This invites an interpretation that our model may have employed structural, rather than semantic, characteristics of tri-grams (i.e., a word-target-word sequence) for the best performance in classification. Given the networks of interlinked clusters of words and symbolic units in human cognition (construct-i-con; Goldberg, 2006), our findings shed light on relations between a polysemous word and an abstract schema including the word (represented as context window) for addressing word-level polysemy.

References

- Baroni, Marco, Georgiana Dinu & Germán Kruszewski. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 238–247.
- Bullinaria, John A & Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. Behavior Research Methods 39(3). 510–526.
- Choo, Miho & Hye-Young Kwak. 2008. Using Korean: A Guide to Contemporary Usage. Cambridge University Press, Cambridge, UK.
- Goldberg, A. E. 2006. Constructions at work: The nature of generalization in language. Oxford: Oxford University Press.

- Goldberg, A. E. 2019 *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton, NJ: Princeton University Press.
- Harris, Zellig S. 1954 *Distributional Structure*. *WORD*. 10(2-3). 146-162.
- Kim, Byoungsoo, Yong-Hun Lee & Jong-Hyeok Lee. 2007. *Unsupervised semantic role labeling for Korean adverbial case*. *Korean Institute of Information Scientists and Engineers*. 32–39.
- Luai, Shalabi Al, Ziyad Shaaban & Basil Al-Kasasbeh. 2006. *Data mining: A preprocessing engine*. *Journal of Computer Science* 2.
- MacDonald, Scott & Michael Ramscar. 2001. *Testing the distributional hypothesis: The influence of context on judgements of semantic similarity*. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. 611–6.
- Turney, Peter D. & Patrick Pantel. 2010. *From frequency to meaning: Vector space models of semantics*. *Journal of Artificial Intelligence Research* 37(1). 141–188.

인공지능의 한국어 접사 처리 연구

비립 · 강예지 · 박서윤 · 김한샘
(연세대)

형태론적인 측면에서 풍부한 접사 시스템을 통해 다양한 어휘를 파생하고 활용하는 것은 교착어인 한국어의 가장 선명한 특징 중 하나이다. 따라서 한국어 인공지능의 언어 이해 및 생성 능력을 제대로 확인하려면 한국어 접사 처리에 관한 식별 및 활용 과정에서 어떠한 성능을 보이는지가 중요한 평가 척도가 될 수 있다. 과거에 해외에서 인도-유럽어족에 속하는 언어를 대상으로 삼아 각종 워드임베딩이나 사전학습모델의 형태론적 굴절 및 파생 능력에 관한 연구가 꾸준히 이어왔으나, 한국어 인공지능 모델의 형태론적 인지 수준, 특히 한국어 접사 시스템의 구분 기준 및 적용 방식 이해도에 대한 검토는 상대적으로 부족한 편이다. 한편, 자연어처리 과정에서 한국어는 다른 언어와 달리 형태소 분석이 확실히 이루어져야 어휘적, 통사적, 의미적 층위의 심층 분석을 정확하게 수행할 수 있다. 그동안 선행연구에서 한국어 인공지능의 형태소 분석 정확률은 95%를 넘긴다는 보고가 수차례 나온 적 있었지만, 실제로 응용하는 과정에서 확인되는 형태소 분석의 수준, 특히 한국어 접사의 인식 수준은 타 범주에 비해 상대적으로 낮다는 것이 확인된다. 신체 장기 폐의 열을 뜻하는 ‘폐열’을 구성하는 명사 ‘폐’를 ‘폐건전지’의 접두사 ‘폐-’로 분석한다든지, 외래어, 인명의 일부를 접사로 인식한다든지 하는 오류를 쉽게 찾아볼 수 있다. 본 연구는 급속도로 발전해 나가는 한국어 인공지능의 형태론적 인지 쟁점으로서의 접사 처리 능력에 초점을 맞춰, 한국어 인공지능 모델의 접사 식별 및 활용 능력 평가 프로세스를 설계하고 실험하였다. 실험을 통해 ‘한국어 인공지능 모델은 한국어 파생어 내 접사와 어기의 경계를 이해하는가? 한국어 인공지능 모델은 한국어 접사와 파생 방식을 인지하는가?’의 연구 문제를 해결하기 위한 기초 연구를 수행하였다.

인공지능의 접사 처리 능력 평가는 크게 두 가지 관점에서 이루어졌다. 접사와 어기의 결합 가능성에 대한 문맥 미고려 실험은 LM 모델의 Fill-in Mask Test 로 진행하였고, 문맥을 고려하여 문장 내의 파생어 사용 적절성은 LLM 모델의 Prompting Test 를 통해 평가하였다. 실험 대상이 된 접사는 총 15 항목으로 관련어 쌍 8 개를 이루도록 다음과 같이 선정하였다.

[반의 관계]

①대_XPN/소_XPN ②신_XPN/구_XPN ③친_XPN/반_XPN ④고_XPN/저_XPN

[유의 관계]

①반_XPN/역_XPN ②심_XSN/감_XSN ③론_XSN/설_XSN ④인_XSN/자_XSN

X-roberta, Kalbert, Klue-Ror-B, Klue-Ror-L, Koelectra, KcELECTRA, Kcbert-B, Kcbert-L 등 7 종의 트랜스포머 기반 언어모델을 대상으로 Fill-in Mask 테스트를 진행한 결과 특정 접사에 결합 가능한 어기와 결합 불가능한 어기를 구분하는 능력은 다국어 모델에 비해 한국어 모델이 뛰어났으며 한국어 모델 사이에서는 보유 어휘 목록이 비슷하기 때문에 큰 차이를 보이지 않았다. 접사의 유의 관계를 기반으로 한 저지 현상 테스트에서는 전반적으로 Distance 값이 높게 나와 유의 관계 접사의 배타적인 결합 패턴을 언어모델이 이미 파생어 단위로 잘 학습한 것으로 보인다. 한편 반의 관계 실험에서는 접사쌍의 의미적 거리를 제대로 파악하지 못한 것으로 파악되었다.

ChatGPT3.5(turbo)를 대상으로 LLM의 접사 처리 능력을 실험한 결과 주어진 접사와 결합 가능한 어기 찾기는 LM보다 LLM 결과가 상대적으로 더 낮은 결과를 보여 한국어 접사에 대한 이해가 부족한 것으로 나타났다. 한편 결합 방향성의 관점에서는 특정 접사와 결합하는 어기를 찾는 것보다 특정 어기와 결합하는 접사 찾기가 더 어려운 태스크였다. 저지 현상을 보이는 유의 관계 접사 쌍 간의 어기 결합 가능성 차이를 잘 파악한 데에 반해, 반의 관계 변별은 50% 이하의 성공률로 성능이 낮게 나타났다. 유의 관계에 비해 반의 관계 인식의 성능이 떨어지는 것은 LM과 LLM에서 공통적으로 나타난 현상이다.

이 연구는 기존 한국어 인공지능 모델의 접사 처리의 한계를 조명하고 앞으로 접사 처리 관련 태스크 성능 개선을 위한 보완책을 탐구하기 위한 기초 연구이다. 한국어 LM & LLM을 대상으로 실험한 결과 나타난 인공지능의 한국어 접사 처리의 한계를 더욱 명확하게 밝히기 위해서 평가용 데이터셋, 평가 지표, 그리고 평가 대상을 확대할 필요성이 있으며, 접사 능력 평가 프로세스에 대한 보강도 지속되어야 한다.

Validating Python Libraries for Text Analysis in Language Research

Sumi Han and Minji Kim
(Hallym University)

With the advent of user-friendly computer analytic tools, such as Python libraries, large datasets of language can be examined automatically, saving significant time and effort (Albrecht, Ramachandran, & Winkler, 2020). However, many of these tools have been developed by computer programmers rather than language researchers. As a result, they remain relatively unknown to the linguistic community, and their validity is rarely tested. To bridge this gap, our study aims to review and validate popular Python libraries for text analysis and offer recommendations for future linguistic research.

We selected Python as our programming language of choice because of its versatility and widespread adoption in recent times (Srinath, 2017). Python's flexibility, readability, and high level of abstraction also enhance user productivity (Srinivasa, 2018). After a comprehensive literature review, we identified Python libraries such as TextDescriptives, Textstat, and Readability for text analysis and complexity assessment. These were evaluated based on their key features and their prominence, as indicated by data from GitHub.

For our review and validation, we used two corpora (lectures and novels) to derive text analytic scores. Our results revealed that textaCy (DeWilde, 2021) and textstat (Ward, 2022) are the most widely recognized tools, as evidenced by their high numbers of stars and forks on GitHub. While many libraries incorporate readability scores, text complexity measures have recently been added to newer libraries. We will further discuss these and other key findings, highlighting their implications for linguistic research.

References

- Albrecht, J., Ramachandran, S., & Winkler, C. (2020). Blueprints for text analytics using Python. O'Reilly Media, Inc.
- DeWilde, B. (2021). Textacy: NLP, before and after spaCy (Version 0.12.0). <https://github.com/chartbeat-labs/textacy>
- Srinath, K.R. (2017) Python—The fastest growing programming language. *International Research Journal of Engineering and Technology (IRJET)*, 4, 354-357.
- Srinivasa-Desikan, B. (2018). *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, SpaCy and Keras*. Packt Publishing Ltd., Birmingham.
- Ward, A. (2022). Textstat. <https://github.com/textstat/textstat>

영어 허사 There 구문의 표찰화

남소영
(부산대학교)

다음의 문장을 Chomsky 표찰화 알고리즘의 관점에서 고려해 보도록 하자.

- (1) a. [γ [There seems [β to be [α a man in the room]]]
b. *It seems to be a man in the room.

위의 α 는 {XP,YP}의 형태로 제휴어와 술어의 관계를 형성하는 구조이다. 표찰화 구조 아래에서는 해당 통사체는 두 요소중 하나의 요소가 상승하거나 가장 주요한 자질을 공유함으로써 표찰화 되는 것이 가능하다. 그러나 α 는 그러한 구조가 가능하지 않으므로 어떻게 α 가 표찰화 되는지 분명하지 않다. 이와 비슷하게 β 는 {H, XP}의 형태이며, 부정사 T 구조를 이룬다. 표찰화 아래에서 그러한 통사체는 H의 최소 탐색 과정에 의해 표찰화 되지만 해당 T는 표찰로 역할을 하기에는 너무 약하며(weak) 이에 어떻게 β 가 표찰화 되는가 역시 분명하지 않다. 더 나아가 γ 는 {there, TP} 형태이며, 허사 술어 구조를 형성하고 있다. 표찰화 아래에서 주술 구조 {NP, TP}는 두 요소가 파이 자질을 공유하거나, 지정어 핵어라는 일치 구조를 형성함으로써 표찰화 되어진다. 그러나 일반적인 NP 구조와는 대조적으로 (1b)에서 처럼 수 자질을 가진 허사 *it*이 사용 되었을때 비분이 되는 것처럼, *there*이 T와 일반적인 수일치 관계를 이루지 못하므로 γ 가 어떻게 표찰화 되는지 또한 분명하지 않다.

그러나 모든 통사체는 표찰화라는 일반적 원리를 충족시키는 것이 바람직하므로 *there* 구문 역시 어떻게 표찰화 되어 이 일반적 원리를 만족시키는지 고려함으로써 *there* 구문에서의 *there*의 외부 병합위치와 적절한 도출과정을 통해 표찰화 구조를 밝혀 보고자 한다.

Transitive *out of -ing* construction: A collocation and constructional approach

Jungsoo Kim and Rok Sim
(Incheon National University, University of South Carolina)

The so-called transitive *out of -ing* construction has the structural pattern, NP subject + V1 + NP object + *out of* VP2[-*ing*] and the construction can give rise to two different types of interpretations depending on the telicity of the predicate introduced by VP2[-*ing*], as exemplified in (1) (Francis et al. 1996; Rudanko 2011; Gawlik 2013; Sim & Kim 2015):

- (1) a. I cheated you all out of being in my life. (COCA-1999-TV)
b. She cheated thousands of people out of getting a rebate. (COCA-2010-NEWS)

Here, (1a) is paraphrased as ‘I caused you to move out of being in my life as if by cheating’. This type of interpretation is referred to as the MOVEMENT/EXTRACTION interpretation, aligning with the idea of persuading someone to leave a particular place or state. On the other hand, (1b) is paraphrased as ‘... by means of cheating she prevented thousands of people from getting a rebate’. This type of interpretation is called the PREVENTION interpretation, involving persuading someone not to do something.

Overall, in literature much less attention has been paid to the transitive *out of -ing* construction as in (1) than the transitive *into -ing* construction (e.g., *I’m actually trying to talk him into putting a little xmas tree in here for us*). In addition, although previous studies on the transitive *out of -ing* construction have made use of attested corpus data, their focus has been rather restricted to the types of verbs in the V1 slot of the construction, disregarding the types of expressions in the other slots of the construction. Furthermore, they have just provided raw frequencies of verb types in the V1 slot, which may not tell us directly about the core meaning(s) associated with the slot.

In order to fill the research gap, we first investigate real-life uses of the transitive *out of -ing* construction on the basis of corpus data from COCA (Davies 2008-). In doing so, we not only examine which (types of) verbs frequently occur in the V1 slot of the construction, but we also look into grammatical properties of the expressions in each of the other slots of the construction, adopting different types of collocation analyses (Stefanowitsch & Gries 2003, 2005; Hilpert 2006, 2012; Perek 2014). The collexeme analysis first reveals that verbs associated with the V1 slot of the transitive *out of -ing* construction often belong to the semantic classes of FORCE and NEGATIVE EMOTION verbs, with an ENTICING verb *talk* being the strongest collexeme, despite not belonging to these semantic classes. The distinctive collexeme analysis also shows that the construction favors animate subjects over inanimate ones, and certain groups of verbs are strongly associated with inanimate subjects. In addition, it is found out that the PREVENTION interpretation is more salient than the MOVEMENT/EXTRACTION interpretation in the transitive *out of -ing* construction, and that the alleged mapping relation between eventuality types and interpretation types can be overridden and the pairs have somewhat different correlation tendencies. Lastly, the covarying collexeme analysis indicates that certain pairs of semantic frames are preferred in describing cause-result relations in this construction.

We then attempt to offer an analysis of the construction from a Construction Grammar (CxG) perspective (Goldberg 1995, 2005; Sag 2012; Kim & Davies 2016; Kim & Michaelis 2020). In particular, we show that its grammatical properties can be well explained via inheritance networks and tight interactions between lexical properties and argument structure constructions.

References

- Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Francis, Gill, Susan Hunston, and Elizabeth Manning (eds.). 1996. *Collins COBUILD grammar patterns 1: Verbs*. London: HarperCollins.
- Gawlik, Oskar. 2013. On the transitive *out of -ing* complementation pattern in contemporary spoken American English: A corpus-based study. *Poznań Studies in Contemporary Linguistics* 49(2): 167-183.
- Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.
- Goldberg, Adele. 2005. *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.
- Hilpert, Martin. 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory* 2(2): 243-256.
- Hilpert, Martin. 2012. Diachronic collostructional analysis: How to use and how to deal with confounding factors. In Kathryn Allan and Justyna Robynson (eds.), *Current methods in historical semantics*, 133-160. Berlin: Mouton de Gruyter.
- Kim, Jong-Bok and Mark Davies. 2016. The INTO-CAUSATIVE construction in English: A construction-based perspective. *English Language and Linguistics* 20(1): 55-83.
- Kim, Jong-Bok and Laura A. Michaelis. 2020. *Syntactic constructions in English*. Cambridge: Cambridge University Press.
- Perek, Florent. 2014. Rethinking constructional polysemy: The case of the English conative construction. In Dylan Glynn and Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 61-85. Amsterdam: John Benjamin Publishing Company.
- Rudanko, Juhani. 2011. *Changes in complementation in British and American English: Corpus-based studies on non-finite complements in recent English*. Basingstoke: Palgrave Macmillan.
- Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. In Hans C. Boas and Ivan A. Sag (eds.), *Sign-based construction grammar*, 69-202. Stanford, CA: CSLI Publications.
- Sim, Rok and Jong-Bok Kim. 2015. The transitive “*out-of V-ing*” construction in English: Corpus-based study [written in Korean]. *The Linguistic Association of Korea Journal* 23(4): 189-210.
- Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.
- Stefanowitsch, Anatol and Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1: 1-43.

Elliptical *why not?*: A discourse-based perspective

Okgi Kim and Jong Bok Kim
(Kyung Hee University)

The elliptical *why not*-question can be used as either an anaphoric information-seeking question or a rhetorical question, as illustrated by the attested data (Merchant 2006, Hofmann 2018, Stockwell 2022):

- (1) a. I can't sleep. Why not? (COCA 1997 TV)
b. Shall we go in? Why not? (COCA 2016 MOV)

One key difference between these two types concerns their licensing contexts (Hofmann 2018): while anaphoric *why not* is only licensed by a negative antecedent, rhetorical *why not* is associated with a modal-ized antecedent (e.g., suggestions/offers).

- (2) a. A: I slept. B: #Why not? (cf. ((1a))
b. A: Won't you care to join me? B: (Sure,) Why not? (COCA 2014 TV)

The ellipsis-based analyses suggested by Hofmann (2018) and Stockwell (2022) agree that the anaphoric *why not* is generated by the neutralization-based analysis suggested by Kramer & Rawlins (2009). Licensing PF-ellipsis of TP under a mutual entailment relation to the negative antecedent, these two analyses assume that the *not* bears an uninterpretable Neg feature to yield a single sentential negation reading. (1a) would thus be derived as follows:

- (3) [CP Why [SP not_[uNeg] [~~TP you can't_[uNeg] sleep~~]]]

The uninterpretable Neg is in a negative concord relation with the interpretable head Neg. This direction seems to be appealing for the proper semantic resolution, but it is first questionable if *not* is really uninterpretable.

We have performed a corpus investigation of the construction using COCA (Corpus of Contemporary American English). We identified 1,000 randomly selected examples and analyzed their grammatical properties. These empirical data show that the previous neutralization analysis is insufficient to capture idiosyncratic behaviors of elliptical *why not*.

For one thing, our corpus search yields examples where the negative antecedent condition is satisfied not just by the sentential head Neg but by a variety of negative expressions including constituent negation or lexical negation:

- (4) a. Look, I'm asking you [not to do this for me]. Why not? (COCA 2007 TV)
b. Computer, bring the subspace transmitter on line. Unable to comply. Why not?
Insufficient power. (COCA 2000 TV)

The ellipsis-based analyses would have difficulties in accounting for the attested data in (5) where a phrasal element occurs after *not*.

- (5) a. Carter replied that they were not always the best. "Why not the best?" (COCA 1994 NEWS)
b. "You can even keep some of the wreckage. None of the coils, though." "Why not the coils?" (COCA 2001 FIC)

The examples in (6) illustrate that anaphoric *why not* can be linked to semantic negation inside a complex NP island in the antecedent.

- (6) a. [Very desirable woman who'll never go out with me]. Why not? One of her friends is one of my exes. (COCA 1995 TV)
 b. [Something that a gentlewoman must not do to a gentleman]. - Why not? - Because men will take love for granted and then not prize it. (COCA 2004 MOV)

This argues against the claim that the negative remnant *not* may be derived via movement at PF, followed by TP ellipsis, just like the remnant in *why*-stripping (Yoshida et al. 2015).

As shown in (7), the putative source of anaphoric *why not* can be semantically associated with an *if*-conditional clause or elliptical *probably not* whose proposition is not presupposed.

- (7) a. Did you report the crime? If not, why not? (COCA 2012 WEB)
 b. Is Fez a suspect? Yeah, probably not, Fez. Why not? (COCA 2007 TV)

This serves as a counterexample to Hofmann's (2018) claim that the negative antecedent requirement is attributed to the factivity of *why*: the prejacent [*not p*] of *why* is presupposed, requiring a background compatible with the negative proposition.

The corpus data in (8) indicate that the ellipsis site of *why not* should be reconstructed from discourse. This further challenges the polarity ellipsis approach that heavily depends upon the syntactic as well as semantic information in the antecedent in order to resolve the elided parts.

- (8) a. Everybody has a name. Everybody but me. Why not? (COCA 2005 MOV)
 b. Once upon a time you could read "B. Monkey" on every tube station. Except the Northern beyond Kennigton. Why not? (COCA 1998 MOV)

In this study, we briefly sketch a semantics/discourse-based approach that can account for regularities as well as idiosyncrasies of the two types of elliptical *why not*-questions. In particular, we argue for LF-focus movement of *why-not*, different from PF-focus movement for negative stripping (e.g., *Kim cooked beans for dinner, not LENTILS*) and *why*-stripping (e.g., *Why Kim?*).

References

- Hofmann, Lisa. 2018. *Why not?* – polarity ellipsis in *why*-questions. Ms., University of California, Santa Cruz.
 Kramer, Ruth & Kyle Rawlins. 2009. Polarity particles: an ellipsis account. In Anisa Schardl, Martin Walkow & Muhammad Abdurrahman (eds.), *Proceedings of NELS*, volume 39, 479-492.
 Merchant, Jason. 2006. Why no(t)? *Style* 40(1 & 2): 20-23.
 Stockwell, Richard. 2022. Elliptical *why not*. Ms., University of Oxford, Christ Church.

뒤섞기 구문에 나타나는 복합명사구제약 효과에 대한 한국어와 일본어 비교 실험 연구

김일규¹ · 손은채² · 지윤주²
(¹강원대학교, ²한국외국어대학교)

섬제약 현상 관련하여 지금까지 개별 언어에 대한 연구는 많이 진행됐지만 언어 간 비교 연구는 거의 없는 실정이다. 본 연구는 두 개의 수용성(acceptability) 판단 실험을 통해 뒤섞기 구문에 나타나는 복합명사구제약 효과에 대해 한국어와 일본어를 비교하여 두 언어 간 공통점과 차이점을 밝혀 궁극적으로 섬제약의 보편성(universality)에 대한 더욱 깊은 이해를 하는데 기여를 하고자 한다.

뒤섞기 구문에 나타나는 복합명사구제약 효과 관련하여 최근 일본어와 한국어를 대상으로 두 개의 실험 연구가 진행되었는데, 두 연구에서 보고한 실험 결과는 섬제약 효과의 수용성(acceptability)에 두 언어 간에 차이가 있음을 보여준다. 본 연구에서는 Fukuda *et al.* (2022)에서 사용한 실험자극을 한국어로 번역하여 한국인을 대상으로 동일한 실험을 진행하고, Kim & Ji (2023)의 실험자극을 일본어로 번역하여 일본인 대상으로 동일한 실험을 진행할 계획이다. 그렇게 함으로써 Fukuda *et al.* (2022)와 Kim & Ji (2023)에서 제시한 뒤섞기 구문에 나타나는 복합명사구제약 효과 관련 상이한 실험 결과가 진정으로 두 언어의 차이에 기인한 것인지 단순히 실험자극의 차이 때문인 것인지를 밝힐 수 있을 것이다.

참고문헌

- Fukuda, Shin & Tanaka, Nozomi & Ono, Hajime & Sprouse, Jon. 2022. An experimental reassessment of complex NP islands with NP-scrambling in Japanese. *Glossa: a journal of general linguistics* 7(1), 1–35.
- Kim, Ilkyu & Yunjoo Ji. 2023. On the nature of CNPC effects in Korean scrambling constructions. *Linguistic Research* 40(2), 151-181.

On the recovery of Korean null objects: Some challenges from experimental evidence Background.

고희정¹ · 김현우² · 강초롱¹
(¹서울대학교, ²연세대학교)

There are two major approaches to Korean null objects: one is the argument ellipsis analysis (Oku 1998; Kim 1999; Saito 2007), and the other is the null pronoun analysis (Ahn and Cho 2020). These two approaches make different predictions regarding the reading preference between two possible interpretations of null objects. The argument ellipsis analysis assumes that null objects contain a full-fledged nominal phrase in their LF structure, leading to the sloppy reading. On the other hand, the null pronoun analysis assumes that the LF structure of null objects includes a pronoun, favoring the strict reading. However, there has been limited quantitative research conducted to test the predictions of each approach (cf. Han et al. 2020). This work focuses on four unresolved questions on the recovery of null objects in Korean and provides quantitative data to advance theoretical debates on them.

Method. The experiment utilized a priming technique. Two prime conditions included either a *self*-NP noun or a pronoun in the object position of the second sentence (1). Each prime item was followed by a target with a null object (2). Twenty-one adult Korean native speakers (age 20-36) read sentences and indicated how they interpreted the object among four options (strict/sloppy/both/“I do not know”).

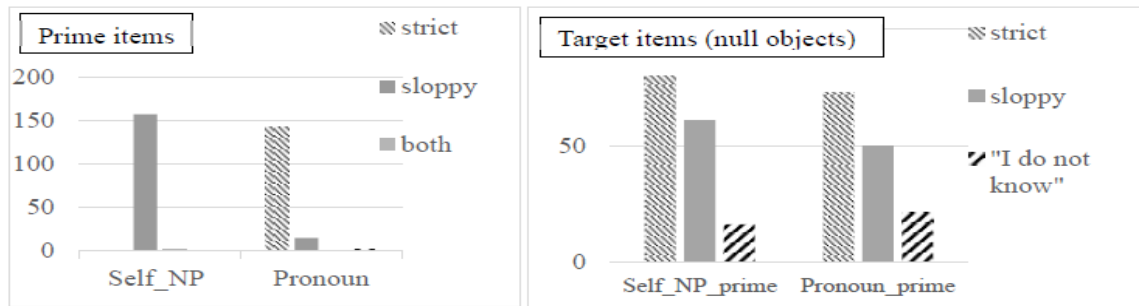
(1) Prime items

Mina-ka cak-i-uy peykay-lul nwulessta. Mina-NOM self-POSS pillow-ACC pushed (lit.) Mina pushed self's pillow.	
Self-NP prime condition	Pronoun prime condition
Arin-ito cak-i-uy peykay-ul nwulessta. Arin-also self-POSS pillow-ACC pushed. (lit.) Arin pushed self's pillow, too.	Arin-ito kukes-ul nwulessta. Arin-also it-ACC pushed. (lit.) Arin pushed it, too.

(2) Target item

Wuyeng-ika cak-i khwusyen-ul nwulessta.
Wuyeng- NOM self cushion-ACC pushed.
(lit.) Wuyeng pushed self's cushion.
Sangmin-ito <e> nwulesstal.
Sangmin-also pushed.
(lit.) Sangmin pushed <e>, too.

Results. The experiment revealed four new observations: 1. A consistent preference for the strict reading for null objects (in interpretation of the target items). 2. An interpretation discrepancy between overt Self-NP/overt pronoun and null objects ($p < 0.01$). 3. A general acceptance of the sloppy reading for null arguments (no significant difference among participants, $p = 0.64$). 4. The absence of priming effects from a full-fledged nominal phrase or a pronoun on the processing of null objects ($p = 0.15$). These four observations challenge both approaches to Korean null objects and suggest that further quantitative research is needed to better understand the syntactic nature of null objects.



Selected references

- Ahn, Hee-Don, and Cho, Sungeun. 2020. "A Pronoun Analysis of Null Arguments in Korean." *Language Research* 56(2):193–223
- Han, Chung-hye, Kim Kyeong-min, Keir Koulton, and Jeffery Lidz. 2020. "Null Objects in Korean: Experimental Evidence for the Argument Ellipsis Analysis." *Linguistic Inquiry* 51(2): 319-340
- Kim, Soowon. 1999. "Sloppy/strict identity, empty objects, and NP ellipsis." *Journal of East Asian Linguistics* 8(4): 255–284
- Oku, Satoshi. 1998. "A theory of selection and reconstruction in the Minimalist perspective." PhD diss. University of Connecticut
- Saito, Mamoru. 2007. "Notes on East Asian argument ellipsis." *Language Research* 43:203-227.

Spoken acceptability judgment on Korean ditransitives: Comparisons with written judgment

Sang-Hee Park and Eunkyung Yi
(국립한밭대학교, 이화여자대학교)

In this presentation, we investigate the impact of presentation modality (spoken vs. written) on the acceptability judgments of native speakers, through a reaction time analysis and comparisons with written judgment. We used Korean ditransitives as a construction of interest, which include two distinct structures that differ in the case marker on the recipient/goal argument: canonical ditransitives (NP1NOM-NP2DAT-NP3ACC-Verb) and double-accusative ditransitives (NP1NOM-NP2ACC-NP3ACC-Verb). While canonical ditransitives are well-accepted by native speakers, double-accusative ditransitives tend to elicit low judgment scores and occur rarely in naturalistic data (Shin and Park 2019; Park and Yi 2021).

We first conducted an auditory perception experiment of Korean ditransitives, in an attempt to shed light on how acceptability scores generally relate to reaction times. Consistent with Nagata (1990), we observed an inverted U-shaped relationship for the entire stimuli set (Figure 1). Upon analyzing each construction independently, we found linear relationships between acceptability scores and reaction times, which are consistent with the lower or higher half of the general relationship. Furthermore, we analyzed the impact of modality of stimuli (spoken vs. written) by conducting a meta-analysis comparing our auditory judgment data with a prior written judgment experiment by Park and Yi (2021). Our findings suggest that the modality of stimuli plays a significant role, but that its influence can vary based on the interplay between construction type and the semantics of the verb (Figure 2). In summary, our results show that presentation modality, construction type, and verb semantics collectively influence sentence perception in Korean. These results suggest it can be beneficial for acceptability judgment studies to use spoken stimuli and take insights from reaction time data.

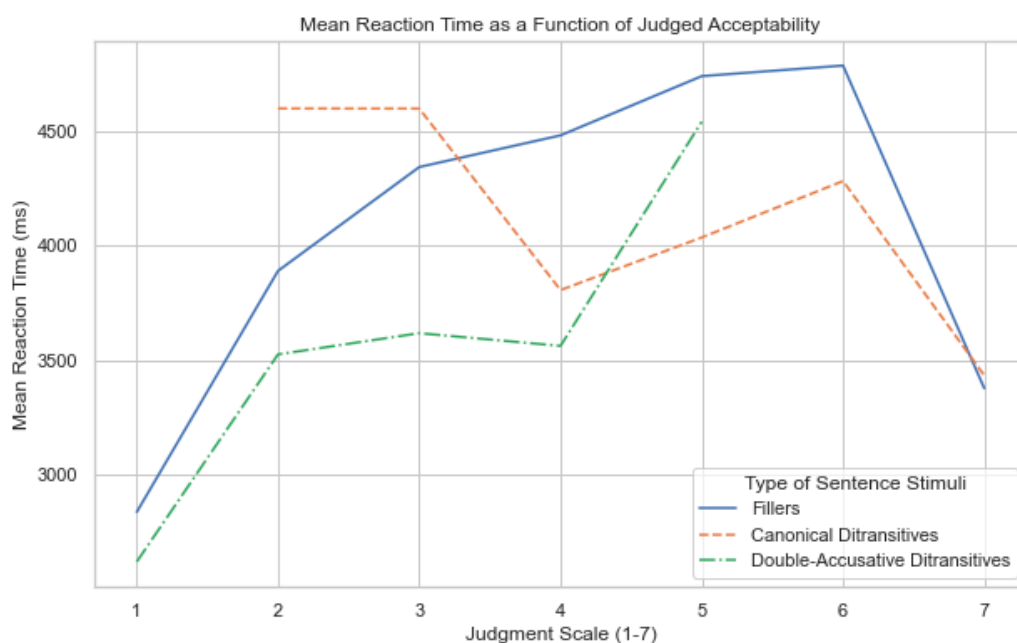


Figure 1. Mean reaction times as a function of judged acceptability

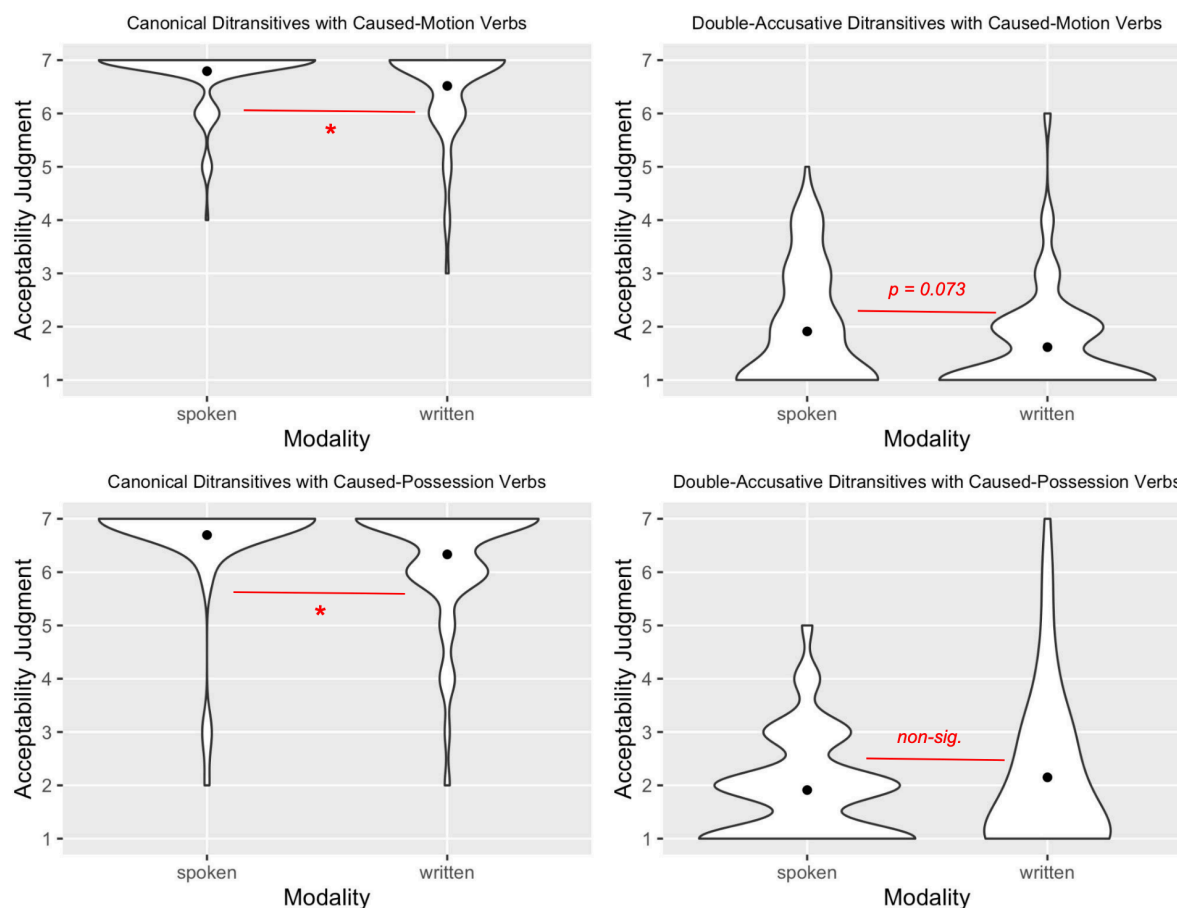


Figure 2. Violin plots of judged acceptability as a function of modality in four ‘structure x verb type’ categories

References

- Nagata, Hiroshi. 1990. On-line judgments of grammaticality of sentences. *Perceptual and Motor Skills* 70: 987-994.
- Park, Sang-Hee and Eunkyung Yi. 2021. Perception-production asymmetry for Korean double accusative ditransitives. *Linguistic Research* 38(1): 27-52.
- Shin, Gyu-ho and Sun Hee Park. 2019. Language use matters: Verb-construction fusion in the comprehension of dative constructions in Korean. Paper presented at the *21st Meeting of the International Circle of Korean Linguistics*. Melbourne, Australia.

한·영 모순어법(oxymoron)용례에 대한 인지의미론적 고찰

김정민
(한국의국어대학교)

본 연구는 한국어와 영어에 나타나는 모순어법(oxymoron)의 용례를 인지의미론적 관점에서 고찰한다. 모순어법(oxymoron)이란 두 개의 상반되는 의미 요소가 결합한 언어적 표현을 말하며, 대표적인 한국어의 예로는 “미운 정,” “찬란한 슬픔,” “소리없는 아우성,” 영어의 예로는 “sweet sorrow,” “living death,” “heavy lightness” 등이 있다. 전통적으로 모순어법은 문학 작품에서 극적인 효과를 주기 위한 수사적 장치로 기술되었으며, 대부분의 선행연구는 문학 작품에서의 그 용례와 기능을 분석하고 있다. 하지만 모순어법은 한국어의 “밀당,” “웃프다,” “시원섭섭하다,” 영어의 “tragicomic,” “frenemy,” “open secret” 등의 소위 일상적인 언어 표현뿐만 아니라 신어에도 종종 발견되는 현상이다. 특히, 일부 용례들은 두 언어에서 공통으로 나타나며 수사적 효과도 미약하기 때문에, 모순어법의 대립관계(opposite relations)가 어떻게 발견되는지 비교 분석할 필요가 있다. 따라서, 본 연구는 두 언어의 모순어법 용례를 형태와 의미적 유형으로 분류하고, 인지의미론적 관점에서 모순어법의 인지적 동기(motivation)를 밝히고자 한다. 결론적으로, 모순어법은 대립의 개념을 인지하는 개념화 양상이며, 개념적 혼성이론(conceptual blending theory)을 비롯한 다양한 인지 기제를 통해 동기화된다.

참고문헌

- 이선영. 2016. “반의어가 결합한 단어의 특징과 의미.” *어문론집* 68: 37-58.
- Flayih, M. 2009. “A Linguistic Study of Oxymoron.” *Journal of Kerbala University* 7.3: 30-40.
- Gibbs, Raymond W., and Lydia R. Kearney. 1994. “When parting is such sweet sorrow: The comprehension and appreciation of oxymora.” *Journal of Psycholinguistic Research* 23: 75- 89.
- Hameed, H. T. 2020. “Oxymoron in Day-to-Day Speech.” *The Asian ESP Journal* 16.4: 140 -152.
- Ruiz, J. H. 2015. “Paradox and Oxymoron Revisited.” *Procedia-Social and Behavioral Sciences* 173: 199-206.
- Shen, Yeshayahu. 1987. “On the Structure and Understanding of Poetic Oxymoron.” *Poetics Today* 8.1: 105-122.

Exploring the Social Meaning of the Usage of *yay/cyay/kyay* for Inanimate Referents in Korean through Frameworks of Sociolinguistics and Conceptual Metaphor

Jinsok Lee
(The Catholic University of Korea)

While a few studies (Kim, 2011; Park 2015; Kim ,2018; and Kim, 2021) have examined (or at least mentioned) the usage of combined form of demonstrative *i/ku/ce* + nominal expression *ai* “child” forming *yay/kyay/cyay* in Korean as an emerging form of third person pronominal expression in modern spoken Korean, it is still true that little was discussed on the usage from a sociolinguistic perspective, especially for inanimate referents (e.g., smartphone, women’s bag, beer, etc.). This study provides a new approach supplementing the understanding of such usage through a combined framework of indexical field (Eckert, 2008), conceptual metaphor (Lakoff, 1993), and sociolinguistic stance (DuBois, 2007) as a follow-up of a preliminary discussion on the phenomenon from Lee (2017). Although Lee (2017) demonstrated that the noun used in the demonstrative + noun construction in Korean was not limited to using “child” but also other human nouns including *chinkwu* “friend” and *nyesek* “brat,” to provide a clearer picture of the use of such construction referring to inanimate referents and the social meaning of such usage, this study examines the realization of *yay/kyay/cyay* and by examining the usages in approximately 20 YouTube clips reviewing the same product.

References

- DuBois, John W. 2007. The stance triangle. In Englebretson, Robert, (Ed.), *Stance taking in discourse: Subjectivity, evaluation, interaction* 139-182. Amsterdam: John Benjamins.
- Eckert, Penelope. Variation and the indexical field. *Journal of Sociolinguistics* 12/4, 453-476.
- Lakoff, George. 1993. The Conceptual Metaphor Theory. In Ortony, A. (Ed.), *Metaphor and Thought* (2nd ed.) 201-251. Cambridge: Cambridge University Press.
- Lee, Jinsok. 2017. The Metaphorical Process Underlying Lexical Variation and its Discourse Effects: Using human nouns to refer to non-human/non-animate referents in Korean. *Language and Information* 21.2, 57-78.
- 김순길. 2011. 조선어 3인칭대명사의 사용제약에 대하여. *중국조선어문* 2011:2, 22-28.
- 김아림. 2018. 한국어 구어 비격식체 3인칭 대명사 ‘애/개/재’. *언어와 언어학* 80, 1-40.
- 김아림. 2021. 명사 ‘아이’에서 3인칭 대명사로의 구문화(Constructionalization)에 대한 용례기반 연구. *담화와 인지* 28-1, 23-57.
- 박재희. 2015. 3인칭 대명사 ‘그’에 관한 소고. *한민족어문학* 71, 102-128.

Bridging the Gap: A Tale of Two Types of Anaphoric Nominals in Korean

이상근¹ · 마정연¹ · 김희경²
(¹ 고려대학교, ²Binghamton University)

This study investigates the distribution patterns of anaphoric nominals in Korean, with a special focus on demonstrative NPs with *ku* ‘that/the’ and bare NPs without *ku* ‘that/the’. Though several scholars have examined a tendency among Korean speakers in the use of Korean anaphoric nominals, these prior investigations either lacked controlled experimental conditions or were limited to professional linguists’ intuitions (Kang 2021, Lee, Jeong, & Jeong 2022). Our study employs a meticulously controlled experimental design encompassing two distinct tasks, rating and cloze tasks. It adequately assesses whether the two types of anaphoric nominals with and without *ku* ‘that/the’ exhibits a parallel dichotomy akin to the strong (for anaphoric familiarity) and weak definites (for situational familiarity) observed in German, an article language (Roberts 2003, Schwarz 2009, 2013).

Employing a sophisticated experimental design comprising both rating and cloze tasks, we explore Korean speakers’ preferences and the underlying factors influencing their choices between the two forms of anaphoric nominals. To be specific, in the rating task, 66 Korean participants were eventually instructed to evaluate the naturalness of 48 complex sentences (32 bridging sentences plus 16 fillers) with demonstrative (anaphoric) NPs, which were sophisticatedly constructed in terms of two conditions, types of bridging references (Condition 1) and types of subject markers (Condition 2). On the other hand, in the cloze task, which was designed to elicit more spontaneous language production, another 76 Korean participants were properly recruited to fill in the blanks with the demonstrative *ku* ‘that/the’ if necessary.

Notably, our findings obtained from mean scores reveal a consistent and intriguing fact: a heightened preference for demonstrative NPs, especially, in [product-producer] bridging contexts, with a significant influence attributed to the presence of topic subject. This preference stands in contrast to [whole-part] bridging contexts, suggesting a subtle interplay between types of bridging references, types of subject markers, and types of anaphoric choices. Statistical analysis through ANOVA underscored a significant effect of Condition 2 on the fill-in-the-blanks as well as the naturalness ratings, even though Condition 1 did not yield statistically significant differences. This divergence adds intricacy to our understanding of the sophisticated interplay between linguistic conditions and anaphoric choices.

We interpret this complex phenomenon within the broader theoretical framework of economy, familiarity, and pragmatic considerations. The implications of our study extend beyond Korean linguistics, contributing valuable insights into the universality of anaphoric strategies and the dynamic interplay between language structure and discourse coherence.

Selected references

- Clark, Herbert. 1975. Bridging. In Schank, R. and B. Nash-Webber. eds., *Theoretical Issues in Natural Language Processing*, 169-174. New York: Association for Computing Machinery.
- Kang, Arum. 2021. “Marking Definiteness in an Articleless Language: The Role of the Domain Restrictor KU in Korean.” *Language and Linguistics* 22-2, 302-337.
- Lee, Sang-Geun, Yoo Jeong Jeong, and Jairim Jeong. 2022. “A Study on the Variations of Korean Demonstrative NPs of Bridging References.” *Korean Journal of Linguistics* 47-2, 317-345.
- Roberts, Craige. 2003. “Uniqueness in Definite Noun Phrases.” *Linguistics and Philosophy* 26, 287–350.
- Schwarz, Florian. 2009. *Two Types of Definites in Natural Language*. Amherst, MA: GLSA Publications.
- Schwarz, Florian. 2013. “Two Kinds of Definites Cross-linguistically.” *Language and Linguistics Compass* 7, 534-559.

Distinction Between Event and Process for Metaphysical-Cum-Semantic Rules of the Progressive

강민지 · 이예식
(경북대학교)

It is well attested in the literature of metaphysics that occurrents are categorized into two types, namely process and event. Such a metaphysical distinction is reflected on the aspectual differences in English. Specifically, the former is expressed in the imperfective or progressive; the latter is the perfective as shown below:

- (1) a. Mary was crossing the street when the truck hit her.
b. Mary crossed the street.

(1a) describes the occurrent of Mary crossing the street which was going on in the reference time was interrupted or ceased in the middle; in contrast, (1b) quantifies an occurrent of Mary getting across the street, which culminated in the reference time. Occurrents like the one referred to by (1a) and by (1b) are called processes and events, respectively, in the literature. In order to deal with processes denoted by the progressive, previous works have made use of the notion of “HOLD” (Parsons 1989), “IP” or “In Progress”(Szabó 2008), “Stage” (of the culminated event, Landman 1992), and many similar yet nuanced notions in the literature of the semantics of the progressive. One of the problems concerning these notions is that they are not specified or elaborated enough. In other words, what do they mean metaphysically?

Without such a metaphysical elaboration would their analyses remain unsatisfied. One of the current study’s aims is to elaborate the notion of process and to reflect it on the semantic rules of the progressive.

To this, the current study will compare two major views of processes to present the major characteristics of processes which make them stand out from events. In Stout (2016), he conceives processes as continuants. In other words, processes are present in their entirety like such normal objects as books, trees, dogs, and like. This means, in turn, they have no comparable temporal parts. As for events, he views them as perdurants having different temporal parts. In his ontology, processes and events are separately exist. Arguing against such views as Stout’s, Steward (2013) treats processes as perdurants with temporal parts like events, but they are different from events in the following sense: processes can get bigger or more complex by adding more temporal parts to the original ones without altering their identity. In this sense, Steward’s view of processes is aligned with the analogical idea that stuff is to things as processes to events. Meanwhile, in Link (1998), similarly to her view, he sees processes as consisting of temporal parts, which are homogeneous in terms of granularity varying contextually. It is worth mentioning that he takes a reductionist view by treating even continuants as consisting of underlying processes or their temporal slices. The current study insists on a more or less similar view: Processes consist of homogeneous temporal parts or actions whose homogeneity is found not in physics but in thought. Furthermore, they participate in generating their corresponding events as underlying processes.

Temporal measures such as “for-time” adverbials would play role as a principle of individuation of processes. People think as “count-quantified” the chunks of processes cut up by the principle of individuation. Accordingly, the current study sees events as supervenient on their underlying processes. Within the framework of the current study, there are no events going on, but only processes can. In this sense, events are like fossils of their underlying processes, which are found as subevents. It is possible,

however, to perceive their presence when they participate in the causation of some state as constituents of the causing process. To set up a model to accommodate the facts about processes and events discussed above, the current study proposes the following building blocks, which constitute the mereological model or structure:

- (2) a. Model: $\langle A, E, P, T, \leq, <, \cup, \triangleright \rangle$, where A: set of objects; E: set of events; P: set of processes, T: set of times, and \leq : a partial order, the “part of” relation; $<$: temporal precedence relation; \cup : sum or fusion relation; \triangleright : constitution relation. In this model, ∞ , the relation of overlap can be derived from the “part of” relation.

Due to the partial ordering between events, processes or times, this model allows for the join semilattices of events, processes, and times. As in Link (1998), these lattices are nonatomic. In other words, the current study will remain neutral concerning atomicity. Furthermore, we are also neutral about the denseness of time. Within this model, it is also possible to derive the following corollaries about processes and events:

- (3) a. $\forall e [e \in E \rightarrow \exists p [p \in P \wedge p \triangleright e]$
 b. $\forall p, p' \exists \theta [p, p' \in P \wedge \theta(p) \wedge \theta(p') \wedge \theta(p) \cup \theta(p') \rightarrow p \cup p']$
 c. $\forall p, e [p \triangleright e \rightarrow \exists \theta, \theta' [\theta(p) \wedge \theta'(e) \wedge \text{Gen}(\theta, \theta') \wedge \tau(p) < \text{Fin}(\tau(e))]]$
 d. $\forall \theta, \theta', e, e' [\text{Gen}(\theta, \theta') \wedge \theta(e) \wedge \theta'(e') \rightarrow \theta = \theta' \vee e \in \text{G-chain}(e')]$
 e. $\forall e', e'' [e', e'' \in \text{G-chain}(e) \rightarrow e', e'' \rightsquigarrow e \wedge [\tau(e') < \tau(e'')] \vee [\tau(e') < \tau(e')]]$, where “ \rightsquigarrow ” stands for a causal flow, a primitive relation in this model.
 f. $\forall p [p \in P \rightarrow \exists E' [E' \subseteq E \wedge \cup E' = p \wedge \forall e, e' \in E \rightarrow \tau(e) < \tau(e') \vee \tau(e') < \tau(e) \vee \tau(e) \infty \tau(e')]]$

(3a,c) state that all events are composed of an underlying process, whose property can generate the property of the event in question. (3b) states that the sum or fusion of homogeneous parts results in the process of the same type. In addition, the last conjunct of (3c), “ $\tau(p) < \text{Fin}(\tau(e))$ ” means that the underlying process ceases at least before the final moment of the event run time.

Armed with the model and the corollaries, the current study defines the semantics of the progressive operator in the progressive sentences as follows:

- (4) $[[\text{PROG}]] = \lambda P \lambda i \lambda C \forall k [k \in R(i, P, C) \rightarrow \exists e' [P'(e') \wedge \text{overlap}(\tau(e'), k) \wedge C \vdash \text{Gen}(P', P)]]$

(4) means that the progressive operator relates three things, a property denoted by the verbal phrase, a reference time, and a context, where the progressive is being asserted. One thing worth mentioning is the relation “R” for a regular partition adopted from Deo (2011), which takes a reference time interval and returns the cells of the partition of the reference time interval. Of course, each cell is characteristic of the property of the ongoing process. The last conjunct “ $\vdash \text{Gen}(P', P)$ ” means the context entails $\text{Gen}(P', P)$.

In this presentation we will show that our ways of viewing processes and events and the metaphysical-cum-semantic rules of the progressive can provide a better analysis of the progressive than previous analyses did. The data given in (5) exemplify the problems of the entailment puzzle and existence failure, and interruption and non-interruption, conjunction puzzle and others:

- (5) a. Mary was building a house. \Rightarrow a'. Mary built a house.
 b. Mary was walking in the park \rightarrow b'. Mary walked in the park.
 c. Mary was swimming across the Atlantic. \Rightarrow c'. Mary swam across the Atlantic.

- d. George is dying from a gunshot and heart disease.
 d'. *George can die from a gunshot and heart disease.

References

- Deo, Ashwini 2011. Unifying the imperfective and the progressive: Partitions as quantificational domains. *Linguistics & Philosophy* 32, 475–521.
- Dowty, David 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Galton, A. 2008. Experience and History: Processes and their Relation to Events. *J. Log. Comput.*, 18, 323–340.
- Steward, H. 2013. ACTIONS AS PROCESSES. *Philosophical Perspectives*, 26, 373–388.
<http://www.jstor.org/stable/23324614>
- Landman, Fred 1992. 'The Progressive', *Natural Language Semantics* 1: 1–32.
- Link, Godehard. 1988. Generalized quantifiers and plurals. In Peter Gardēnfors (ed.), *Generalized Quantifiers: Linguistic and Logical Approaches*, Dordrecht: Kluwer
- Mourelatos, A. P. D. 1978. Events, Processes, and States. *Linguistics and Philosophy*, 2(3), 415–434.
<http://www.jstor.org/stable/25000995>
- Parsons, T. 1989. The Progressive in English: Events, States and Processes. *Linguistics and Philosophy*, 12(2), 213–241. <http://www.jstor.org/stable/25001338>
- Stout, R. 2016. The Category of Occurrent Continuants. *Mind*, 125(497), 41–62.
<http://www.jstor.org/stable/43904705>
- Szabo, Z. G. 2008: *Things in progress*. *Philosophical Perspectives*, 22, 499 – 525.

영어 전치사 to 에 대한 단상 - 전치사 여격 구문을 중심으로 -

장경철
(부산대학교)

현대 표준영어에서 전치사 여격 구문(prepositional dative construction)의 가장 두드러진 특징은 전치사의 사용과 그에 따른 어순일 것이다. 예를 들면, 아래 (1)과 같이 이중 목적어 구문(double object construction)과 대조해보면 전치사 여격 구문에서는 전치사 to가 (간접 목적어) Mary와 결합하여 전치사구(PP)를 이루고 문미에 온다. 이와 함께, (직접목적어) candy는 동사(give) 바로 다음에 위치한다.

- (1) a. John gave candy to Mary. (전치사 여격 구문)
b. John gave Mary candy. (이중 목적어 구문)

이와는 반대로 일부 선행연구에서는 이 두 구문의 유사한 점을 논항구조에서 찾는다 (Goldberg 1995; Kim & Michaelis 2020). 아래 (2)와 같이 모두 세 개의 논항이 필요하고 각각의 논항은 행위자(agent), 수혜자(recipient), 대상(theme)의 의미역을 갖는다는 것이다.

- (2) a. give: <NP:AGENT, NP:THEME, PP:RECIPIENT> ((1a)의 논항구조)
b. give: <NP:AGENT, NP:RECIPIENT, NP:THEME> ((1b)의 논항구조)

하지만, 문제는 수혜자(recipient) 논항의 통사범주가 상이하다는 점에 있다. (2a)에서는 전치사구(PP)가 되고, (2b)에서는 명사구(NP)가 된다. 수혜자가 명사구가 의미하는 사람이나 그에 준하는 개체라면 납득이 되지만 전치사구가 뜻하는 방향이나 경로 자체가 직접적인 수혜자가 될 수는 없을 것이다. 이 점을 고려하면 전치사는 해당 구문에서 무슨 역할을 하고 있는지 묻지 않을 수 없다. 본 발표에서는 Chang (2019)의 입장에서 아래 (3)과 같이 전치사 to가 동사와 함께 복합술어(complex predicate)를 구성한다고 주장할 것이다.

- (3) give to: <NP:AGENT, NP:THEME, NP:RECIPIENT/GOAL> ((2a)와 대조)

그리고 이런 입장을 지지하는 방향으로 논의할 것이다.

참고문헌

- Chang, Kyungchul. 2019. A Complex Predicate Approach to the Caused-Motion Construction in English. *현대문법연구* 104: 77-92.
- Goldberg, Adel. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: The University of Chicago Press.
- Kim, Jong-Bok and Laura A. Michaelis. 2020. *Syntactic Constructions in English*. Cambridge: Cambridge University Press.

베트남 화자가 산출한 한국어 단모음에 대한 한국어 모어 화자의 지각연구

이주희
(경희대학교)

본 연구는 베트남인 한국어 초급 화자가 발화한 한국어 단모음에 대한 한국인 모어 화자의 지각 연구로써, 베트남인 초급 화자의 한국어 모음 산출(Lee 2023)과 한국어 모음에 대한 베트남인 화자의 모음 지각(이주희 2023)과 관련된 후속 연구이다. 선행연구에서는 한국어를 목표 언어로 베트남인 초급 화자가 산출한 모음의 음성적 특성(Lee 2023)과 한국어 모어 화자가 산출한 한국어 모음에 대하여 베트남인 초급 화자가 어떻게 지각하고 베트남어의 모음에 범주화(이주희 2023)하는지에 집중했다면, 본 연구에서는 베트남인 초급 화자의 한국어 모음 산출을 자극으로 설정하고 한국인 모어 화자가 이 자극들에 대하여 어떻게 지각하는지 알고자 한다.

모국어(L1)의 음운적 범주와 관련된 음성적 인지는 목표어(L2)의 학습에 지대한 영향을 준다는 것은 널리 밝혀진 사실이며(Flege 1987, 1991, 1995, Best 1995, Best and Tyler 2007) 이와 관련된 다수의 연구가 있었지만 주로 영어를 포함한 서구어를 목표 언어로 설정하고 학습자 모어의 관점에서 두 언어 사이의 음운 및 음성적 차이에 기인한 지각 연구가 수행되었다(Tsukada et al. 2005, Yun 2005, Hwang and Lee 2012, Heo and Park 2012, Lee and Shin 2015). 최근에는 한국어를 목표어(L2)로 베트남인 한국어 학습자를 대상으로 한 다양한 산출 및 지각 연구가 보고되었다(박시균, 이지영 2017, 정민환, 이나라 2017, 장혜진 2018, 권기현 2018, 류연이 2019, 박시균, 류연이 2019, 김라엘 2020, 이주희 2023a, 2023b, Lee 2023). 이들의 연구는 베트남인 화자를 대상으로 한국어와 베트남어의 단모음 산출과 지각 실험을 바탕으로 두 언어 사이에서 드러난 모음의 음운적 차이와 음향음성학적인 특성을 함께 논의하며, 베트남어 화자가 한국어 모음을 듣고 베트남어 모음의 범주에서 어떻게 지각하고 범주화하는지 논의되었다. 그러므로 한국어 학습자 측면의 산출과 지각에 논의가 집중되어 있다고 볼 수 있다.

본 연구의 목표는 한국어 모어 화자를 대상으로 베트남인 초급 화자가 산출한 한국어 모음에 대한 모어 화자의 인지적 범주를 알고자 하는 것이다. 따라서 본 실험 결과를 통하여 한국어 모어 화자가 인지할 수 있는 각 단모음의 음향음성학적인 범주를 추가적으로 논의할 수 있을 것이다.

참고문헌

- 김라엘. 2020. 「숙달도에 따른 베트남인 학습자의 한국어 단모음에 대한 산출 양상 연구」, 고려대학교 교육대학원 석사학위논문.
- 권기현. 「한국어-베트남어 단모음 대조를 통한 난이도 위계 설정」, 『한국언어문화학』 15(2), 한국언어문화학회, 2018, 93-123 면.
- 류연이. 2019. 「베트남인의 학습자의 한국어 단순모음 습득에 대한 실험음성학적 연구」, 군산대학교 교육대학원 석사학위논문.

- 박시균, 류연이. 「베트남인 학습자의 한국어 단모음 인지와 산출 연구」, 『언어학』 85, 2019, 139-72 면.
- 이나라, 정환민. 「베트남인 초급 학습자가 산출한 단모음에 대한 한국어 모어 화자의 지각」, 『이중언어학』 78, 이중언어학회, 2020, 87-116 면.
- 이주희. 「베트남 화자의 한국어 단모음에 대한 음성학적 연구: 학문 목적 수준의 여성 화자를 중심으로」, 『언어』 48(1), 한국언어학회, 2023, 129-58 면.
- 이주희. 「베트남 화자의 한국어 모음 지각 연구」, 『음성·음운·형태론 연구』 29(1), 한국음운론학회, 2023, 105-36 면.
- 장혜진. 「베트남인 고급 학습자의 한국어 단모음에 대한 실험음성학적 연구」, 『한국어학』 80, 한국어학회, 2018, 211-34 면.
- 정민환, 이나라. 2017. 「베트남인 초급 한국어 학습자의 단모음 발음에 대한 실험음성학적 연구 -여성 화자를 중심으로-」, 『이중언어학』 67, 이중언어학회, 2017, 285-307 면.
- Heo, Younhyon and Hanyong Park. 「American Learners' Identification of Korean Vowel Categories」, 『Harvard Studies in Korean Linguistics XIV』 MA: Harvard University, Cambridge, 2012, 77-84.
- Hwang, Hyosung and Ho-Young Lee. 2015. 「The effect of high variability phonetic training on the production of English vowels and consonants」, Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK.
- Lee, Juhee. 「Vowel production from Vietnamese learners of Korean: with reference to the elementary level」, 『International Language and Literature』 54, The International Society of Language and Literature, 2023, 255-280.
- Lee, Shinsook and Dong-Jin Shin. 「Estimating Korean EFL listeners' perception of English vowels with reference to cross-language labelling」, 『Studies in Phonetics, Phonology and Morphology』 21(2), The Phonology-Morphology Circle of Korea, 2015, 297-321.
- Tsukada, Kimiko, David Birdsong, Ellen Bialystock, Molly Mack, Hyekung Sung and James Flege. 「A developmental study of English vowel production and perception by native Korean adults and children」, 『Journal of Phonetics』 33, 2005. 263-290.
- Yun, Yungdo. 「Korean listeners' perception of English /i/, /ɪ/, and /e/」, 『Speech Sciences』 12(1), Korean Society of Speech Sciences, 2005, 75-87.

일본인 한국어 학습자의 의문문 문말 억양에 나타나는 모어 간섭

우에다 아스카 · 주현희
(부경대학교)

일본어와 한국어는 통사적으로 유사하기 때문에 일본인 한국어 학습자는 다른 언어권 학습자들에 비해 한국어 습득이 비교적 뛰어나다. 그러나 학습자들이 고급 수준에 도달하였음에도 억양은 여전히 부자연스러우므로 보다 한국어다운 한국어를 구사하기 위해서 운율적인 요소의 교육이 필요하다.

기존의 연구들이 의문문 문말의 상승 여부, 또는 상승의 정도를 보고 일본인 한국어 학습자의 억양이 부자연스러운 원인을 찾는 것이었다면, 본고에서는 학습자 억양의 부자연스러움의 원인을 모어 간섭 즉, 단어 악센트에서 찾아보고자 한다.

악센트나 억양은 문장 단위로 발음을 교육할 때 고려해야 할 중요 요소이다. 일본어의 경우 각각의 단어는 고유의 악센트 유형을 가지며, 예컨대 ‘雨(ame), 鯨(ame)’처럼 같은 음소 계열 단어라고 하더라도 악센트에 따라 다른 의미를 나타낸다. 한편, 문장은 문말, 구말 억양에 따라서 문장의 의미나 역할이 변화된다. 악센트나 문말 억양이 틀릴 경우, 듣는 사람에게 부자연스러운 느낌을 줄 뿐만 아니라 의미 혹은 의도가 달라질 수도 있다.

이러한 맥락에서 본고에서는 일본인 한국어 학습자의 문말 억양 실현 양상을 질적인 측면에서 살펴보고자 한다. 피험자는 한국 거주 기간이 7~10년 이고, 고등 교육기관에서 학습한 30대 일본인 여성 학습자들이다. 표준 발화 제공자는 서울 출신의 20대 한국인 여성이다. 문말 억양에서도 모어의 간섭이 나타나는지, 문말에 오는 단어의 악센트를 중심으로 살펴볼 것이다. 이를 위해 악센트 차이가 있는 일본어 단어로 구성된 한국어 의문문을 읽었을 때 실현되는 악센트의 특징을 살펴보고 교육 방안을 제안하고자 한다.

악센트 차이를 살펴본 결과 일본어 HLL 단어도 LHH 단어도 한국어로 발음할 때는 악센트가 전이되는 모습을 관찰할 수 있었다. 특히 일본어 HLL 인 한국어 단어에서는 저성조군에서 시작한 단어들도 일본어와 같이 고조로 시작한 것을 볼 수 있었다. 그러나 일본어 악센트가 긍정적 전이를 하여 한국어 모어 화자와 같은 유형으로 나타나지만, 단어 내의 음높이 폭에서 차이가 나타났다. 한국인 모어 화자의 경우 완만한 곡선으로 나타나는 반면, 일본인 한국어 학습자의 경우 단어 내에서 음높이가 급격하게 달라졌다.

이처럼 일본어 단어의 악센트가 긍정적으로 전이가 되었다고 하더라도 단어 내의 급격한 음높이 차이로 인해 부자연스럽게 들릴 수 있음을 알 수 있었다. 또한 한국에서 오래 거주하거나 한국어 고급 수준까지 도달한 학습자라 하더라도 일본어의 악센트가 그대로 남아 있는 경우가 있음을 알 수 있었다. 의문문을 발화할 때 문말 억양의 상승 정도 뿐만 아니라 단어 악센트 차이도 외국인 같은 발음, 어색함을 느끼게 하는 원인이 될 수 있다.

따라서 이 결과를 토대로 한국어 교육에 있어서 세 가지 방안을 제안하고자 한다. 첫째로 한국어는 음의 높낮이로 단어를 분별하는 것이 아니라 음의 길이로 분별한다는 것을 알려준다. 두 번째로 한국어 강세는 주로 첫 음절에 오지만 3음절 이상인 낱말에서 첫 음절이 개음절인 경우 둘째 음절에 강세가 온다는 것을 제시한다. 한국어에는 강약형과 약강형의 두 가지 형태의 리듬이 있으므로 예를 제시하면서 문장 당위로 강약 리듬을 연습한다. 세 번째로는 일본인 학습자의 의문문 억양은 의문문이라는 것을 의식한 탓에 마지막에 급격히 상승하는 경향이 나타난다. 마지막을 급격히 올리는 것으로 강세구의 음절마다의 높낮이가

현저해지는 것도 역량이 부자연스럽게 느껴지는 원인의 하나라고 할 수 있다. 따라서 학습자에게 음의 높이가 왔다갔다 하지 않도록 주의하는 것도 필요하다.

The Language of Modesty and Confidence in Doctoral Dissertations: A corpus-based cross-disciplinary comparison between Korean L2 and English L1 writers

Hoe Kyeung Kim and Sang-Geun Lee
(Binghamton University, Korea University)

Despite the emergence of doctoral dissertations as a distinct research genre, there is limited understanding of how dissertation writers from different language backgrounds express levels of modesty and confidence. Notably, comparative studies on the utilization of English hedges and boosters by Korean EFL learners have predominantly focused on a single academic discipline. This study aims to bridge this gap by examining the patterns of hedges and boosters employed by Korean-speaking (KS) and English-speaking (ES) dissertators in two distinct disciplines: English Education and Biology. To achieve this, we analyzed a total of 120 doctoral dissertation discussions, with 60 written by ES and 60 by KS, distributed across the two fields. The hedge and booster patterns between the KS and the ES in each discipline showed differences. Our findings revealed significant distinctions in hedge and booster usage between Korean-speaking and English-speaking writers within each discipline. Moreover, cross-disciplinary comparisons revealed disparities not only between the two language groups but also within each group. It was observed that ES in Biology employed hedges more frequently than their counterparts in English Education, challenging the claim that hedges are primarily utilized in the field of humanities. Qualitative analysis of the hedging practices within each discipline further revealed a clear difference in the deployment of hedges and boosters between the two language groups. From the perspective of L2 education, these findings suggest no overgeneralization of hedging strategies in understanding Korean EFL learners' dissertation writing, particularly within the scientific domains. Further studies on the use of hedging and booster devices across diverse academic disciplines are recommended.

Semantics of people and places: A corpus-assisted assisted NSM analysis of Tagalog affix *taga-*

Nicole Charis M. Pontanilla and Eugene Chung
(Korea University)

This study investigates the polysemy of the Tagalog affix *taga-* from a cognitive semantic perspective. Specifically, the study aims to 1) uncover the various meanings of *taga-*, 2) identify its prototypical meaning, and 3) determine how and why the peripheral senses emerged from the core meaning. By employing a Natural Semantic Metalanguage (NSM) analysis on the data collected from Sketch Engine's Tagalog Web 2019 (tITenTen19), Tagalog.com, and Tagalog-Korean Corpus of National Institute of Korean Language's 모두의 말뭉치, five senses of *taga-* were identified:

- (1) Demonym Taga-1: "Di ko alam may **taga Pilipinas** palang di alam yung turon." (*I didn't know that there are **people from the Philippines/Filipinos** who don't know what turon is.*)
- (2) Residential Taga-2: "...ang biktima na si Lemuel Bellaza, 61 anyos, may asawa, magsasaka at **taga Barangay Crossing Rubber** sa nasabing bayan." (*...the victim, Lemuel Bellaza, 61 years old, married, a farmer, and a resident of **Barangay Crossing Rubber** in the said town.*)
- (3) Workplace Taga-3 : "Sa pananaw ni ACT Teachers Rep. France Castro mukhang hindi ginamit ng mga **taga-BSP** (Bangko Sentral ng Pilipinas) ang kanilang common sense sa pagpapalabas ng 5 peso coin na halos kaparehas ng piso." (*In the perspective of ACT Teachers Rep. France Castro, it seems that the **people from/working at the Central Bank of the Philippines** did not use their common sense in releasing the 5 peso coin that closely resembles the one peso coin.*)

In addition to these, *taga-* can also express membership to a certain group (Membership-based Taga-4) and occupation (Occupational Taga-5) as demonstrated by the forms "taga-simabahan" (*church people*) and "tagaluto" (*a cook*), respectively. Overall, the findings reveal that Taga-1 functions as the prototype sense as evidenced by the presence of the semantic molecule "place" across all the semantic explications of the peripheral senses. In essence, Taga-2 to Taga-5 successfully convey their unique meanings of residence, workplace, group membership and occupation by drawing from the central idea of belonging to a place that Taga-1 primarily carries. Notably, the workplace sense of *taga-* (Taga-3) plays a vital role in connecting the demonym and occupational sense of *taga-*. Motivated by conceptual coherence, conceptual extension, specificity, and expression of different semantic roles, the polysemous nature of *taga-* reflects the human mind's tendency to categorize concepts based on shared features and relationships.

K-pop 외국 팬들의 한국어 외래어 사용 현상 연구

황경진
(고려대학교)

2010년대 중후반 이후 k-pop이 전 세계적으로 선풍적인 인기를 끌면서 k-pop의 중심 언어 중 하나인 한국어 역시 전 세계적으로 널리 알려졌다. 많은 k-pop 팬들이 한국어로 된 노래를 따라 부르고 한국어로 된 가수들의 동영상을 보면서 한국어의 영향을 많이 받게 되었다. 그리고 그에 따라 한국어를 배우지 않았더라도 한국어로 된 단어를 외래어로 사용하는 빈도가 증가하였다. 본 연구는 트위터를 중심으로 하여 영어 및 스페인어 화자들이 작성한 k-pop에 대한 트윗을 수집하고 그를 분석하여 한국어 단어의 사용 양상에 대해 연구하고자 한다. 본 연구는 우선 파이썬 크롤링을 통해 2022년과 2023년에 작성된 세븐틴, 방탄소년단, 엔시티 드림, 몬스타엑스, 블랙핑크, 아이브, 트와이스, 에스파와 같은 아홉 아이돌 그룹 관련된 트윗 중 영어와 스페인어로 작성된 것을 추출하여 데이터를 수집하였다. 그리고 그 수집한 데이터 중 한국어가 외래어처럼 사용된 사례들을 모아 분석하였다.

데이터 분석 결과 한국어 외래어 사용은 크게 세 가지 유형으로 나뉘어졌다. 가장 많은 유형 중 하나는 한국어 발음을 영어로 써 그대로 사용하는 것이다. “ssibal(시발)”과 같은 욕설이나 “sul(술)”과 같이 한국어 단어의 원형을 영어로 작성하여 사용할 뿐만 아니라 “sming(스밍, 스트리밍의 준말)”, “homema(홈마, 홈페이지 마스터의 준말)”, “akgae(악개, 악성 개인팬의 준말)”, “saseng(사생, 사생팬의 준말)” 등 한국인 k-pop 팬들이 사용하는 은어를 그대로 영어로 차용하여 사용하는 경우도 발견되었다. 두 번째 유형은 타 문화권에는 없는 한국어 문화에서 존재하는 언어를 그대로 사용하는 것이다. 주로 호격에서 많이 나타나는데 “hyung(형)”, “unnie(언니)”, “ya(야)”, “-ah(-아)”, “-nim(-님)”, “sajangnim(사장님)” 등과 같이 한국어에서만 쓰이는 호칭어가 해당되었다. 뿐만 아니라 “sunbae(선배)”, “hoobae(후배)” 등 한국 문화에만 존재하는 사회적인 용어 역시 그대로 사용되었다. 마지막 유형은 한국어 단어를 영어로 번역하여 사용하는 경우로, 그를 의미하는 별도의 영어 단어가 있음에도 불구하고 한국어 단어를 직역한 결과를 사용하는 것이다. 대표적으로 “media play(언론 플레이의 번역)”가 있다. 영어로 “media manipulation”이라는 단어가 있음에도 불구하고 해당 단어를 사용한다. “do comeback(컴백하다)” 역시 “release new album”이라는 단어가 있음에도 불구하고 한국식 표현을 번역한 단어를 사용한다.

위와 같은 한국어 외래어 사용에서 특징적인 점은 이러한 사용을 국적과 모국어에 상관없이 사용한다는 점이다. 그리고 이러한 한국어 외래어 사용이 의사소통의 수단으로 기능한다. 스페인어 화자 뿐만 아니라 인도네시아, 일본, 브라질 등 다른 국가의 k-pop 팬들도 위와 같은 표현을 사용함으로써 서로 소통하는 것이 관찰되었다. 즉, k-pop의 영역에서 한국어가 일종의 공용어처럼 사용되는 것이다. “no media play, please”라는 인도네시아 팬의 트윗에 “wae media play? sming hwaiting”이라고 답글을 단 영어 화자 팬의 멘션이 대표적인 예이다. 뿐만 아니라 “media play”나 “do comeback”과 같은 예시의 경우, 언어권에 상관없이 영어로 직역한 단어를 사용한다는 점 역시 특징적이다.

본 연구는 이러한 분석을 통해 k-pop 문화의 확산으로 인한 한국어의 사용 증가라는 긍정적인 측면을 확인했다는 점에서 의의가 있다. 뿐만 아니라 k-pop의 문화 내에서 한국어 표현들이 공용어처럼 쓰여 k-pop을 즐기는 외국 팬들이 소통의 수단으로 사용한다는 점을 확인했다는 점에서 역시 의의가 있다.

On the Lexical Meaning of *Seltukha*-‘Persuade’ in Korean

이주원
(전주대학교)

This paper offers a unified analysis of the Korean verb *seltukha*-‘persuade’ used in two constructions. Challenging traditional assumptions, I propose a non-control analysis for the *seltukha*-construction with a *tolok*-clause, considering it as an adjunct. Additionally, I investigate the *seltukha*-construction with a *ta-ko*-clause, which is argued to function as a complement. It is also demonstrated that both of these clauses can coexist within a *seltukha*-construction, supporting the idea of a single lexical item for the non-control verb *seltukha*-‘persuade’ in Korean, unlike English. The proposed semantic analysis reveals *seltukha*-‘persuade’ as a lexical causative, wherein the object’s belief in the content expressed by a *ta-ko*-clause is the inherent result, and a *tolok*-clause describes the subsequent outcome. I believe this study contributes to our better understanding of Korean grammar, challenges our understanding of “control” verbs, and establishes a foundation for exploring similar constructions in other languages.

Selected References

- Bresnan, Joan. 1982. Control and complementation. *Linguistic Inquiry* 13(3): 343-434.
- Choe, Hyon Sook. 2006. On (backward) object control in Korean. *Harvard studies in Korean linguistics* XI, 373-386. Kyunggi: Hanshin.
- Culicover, Peter W. and Ray Jackendoff. 2001. Control is Not Movement. *Linguistic Inquiry* 32(3): 493-512.
- Monahan, Philip J. 2003. Backward object control in Korean. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, Gina Garding & Mimu Tsujimura (eds), 356-369. Somerville MA: Cascadia Press.
- Polinsky, Maria. 2007. Object control in Korean: Obligatory vs. non-obligatory control. *ZAS Papers in Linguistics* 47, 197-217.
- Pollard, Carl J. and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications.
- Sag, Ivan A., Thomas Wasow and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*, (2nd edition). Stanford, CA: CSLI Publications

Acoustic and articulatory characteristics of Korean vowels /we/*, /wɛ/, /we/: A preliminary report

Dae-yong Lee¹, Jiyeon Song¹, Sahyang Kim^{2,1}, and Taehong Cho¹
(¹Hanyang Institute for Phonetics & Cognitive Sciences of Language (HIPCS),
Hanyang University, ²Hongik University)

This study examines the acoustic-articulatory features of three Korean diphthongal vowels—/we/* (‘외’), /wɛ/ (‘외’), and /we/ (‘외’)—within monosyllabic contexts, with and without the onset /h/, spoken by young Seoul speakers in their twenties. (/we/* refers to the diphthongized vowel stemming from the previous monophthong /ø/.) The findings suggest that these three diphthongs do not significantly differ in various phonetic aspects, indicating a substantial merger. However, there is evidence that these young speakers still maintain a subtle phonetic contrast between the ‘new’ diphthong (/we/) and one of the ‘old’ diphthongs (/wɛ, we/). The distinctions are further influenced by the onset context, making the results intricate. For instance, a difference between /we/* and /wɛ/ is noticed only in articulatory movement duration with the onset /h/, whereas a similar difference is observed between /we/* and /we/ only without the onset /h/. Furthermore, there is an asymmetry between acoustic and articulatory measures. Specifically, a more distinct contrast between /we/* and one of the ‘old’ diphthongs emerged in articulatory dimensions like movement duration and tongue position. Notably, unlike some reported cases of sound change in Seoul Korean, there is no indication of gender-related differences suggesting female speakers lead the merger. If anything, female speakers exhibit more distinctions between the ‘new’ and ‘old’ diphthongs compared to male speakers. These results are further discussed in their implications for the importance of multi-dimensional phonetic studies and the role of gender in understanding the nuanced extent of sound change.

The Effect of Age, Morphological Structure, and Syllable-Count on the Realization of Korean Nasal-Liquid Sequences

Rok Sim¹, Drew Crosby², and D. Eric Holt¹
(¹University of South Carolina, ²Korea University)

This paper investigates variability in the realization of Korean nasal-liquid (NL) sequences at stem boundaries and accounts for this variation using a Vestige Theory approach (Kostakis 2010, 2017), an offshoot of Optimality Theory (OT) (McCarthy & Prince 1995; Prince & Smolensky 2004). We show experimental evidence for frequency, syllable count, and age effects on the realization of NL sequences. We then present a theoretical model to account for these findings.

The direction of assimilation is complex (Kang 2002; Sohn 2006). Both (1) and (2) involve NL sequences; however, (1), a word composed of two bound morphemes, is realized as [ll] (i.e., regressive assimilation), while (2), a word consisting of a free morpheme and a bound morpheme, is realized as [nn] (i.e., progressive assimilation).

- (1) /nl/ → [ll]
- | | | | | |
|----------|-------------------|--------------|----------|-------------------|
| a. /kɪn/ | ‘diligent’ + /lo/ | ‘work’ | [kɪllo] | ‘labor’ |
| b. /pan/ | ‘counter’ + /lon/ | ‘discussion’ | [pallon] | ‘counterargument’ |
- (2) /nl/ → [nn]
- | | | | | | |
|------------|----------|--------|--------------|-----------|-----------------------------|
| a. /nokɪn/ | ‘Nokeun’ | + /li/ | ‘small town’ | [nokɪnni] | ‘small town (Li) of Nokeun’ |
| b. /dæhan/ | ‘Daehan’ | + /lo/ | ‘street’ | [dæhanno] | ‘Daehan Street’ |

Interestingly, (2), which is expected to only allow progressive assimilation, allows assimilation in both directions as in (3), although the degree of acceptability between (3a) and (3b) differs; [ll] in (3b) is less acceptable than in (3a), perhaps linked with frequency-effects.

- (3)
- | | | | |
|---------------|-----------|------------|-----------------------------|
| | (i) [nn] | (ii) [ll] | |
| a. /nokɪn-li/ | [nokɪnni] | [nokɪlli] | ‘small town (Li) of Nokeun’ |
| b. /dæhan-lo/ | [dæhanno] | ?[dæhallo] | ‘Daehan Street’ |

Variation in the directionality of assimilation is also observed in loanword compounds as in (4). Canonical progressive [nn] assimilation in loanword compounds is shown in the (i) forms (Sohn 2006), with regressive [ll] assimilation in the (ii) forms. We again observe differential acceptability between (4a) and (4b), perhaps due to the syllable count of the initial morpheme.


- (4) Loanword compounds (/n + l/)
- | | | | |
|----------------------------------|-------------|-------------|--------------|
| | (i) [nn] | (ii) [ll] | |
| a. /wən/ ‘one’ + /lum/ ‘room’ | [wənnum] | [wəllum] | ‘studio apt’ |
| b. /taun/ ‘down’ + /loti/ ‘load’ | [taunnodei] | ?[taullodi] | ‘download’ |

We conducted a production experiment with 22 native Korean speakers. To examine social and morphological effects, we built two logistic regression models with *realization* ([nn]/[ll]) as the dependent variable and either *frequency of the first morpheme* or *syllable count of the first morpheme* (one/two syllables) and their interactions with *age* as independent variables. Results reveal that words with high-frequency initial stems (e.g., *Daehan*, 3b) favored realization as [nn] ($p < 0.01$), as did words with two-syllable morphemes (e.g., /taun/, 4b) ($p < 0.01$). Additionally, older participants exhibited

more [nn] assimilation than younger participants ($p < 0.01$). In particular, loan word compounds with a one-syllable initial morpheme exhibited significantly greater age-based differences than two-syllable initial morphemes ($p < 0.01$).

To formally model this, we employ Kostakis' (2010, 2017) Vestige Theory model. Vestige Theory posits that, in accordance with language change, constraints are demoted in the grammar, but a vestige constraint is left behind that flashes on/off according to factors such as social context and word frequency. Specifically, we propose that the formerly high-ranked constraint, MAX-(nasal/coda), is demoted below MAX-(lateral). This constraint demotion leaves behind an output-sensitive vestige constraint. This output-sensitive vestige constraint probabilistically activates based on factors such as age: for older individuals, the probability of the vestige constraint being on is higher, and thus, [nn], is more likely to surface than for younger individuals whose vestige constraint is more likely to be off, and thus to realize [ll]. By employing this approach, we are able to successfully model the variation in NL sequences using the machinery of formal phonology and the methodology of variationist sociolinguistics.

(5)



Input: /nokɪn.li/	SYLLCON	MAX- (nasal/coda) _{vestige}	MAX- (lateral)	MAX- (nasal/coda)	MAX- (nasal)
Output: [nn] ~ [ll]					
a. nokɪn.li	*!				
≡ b. nokɪl.li		(*!)		*	*
≡ Vestige c. nokɪn.ni			*!		

References

- Davis, Stuart and Seung-Hoon Shin. 1999. The Syllable Contact Constraint in Korean: An Optimality Theoretic Analysis. *Journal of East Asian Linguistics* 8: 285–312.
- Kang, Hyunsook. 2002. On the optimality-theoretic analysis of Korean nasal-liquid alternations. *Journal of East Asian Linguistics* 11(1): 43–66.
- Kostakis, Andrew. 2010. Vestige Theory: Sociolinguistic evidence for output-output constraints. *Lingua* 120: 2476–2496.
- Kostakis, Andrew. 2017. Extrametricality in Latin. *Transactions of the Philological Society*, 115: 336-364. <https://doi.org/10.1111/1467-968X.12103>
- Kwon, Iksoo. 2014. Arbitrary assimilation of /nl/ in Korean. *Language and Linguistics* 64, 1-30.
- McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative identity. In Beckman, Walsh Dickey, and Urbanczyk (eds.), *Papers in Optimality Theory*, 249–384.
- Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.
- Sohn, Hyang-Sook. 2006. Morphological basis of sonorant assimilation in Korean. *Studies in Phonetics, Phonology and Morphology* 12(2): 321–338.

드라마 자료를 통해 살펴본 {네}, {예} 사용의 변이와 변화*

장교진 · 신지영
(고려대학교)

본 연구는 1990 년대와 2010 년대의 드라마 자료를 통해 응답 표현으로 사용되는 {네}와 {예}의 사용 양상을 살펴서 이 응답 표현이 시대에 따라 어떠한 변화 양상을 보이는지, 이러한 변화가 성별에 따라 어떠한 차이를 보이는지를 확인하고자 한다.

선행 연구에 따르면 {네}와 {예}는 성별어적 속성이 있고, 시기에 따라 사용 양상에 변화가 있는 것으로 보고된 바 있다. 대체로 남성은 {예}를, 여성은 {네}를 더 많이 사용하고 있으며, 과거에 비해 현재로 올수록 남성의 {네} 사용 비율이 늘면서 성별의 차이가 줄어들다는 것이 기존의 관찰이었다. 하지만 이 연구들은 실제 발화 자료보다 출석에 대한 응답 자료나 문자 메시지 자료를 기반으로 하고 있다. 이에 본 연구는 실제 발화에 가까운 드라마 자료를 분석해 시대별, 성별에 따른 응답 표현 {네}와 {예}의 사용 양상을 확인하고자 한다.

이를 위해 1990 년대 드라마와 2010 년대 드라마를 대상으로 각 시대를 대표하는 드라마를 4 편씩, 총 8 편의 드라마를 선정하였다. {네}와 {예}가 사용되는 장면을 두루 관찰하기 위해 각 시대를 대표하는 드라마에는 격식적인 장면을 주로 관찰할 수 있는 회사 드라마 2편과 비격식적인 장면을 주로 관찰할 수 있는 가족 드라마 2편이 포함되도록 하였다. 선정한 드라마의 5~9 회, 총 40 회 분량의 전사 자료를 분석하였다.

지금까지의 분석 결과를 정리하면 다음과 같다. 첫째, 1990 년대 격식적인 장면에서 남성과 여성 모두 {네}보다 {예}를 더 많이 사용하였다. 비격식적 장면에서 남성은 {예}를 더 많이 사용하였지만, 여성은 {네}를 더 많이 사용하는 경향이 관찰되었다. 이처럼 여성은 격식적 상황과 비격식적 상황에 따라 {네}와 {예}의 사용을 다르게 하고 있음에도 불구하고 {네}만이 여성어로 논의되어 온 것은 격식적인 장면에서의 여성의 발화에 덜 주목했기 때문으로 보인다. 둘째, 2010 년대에서는 격식적인 장면에서는 남성이 {예}를 여성이 {네}를 주로 사용하였다. 비격식적 장면에서는 남성과 여성 모두 {네}를 더 많이 사용하였다. 이는 남성이 {네}와 {예}를 모두 사용하되 격식적인 장면에서는 여성과 다르게 남성이 {예}를 선택하여 사용한다는 선행 연구의 결과와도 일치한다. 셋째, 과거 여성은 격식성 여부에 따라 전략적으로 {네}와 {예}를 다르게 사용하고 있었지만, 현재에는 오히려 남성이 격식성 여부에 따라 전략적으로 {네}와 {예}를 다르게 사용하고 있는 양상을 보였다.

본 연구는 드라마 자료를 통해 {네}와 {예}의 사용이 시대에 따라 변화하였으며 이러한 변화가 성별에 따라 차이가 있음을 확인하였다. 기존의 논의 들은 현장 시간만을 고려하여 {네}와 {예}의 공시적 변이를 밝히는 데에 집중하였다. 그러나 본 연구에서는 1990 년대와 2010 년대의 드라마 자료를 바탕으로 실제 시간을 고려하여 {네}와 {예}의 변화 양상을 직접적으로 살펴었다는 점에서 연구의 차별성과 의의를 찾을 수 있다.

*이 논문 또는 저서는 2022 년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5A2A03053537).

회사 내 호칭어 변화에 대한 연구: 1980 년대와 2020 년대 드라마 자료 분석을 바탕으로

박지은 · 신지영
(고려대학교)

본 연구는 1980 년대 직장 드라마와 2020 년대의 직장 드라마에서 사용되는 회사 내 호칭어를 관찰하고 비교함으로써, 호칭어의 사용이 직급이나 성별 등과 관련된 사회언어학적 변수임을 밝히고, 사회의 변화가 이러한 호칭어 사용에 영향을 미쳤음을 확인하고자 한다.

청자 대우법의 사용을 통해 나와 상대방과의 관계에 대한 인식을 드러내는 것처럼, 한국어 화자들은 호칭어를 통해서도 상대방과의 관계에 대한 인식을 드러낼 수 있다. 이때, 호칭어의 사용은 존대법보다도 더 유의미한 사회적 화 시일 수 있다. 청자 대우법의 사용은 문법적으로 필수적이지만, 호칭어의 사용은 그렇지 않기 때문이다. 꼭 사용하지 않아도 되는 호칭어를 사용한다는 것은 나와 상대방과의 관계를 언어라는 장치를 통해 적극적으로 드러내려는 화자의 사회적 의지의 표명이라고 할 수 있다. 누가 누구를 어떤 호칭어로 부르며, 호칭어를 많이 사용하는지를 관찰함으로써, 그 안에 담긴 화자들의 가치관과 사회적 인식을 엿볼 수 있을 것이다. 그리고 사회 변화로 인해 사회 구성원들의 가치관이 변화하였다면, 이는 여러 사회적 직시 표현들의 사용 양상에 그대로 반영될 것이다.

본 연구에서는 여러 호칭어 중 회사 내 호칭어에 초점을 두어, 1980 년대 직장 드라마와 2020 년대 직장 드라마에서 화청자의 직급, 성별 위계에 따라 호칭 표현이 얼마나 많이 사용되며, 이때 어떠한 호칭 표현이 사용되는지 비교 분석하고자 한다. 이를 통해 회사 내에서 상위직급자나 하위직급자에 대한 인식이 어떻게 변화되었으며, 특정 성별에 대한 인식은 어떠한 변화 양상을 겪었는지 엿보고자 한다. 가족 호칭어 등 여러 호칭 체계가 존재하겠지만 회사 내 호칭어를 연구 대상으로 삼은 것은 회사는 직급이 존재하는 만큼 위계 구조가 분명히 상정된 집단이기 때문이다. 회사 집단 내 여러 관계들의 호칭어 사용 양상이 어떻게 변화하였는지 관찰함으로써, 한국 사회 전반에 미친 수평화 및 평등의식의 물결이 ‘상사-부하’라는 사내 위계 관계에는 어떠한 영향을 주었으며 여성 직원에 대한 인식에는 어떠한 영향을 미쳤는지 관찰할 수 있을 것이다.



Presentation Materials

L3 영어 관계절과 Wh-이동 연구

인천대학교
발표일 : 2023.12. 9.
발표자 : 박유순

L1, L2, L3 통사의 변별적 자질

	L1 러시아어 L1 중국어	L2 한국어	L3 영어
어순	SVO	SOV	SVO
관계절	해전위 해후위	해후위	해전위
Wh-의문문	Wh-이동 Wh-원위치	Wh-원위치	Wh-이동



모국어와 제2언어가 제3언어의 통사습득에 미치는 영향: 영어의 관계절과 Wh-의문문 연구

1.2. 연구의 목적

- 본 연구에서 제3언어를 습득할 때 어느 모델이 영향을 끼치는가?
- 누적 향상 모델** (Cumulative Enhancement Model, Flynn, Foley, & Vinnitskaya, 2004)
L3와 같은 통사 구조를 가진 이전에 배운 언어 경험이 영향을 미치는지를 알아본다.
누적 향상 모델이 맞다면 L3와 비슷한 통사구조를 가진 L1이 L3 습득에 긍정적인 영향을 미칠 것이다.
- L2 지위 요소 모델** (L2 Status Factor Model, Bardel & Falk, 2007)
L2 한국어가 통사 구조가 다른 L3 영어 습득에 방해가 되어 L1이 긍정적인 영향을 미칠 수 있는지 알아본다.
L2 지위 요소 모델이 맞다면 L1과 L3가 유행론적으로 같더라도 L1이 L3 습득에 긍정적인 영향을 미칠 수 없을 것이다.

4. 연구 내용 및 방법

• 기존 및 본 연구의 실험 언어 조합

- 누적향상모델 **가. L1 ≠ L2 = L3 ≠ L1**
Flynn 외가 제안 **나. L1 ≠ L2 ≠ L3 = L1**
- L2 지위요소모델 **가. L1 ≠ L2 = L3 ≠ L1 (D/G 그룹)**
 나. L1 ≠ L2 ≠ L3 = L1 (E/N 그룹)
- 본 연구 **가. L1 ≠ L2 ≠ L3 = L1 (A 그룹)**
 다. L1 = L2 ≠ L3 ≠ L1 (B 그룹)

4. 연구 내용 및 방법

연구 대상자: 중도입국 청소년, 23명, 11 - 24세

관계절 실험

A 그룹: 6 / B 그룹: 6 = 12명

Wh-의문문 실험

A 그룹: 14 / B: 그룹: 9 = 23명

4. 연구 내용

A 그룹 : L1 ≠ L2 ≠ L3 = L1
 B 그룹 : L1 = L2 ≠ L3 ≠ L1

관계절과 Wh-의문문 실험

A 그룹 : 러시아어 ≠ 한국어 ≠ 영어 = 러시아어
 B 그룹 : 중국어 = 한국어 ≠ 영어 ≠ 중국어

4.2. 관계절의 연구 방법

가설과 예측

1. 누적 향상 모델을 따른다면, 이전에 습득한 언어 지식이 누적되어 **새로운 언어를 습득할 때 긍정적인 영향**을 미칠 것이다. A와 B 그룹의 모든 조건이 동일하다면 L3 영어와 해당 항목 개변수가 같은 **A 그룹이 L3를 습득하는데 B 그룹보다 촉진적인 역할을 할 것이다.**
1. L2 지위 요소 모델을 따른다면, **L2 한국어가 L1 전이를 차단하여 L3를 습득하는데 어려움을 겪을 것이다.** A와 B 그룹의 모든 조건이 동일하다면 L2와 L3의 해당 항목 개변수가 다르므로 **A와 B 그룹의 습득 결과는 비슷할 것이다.**

4.3. Wh-의문문의 연구 방법

가설과 예측

1. 누적 향상 모델을 따른다면, 이전에 습득한 모든 언어 지식이 누적되어 **새로운 언어를 습득할 때 중립을 유지하거나 새로운 언어 구조의 습득이 수월해진다.** A와 B 그룹의 모든 조건이 동일하다면 **A 그룹이 L3를 습득하는데 B 그룹보다 수행능력이 높을 것이다.**
2. L2 지위 요소 모델을 따른다면, L2 한국어가 L1이 L3에 전이되는 것을 방해하여 L3를 습득 하는데 어려움을 겪을 것이다. A와 B 그룹의 모든 조건이 동일하다면 **A와 B 그룹의 습득 결과는 비슷할 것이다.**

4.2. 관계절의 연구 방법

4.2.2.1. 틀고 따라 하기 과업 - 수동적

- 3가지 유형: a. 의미가 있는 어휘핵
b. 의미가 없는 어휘핵
c. 자유관계절
- 4가지 기능: S/S, S/O, O/S, O/O

4.2.2.2. 그림보고 단어 배치하기 - 능동적인 문법성 판단을 요구

- 3가지 유형 4가지 기능
- 실험 참가자들이 핵 위치를 정확하게 판단할 수 있는지 여부

4.3. Wh-의문문의 연구 방법

4.3.2. Wh- 유도 과업

- Wh- 유도 과업 1:
who(m), what, where, when, why, how
참가자들이 주어진 단어를 모두 사용하여 의문문 만들기
- Wh- 유도 과업 2: (논문 77쪽 참고)
참가자들이 그림과 B의 대답을 보고 A의 질문(의문문) 만들기

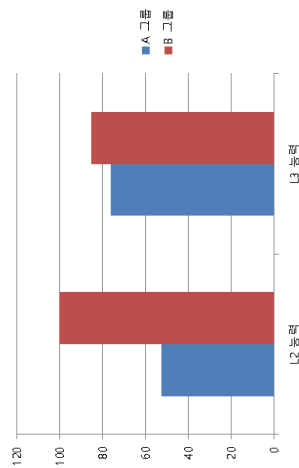
4.3. Wh-의문문의 연구 방법

4.3.2. Wh- 유도 과업

- 6가지 형태 (Kim, 2000, 2002)
 - a. Wh-(is/'s)-NP : L1 어린이 영어(암기된 형태)
 - b. Wh-S-V : L1 어린이 영어
 - c. Wh-Aux/do-S-V : 완성된 문장
 - d. Wh-V-S : L1 한국어, L2 영어
 - e. S-V-Wh : L1 한국어, L2 영어
 - f. S-Wh-V : L1 한국어, L2 영어

관계절: L2 / L3 능력

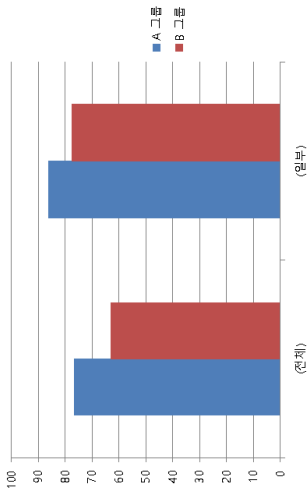
4.2.1. 연구 설계



- L2 능력: A 그룹(6명) 52.5점 / B 그룹(6명) 100점 (47.5점 차)
- L3 능력: A 그룹(6명) 75.83점 / B 그룹(6명) 85점 (9.17점 차)
- Bardel과 Sánchez(2017): 명시적 문법교육 → 메타언어 지식 발달

5. 실험 연구 결과 분석 및 논의 : 관계절

5.1.1. 듣고 따라 하기 결과 분석



- 전체: A 그룹(6명) 76.5점 / B 그룹(6명) 62.83점 (13.67점 차)
- 일부: A 그룹(2명) 86점 / B 그룹(3명) 77.33점 (8.67점 차)
- B 그룹: 공교육과 사교육에서 진중적인 문법 훈련을 받았음에도 불구하고 A 그룹보다 뛰어난 성과를 보이지 않았다.

5. 실험 연구 결과 분석 및 논의 : 관계절

5.1.2. 그림 보고 단어 배치하기 결과 분석:

해방향 매개변수의 민감성 : 주어 + 동사 + what

(48) 정문: **What** pushes Anna is a cute robot. P-SSc

문형: () () is a cute robot.

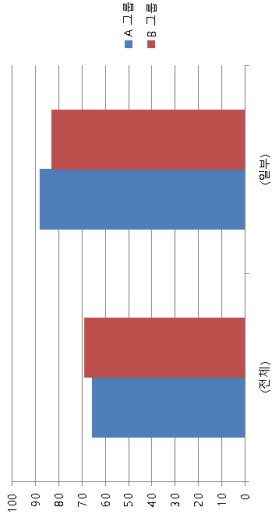
오류: Anna pushes **what** is a cute robot.

A 그룹 : 전체 오류 30개 중 8개, **26.67%**

B 그룹 : 전체 오류 26개 중 12개, **46.15%**

5. 실험 연구 결과 분석 및 논의 : 관계절

5.1.2. 그림 보고 단어 배치하기 결과 분석



- 전체: A 그룹(6명) 65.50점 / B 그룹(6명) 69점 (3.5점 차)
- 일부: A 그룹(2명) 88점 / B 그룹(3명) 83점 (5점 차)
- Bardele과 Sánchez(2017): 명시적 문법교육 → 메타언어 지식 발달

5. 실험 연구 결과 분석 및 논의 : 관계절

5.1.2. 그림 보고 단어 배치하기 결과 분석:

해방향 매개변수의 민감성 - 관계대명사의 위치는 정확

(49) 정문: Cookie Monster pulls **what** Big Bird pushes. P-Ooc

문형: Cookie Monster pulls () () () .

오류: Cookie Monster pulls **what** pushes Big Bird.

A 그룹 : 전체 오류 30개 중 7개, **23.33%** - 내재적 지식 작용

B 그룹 : 전체 오류 26개 중 1개, **3.85%**

5. 실험 연구 결과 분석 및 논의: Wh-의문문

5. 실험 연구 결과 분석 및 논의: Wh-의문문

5.2.1. Wh-의문문 결과 분석

5.2.1. Wh-의문문 결과 분석

- Wh-어 문두로 이동

표 5.8. Wh-의문문 (전체)

패턴	A 그룹 (14명)		B 그룹 (9명)		합계 (23명)	
	문장 수	백분율	문장 수	백분율	문장 수	백분율
1 Wh-S-V	47	27.98	31	28.70	78	28.26
2 Wh-V-S	43	25.60	31	28.70	74	26.81
3 Wh-Aux-S-V	31	18.45	17	16.67	48	17.39

A 그룹: 76.2%, B 그룹: 75% → L2의 영향으로 볼 수 없음

5. 실험 연구 결과 분석 및 논의: Wh-의문문

5. 실험 연구 결과 분석 및 논의: Wh-의문문

5.2.1. Wh-의문문 결과 분석

5.2.1. Wh-의문문 결과 분석

표 5.10. L2와 L3 수준이 각각 '상'인 경우

문장 패턴	A 그룹 (3명)		B 그룹 (5명)		합계 (8명)	
	문장 수	백분율	문장 수	백분율	문장 수	백분율
1 Wh-S-V	9	25	22	36.67	31	32.29
2 Wh-V-S	3	8.33	14	22.33	17	17.71
3 Wh-Aux-S-V	16	44.44	14	23.33	30	31.25

표 5.13. L2 한국어 수준별 wh-의문문

문장 패턴	하 (%)		중 (%)		상 (%)	
	A(4명)	B(0명)	A(5명)	B(2명)	A(5명)	B(7명)
1 Wh-S-V	16.67	-	30	20.83	35	29.76
2 Wh-V-S	37.50	-	33.33	54.17	8.33	21.43
3 Wh-Aux-S-V	18.75	-	8.33	4.17	28.33	20.24

5. 실험 연구 결과 분석 및 논의 : Wh-의문문

5.2.1. Wh-의문문 결과 분석

표 5.13. L2 한국어 수준별 wh-의문문 결과

▶ **Wh-Aux-S-V:**

- Wh-의문문 완성도의 수행성과가 비슷
- 사고육과 공교육을 충분히 받은 B 그룹의 수행성과가 높아야 함에도 불구하고 두 그룹이 비슷한 것은 A 그룹의 L1 매개변수가 L3와 같아서 L1이 L3 습득에 촉진적인 역할을 함

5. 실험 연구 결과 분석 및 논의 : Wh-의문문

5.2.1. Wh-의문문 결과 분석

표 5.15. L3 수준별 wh-의문문

	문장 패턴	하 (%)		중 (%)		상 (%)	
		A(1명)	B(1명)	A(8명)	B(3명)	A(5명)	B(5명)
1	Wh-S-V	33.33	8.33	30.21	25	23.33	35
2	Wh-V-S	50	83.33	25	11.11	21.66	28.33
3	Wh-Aux-S-V	0	0	6.25	5.56	41.67	25

5. 실험 연구 결과 분석 및 논의 : Wh-의문문

5.2.1. Wh-의문문 결과 분석

표 5.15. L3 영어 수준별 wh-의문문 결과:

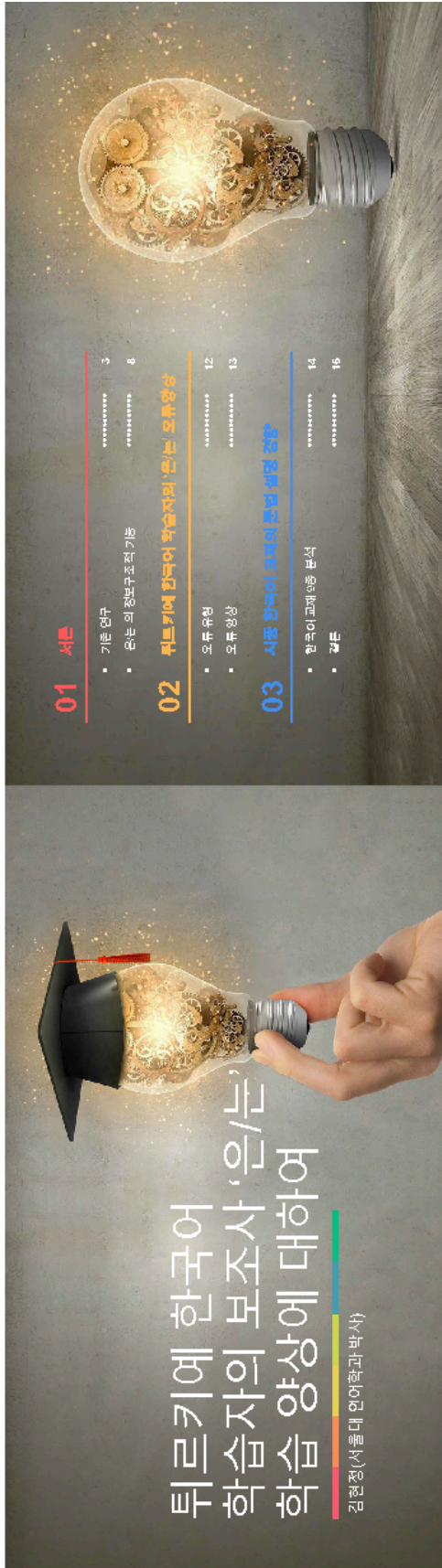
▶ **Wh-Aux-S-V: '상' 수준**

- A 그룹이 B 그룹보다 문장 완성도가 더 높음
- L2가 중요한 역할을 하지 않는 것으로 보이며,
- L3와 wh-매개변수가 같은 A 그룹의 L1이 언어를 습득할 때 더 긍정적인 영향을 미치는 것으로 판단됨.

6. 결론

• 누적항상 모델 지지

Flynn 외는 3개 언어의 한 그룹만 실험했으나 실질적으로, 이전에 경험한 언어의 매개변수가 유사하다면 L1, L2에 관계 없이 언어지식이 '누적'된다는 것을 염두에 두고 실험한 것으로 유추됨



들어가며

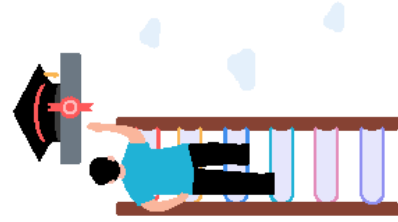
한국어를 배우는 외국인 학습자가 자주 오류를 일으키는 문법 사항 중의 하나가 보조사 '은/는'의 사용이다. 이는 한국어 보조사 '은/는'의 정보구조적 기능에 대해 완전한 속지가 이뤄지지 않아서 초래되는 오류이다. 국어학계에서는 한국어 보조사 '은/는'의 정보구조적 기능에 대해 많은 연구를 축적해 왔다. 큰 틀에서는 대조와 화제(topic)의 측면에서 '은/는'을 다룬다. 더 나아가서는 화제대조, 초점대조 등으로 분석을 심화하기도 한다.



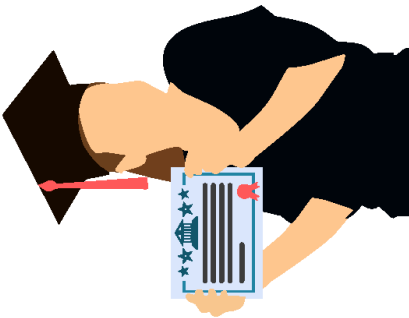
들어가며

'은/는'의 오류 양상은 한국어를 배우는 외국인 학습자가 모국어와 상관 없이 언어 보편적으로 비슷한 모습을 보인다고 간주할 수도 있으나, 분명히 개별 국가의 언어적 특징과 한국어외의 유사성 또는 차이점으로 인해 그 나라 학습자 특유의 오류를 발생시키기도 한다. 보고에서는 튀르키예의 한국어 학습자가 주로 일으키는

'은/는'의 오류 양상을 살펴보고 한다. 아울러 '은/는'의 사용 오류 양상을 밝히고 시중 한국어 교재에서 이에 대한 문법 설명을 어떻게 하고 있는지 점검할 것이다.



기존 연구



- 오한정(2011), 인지적 도식 사용
- 김경은·이선웅(2017), 문법 교재 설명 불충분 지적
- 허남영(2010), 독일어권 학습자 대상
- 이훈(2006), 중국어권 학습자 대상

‘은/는’의 정보구조



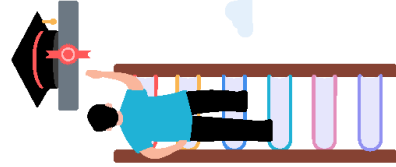
화제 예시

영희는 이제 백화점에 갔어.
영희가 어제 백화점에 간 사실
화제대조
예를 어디 있어?
질수는 토비에 있어.

관계적 신정보

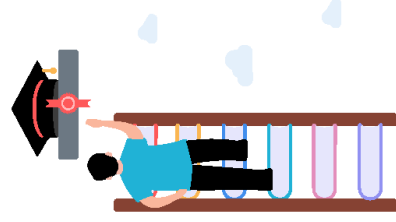
지금부터 현 경제상황을 살펴볼 것입니다.
현 경제상황은 매우 어려운 자경에 처해 있습니다.
조전대조
늦가 왔니?
활수는 왔어.

튀르키예 학습자의 ‘은/는’ 오류 양상



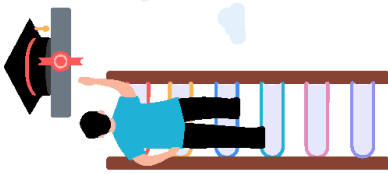
- 튀르키예 한국어 학습자의 ‘은/는’ 오류 양상은
- 주격조사 ‘이/가’ 등 다른 조사와 헷갈려서 이를 잘못 사용하거나,
 - 저격 및 저소격조사 ‘에’와 ‘에서’에 보조사 ‘은/는’을 결합하지 못해 생기는 오류,
 - 튀르키예어 소유격조사 In(이하 in, in, un, ün을 대표)과의 혼동
- 으로 크게 나눌 수 있다.

튀르키예 학습자의 ‘은/는’ 오류 양상



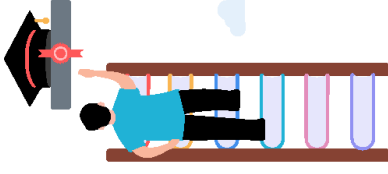
- 주격조사 ‘이/가’ 등 다른 조사와 헷갈려서 이를 잘못 사용한 경우
 - 제가 좋아하는 배우(*가/는) 남주형(이)예요.
 - 거기에는 바다가 있어서 우리(*가/는) 주말마다 바다에 가서 수영했어요.
- 저는 주말마다 시장에 가요. 그 시장(이/은) 우리 집 앞에 있어요.
 - 보드틀에 가봤어요. 여기(이/가/는) 너무 예뻐요.

튀르키예 학습자의 '은/는' 오류 양상



a. 제주도(*에/는) 공기도 좋고 경지도 아름다워요.

튀르키예 학습자의 '은/는' 오류 양상



2) 저격 및 저소격조사 '에'와 '에서'에 보조사 '은/는'을 결합하지 못해 생기는 오류

- a. 온라인 쇼핑에서 저물은 아주 쉽지만 백화점(*에서/에서는) 지불하는 것이 때때로 피곤할 수 있습니다.
- b. 그리고 마트(*에서/에서는) 여러 물건을 찾는게 어려워요.
- c. 백화점(*에서/에는) 가게가 많아서 여러 가지 물건을 찾을 수 있어요.

튀르키예 학습자의 '은/는' 오류 양상

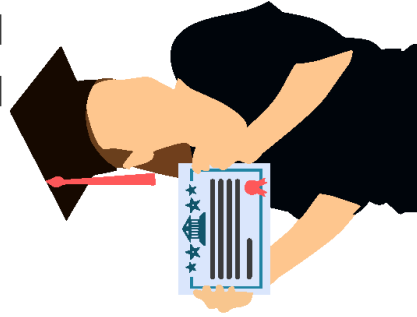


3) 튀르키예어 소유격조사 In(이하 in, in, un, ün을 대표) 과의 혼동

a. 어머니(*는/은) 그 사람들이 권했다고 할 것 같아요.

<튀르키예 명사 수식법> (김효정, 1994: 58)	의미
수식명사	피수식명사
okul-unön-ü	학교의 앞
simf-in	교실의 문
araba-nin	자동차의 색

'은/는' 사용의 경향성



오류 유형	백분율
'은/는'과 '이/가'의 혼동	18%
'은/는'과 '을/를'의 혼동	13.6%
주격 조사 생략	9%

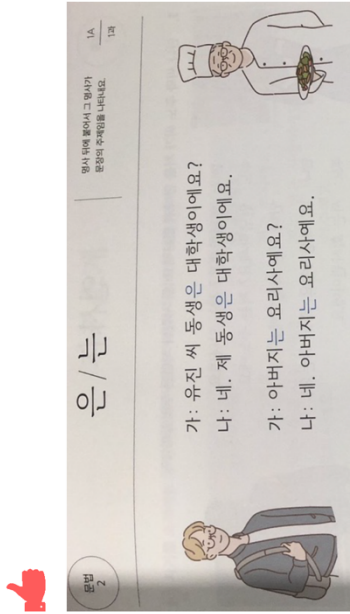
에르제네스대학교 1~2학년 22명 대상
2021년 봄학기 한국어 말하기 수업 과제

한국어 교재의 '은/는' 문법설명 경향

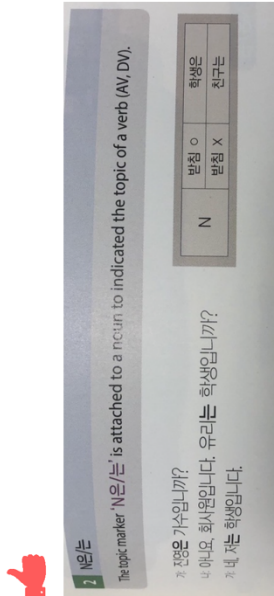


시중 한국어 교재 9종 대상

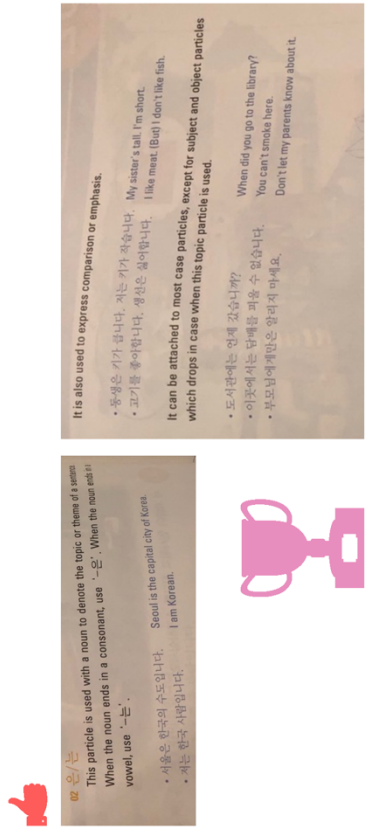
시중 한국어 교재	시용어	연서대	교과대	서양대	성공대	건국대	외대	중앙대	합계
재시 아부	0	0	X	0	0	0	0	X	6종
재시 단원	1과	1과	.	.	1과	1과	5과	.	1과



세종한국어 1A, 1과 '자기소개'



외국인을 위한 한국어 1-1(한국인대), 5과 '안녕하세요?'



연세한국어 1-1, 1과 '인사'

결론



시중 한국어 교재 9종 대상조사 '은/는'을 조금 단
계의 학습자에게 자세한 문법 설명을 하면서 가르치
는 것은 적절하지 않을 것이다. 조금 학습자에게 학습
부담을 주고 자칫 학습 의욕을 떨어뜨릴 수 있기 때
문이다. 그렇지만 한국어 회화를 어느 정도 구사하는
중급 이상의 학습자에게는 '은/는'에 대한 자세한 문
법 지식을 설명해줄 필요가 있을 것이다. 시소해보일
수 있지만 자연스러운 한국어 발화를 돕기 위해 필요
한 교수 단계이다.

결론



특히 외국인 학습자는 한국어를 자연스럽게 잘 구사
한다는 한국인의 정반에 따라 학습 의욕이 고차되고
고급 수준의 한국어 구사에게까지 이를 동기로 삼을 수
있으므로, 교수자가 이러한 심화 문법 지식을 잘 가르
쳐서 학습자의 한국어 구사 능력을 끌어올리는 것이
중요할 것이다.

Introduction	Corpus	Methods	Results	Conclusion
○○○○○	○○○	○○○○	○○○	○○○
○○○○○	○○○	○○○○	○○○	○○○
○○○○○	○○○	○○○○	○○○	○○○

Outline

How does context window size address polysemy of adverbial postposition -(u)lo in Korean?

Introduction
Polysemy in Korean
Distributional semantic models (DSMs)

Corpus
Sejong corpus
A hand-coded corpus

Methods

Results

Conclusion

Seongmin Mun

Ajou University

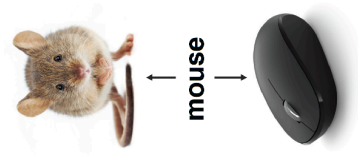
9th December 2023



Introduction	Corpus	Methods	Results	Conclusion
●○○○○	○○○	○○○○	○○○	○○○
●○○○○	○○○	○○○○	○○○	○○○
●○○○○	○○○	○○○○	○○○	○○○

Introduction

Polysemy, one type of ambiguity, occurs when one form delivers multiple meanings/functions (Glynn and Robinson, 2014).

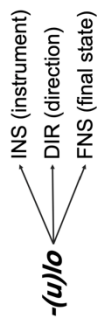


Introduction	Corpus	Methods	Results	Conclusion
●○○○○	○○○	○○○○	○○○	○○○
●○○○○	○○○	○○○○	○○○	○○○
●○○○○	○○○	○○○○	○○○	○○○

Introduction	Corpus	Methods	Results	Conclusion
○○○	○○	○○○○	○○○	○○○
●○○○	○○	○○○○	○○○	○○○
○○○○	○○	○○○○	○○○	○○○
Polysemy in Korean				

Korean language

Korean is a Subject-Object-Verb language, which marks grammatical information with dedicated postpositions (Sohn, 1999).



Polysemy in Korean adverbial postposition

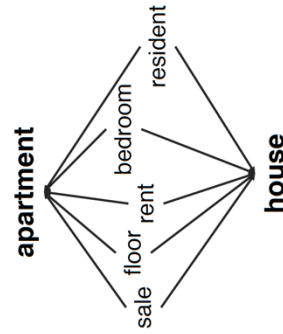
범인은 어두운 골목으로 달아났다.
 pemi-nun etwuwun kolmok-ullo talan-ass-ta.
 criminal-NOM dark alley-DIR flee-PST-DECL
 'The criminal fled into a dark alley.'

Figure: An example sentence involving the postposition -(u)lo as a function of DIR (direction)

How does context window size address polysemy of adverbial postposition -(u)lo in Korean?				
Introduction	Corpus	Methods	Results	Conclusion
○○○	○○	○○○○	○○○	○○○
●○○○	○○	○○○○	○○○	○○○
○○○○	○○	○○○○	○○○	○○○
Polysemy in Korean				

Question: How can a speaker understand the functions of postpositions?

Construal of a polysemous word occurs in conjunction with a series of words, delivering various framesemantic meanings (Goldberg, 2006) and yet purporting similar interpretations (Harris, 1954).

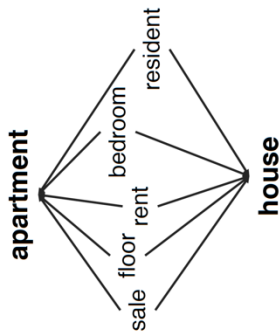


How does context window size address polysemy of adverbial postposition -(u)lo in Korean?				
Introduction	Corpus	Methods	Results	Conclusion
○○○	○○	○○○○	○○○	○○○
●○○○	○○	○○○○	○○○	○○○
○○○○	○○	○○○○	○○○	○○○
Polysemy in Korean				

Introduction ○○○○ ●○○○
 Corpus ○○ ○○ ○○
 Methods ○○○○ ○○○○
 Results ○○○ ○○○
 Conclusion ○○○ ○○○
 Distributional semantic models (DSMs)

Concept of DSMs

The concept of distributional semantic models (DSMs) is that a **word meaning is closely tied to a context** that is created by a group of neighborhood words, dubbed the **distributional hypothesis** (Firth, 1957; Harris, 1954).



Previous studies on Korean adverbial postpositions

Study	Corpus type	Data size	Method	Accuracy
Bae et al. (2015)	Korean PropBank	4,882 sentences	One-hot encoding + Structural SVM & FFNN (Feed-Forward Neural Network)	0.75
Kim & Ock (2016)	Sejong corpus	59,220 sentences	One-hot encoding + CRF (Conditional Random Fields Model)	0.83
Lee et al. (2015)	Korean PropBank	4,882 sentences	Word2vec (SGNS) + Structural SVM (Support Vector Machine)	0.77
Mun & Shin (2020)	Sejong corpus	2,100 sentences	PPMI & SVD + Similarity-based estimate	0.74
Park & Cha (2017)	Sejong corpus	14,335 sentences	Word2vec (SGNS) + CRF	0.77
Shin et al. (2005)	Sejong corpus	4,355 sentences	Word token-based embedding + SVM	0.71
Yoon et al. (2016)	Korean PropBank	4,714 sentences	One-hot encoding + Bidirectional LSTM-CRFs	0.66

Introduction ○○○○ ●○○○
 Corpus ○○ ○○ ○○
 Methods ○○○○ ○○○○
 Results ○○○ ○○○
 Conclusion ○○○ ○○○
 Distributional semantic models (DSMs)

How does context window size address polysemy of adverbial postposition -(u)lo in Korean?

Context window: a range of words surrounding a target word, affecting the determination of its characteristics (Lison and Kutuzov, 2017).

Question: How does context window address polysemy interpretation in Korean?

Introduction ○○○○○	Methods ○○○○○	Results ○○○	Conclusion ○○○
Corpus ●○○○	Corpus ●○○○	Methods ○○○○○	Results ○○○
Introduction ○○○○○	Introduction ○○○○○	Introduction ○○○○○	Conclusion ○○○
Sejong corpus			

What is Sejong corpus?

- ▶ Sejong corpus was created by the 21st Century Sejong Project, a ten-year-long project that was launched in 1998.
- ▶ Sejong corpus is a representative large-scale corpus in Korean (Shin, 2008).
- ▶ Previous studies often used this corpus as a linguistic resource (e.g., Kim & Ock, 2016; Park & Cha, 2017; Shin et al., 2005).

Corpus

How does context window size address polysemy of adverbial postposition -(u)lo in Korean?			
Introduction ○○○○○	Methods ○○○○○	Results ○○○	Conclusion ○○○
Corpus ●○○○	Corpus ●○○○	Methods ○○○○○	Results ○○○
A hand-coded corpus			

Description for input

- ▶ A portion of Sejong corpus (Shin, 2008), with semantic annotations of -(u)lo cross-verified by three native speakers of Korean ($k=0.95$).
- ▶ Data: 2,100 sentences
 - ▶ -(u)lo: Final state(700), Instrument(700), Direction(700)

Creating training and test sets

Training set

이것/NP 이/JKS 닦무리/NING (으)로/KE/FNS 나가/AVV _다/EF /SF
달_05/NING 이/JKS 어느새/MAG 서쪽/NING (으)로/KE/DIR 가들/AVV 고/EC 있/XX 있/EP 습니다 /EF /SF

Test set

해속/NNP 이/JKS 북도_04/NING (으)로/KB 나가/AVV 있/EP 다/EF /SF

Figure: Example sentences used in the model training and testing
(-(u)lo)

How does context window size address polysemy of adverbial postposition -(u)lo in Korean?			
Introduction ○○○○○	Methods ○○○○○	Results ○○○	Conclusion ○○○
Corpus ●○○○	Corpus ●○○○	Methods ○○○○○	Results ○○○
A hand-coded corpus			

Introduction 00 0000 0000
 Corpus 00 00 00
 Methods 0000
 Results 000
 Conclusion 000
 A hand-coded corpus

Creating training and test sets

Training set

이것(NP 이/JKS 뒤무리/NGG (으)로/KB_FNS 나타나/VV _다/EF /SF
 달_05/NGG 이/JKS 어느새/MAG 서쪽/KB_DIR 기물/AV 고/EC 왔/XX 왔/EP 습니다
 /EF /SF

Test set

해속(NNP 이/JKS 북도_04/NGG (으)로/KB_가/가/VV 왔/EP 다/EF /SF

Figure: Example sentences used in the model training and testing (-(u)lo)

Introduction 00 0000 0000
 Corpus 00 00 00
 Methods 0000
 Results 000
 Conclusion 000
 How does context window size address polysemy of adverbial postposition -(u)lo in Korean?

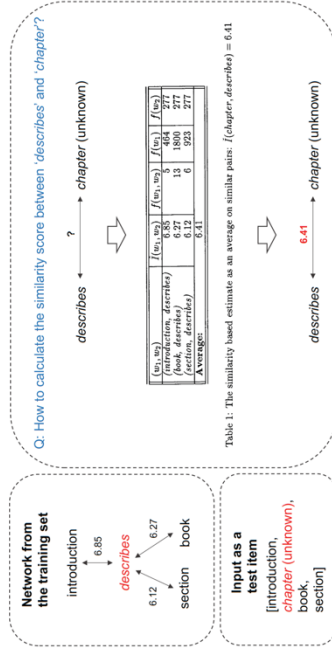
Methods

Word embedding model: PPMI-SVD

- ▶ **Model training:** Adapting a distributional semantic model (Harris,1954), an unsupervised learning algorithm was devised by combining Singular Value Decomposition with Positive Pointwise Mutual Information (i.e., PPMI-SVD).
- ▶ **Classification:** similarity-based estimate (Dagan et al., 1993) by calculating cosine similarity scores between -(u)lo and its co-occurring content words.

Similarity-based estimate (Dagan et al., 1993)

Similarity-based estimate (Dagan et al., 1993)

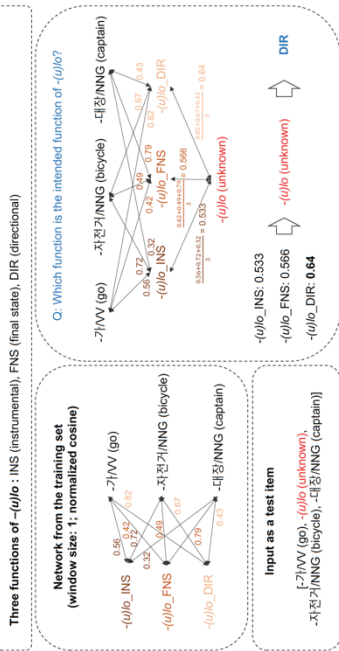


Introduction 00 0000 0000
 Corpus 00 00 00
 Methods 0000
 Results 000
 Conclusion 000
 How does context window size address polysemy of adverbial postposition -(u)lo in Korean?

Introduction	Corpus	Methods	Results	Conclusion	Introduction	Corpus	Methods	Results	Conclusion
○○○○○	○○○	○○○○○	○○○	○○○	○○○○○	○○○	○○○○○	○○○	○○○

Our approach (adapted from Dagan et al., 1993)

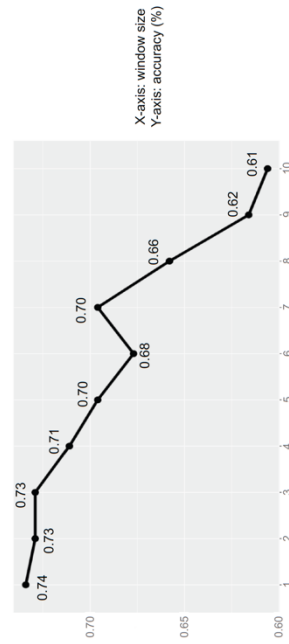
Our approach (adapted from Dagan et al., 1993)



Results

Introduction	Corpus	Methods	Results	Conclusion	Introduction	Corpus	Methods	Results	Conclusion
○○○○○	○○○	○○○○○	○○○	○○○	○○○○○	○○○	○○○○○	○○○	○○○

Classification: PPMI-SVD



Our model achieved the highest classification accuracy rate in the window size of one, and the accuracy rates decreased as the window size increased.

Evaluation

ICCGT1: How does context window size address polysemy of adverbial postposition -(u)lo in Korean?

Similarity Based Estimation: -(u)lo

Context window size: window 1

Input Sentence: Input your sentence ...

Analyze

Introduction	Corpus	Methods	Results	Conclusion	Introduction	Corpus	Methods	Results	Conclusion
○○○○○	○○○	○○○○○	○○○	○○○	○○○○○	○○○	○○○○○	○○○	○○○

Introduction ○○ ○○○ ○○○○	Corpus ○ ○○ ○○	Methods ○○○○	Results ○○○	Conclusion ●○○	Introduction ○○ ○○○ ○○○○	Corpus ○ ○○ ○○	Methods ○○○○	Results ○○○	Conclusion ○○○
-----------------------------------	-------------------------	-----------------	----------------	-------------------	-----------------------------------	-------------------------	-----------------	----------------	-------------------

Conclusion

- ▶ **Classification**
 - ▶ The result aligns with the small-window-size advantage (Bullinaria Levy, 2007).
 - ▶ Considering that a narrower range of context window relates more to syntactic than to semantic information (Patel et al., 1997), our model may have employed structural, more than semantic, characteristics of tri-grams (word-target-word) for the best classification performance.
- ▶ **Evaluation**
 - ▶ The size of the window affects the accuracy of polysemy interpretation.

Introduction ○○ ○○○ ○○○○	Corpus ○ ○○ ○○	Methods ○○○○	Results ○○○	Conclusion ●○○	Introduction ○○ ○○○ ○○○○	Corpus ○ ○○ ○○	Methods ○○○○	Results ○○○	Conclusion ○○○
-----------------------------------	-------------------------	-----------------	----------------	-------------------	-----------------------------------	-------------------------	-----------------	----------------	-------------------

Introduction ○○ ○○○ ○○○○	Corpus ○ ○○ ○○	Methods ○○○○	Results ○○○	Conclusion ○○○
-----------------------------------	-------------------------	-----------------	----------------	-------------------

Thank you for listening.

Introduction ○○ ○○○ ○○○○	Corpus ○ ○○ ○○	Methods ○○○○	Results ○○○	Conclusion ○○○
-----------------------------------	-------------------------	-----------------	----------------	-------------------

목차

- 서론
- 관련 연구
- 실험 분석
- 결론

인공지능의 한국어 접사 처리 연구

비립, 강예지, 박서윤, 김한샘

서론 - 문제 제기

- 현재까지의 한국어 형태소 분석기 성능:

"Kwii는 빠른 속도와 범용적인 성능을 자랑하는 한국어 형태소 분석기 라이브리입니다... 웹 텍스트의 경우 약 87%, 문어 텍스트의 경우 약 94% 정도의 정확도로 한국어 문장의 형태소를 분석해 낼 수 있습니다."
출처: <https://github.com/bab2mni/Kwii> ---

→ 그러나 나머지 5%의 한계에는 어떤 것이 있는가?

- 한국어 형태소 분석기의 5% 한계:

- 한국어 형태소 분석기 성능 보고 시, 보통 전체 정확률만 보고하고 태그별 분석 정확률은 밝히지 않음.
- 접사를 포함하는 일부 희소 품사는 전체 평가 데이터셋에 비율이 적어서 학습 불충분 가능성.

→ 형태소 자동 분석 과정에서 접사와 어기 같은 소분류 어휘에 관한 인공지능 모델 처리 능력 부족이 우려됨.

서론 - 접사 자동 분석 오류 유형

- 접사 태깅 기준 애매:
 - 어고수회 → "어-" 때로 NNG로 때로 XPN로 태깅
 - 명사 등을 접사로 오인식:
 - 폐열 → 폐-열, 주성분 → 주-성분, 남강도짓 → 남강도-짓
 - 외래어 일부를 접사로 오인:
 - 램(-덜); 캅(-스톤)
 - 지명, 인명 등 고유명사를 때때로 파생어로 오인
 - (정경)-심; (박준)-명
-

서론 – 연구 문제

- 연구 문제:
 - 한국어 인공지능 모델은 한국어 파생어 내 접사와 어기의 경계를 이해하는가?
→ 접사 식별 능력
 - 한국어 인공지능 모델은 한국어 접사와 파생 방식을 인지하는가?
→ 접사 활용 능력
- 연구 방법:
 - 한국어 인공지능 모델의 접사 식별 및 활용 능력 평가 프로세스 설계
 - 위 프로세스를 바탕으로 한국어 인공지능 모델의 접사 처리 성능 및 한계 분석

관련 연구 – 인공지능 모델의 “파생 지식”

- “파생 지식”은 교착어(한국어) 어휘 시스템 이해 및 접근에 필수적이며, 이는 결국 고성능 인공지능 모델의 훈련 및 강화에 반드시 필요한 학습 자료임.
- 인공지능의 파생 지식은 기본적으로 파생 어휘 간의 형태적인 연결 고리 특성을 의미하며, 이러한 특성은 대규모 말뭉치에서 추출한 후에 일종의 기계 학습용 파생 어휘 사전 DB를 구축할 수 있음.
- 다만 언어마다 파생 지식의 범위와 종류 그리고 이를 전산화할 수 있는 텍스트 데이터 양이 많이 다를 수 있음.
- 아직까지 파생 지식과 인공지능 모델 간의 연계 표준화 프로세스는 부재

관련 선행 연구: Weischedel, 1986; McShane & Nirenburg, 2003; Beesley, 2004; Edmiston & Stratos, 2018; Musil et al., 2019; Park et al., 2021

관련 연구 – 언어모델 블랙박스의 조명법

- 여태까지 인공지능 기반 언어모델에 의심의 여지가 계속 남을 수밖에 없는 것은 모델의 불투명성 때문임
- 즉, 파생 지식을 포함한 인공지능의 모든 학습 결과는 인간이 직접 관찰할 수 없는 블랙박스에 갇혀 있음
- 이에 따라 연구자들은 언어모델의 지식 확인 및 모델과의 교류 방법을 계속 모색해 왔음. 이 과정에서 언어학적 지식을 토대로 증명하는 방식이 유행함
- 최근 자유로운 텍스트 생성이 가능한 초거대 언어모델이 유행 후, 언어모델과의 상호작용은 한층 수월해졌으며, 언어모델에게 던지는 질문인 프롬프트를 일차나 잘 설계하느냐에 따라 언어모델의 내부 저장 지식을 더 정확하게 신속하게 찾을 수 있음

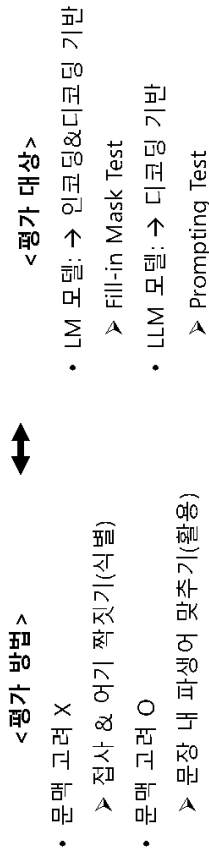
관련 선행 연구: Vulić et al., 2020; Miaschi et al., 2020; Mikhailov et al., 2021; Liu et al., 2023; Cao et al., 2023

관련 연구 – 복잡한 한국어 접사 체계

- 한국어는 교착어적 특성으로 인해 한자어, 고유어 접미사 및 접두사가 다수 존재하며, 이에 따라 사전 편찬 시 접사 표제어 관련 연구 및 조사가 장기적으로 이어져 왔음. 즉, 사전에는 한국어 접사 체계의 매우 방대한 파생 지식이 저장되어 있음.
- 매년 새로 탄생하는 신조어 출성량 상당 부분은 접사에서 파생된 것임. 다른 언어에 비해 서 한국어 접사의 어휘 생산성은 아주 강한 편.
- 사전 등재 접사 및 관련 파생어, 그리고 신조어 접사 파생은 모두 파생 지식에 해당하며 이는 언어모델 뿐만 아니라 외국인 학습자나 국내 초중고생에게도 매우 큰 학습 부담이 있음. 따라서 학습 효율성을 끌어올리기 위해, 한국어 접사 인식 관련 테스트 기준 및 방법 정비가 시급함.

관련 선행 연구: 장경현, 2019; 정려란, 2019; 조남호 & 노명희, 2021; 이주미 & 이연경, 2022

본론 - 접사 처리 능력 평가 프로세스



본론 - 실험 준비

- 접사 & 어기 리스트 추출
 - 소스: 국립국어원 신문 말뭉치 NIKL_NEWSPAPER_v2.0
 - (2009-2018, 350만 건 기사, 문어 정제 텍스트)
 - 방법: 1차 자동 형태소 분석 + 2차 수동 검증
 - (1차: Kivi 형태소 분석기, 2차: 언어학 전공자 3명)
 - 추출 대상:
 - 접사 범위: 총 15개
 - 반의 관계(㉠_대_XPN/소_XPN ㉡_신_XPN/구_XPN ㉢_진_XPN/반_XPN ㉣_고_XPN/저_XPN)
 - 유의 관계(㉤_반_XPN/여_XPN ㉥_심_XSN/감_XSN ㉦_론_XSN/설_XSN ㉧_인_XSN/자_XSN)
 - 어기 범위: 2글자 이상 일반명사(일부 외래어,의성·의태어 등 제외)
 - (말뭉치 전체 범위) 고빈도 어기 리스트(상위 30개)
 - (접사 범위 내) 상대 고빈도 어기 리스트(접사별 상위 100개)

본론 - LM 실험: 어휘 단위 - 어기 찾기

- 실험 절차:
 - 마스크 템플릿 설계:
 - 접두사인 경우: "접두사+[MASK]" → "대[MASK]"
 - 접미사인 경우: "[MASK]+접미사" → "[MASK]심"
 - [MASK] ∈ (접사별 어기 리스트| 어기는 해당 모델의 vocab에 있어야 함)
 - Fill-in Mask 테스트 실행:
 - fill_in_masker = (model=model_name, task='fill-in-mask', batch_size=8, top_k=30)
 - 접사별 출현확률 구하기 → "fill_in_masker("접두사+[MASK]", targets= 어기들)"
 - Targets에 들어갈 어기들은 "결합 가능(right)" 및 "결합 불가능(wrong)" 그룹으로 구분하여 실험
 - 추론 통계 분석
 - 차이 검증 U test: → U test 통과 시 right과 wrong 그룹 간에 차이가 확실한 것으로 간주

본론 - LM 실험

- 실험 대상:
 - 7개 트랜스포머 기반 언어 모델
- 실험 도구:
 - Huggingface Fill-in Mask Pipeline
- 평가 기준:
 - 통계 검증 및 자체 정량 지표

모델명	모델 카드 정보	파라미터 수	언어
X-roberta	xlm-roberta-large	355M	다국어
Kalbert	bongsoo/albert-small-k-or-v1	47M	한국어
Klue-Ror-B	klue/roberta-base	125M	한국어
Klue-Ror-L	klue/roberta-large	355M	한국어
Kolectra	monologg/kolectra-base-e-v3-discriminator	110M	한국어
KGELECTRA	beomi/KGELECTRA-base-v2022	110M	한국어
Kcbert-B	beomi/kcbert-base	110M	한국어
Kcbert-L	beomi/kcbert-large	340M	한국어

본론 - LM 실험: 어휘 단위 - 어기 찾기

	*sig_num (p-value < 0.1)	non-sig_num (p-value >= 0.1)	sig_num (p-value < 0.05)	non-sig_num (p-value >= 0.05)
X-roberta	4	11	3	12
Kalbert	1	14	0	15
Klue-Ror-B	12	3	12	3
Klue-Ror-L	13	2	13	2
Koelectra	12	3	12	3
KLECTRA	12	3	12	3
Kcbert-B	12	3	10	5
Kcbert-L	12	3	10	5

*Sig_num: p-value가 유의준을 넘은 횟수

- sig_num 횟수가 많을수록 해당 LM이 특정 접사에 결합 가능한 어기와 결합 불가능한 어기를 더 확실히 구분하는 것을 의미
- 전반적으로 한국어 모델은 다국어 모델보 다 성능 우수 → 다국어 모델보다 한국어로 된 vocab list가 더 큰 것이 영향.
- 한국어 모델 중 같은 모델에서 파생한 시리 즈별 모델들은 파라미터 규모 차이가 크지 만 어기 찾기에 성능 차이는 크지 않음 → vocab list가 비슷하기 때문인 것으로 추정
- 경량화 모델 albert은 제일 하위권 → 파생 지식을 충분히 학습하기 위해서는 파라미 터 저장 공간이 필요.

본론 - LM 실험: 어휘 단위 - 접사 찾기

실험 절차:

- 마스크 템플릿 설계:
 - 접두사인 경우: "[MASK]+어기" → "[MASK]기엀"
 - 접미사인 경우: "어기+[MASK]" → "가각[MASK]"
 - [MASK] ∈ {어기별 접사 리스트| 접두사와 접미사 2그룹으로 나눔}
- Fill-in Mask 테스트 실험:
 - fillin_masker = (model=model_name, task=fill-mask, batch_size=8, top_k=15)
 - 접사별 출현확률 구하기: "fillin_masker("[MASK]+어기)" 혹은 "어기+[MASK]" , targets= 접두사 혹은 접미사"
 - Targets에 들어갈 접사들은 "9개 접두사" 혹은 "6개 접미사" 총 15개 접사로 설정
- 추론 통계 분석
 - point-biserial 상관성: → 주어진 어기에 총 15개 접사가 결합 후보가 되며, Fill-in Mask로 예측한 출 현확률(연속성 데이터)과 기존의 주석자 직관 판단("0" & "1" 이분 데이터) 간 상관성을 측정.

본론 - LM 실험: 문장 단위 - 유의 관계

실험 절차:

- 문장 마스크 템플릿 설계:
 - 앞 문맥 "[MASK]+어기" 뒤 문맥 → "... 다음 [MASK]원칙 같은 ..."
 - [MASK] ∈ {유의관계 접사 '정', '심/감', '덕/관', '통/생', '인/자'}
 - 한 접사당에 총 10쌍 교대 불가능 어기 → '심/감', '소비(심)/개방(감)
 - 따라서 한 접사당에 총 20개 마스크된 문장 생성 가능.
- Fill-in Mask 테스트 실험:
 - fillin_masker = (model=model_name, task= 'fill-mask', batch_size=8, top_k=20)
 - 유의 관계 접사 쌍 출현확률 구하기 → "fillin_masker(마스크된 문장, targets= 접사 쌍)"
 - 유의 관계 접사 쌍 중 하나는 정답, 하나는 오답 → 정답의 출현 확률은 오답보다 훨씬 더 커야 함.
- 거리 비교:
 - 정답-오답 접사 간 거리(Distance) → R0 = 정답 접사의 출현 확률, R1 = 정답 접사의 출현 확률, Distance = (R0-R1)/R0

본론 - LM 실험: 어휘 단위 - 접사 찾기

	PB_cor (>0.7)	PB_cor (0.5~0.7)	PB_cor (<0.5)	PB_cor average
X-roberta	0	0	30	-0.18
Kalbert	0	0	30	-0.19
Klue-Ror-B	1	4	25	0.10
Klue-Ror-L	0	5	25	0.12
Koelectra	0	4	26	0.08
KLECTRA	1	5	24	0.05
Kcbert-B	0	4	26	0.08
Kcbert-L	0	5	25	0.10

- point-biserial 값이 클수록 해당 모델이 주어 진 어기에 결합 가능한 접사를 더 잘 골라 줄 수 있음.
- 전반적으로 매우 낮은 수치를 보임. 이는 최근 LLM에서 드러난 'reversal curse' 현상과 비슷한 면이 있음. 즉 접사를 알려주면 어기를 잘 찾 을 수 있지만, 반대로 어기를 알려줄 경우 오 히려 접사는 찾지 못함
- 접사와 결합하는 어기를 찾기가 쉬운 이유는 관련 어기가 이미 충분히 학습되었기 때문. 반 대로, 어기와 결합하는 접사 찾기를 수행하기 위해서는, LM은 훨씬 더 많은 '파생 지식'을 배워야 함.

본론 - LM 실험: 어휘 단위 - 유의 관계

- 유의관계에서 Distance 값이 클수록 해당 모델이 접사상의 의미적 거리 즉, 유의 관계 가능성을 더 잘 파악한다고 볼 수 있음.
- 전반적으로 결과가 상당히 좋은 편 이는 전통적인 인공어-다국어 기반 모델이 위치 정보를 훨씬 더 잘 인식하기에 접사 한상 패턴, 즉 어휘 접사와 어떤 어휘만 쓸 수 있는지를 이미 잘 학습한 것으로 보임.

	distance >0.7	distance 0.5~0.7	distance 0~0.5	distance average
X-roberta	72	4	4	0.91
Kalbert	64	5	11	0.82
Klue-Ror-B	61	9	10	0.81
Klue-Ror-L	67	4	9	0.85
Koelectra	71	5	4	0.82
KcELECTRA	45	16	19	0.68
Kcbert-B	66	6	8	0.85
Kcbert-L	64	6	10	0.84

본론 - LM 실험: 문장 단위 - 반의 관계 실험 절차:

- 문장 마스크 토큰릿 설계:
 - 기본 토큰릿: {word}의 뜻과 반대되는 단어는 {word2}이다;
 - word1=반의 관계 접사+어휘; word2=[MASK+어휘] → word1=신세력, word2= [MASK]세력
 - 반의 관계 접사상 총 8개 각 접사 쌍에 총 10개씩 어휘가 있음.
- Fill-in Mask 테스트 실험:
 - fillin_masker = (model=model_name, task= fill-mask , batch_size=8, top_k=1)
 - 반의 관계 접사2 출현 확률 구하기 → "fillin_masker(마스크된 문장, targets= 접사2)"
 - 무작위 1위 출현 확률 구하기 → "fillin_masker(마스크된 문장)"
 - R0 = 무작위 1위 후보의 출현 확률 값; R1 = 반의 관계 접사2의 출현 확률 값
- 거리 비교:
 - 무작위 1위 반의 관계 접사2 간 거리(Distance): Distance = (R0-R1)/R0
 - 무작위 1위를 뽑을 때 반의 관계 접사2가 나오는 경우가 가장 이상적.
 - 따라서 Distance가 작아야 반의 관계 접사를 더 잘 찾는 것으로 간주할 수 있음.

본론 - LM 실험: 어휘 단위 - 반의 관계 실험 절차:

- 문장 마스크 토큰릿 설계:
 - 기본 토큰릿: {word}의 뜻과 반대되는 단어는 {word2}이다;
 - word1=반의 관계 접사+어휘; word2=[MASK+어휘] → word1=신세력, word2= [MASK]세력
 - 반의 관계 접사상 총 8개 각 접사 쌍에 총 10개씩 어휘가 있음.
- Fill-in Mask 테스트 실험:
 - fillin_masker = (model=model_name, task= fill-mask , batch_size=8, top_k=1)
 - 반의 관계 접사2 출현 확률 구하기 → "fillin_masker(마스크된 문장, targets= 접사2)"
 - 무작위 1위 출현 확률 구하기 → "fillin_masker(마스크된 문장)"
 - R0 = 무작위 1위 후보의 출현 확률 값; R1 = 반의 관계 접사2의 출현 확률 값
- 거리 비교:
 - 무작위 1위 반의 관계 접사2 간 거리(Distance): Distance = (R0-R1)/R0
 - 무작위 1위를 뽑을 때 반의 관계 접사2가 나오는 경우가 가장 이상적.
 - 따라서 Distance가 작아야 반의 관계 접사를 더 잘 찾는 것으로 간주할 수 있음.

본론 - LM 실험: 문장 단위 - 반의 관계

- 반의 관계에서 Distance 값이 작을수록 해당 모델이 접사상의 의미적 거리 즉, 반의 관계를 더 잘 파악한다고 볼 수 있음.
- 전반적으로 매우 낮은 상응의 결과로, 유의 관계 결과와 함께 살펴보면 때 전통적인 LM들은 한국어 접사에 관한 지식은 의미적인 것보다 형태/위저적인 정보에 편중되어 있음을 알 수 있음.
- 즉 기존 LM이 파악한 한국어 파생지식 수준은 형태/위저에 머물러 있어 그 수준이 감지 않은 것으로 판단됨.

	distance <0.3	distance 0.3~0.5	distance > 0.5	distance average
X-roberta	0	0	80	1.00
Kalbert	0	0	80	1.00
Klue-Ror-B	6	2	72	0.86
Klue-Ror-L	2	0	78	0.94
Koelectra	0	0	80	1.00
KcELECTRA	0	0	80	1.00
Kcbert-B	12	4	64	0.76
Kcbert-L	20	2	58	0.69

본론 - LLM 실험: 어휘 단위 - 어휘 찾기

- 실험 대상:
 - ChatGPT3.5(turbo)
- API Request


```
client.chat.completions.create(
    model="gpt-3.5-turbo",
    messages=[{"role": "system", "content": "당신은 한국어를 가르치는 선생입니다."},
              {"role": "user", "content": prompt}]
```
- 평가 기준:
 - 정량 지표(accuracy, precision, recall, f1)

본론 – LLM 실험: 어휘 단위 – 어기 찾기

- 프롬프트:
 - 지시문: (2개 준비)
 - 점두사용: "아래 어휘 리스트 중에, '{affix}' + X 식으로 새로운 단어를 만들 수 있는 것을 전부 다 골라주세요. #n"
 - 점미사용: "아래 어휘 리스트 중에, X + '{affix}' 식으로 새로운 단어를 만들 수 있는 것을 전부 다 골라주세요. #n"
 - 어기서 affix는 총 15개 선정된 점사 전부 해당.
- 인풋: "어휘 리스트: {root}"
 - root는 점사마다 해당 점사와 결합 가능한 root 15개, 불가능한 것 15개, 총 30개로 구성
- 프롬프트 예시:
 - 아래 어휘 리스트 중에, '대'+X 식으로 실제로 사용 가능한 단어를 만들 수 있는 것을 전부 다 골라주세요.

어휘 리스트: [원칙, 가족, 대화, 생활,]

본론 – LLM 실험: 어휘 단위 – 어기 찾기

- ChatGPT 답안 예시:
 - “어휘 리스트 중에서 X + ‘감’으로 새로운 단어를 만들 수 있는 것은 다음과 같습니다:
 - 생동감, 수분감, 실망감, 박자감, 균형감, 밀착감, 희의감, 저항감, 구약감, 연결감, 산상감, 증정감, 투쟁감, 안전감, 효행감, 안락감, 기력감, 애착감”
- 결과 분석:
 - 주어진 점사와 결합 가능한 어기 찾기는 LM보다 LLM결과가 상당히 더 낮은 결과를 보임.
 - “용량” 등 극수어의 어기를 제외하고 전반적으로 한국어 점사에 대한 개념 이해 부족.

기법	accuracy	precision	recall	f1
기준	0.53	0.45	0.44	0.44
규모	0.60	0.75	0.67	0.58
관성	0.40	0.20	0.50	0.29
부분	0.27	0.13	0.50	0.21
도시	0.60	0.30	0.50	0.38
보통	0.33	0.50	0.55	0.31
가분	0.40	0.51	0.50	0.40
성향	0.40	0.50	0.50	0.39
상황	0.33	0.18	0.42	0.25
규리	0.40	0.68	0.55	0.35
상용인	0.20	0.57	0.54	0.20
재정	0.53	0.55	0.54	0.52
법관	0.07	0.04	0.25	0.06
기술	0.73	0.37	0.50	0.42
기분	0.44	0.50	0.42	0.42
다귀	0.13	0.07	0.50	0.12
주조	0.33	0.35	0.36	0.33
행안	0.60	0.67	0.75	0.58
홍산	0.40	0.42	0.42	0.40
성능	0.13	0.07	0.50	0.12
다수	0.53	0.68	0.68	0.53
극강	0.47	0.69	0.60	0.44
민소	0.47	0.64	0.67	0.46
지인	0.27	0.13	0.50	0.21
부흥	0.47	0.37	0.50	0.31
유명	0.67	0.82	0.82	0.83
충원	0.47	0.50	0.50	0.46
산인	0.33	0.18	0.42	0.25
평가	0.47	0.23	0.50	0.32

본론 – LLM 실험: 어휘 단위 – 점사 찾기

- 프롬프트:
 - 지시문: (2개 준비)
 - 점두사용: "아래 어휘 리스트 중에, X + '{root}' 식으로 실제로 사용 가능한 단어를 만들 수 있는 것을 전부 다 골라주세요. #n"
 - 점미사용: "아래 어휘 리스트 중에, '{root}' + X 식으로 실제로 사용 가능한 단어를 만들 수 있는 것을 전부 다 골라주세요. #n"
 - 어기서 root는 준비된 밑용지 전체에서 고빈도 어기 30개로 순화
- 인풋: "어휘 리스트: {affix}"
 - 점두사와 점미사 그룹을 따로 나눠서 '{affix}'를 구성
 - 점두사는 총 9개, 점미사는 총 6개
- 프롬프트 예시:
 - 아래 어휘 리스트 중에, '평가'+X 식으로 실제로 사용 가능한 단어를 만들 수 있는 것을 전부 다 골라주세요.

어휘 리스트: [감, 산, 샬, 문, 인, 자]

본론 – LLM 실험: 어휘 단위 – 점사 찾기

- ChatGPT 답안 예시:
 - “태산업, 소산업, 석산업, 구산업, 반산업, 전산업, 고산업, 저산업, 예산업 어휘 리스트 중에서 ‘산’과 결합하여 사용 가능한 단어는 다음과 같습니다:
 - 산업장, 산업에 대한 명칭 또는 관심
 - 산업업: 산업의 현재 상태 또는 운영에 대한 평가
 - 산업물: 산업에 대한 이론 또는 연구
 - 산업질: 산업에 대한 설명 또는 분석
 - 산업인: 산업 분야에서 일하는 사람”
- 결과 분석:
 - 주어진 어기에 대해 점사를 결합하는 실험에서 LLM의 답안은 어기에 거의 모든 점사를 결합하는 경향을 보임.
 - 상대적으로 ChatGPT는 어기 찾기보다 ‘점사 찾기’를 더 어려워 하는 것으로 보임.
 - 이는 LM 실험 결과에서 확인한 reversal curse와 같은 맥락으로 볼 수 있음.

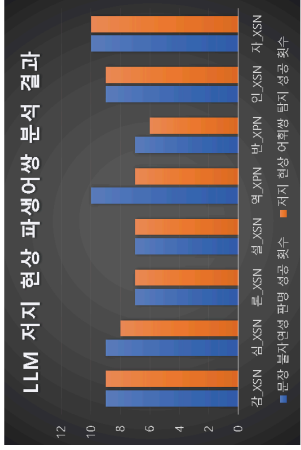
	accuracy	precision	recall	f1
강_XSN	0.53	0.53	0.53	0.53
고_XPN	0.50	0.25	0.50	0.33
구_XPN	0.50	0.25	0.50	0.33
론_XSN	0.50	0.25	0.50	0.33
질_XSN	0.67	0.68	0.67	0.66
역_XPN	0.43	0.36	0.43	0.35
신_XPN	0.50	0.25	0.50	0.33
인_XSN	0.50	0.25	0.50	0.33
저_XPN	0.50	0.25	0.50	0.33
자_XSN	0.57	0.77	0.57	0.47
전_XPN	0.50	0.25	0.50	0.33
대_XPN	0.50	0.25	0.50	0.33
소_XPN	0.50	0.25	0.50	0.33
반_XPN	0.50	0.25	0.50	0.33
셀_XSN	0.67	0.70	0.67	0.65

본론 - LLM 실험: 문장 단위 - 유의 관계

- 프롬프트:
 - 지시문:
 - 아래 2개 문장 중에 부자연스러운 문장이 있는가? 있다면 그 문장에서 어떤 부분(단어)이 문제인가? why
 - 인풋:
 - '문장 1:' + [sentence1] + 'why' + '문장 2:' + [sentence2]
 - sentence1은 유의관계 점사상 중, "정답 점사 + 어기가 들어가 있음."
 - sentence2는 유의관계 점사상 중, "오답 점사 + 어기가 들어가 있음."
 - 프롬프트 예시:
 - 아래 2개 문장 중에 부자연스러운 문장이 있는가? 있다면 그 문장에서 어떤 부분(단어)이 문제인가?
- 문장1: ...를 피부에 대해 긴장감을...
- 문장2: ...를 피부에 대해 긴장심을...

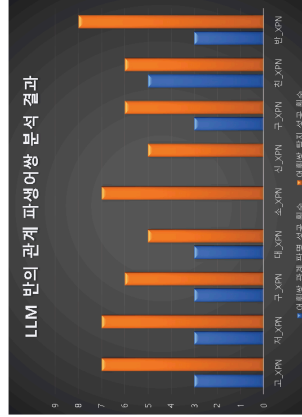
본론 - LLM 실험: 문장 단위 - 유의 관계

- ChatGPT 답안 예시:
 - 문장 2에서 "긴장심"이라는 단어가 부자연스럽습니다. 올바른 표현은 "긴장감"입니다.
- 결과 분석:
 - 대체로 지지 현상에 속한 유의 관계 점사 상 간의 어기 결합 가능성을 잘 파악
 - 다만 어휘 단위에서의 낱은 성능과 대조가 된다는 점에서 LLM이 생성에 특화된 모델임을 확인할 수 있으며, 이러한 특성상 상세한 정보(문맥 정보 포함)를 입력으로 받을 경우 점사 처리 능력이 더욱 상승할 수 있음을 관찰.



본론 - LLM 실험: 문장 단위 - 반의 관계

- 프롬프트:
 - 지시문:
 - 아래 2개 문장 중에 뜻이 정반대되는 단어 쌍은 있는가? 있다면 그 단어 쌍은 무엇인가? why
 - 인풋:
 - '문장 1:' + [sentence1] + 'why' + '문장 2:' + [sentence2]
 - sentence1은 반의 관계 점사 상 중, "1번 점사 + 어기가 들어가 있음."
 - sentence2는 반의 관계 점사 상 중, "2번 점사 + 어기가 들어가 있음."
 - 프롬프트 예시:
 - 아래 2개 문장 중에 뜻이 정반대되는 단어 쌍은 있는가? 있다면 그 단어 쌍은 무엇인가?
- 문장1: ... 한일 구시대를 열려면 ...
- 문장2: ... 한일 신시대를 열려면 ...



본론 - LLM 실험: 문장 단위 - 반의 관계

- ChatGPT 답안 예시:
 - 예상과 예상에 사용된 단어 "신세력"과 "구세력"의 뜻이 정반대됩니다. "신세력"은 '새로 성장되거나 부상한 힘을 의미하고 "구세력"은 '이전부터 갖추고 있던 힘을 의미합니다.
- 결과 분석:
 - 반의 관계 변별에 있어서, LLM은 여전히 50% 이하의 성공률을 보임.
 - 다만 두 개 문장 중 반의 관계 단어를 찾는 데 대부분 성공
 - 이것은 역시 앞서 LM 실험 결과와 일맥상통한 부분. 즉 현재 LLM은 LM보다 파생 지식의 이해 수준이 많이 상승했으나, 한국어 점사체계의 의미적인 이해도가 아직 떨어지는 상태로 보임.

참고 문헌

Balaven, K., Bella, G., & Giurcibuda, F. (2021, August). MorphNet: a large multilingual database of derivational and inflectional morphology. In *Proceedings of the 18th signa symposium workshop on computational research in phonetics, phonology, and morphology* (pp. 39-46).

Beesley, K. R. (2004, April). Morphological analysis and generation: A first step in natural language processing. In *First Steps in Language Documentation for Minority Languages: Computational Linguistic Tools for Morphology, Lexicon and Corpus Compilation, Proceedings of the SACLING Workshop at LREC* (pp. 1-8).

Cao, B., Lin, H., Han, X., Liu, F., & Sun, L. (2022). Can prompt probe pretrained language models' understanding the invisible risk from a causal view. *arXiv preprint arXiv:2202.12259*.

Edition, D., & Shades, K. (2018, July). Compositional morpheme embeddings with affixes as functions and stems as arguments. In *Proceedings of the Workshop on the Role of Language Structure in Neural Architecture for NLP* (pp. 1-5).

Hofmann, V., Piterouhant, J. B., & Schütze, H. (2020). DaGoBERT: Generating derivational morphology with a pretrained language model. *arXiv preprint arXiv:2006.02672*.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(3), 1-35.

McShane, M., & Nienburg, S. (2023). Blasting open a choice space: Learning inflectional morphology for NLP. *Computational Intelligence*, 15(2), 111-135.

Mizuchi, A., Burato, D., Dell'Orletta, F., & Venturi, G. (2020). Uniquistic profiling of a neural language model. *arXiv preprint arXiv:2010.01689*.

Mikhilov, V., Senkov, O., & Arsenova, E. (2021). Morph call: Probing morphosyntactic content of multilingual transformers. *arXiv preprint arXiv:2104.12947*.

Mullik, T., Veda, J., & Marasick, D. (2019). Derivational morphological relations in word embeddings. *arXiv preprint arXiv:1906.02510*.

Shi, H. H., Zheng, K. J., Wang, C., Steiner, K., Liu, H., & Schwartz, L. (2021). Morphology matters: a multilingual language modeling analysis. *Transactions of the Association for Computational Linguistics*, 9, 1-15.

Vuic, I., Pantić, E. M., Litichev, R., Grawak, G., & Komonen, A. (2020). Probing pretrained language models for lexical semantics. *arXiv preprint arXiv:2010.05773*.

Waschewski, R. H. (1998). Knowledge representation and natural language processing. *Proceedings of the IEEE*, 74(7), 805-820.

이규태 (2023) (2023) 대학원 석사 학위논문, 고려대학교 대학원 석사 학위논문, 21, 131(3), 1427-1442.
 정재현 (2019) (2019) 인공어 생성을 위한 자연어 처리 연구, 연세대학교 석사학위논문, 20(2), 311-335.
 권태현 (2019) (2019) 인공어 생성을 위한 자연어 처리 연구, 연세대학교 석사학위논문, 20(2), 311-335.
 권태현 (2019) (2019) 인공어 생성을 위한 자연어 처리 연구, 연세대학교 석사학위논문, 20(2), 311-335.
 권태현 (2019) (2019) 인공어 생성을 위한 자연어 처리 연구, 연세대학교 석사학위논문, 20(2), 311-335.

결론

- 한국어 인공지능 모델의 성능 향상에는 한국어 접사 시스템의 구분 기준 및 적용 방식 이해도 제고가 필수임.
- 한국어 인공지능 모델의 한국어 접사 처리 능력에 파생 지식 수준이 매우 중요한 영향을 발휘
- 파생 지식 수준 평가 시, 표층적인 패턴 외에 파생어 내 외부 의미적 연계성과 관련되는 심층적인 패턴 이해도 테스트에 집중할 필요가 있음.
- 현재 대표적인 한국어 LM & LLM을 대상으로 실험한 결과, 한국어 접사 시스템 이해에 한계가 있으며, 이를 개선하기 위해 평가용 데이터셋, 평가 지표, 그리고 평가 대상을 확대할 필요성이 있음
- 또한 개발한 기존 접사 능력 평가 프로세스에 대한 지속적인 보강 역시 필요.

Transitive out of -ing construction: A collostructional and constructional approach

Jungsoo Kim & Rom Sim
 jungsookim@khu.ac.kr; RSIM@email.sc.edu

Incheon National University; University of South Carolina

2023 LSK Winter Conference
 (Hosted by Kyung Hee University, Seoul)
 December 9, 2023



- 1 Introduction
- 2 Methods and data
- 3 Results
 - Typical transitive out of -ing construction
 - Atypical variants
- 4 A construction grammar perspective account
 - Key tenets of construction grammar
 - The transitive out of -ing construction
 - Two atypical variants
- 5 Conclusion

Introduction

- Transitive out of -ing construction (Francis et al. 1996; Rudanko 2011; Gawlik 2013, Sim & Kim 2015)
 - (1) a. I cheated you all out of being in my life. (COCA-1999-TV)
 - b. She cheated thousands of people out of getting a rebate. (COCA-2010-NEWS)
- Form: NP subject + V1 + NP object + out of VP2[-ing]
- Two possible meanings depending on the telicity of the predicate introduced by VP2[-ing]: 1) movement/extraction and 2) prevention interpretations

Transitive out of -ing construction

Research gaps and motivations

- Much less attention paid to the transitive *out of -ing* construction than the transitive *into -ing* construction (see, among others, Rudanko 1991, 2002, 2011; Hunston & Francis 2000; Gries & Stefanowitsch 2004; Kim & Davies 2016; Davies & Kim 2019)
 - (2) a. I'm actually trying to talk him into putting a little xmas tree in here for us. (COCA-2012-WEB)
 - b. They didn't try to fool you into thinking they were home. (COCA-2016-FIC)
- Main focus restricted to the types of verbs in the V1 slot of the construction, disregarding the types of expressions in the other slots of the construction (Rudanko 2011; Gawlik 2013)
- Corpus-based studies only with raw frequencies of verb types in the V1 slot, which may not tell us directly about the core meaning(s) associated with the slot (Stefanowitsch & Gries 2003)

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 5 / 34

Research goals

- To investigate real-life uses of the transitive *out of -ing* construction on the basis of authentic corpus data from COCA (Davies 2008-)
- To examine which (types of) verbs frequently occur in the V1 slot of the construction, and to look into grammatical properties of the expressions in each of the slots of the construction, adopting different types of collocutional analyses
- To attempt to offer an analysis of the construction from a Construction Grammar (CxG) perspective (in particular, via inheritance networks and tight interactions between lexical properties and argument structure constructions) (Goldberg 1995, 2005; Sag 2012; Kim & Michaels 2020)

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 6 / 34

Methods

- Collexeme analysis: which verbs occur in the V1 slot of the construction more frequently than expected
- Distinctive collexeme analysis: whether certain groups of verbs in the V1 slot are more or less likely to be associated with particular subject types (animate vs. inanimate), and object types (canonical vs. one's way vs. reflexive)
- Covarying collexeme analysis: interaction between verbs in the V1 slot and those in the V2[-ing] slot
- To conduct these different types of collocutional analyses: *coll.analysis 4.0* package for R Studio; log-likelihood ratio (LLR or G2) as an index of collocation strength (3.8415 or higher corresponding to $p < 0.05$ and 10.8276 corresponding to $p < 0.001$)

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 8 / 34

Methods and data

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 7 / 34

Data

- Data from COCA (Corpus of Contemporary American English): largest structured corpus of Contemporary American English

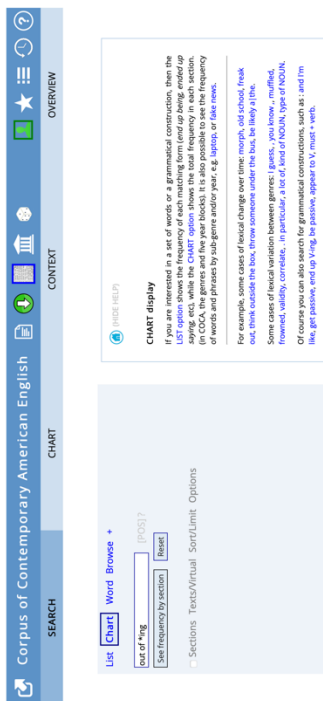


Figure 1: Screenshot of the string search in the search box via the COCA web interface

- 7,581 matching tokens!

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 9 / 34

Verbs in the V1 slot: verb types

- Fifty-three different verbs from 620 typical transitive out of -ing construction
 - talk (408), get (33), scare (24), cheat (15), keep (14), shut (11), lock (9), price (9), intimidate (8), pull (6), drive (5), frighten (5), rule (5), screw (5), bully (4), leave (4), take (4), train (4), force (3), kick (3), shame (3), bluff (2), browbeat (2), coax (2), knock (2), push (2), trick (2), annoy (1), argue (1), bigfoot (1), bribe (1), charm (1), con (1), counsel (1), cut (1), edge (1), finagle (1), grandfather (1), harass (1), hinder (1), jawbone (1), laugh (1), legislate (1), lobby (1), move (1), persuade (1), rock (1), snap (1), spell (1), spook (1), terrify (1), want (1), yank (1)
 - ... the Clinton administration talked the Republican Congress out of passing it back in 2000, or some time. (COCA-2007-SPOK)
Playing hooky for one day isn't gonna get you out of doing your big interview. (COCA-2019-TV)
It's almost like they set this whole thing up to scare me out of smoking. (COCA-1996-TV)
 - talk with the highest frequency (408 instances), followed by get (33 instances), scare (24 instances), cheat (15 instances), keep (14 instances), and shut (11 instances)
 - The rest 47 verbs each occurring fewer than 10 times

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 11 / 34

Results

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 10 / 34

Verbs in the V1 slot: semantic classification

- Classification of verbs found in the V1 slot of the construction depending on their meanings
 - By means of deception or trickery (6 types): *bluff, browbeat, cheat, con, finagle, trick*
 - By means of exerting force or pressure, sometimes understood metaphorically (22 types): *bigfoot, bully, cut, drive, edge, force, grandfather, harass, kick, knock, lock, move, price, pull, push, rock, rule, screw, shut, snap, take, yank*
 - By means of arousing fear, irritation, anger, annoyance, confusion, or surprise (7 types): *annoy, frighten, intimidate, scare, shame, spook, terrify*
 - By means of enticing, flattering, or verbal persuasion (8 types): *argue, bribe, coax, counsel, jawbone, lobby, persuade, talk*
 - By other specific means (5 types): *charm, laugh, legislate, spell, train*
 - By nonspecific means (5 types): *get, hinder, keep, leave, want*
- The most prominent semantic class: "exertion of force or pressure" verbs (22 different verb types)
- The rest with similar different verb type numbers

Verbs in the V1 slot: collexeme analysis

- The raw frequencies of the verbs in (5) may not tell us directly about how strongly the verbs are associated with the V1 slot of the construction!
- Table 1: 10 strongest collexeme verbs in the V1 slot of the typical transitive out of -ing construction in COCA

Rank	Collexeme	Freq.	Collstrength (LLR)
1	talk	635,614:408	3448.33
2	scare	24,484:24	202.99
3	cheat	21,523:15	116.48
4	price	10,965:9	72.73
5	intimidate	6,613:8	70.81
6	shut	85,625:11	48.88
7	lock	43,093:9	48.43
8	frighten	7,953:5	37.72
9	bully	8,490:4	27.90
10	screw	25,635:5	26.21

- talk being the strongest collexeme, followed by scare and cheat; the other collexemes here with relatively high collocation strength values, with their values in between 20 and 100
- Interestingly, only one collexeme verb in common (i.e., bully) between the 10 strongest collexemes in the V1 slot of the transitive into -ing construction in Stefanowitsch & Gries (2003) and those 10 strongest collexemes in the V1 slot of the typical transitive out of -ing construction here
- Two major semantic classes of collexemes in the V1 slot: "force" and "negative emotion verbs"

Subject types

- Predominance of active voice examples: 546 out of 620 examples (88.1%)
- Nonetheless, 74 passive voice examples (11.9%)
 - He has to be **talked** out of doing this by his comrades. (COCA-2012-MAG)
 - And how many Americans will be **priced** out of owning a car as a result? (COCA-2012-WEB)
- Predominance of animate subject examples: 495 examples out of 546 active voice examples (90.7%)
- 51 inanimate subject examples (9.3%)
 - Have **Republicans** finally embarrassed themselves out of calling themselves the family values party? (COCA-2007-SPOK)
Besides, **his story** could scare kids out of using steroids. (COCA-1993-MAG)

Subject types (cont'd)

- 10 strongest distinctive collexemes in the V1 slot of the typical transitive out of -ing construction with an animate subject and the one with an inanimate subject in COCA

Animate subject (N = 495)		Inanimate subject (N = 51)					
Rank	Collexeme	Freq.	LLR	Rank	Collexeme	Freq.	LLR
1	talk	385:3	109.71	1	get	15:18	48.51
2	cheat	7:0	1.38	2	scare	15:9	14.94
3	drive	5:0	0.99	3	lock	1:2	5.93
4	bully	4:0	0.79	4	hinder	0:1	4.76
5	screw	4:0	0.79	5	spook	0:1	4.76
6	shame	3:0	0.59	6	terrify	0:1	4.76
7	bluff	2:0	0.39	7	yank	0:1	4.76
8	shut	2:0	0.39	8	rule	2:2	4.38
9	trick	2:0	0.39	9	frighten	3:2	3.39
10	annoy	1:0	0.19	10	pull	3:2	3.39

- Walking back after Mass, she tried to make an **excuse** that would get her out of eating with her aunts, but Pauline wouldn't hear of it. (COCA-1991-FIC)
Besides, **his story** could scare kids out of using steroids. (COCA-1993-MAG)
- Major semantic classes of verbs in the V1 slot with inanimate subject: 1) "nonspecific means", 2) "negative emotion", and 3) "force" verbs

Eventuality and interpretation types

- Classification of the identified typical transitive *out of -ing* construction examples based on their interpretation types
 - Movement/extraction interpretation (98 instances): For example, when Bill writes that Paterno fired Sandusky, he's either reading the Fresh Report incorrectly or he read something in Joe's book that suggests Paterno did actually force Sandusky out of coaching.
 - Prevention interpretation (468 instances): Two weeks ago I saw Cameron Diaz at Fred Segal... and I talked her out of buying... **this truly heinous angora sweater.** (COCA-2001-MOV)
 - Unclear/Ambiguous (54 instances): If this is about money... Why? You planning on bribing me out of **writing the book?** (COCA-2009-TV)
- Mismatch cases where the canonical mapping relation between eventuality types and interpretation types breaks down
 - Vera and several of his friends tried to talk him out of **flying on New Year's Eve**, but Clemente insisted he had to go. (COCA-1992-MAG) (atelic predicate + prevention interpretation)
- Clear movement/extraction interpretation: 96 atelic predicate examples (98%) vs. 2 telic predicate examples (2%)
- Clear prevention interpretation: 275 telic predicate examples (58.8%) vs. 193 atelic predicate examples (41.2%)
- Much stronger tendency for the atelic predicate-movement/extraction interpretation pair than for the telic predicate-prevention interpretation pair

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 17/34

Verbs in the V1 slot: verb types

- Matrix verbs found in the identified atypical transitive *out of -ing* construction examples
 - With a reflexive NP object (15 types, 63 instances): *talk* (41), *argue* (3), *get* (3), *reason* (3), *cheat* (2), *snap* (2), *buy* (1), *cut* (1), *embarrass* (1), *intimidate* (1), *leave* (1), *pull* (1), *rationalize* (1), *take* (1), *train* (1)
 - With a possessive way NP object (23 types, 42 instances): *talk* (9), *weasel* (4), *buy* (3), *find* (2), *negotiate* (2), *pay* (2), *work* (2), *worm* (2), *bargain* (1), *blitz* (1), *bluff* (1), *blurb* (1), *bungle* (1), *charm* (1), *cheat* (1), *claw* (1), *dance* (1), *magic* (1), *rationalize* (1), *reason* (1), *save* (1), *weave* (1), *wiggle* (1)
- I probably talked myself out of being a great champion more than I talked myself into it. (COCA-2008-MAG) (reflexive NP object)
 - Sometimes the lizard can't just run away. When that happens, it tries to bluff its way out of becoming a tasty treat. (COCA-2005-MAG) (possessive way NP object)
- The most frequent matrix verb in each of these atypical variants: *talk*
- The other verbs with low frequencies (lower than 4)

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 19/34

V1-V2[-ing] pairs

- 10 strongest covarying collexeme pairs in the matrix verb slot and the VP[-ing] predicate slot of the typical transitive *out of -ing* construction in COCA

Rank	Word1	Word2	Freq.W1	Freq.W2	Obs.W1.W2	LLR
1	screw	selling	5	2	2	20.18
2	cheat	witnessing	15	2	2	15.13
3	price	owning	9	3	2	13.58
4	scare	smoking	24	2	2	13.15
5	scare	using	24	7	3	10.55
6	bully	entering	4	1	1	10.35
7	frighten	bidding	5	1	1	9.85
8	screw	voting	5	12	2	9.47
9	intimidate	proceed ing	8	1	1	8.82
10	lock	broadcasting	9	1	1	8.57

- When we're creative directors, we'll **screw** other people out of **selling** their clearly superior campaigns. (COCA-2009-TV)
 - If I could get past it, I figured I could **cheat** millions of viewers out of **witnessing** the first 2,000-foot plummet of a TV host. (COCA-1996-MAG)
- Well-known TRICKERY-BELIEF frame in the transitive *into -ing* construction; but FORCE/TRICKERY/NEGATIVE EMOTION-ECONOMIC/POLITICAL/HEALTH-RELATED ACTIVITY frames here
 - Another difference between the transitive *into -ing* construction and the transitive *out of -ing* construction

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 18/34

Verbs in the V1 slot: semantic classification

- Matrix verbs found in the V1 (atypical one with a reflexive NP object):
 - By means of deception or trickery (1 type): *cheat*
 - By means of exerting force or pressure, sometimes understood metaphorically (4 types): *cut*, *pull*, *snap*, *take*
 - By means of arousing fear, irritation, anger, annoyance, confusion, or surprise (2 types): *embarrass*, *intimidate*
 - By means of enticing, flattering, or verbal persuasion (5 types): *argue*, *buy*, *rationalize*, *reason*, *talk*
 - By other specific means (1 type): *train*
 - By nonspecific means (2 types): *get*, *leave*
- Matrix verbs found in the V1 (atypical one with a possessive way NP object)
 - By means of exerting force or pressure, sometimes understood metaphorically (3 types): *blitz*, *wiggle*, *worm*
 - By means of enticing, flattering, or verbal persuasion (7 types): *bargain*, *buy*, *negotiate*, *pay*, *rationalize*, *reason*, *talk*
 - By other specific means (7 types): *blurb*, *bungle*, *charm*, *claw*, *dance*, *magic*, *weave*
 - By nonspecific means (3 types): *find*, *work*, *save*
- Prominent "enticing" and "trickery" verbs in (16a): prominent "enticing" and "specific means" verb in (16b)
 - Quite different from those found in the V1 slot of the typical transitive *out of -ing* construction

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 12-9-2023 20/34

Verbs in the V1 slot: distinctive colloceme analysis

- 10 strongest distinctive collocemes in the V1 slot of the typical transitive out of -ing construction and its two atypical variants depending on the NP object types in COCA

Rank	Typical (N = 620)		Reflexive (N = 63)		Possessive way (N = 42)	
	Colloceme	LLR	Colloceme	LLR	Colloceme	LLR
1	talk	406:41.9	argue	1:3.0	wasel	0:0.4
2	scare	24:0.0	reason	0:3:1	worm	0:0:3
3	keep	14:0:0	snap	1:2:0	buy	0:1:3
4	shut	11:0:0	embarrass	0:1:0	find	0:0:2
5	lock	9:0:0	cut	1:1:0	negotiate	0:0:2
6	price	9:0:0	rationalize	0:1:1	pay	0:0:2
7	get	33:3:0	buy	0:1:3	work	0:0:2
8	drive	5:0:0	leave	4:1:0	bargain	0:0:1
9	frighten	5:0:0	take	4:1:0	blitz	0:0:1
10	rule	5:0:0	train	4:1:0	blurb	0:0:1

- No particularly preferred semantic class(es) of verbs for the typical transitive out of -ing construction
- Preference for “verbal persuasion” verbs for the atypical variant with a reflexive NP object
- Preferences for some “action” verbs for the atypical variant with a possessive way NP object

Some distinctive properties of the two atypical variants compared to the typical one

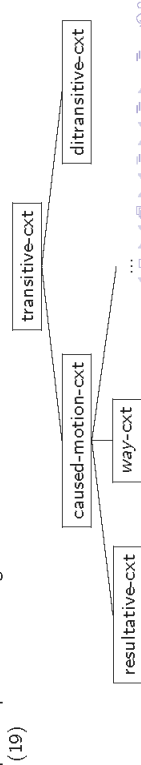
- No passive verbs in the two atypical variants
(17) a. *{Himself/His way} has to be talked out of doing this by his comrades.
b. *{Themselves/Their way} will be priced out of owning a car as a result.
- No occurrence with an inanimate subject in the two atypical variants
- No meaningful results from the covarying colloceme analysis; almost all pairs found only once

Key tenets of CG

- In the CxG framework, argument structures are also taken to be “constructions”, with a verb’s inherent core meaning being distinctive from the semantics linked to the argument structure constructions.
(18) a. Pat kicked the ball. (transitive construction)
b. Pat kicked his partner the puck. (ditransitive construction)
c. Pat kicked the car down the road. (caused-motion construction)
d. Pat kicked the door open. (resultative construction)
e. Pat kicked his way around the garden (way-construction)
- Related constructions and semantic properties

Construction type	Form	Semantic property
transitive-ctx	V Obj	X ACTS ON Y or X EXPERIENCES Y
ditransitive-ctx	V Obj1 Obj2	X CAUSES Y TO RECEIVE Z
caused-motion-ctx	V Obj Oblique	X CAUSES Y TO MOVE Z
resultative-ctx	V Obj Pred	X CAUSES Y TO BECOME Z
way-ctx	V one’s way Oblique	X CAUSES Y TO GO ALONG PATH Z

Hierarchical inheritance networks in CxG allow us to capture both generalizations and specific patterns among constructions.



A construction grammar perspective account

Transitive out of -ing construction

- (20) The transitive out of -ing construction
 [transitive-*cx*t & caused-motion-*cx*t & resultative-*cx*t]
 SYN: [NP1_x V NP2_y out of VP[-ing]_z]
 SEM/PRAG: x causes y to become not z
 & entail 'to become not z' actually happened
- (21) a. She cheated thousands of people out of getting a rebate. (prevention interpretation)
 transitive-*cx*t: She acted on thousands of people by cheating them.
 b. She caused thousands of people not to get a rebate by cheating them.
 caused-motion-*cx*t: She caused thousands of people not to get a rebate by cheating them and the event of their not getting a rebate actually happened.
 transitive-out of-ing-*cx*t: She caused thousands of people not to get a rebate by cheating them and the event of their not getting a rebate actually happened.
- (22) a. They shamed John out of crying. (movement/ extraction interpretation)
 transitive-*cx*t: They acted on John by shaming him.
 b. They caused John not to cry by shaming him.
 caused-motion-*cx*t: They caused John not to cry by shaming him.
 c. They caused John not to cry by shaming him and the event of his John's not staying crying actually happened.
 transitive-out of-ing-*cx*t: They caused John not to cry by shaming him and the event of his John's not staying crying actually happened.
- No postulation of the one-to-one mapping relation between the eventuality type of the VP[-ing] and the interpretation type.
- (23) You planning on bribing me out of writing the book.

Transitive out of -ing construction (cont'd)

- The construction-based approach proposed here suggests that a verb which can induce a causation reading can be used in the transitive out of -ing construction, although it is not commonly used in this way, just like the case with *talk*.
- (24) a. Got 73 wpm, and wish I was faster. Was homeschooled, and argued my parents out of making me put my hands in the correct positions. (COCA-2012-WEB)
 b. He could have told his roommates, and they'd have laughed him out of taking an old folktale to heart. (COCA-1993-FIC)
 c. My anger, honestly, has turned toward-toward legislators who have legislated me out of carrying my gun for my protection. (COCA-1992-SPOK)

Key properties of the way-construction

- The way-construction: 1) the possessive pronoun is coreferential with the subject; 2) the construction involves movement of the subject along a path expressed by a directional PP; 3) the directional PP modifies the path indicated in the possessive way
- (25) a. Rasselas dug his way out of the Happy Valley.
 b. The wounded soldiers limped their way across the field.
 c. Convulsed with laughter, she giggled her way up the stairs.
- The way-construction is also an extension of the caused-motion construction.

The atypical variant with a possessive way NP object

- Prediction: The way-construction would interact with the transitive out of -ing construction.
- (26) a. I was able to talk my way out of being killed, but it never should have gotten to that point. (COCA-1997-MAG)
 b. This is probably the biggest way companies weasel their way out of quoting you a full price. (COCA-2012-BLOG)
- While having similar properties to the typical transitive out of -ing construction, its atypical variant with a possessive way NP object has a distinctive characteristic on its own. Certain verbs are exclusively or more favorably used in the atypical construction.
- (27) a. He know they will ask the hard question and not let him dance his way out of answering the question. (COCA-2012-WEB)
 b. When it comes to cybersecurity, the worst thing you can do is run out and try to buy your way out of worrying about it. (COCA-2012-BLOG)
- It is rather difficult to come up with a situation where the someone's dancing or weaseling causes someone else but himself/herself not to be a resultant state described by the VP[-ing].
 - Verbs like *buy* and *find* are frequently observed in the general way-construction.

The atypical variant with a reflexive NP object

- The *way*-construction has several subtypes, including the one with a reflexive NP object (Mondorf 2011; Kim & Davies 2016).
- Prediction: The transitive *out of -ing* construction would license a variant with a reflexive NP object as well.
 - (28) a. She tried to **talk** herself out of doing it, unsure if her timing was right or wrong. (COCA-2004-FIC)
 - b. I was trying to **argue** myself out of snapping him up over the counter and beating some morals into him. (COCA-2004-FIC)
 - c. What can I do to **get** myself out of going? (COCA-2018-SPOK)

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 29 / 34 12-9-2023

Conclusion

Some differences between the two atypical variants

- Quite a few verbs found in the atypical variant with a possessive *way* NP object describe some type of motion.
- Quite a few verbs found in the one with a reflexive NP object depict some kind of mental activity or emotion.
 - (29) a. He's not Bush trying to **worm** his way out of explaining why thousands of Americans died in a pointless war. (COCA-2012-BLOG)
 - b. Well, I barely **clawed** my way out of being poor. (COCA-1992-SPOK)
- It helps you avoid that whole process of **rationalizing** yourself out of taking action on something you instinctively know is worth a try. (COCA-2012-WEB)
 - b. Have Republicans finally **embarrassed** themselves out of calling themselves the family values party? (COCA-2009-SPOK)

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 30 / 34 12-9-2023

Conclusion

- In this work, we presented an analysis of the grammatical properties of the transitive *out of -ing* construction based on data from COCA, performing different types of collocational analyses.
- Verbs associated with the V1 slot of the transitive *out of -ing* construction often belong to the semantic classes of 'force' and 'negative emotion' verbs, with an 'enticing' verb *talk* being the strongest collexeme.
- The construction favors animate subjects over inanimate ones and certain groups of verbs are strongly associated with inanimate subjects.
- The prevention interpretation is more salient than the movement/extraction interpretation in the transitive *out of -ing* construction and the alleged mapping relation between eventuality types and interpretation types can be overridden and the pairs have somewhat different correlation tendencies.
- Certain pairs of semantic frames are preferred in describing cause-result relations in this construction.
- The constructional analysis proposed here enables us to provide a neat and systematic explanation for the general grammatical properties of the transitive *out of -ing* construction as well as the idiosyncrasies of the two variants.

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 31 / 34 12-9-2023

JS Kim & R. Sim (INU & U of SC) Transitive out of -ing construction 32 / 34 12-9-2023

References

- Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Davies, Mark & Jong-Bok Kim. 2019. Historical shifts with the INTO-CAUSATIVE construction in American English. *Linguistics* 57(1): 29-58.
- Francis, Gill, Susan Hunston & Elizabeth Manning (eds.). 1996. *Collins COBUILD grammar patterns 1: Verbs*. London: HarperCollins.
- Gawlik, Oskar. 2013. On the transitive out of -ing complementation pattern in contemporary spoken American English: A corpus-based study. *Poznan Studies in Contemporary Linguistics* 49(2): 167-183.
- Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.
- Goldberg, Adele. 2005. *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Co-varying collexemes in the into-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind*, 225-236. Stanford: CSLI Publications.
- Hunston, Susan & Gill Francis. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Kim, Jong-Bok & Mark Davies. 2016. The INTO-CAUSATIVE construction in English: A construction-based perspective. *English Language and Linguistics* 20(1): 55-83.
- Kim, Jong-Bok & Laura A. Michaelis. 2020. *Syntactic constructions in English*. Cambridge: Cambridge University Press.
- Mondorf, Britta. 2011. Variation and change in English resultative constructions. *Language Variation and Change* 22: 397-421.
- Rudanko, Juhani. 1991. On verbs governing in -ing in present-day English. *English Studies* 72(1): 55-72.
- Rudanko, Juhani. 2002. *Complements and constructions*. Lanham: University Press of America.
- Rudanko, Juhani. 2011. *Changes in complementation in British and American English: Corpus-based studies on non-finite complements in recent English*. Basingstoke: Palgrave Macmillan.
- Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-based construction grammar*, 69-202. CA: CSLI Publications.
- Sim, Rok & Jong-Bok Kim. 2015. The transitive "out-of V-ing" construction in English: Corpus-based study [written in Korean]. *The Linguistic Association of Korea Journal* 23(4): 189-210.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.

The semantics of people and places: A corpus-assisted NSM analysis of the Tagalog affix *Taga-*

NICOLE PONTANILLA, EUGENE CHUNG
Department of Linguistics, Korea University

LSK 2023 Winter Conference
Kyunghee University
December 9, 2023

2

OVERVIEW

- I. Introduction
- II. Previous Research
- III. Methodology
- IV. Result: Distribution of *Taga-*
- V. Discussion: The Different Senses of *Taga-*
- VI. Summary and Conclusion

HUMAN SOCIAL CATEGORIES

- Demonyms
 - 'the name commonly given to the residents of a place' (Roberts, 2017)
- Occupation words
 - Then: 'how individuals spent their time' (OED, 1989)
 - Now: 'what people do for a living'

4

I. Introduction

THE TAGALOG AFFIX TAGA -

- Demyonym
 - taga-Manila 'a person from Manila'
 - taga-Pilipinas 'a person from the Philippines'
- Occupation word
 - tagaluto 'cook'
 - tagasalin 'translator'

5

THE TAGALOG AFFIX TAGA -

- Rubino (2002)
 - 'denotes origin'
 - 'may express occupation'
- Tagalog Reference Grammar (Schachter and Otones, 1983):
 - *taga-* + base form
 - 'a person from the place designated by the base'
 - 'a person employed or delegated to perform the action designated to perform the action designated by the related verb'

6

THE TAGALOG AFFIX TAGA -

- (1) hindi ba kayo nagtataka kung bakit Sinasabi ng mga **taga INC**
 'Don't you wonder why **members of the Iglesia ni Cristo** claim
 na sila ang tunay na relihiyon...
 that they are the true religion...'

7

THE TAGALOG AFFIX TAGA -

- (2) Sa pananaw ni ACT Teachers Rep. France Castro
 'In the perspective of ACT Teachers Rep. France Castro,
 mukhang hindi ginamit ng mga **taga-BSP**
 it seems that the **people from/working at BSP** did not use
 (Bangko Sentral ng Pilipinas) ang kanilang common sense...
 (Central Bank of the Philippines) their common sense...'

8

COGNITIVE SEMANTICS

- Prototype Theory
 - Eleanor Rosch (1973)
 - Prototype refers to the most typical example of a category.
 - The prototype sense represents the most central and prototypical meaning of a word, while other related senses exhibit varying degrees of resemblance to this prototype.

9

COGNITIVE SEMANTICS

- Radial Categories
 - Category structure as a radial network
 - The prototype, situated at the center, embodies the most prototypical members, while moving towards the periphery entails a decrease in prototypicality and shared defining features among members.

10

RESEARCH OBJECTIVES

1. to uncover the different senses of the polysemous Tagalog affix *taga-*
2. to identify which among those is the prototypical meaning of *taga-*
3. to determine how and why the peripheral senses emerged from the core meaning

11

II. Previous Research

CORPUS - ASSISTED NSM ANALYSIS

- Wierzbicka (1992, 1996); Goddard (1998):
 - Natural Semantic Metalanguage
 - from circular definition of words to decomposition to smaller concepts
 - semantic primes → semantic explication
 - NSM studies: emotion, body parts, kinship, etc.
- Roberts (2017)
 - Categorization of demonyms
 - semantic molecule “country”

13

CORPUS - ASSISTED NSM ANALYSIS

- Goddard (2020):
 - ‘country,’ ‘land,’ and ‘nation,’
 - people in specific geographical locations
- Roberts (2010):
 - occupational categories: doctors, lawyers, teachers, plumbers, electricians

14

III. Methodology

CORPUS - ASSISTED NSM ANALYSIS

- Corpus
 - Sketch Engine’s Tagalog (Filipino) Web 2019 (ttTenTen19) Corpus with around 198,300,000 entries
 - Tagalog.Com’s TDC Corpus Tool with around 24,800,000 entries
 - 모두의 말뭉치’s Korean-Tagalog Parallel Corpus with around 1,000,000 phrases
- Tool: AntConc

16

CORPUS - ASSISTED NSM ANALYSIS

- Natural Semantic Metalanguage Analysis
 - Goddard and Wierzbicka (1994, 1995, 2002, 2014)
 - The meaning of any linguistic expression can be explained using a set of universal semantic primes.
 - 65 semantic primes (+ semantic molecules [m])
 - I, YOU, SOMEONE, PEOPLE, ONE, TWO, SOME, THINK, KNOW, DO, HAPPEN, NOT, MAYBE, CAN, etc.

17

IV. Results

DISTRIBUTION OF TAGA -

DISTRIBUTION OF TAGA -

Table 1. 10 most frequent noun collocates of *taga-* in Sketch Engine

Noun	Cooccurrences	Relative Frequency	L-score
1 Corinto 'Corinth'	660	2.83	25.67
2 Roma 'Rome'	386	1.66	19.61
3 Efeso 'Ephesus'	282	1.21	16.78
4 Sait	203	0.87	14.11
5 Mindanao	196	0.84	13.98
6 Davao	180	0.77	13.40
7 Filipos	146	0.63	12.08
8 UP	141	0.61	11.80
9 Galicia 'Galicia'	140	0.60	11.82
10 Tesalónica 'Thessalonica'	138	0.59	11.74

19

Table 2. 10 most frequent verb collocates of *taga-* in Sketch Engine

Verb	Cooccurrences	Relative Frequency	L-score
1 <i>bilang</i> 'advise'	233	1.00	15.25
2 <i>sinubay</i> 'to monitor'	182	0.78	13.49
3 <i>suporta</i> 'support'	152	0.65	12.22
4 <i>sinood</i> 'to follow'	117	0.50	10.71
5 <i>bantay</i> 'guard/watch'	81	0.35	8.96
6 <i>linis</i> 'to clean'	79	0.34	8.86
7 <i>gawa</i> 'to make/do'	68	0.29	7.85
8 <i>alaga</i> 'to take care of'	66	0.28	7.95
9 <i>labas</i> 'to take out'	66	0.28	7.57
10 <i>luto</i> 'to cook'	53	0.23	7.23

20

DISTRIBUTION OF TAGA -

- *Taga-* + noun
 - Combines with categories of the same level
- (3) Tatlong **taga-Mindanao, dalawang taga-Luzon, at**
'The three contestants from Mindanao, two from Luzon, and
isa mula Visayas ang magbabakbakan para sa pagkakataong
one from Visayas will battle it out for a chance
 masungkit ang matupad ang kanilang pangarap...
to fulfill their dreams...'

21

DISTRIBUTION OF TAGA -

- *Taga-* + verb
 - Each form is considered distinct
- (4) Pero as much as possible, kung kaya ko magcontribute,
'But as much as possible, if I can contribute,
 magbibigay ako. Sa chores, all-around,
I will. With chores, I'd (help) all-around.
Taga-luto, taga-linis, taga-hugas, at taga-laba.
The one to cook, to clean, to wash (dishes), and to do laundry.'

22

DISTRIBUTION OF TAGA -

- Distributive property of *taga-*
 - Preceding clause indicates shared common ground
 - Does not apply to *taga-* + verbs
- (5) May nag-lu-lunch na **taga-UP, Ateneo at La Salle.**
'Someone from UP, Ateneo, and La Salle are having lunch.'

23

DISTRIBUTION OF TAGA -

- Summary:
- *Taga-* + noun
 - Combines with categories of the same level
 - *Taga-* + verb
 - Each form is considered distinct
 - Distributive property of *taga-*
 - Preceding clause indicates shared common ground
 - Does not apply to *taga-* + verbs

24

V. Discussion

THE DIFFERENT SENSES OF TAGA -

TAGA -1: DEMONYM TAGA -

'a person who is from the place designated by the base'

1.A Nationality-based taga-

[A] *Taga-Pilipinas* (*Tsina*, etc.)

- People of one kind
- Many people of this kind live in one country [m] at some time
- This country is called [m] *Pilipinas* (*Tsina*, etc.)
- Many people of this kind lived in this place for a long time before
- Because of this, people can think like this: people of this kind are like part of this

26

TAGA -1: DEMONYM TAGA -

- (6) Di ko alam may **taga Pilipinas** palang
'I didn't know that there's **someone from the Philippines/Filipinos**
di alam yung turon.
who don't know what turon is.'

27

TAGA -1: DEMONYM TAGA -

1.B Continent-based taga-

[B] *Taga-Europa* (*Asya*, etc.)

- people of one kind
- many people of this kind live in a one big place
- many countries [m] are parts of this place**
- people can say what place this is with the word: *Europa* (*Asya*, etc.)
- many people of this kind live in this place for a long time

28

TAGA -1: DEMONYM TAGA -
1.C Province-based taga-

- [C] *Taga-Cavite (Cebu, etc.)*
- people of one kind
 - many people of this kind live in one place
 - a place of this kind is one part of a country [m]**
 - this part of the country [m] is called *Cavite (Cebu, etc.)*
 - many people of this kind lived in this place for a long time before
 - because of this, people can think: people of this kind are like part of this place

29

TAGA -1: DEMONYM TAGA -
1.D City-based taga-

- [D] *Taga-Manila (Makati, etc.)*
- people of one kind
 - many people of this kind live in one city [m]**
 - there are many places of this kind in many countries [m]**
 - this city is called *Manila (Makati, etc.)*
 - many people of this kind lived in this place for a long time before
 - because of this, people can think like this: people of this kind are like part of this place

30

TAGA -2: RESIDENTIAL TAGA -

'a person who resides at the place denoted by the base'

- (7) maisasalba pa ang buhay ng biktimgang si Elias Bonifacio,
'*The life of the victim, Elias Bonifacio,*
may asawa, walang hanapbuhay, at taga-No. 138 CP Garcia St.,
who is married, unemployed, and resides at No. 138 CP Garcia St.,
Tondo, Manila.
Tondo, Manila, can still be saved.'

31

TAGA -2: RESIDENTIAL TAGA -

- [E] *Taga-(Specific address)*
- someone not like others
 - someone can live in one place at a time
 - this place is part of a city [m] or province [m]
 - this place is called (specific address)

32

TAGA -3: WORKPLACE TAGA -

'a person working at a place denoted by the base'

- (8) Ganyan ginagawa ko para sure na
'That's what I do to make sure that the caller
taga-bangko or agency talaga yung tumatawag.
is really **from the bank or agency.**'

33

TAGA -3: WORKPLACE TAGA -

'a person working at a place denoted by the base'

- (9) ...pero maaring kung sino man ang mananalo rito eh
'...but whoever wins here could potentially
ang i-re-rehaul ng mga taga-It's Showtime bilang
be rehauled by the **people at (staff of) It's Showtime** as the
"heiress apparent" ni Vice Ganda.
"heiress apparent" of Vice Ganda.'

34

TAGA -3: WORKPLACE TAGA -

[F] *Taga-bangko* (workplace names)

- people of one kind
- people of this kind are in a place for a time
- there are many places of this kind in one city [m] or province [m]
- people in this place think the same
- people in this place do the same kind of things
- this place is called *bangko* (workplace names)
- some people become part if this place because it is like this:
 - they know how to do some kind of things
 - they are good at it
 - someone said to them "You do it."

35

TAGA -3: WORKPLACE TAGA -

[F] *Taga-bangko* (workplace names)

- other people become part of this place because it is like this:
 - they do not know how to do it
 - someone thinks:
 - "They can do it after knowing how to do it for some time."
 - after doing something for some time, they know how to do it
- when they do these things in this place, something good will happen to them

36

TAGA -4: MEMBERSHIP-BASED TAGA -

'a person who is part of a group denoted by the base'

- (10) Si F. Manalo pa nga ang NAG UUTOS sa mga **taga INC** na 'it is even F. Manalo who ORDERS the **members of the INC**
 HUWAG KUMAIN NG DUGO KAHIT KAILAN.
 NOT TO EAT BLOOD UNDER ANY CIRCUMSTANCES.
 Wala sa Catholic Church ang gamitong Doktrina.
 This doctrine is not present in the Catholic Church.'

37

TAGA -4: MEMBERSHIP-BASED TAGA -

[G] *Taga-INC* (group name)

- a. people of one kind
- b. people of this kind think the same
- c. people of this kind know the same things
- d. people of this kind do the same kind of things
- e. sometimes, people of this kind are in one place
- f. these people are called: *INC* (group name)
- g. people who are part of this thinks:
 "Something good will happen if I am a part of this."

38

TAGA -5: OCCUPATIONAL TAGA -

'a person assigned to perform the task denoted by the base'

- (11) Sa aking napansin dati wala pa clang youtube channel
 'From what I noticed before, they didn't have a YouTube channel yet,
 cla Tita Benben & Yahzel, parang nagpatulong yata
 Tita Benben & Yahzel. It seemed like Yazhel was asked
taga bantay ni Jodalyn kay Yahzel...
to look after Jodalyn...'
- (12) Nag apply Na **Taga linis** Sa Opisina
 'applied as an office **janitor**.'

39

TAGA -5: OCCUPATIONAL TAGA -

[H] *Taga-luto* (*bantay*, etc.)

- a. people of one kind
- b. people of this kind do something
- c. they do this thing because it is like this:
 -they know how to do it
 -no one else will do it
 -someone else above them said: "Do this."
- d. when they do this thing, it is like this:
 -they can get something good from it
 -they can do something good for other people
- e. they do this thing in a place

40

VI. Summary and Conclusion

SUMMARY AND CONCLUSION

Different senses of *Taga-*

- A. *Taga-1* Demonym *Taga-*
- B. *Taga-2* Residential *Taga-*
- C. *Taga-3* Workplace *Taga-*
- D. *Taga-4* Membership-based *Taga-*
- E. *Taga-5* Occupational *Taga-*

42

SUMMARY AND CONCLUSION

- Prototypical sense: Demonym *Taga-1* as evidenced by the presence of semantic molecule “place” across all explications
- Residence *Taga-2* highlights individuality over shared identity (hyponym of *Taga-1*).
- Workplace *Taga-3* emphasizes occupation rather than residence.
- Membership-based *Taga-4*
 - similarity with *Taga-3*: “people of this kind think the same” and “people of this kind do the same kind of things”
 - difference with *Taga-3*: “these people are called...” rather than “this place is called...”.

43

SUMMARY AND CONCLUSION

- Occupational *Taga-5* seems the least prototypical in relation to *Taga-1*.
- *Taga-3* establishes a bridge between the geographical association of *Taga-1* and the task-based association of *Taga-5*.

44

SUMMARY AND CONCLUSION

- Motivations of *Taga*-s polysemy
 - Conceptual coherence
 - Conceptual extension
 - Specificity (Hyponymy)
 - Expression of semantic roles
- Utilization of corpus and NSM as tools for analyzing the polysemy of *taga*- proved to be highly productive.

Thank you.

LSK 2023 Winter Conference
Kyunghee University
December 9, 2023

45

K-pop 외국 팬들의 한국어 외래어 사용 현상 연구

황경진
고려대학교
서어서문학과

목차

<p>1. 서론 연구 배경 및 목적</p> <p>2. 연구 방법 연구 데이터의 모 분류 방법</p>	<p>3. 데이터 분석 결과 KPOP 팬들의 한국어 사용 현황 및 유형 분석</p> <p>4. 결론 본 연구의 의의 및 후속 연구</p>
---	--

서론

한국국제교류재단 2020년, 2022년 지구촌 한류연감
98개국 재외공관을 통해 받은 현지 한류 동호회 등락인 회원 현황
자료를 합산

1억 7800만명
2022년

1억 477만 7808명
2020년

9932만 8237명
2019년

서론

K-pop 팬들의 주요 소통 및 정보 교환 채널:
Twitter

#KpopTwitter 2021
Countries Tweeting most about K-pop

#KpopTwitter
2021년 KPOP 팬 토크 플랫폼 현황

전 세계 K-POP 인기 국가

전 세계

케이팝 레이다

225명의 유튜브 아티스트 자트 영상 조회 수 등을 집계

서론

- ✔ Check point K-pop 팬들이 K-pop에 관한 주제로 트윗을 할 때 나타나는 언어의 양상은 어떠한가?
- ✔ Check point K-pop 팬들에게 미치는 한국어의 영향은 어떠한가?
- ✔ Check point K-pop 팬들이 사용하는 한국어의 양상은 어떠한가?

연구 방법

1. 파이썬 크롤링을 통해 트위터 데이터를 수집
 - 기간: 2022년 1월 ~ 2023년 11월
 - 지역: 북미, 중미, 남미
 - 언어: 영어, 스페인어
 - 키워드: 세븐틴, 방탄소년단, 엔시티, 드림, 몬스타엑스, 블락핑크, 아이브, 트와이스, 에스파
2. 수집한 데이터의 특징 분석 (지역별, 언어별)
3. 수집한 데이터에서 한국어 사용이 관찰되는 트윗을 추출 후 유형 분류 (한국어 사용 양상 분석)




그림 4


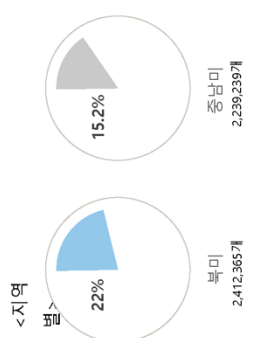


그림 5

데이터 분석 결과 - 지역별, 언어별 데이터 특징

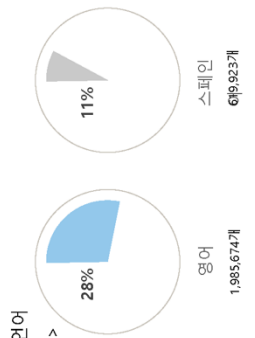
지역별 K-pop 팬들의 한국어 사용 특징

<지역>



지역	비율	개수
북미	22%	2,412,365개
중남미	15.2%	2,239,239개

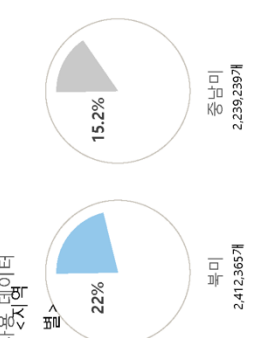
<언어>



언어	비율	개수
영어	28%	1,985,674개
스페인	11%	619,923개

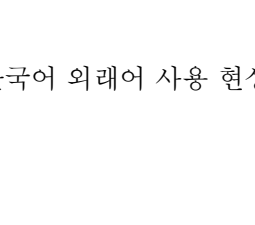
파이썬 크롤링 수집 전체 데이터와 한국어 사용 데이터

<지역>



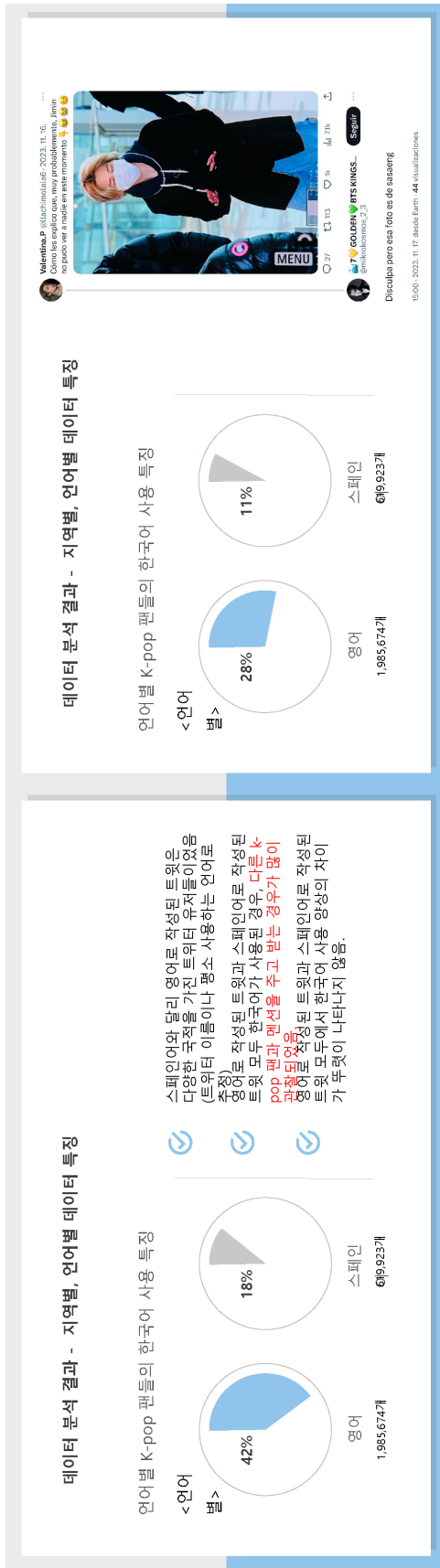
지역	비율	개수
북미	22%	2,412,365개
중남미	15.2%	2,239,239개

<언어>



언어	비율	개수
영어	28%	1,985,674개
스페인	11%	619,923개

북미보다 중남미 지역에서 k-pop 관련 트윗의 수와 그 중 한국어 사용의 빈도가 높게 나타난다.
그러나 중남미에서 관찰된 한국어 사용 트윗의 경우, 중남미에서 쓰여졌지만 영어로 작성된 트윗이 많았음.
북미에서 작성된 트윗의 경우 영어로 작성된 트윗 뿐만 아니라 인도네시아어, 한국어, 스페인어 등 다양한 언어가 관찰됨.
중남미에서 작성된 트윗의 경우 스페인어 뿐만 아니라 포르투갈어로 작성된 트윗도 30%정도 차지함.



데이터 분석 결과 - K-pop 팬들이 사용한 한국어의 양상

- "media play(인플레이의 변역) - 영어로 "media manipulation"이라는 단어가 있음.
- "do comeback(comeback하다)" - 영어로 역시 "release new album"이라는 단어가 있음.
- "dispatch(디스패치, 한국 언론사, 문맥상 활석연분을 뜻하여 사용) - 영어로 "yello journalism"이라는 단어가 있음.

유형 3

1%

영향도 비도수
각도로 작은 비중을 차지

- 별도의 영어 단어가 있음에도 불구하고 한국어 단어를 직역한 결과를 사용
- 스페인어 화자들도 한국어 단어의 한국어 발음을 쓸 때 스페인어가 아니라 영어의 발음 규칙 대로 표기하는 현상이 관찰됨.

결론

- 영어, 스페인어 및 기타 언어 화자 및 트윗이 작성된 지역과 상관없이 일반적으로 한국어는 영어의 발음의 규칙을 따라 사용 됨.
- 트위터 상에서 멘션을 서로 주고받을 때 기본 공용어는 영어. 그러나 k-pop 과 관련된 단어는 영어에 있는 단어에도 불구하고 한국어의 단어를 사용하는 것이 관찰됨.
- K-pop 문화를 향유하는 사람들에게 한국어의 단어들은 일종의 공용어이자, 언어로 기능함.

연구의 의의 및 후속 연구

- k-pop이 한국어 및 한국 문화 전파에 기여하고 있으며 k-pop내 공용어처럼 기능하고 있음을 확인.
- 향후 k-pop 관련 번역에 있어서, 번역의 질을 더 높일 수 있음.

박소영 (2019). 이베로아메리카 한류와 번역. 방탄소년단 콘텐츠의 스페인어 팬 번역을 중심으로. 한국언어문화, 69(9), 91-116.

>> 자문번역보다 팬들이 한국어 k-pop 콘텐츠를 직접 번역한 것이 흥미어. 한국어 언어 등을 더 효과적으로 번역하고 있음.

- 추기적인 한국어 사용에 대한 연구가 필요하며 이를 교육 및 번역에 활용할 수 있는 방안을 찾는 연구가 필요함.

참고 문헌

박소영. (2019). 이베로아메리카 한류와 번역. 방탄소년단 콘텐츠의 스페인어 팬 번역을 중심으로. 한국언어문화, 69(9), 91-116.

그림 출처

- <그림 1> 전자신문 < 글로벌 웹 인기를 한눈에...2022 케이팝 세계지도, 공개 >, <https://www.wtnews.com/20221123000188>
- <그림 2> 트위터 공식 블로그 <#KpopTwitter 2021: 진 세계서 78억 트윗 발생+기록 경신! >, https://blog.twitter.com/ko_kj/topics/insights/2021/kpoptwitter-2021
- <그림 3> 트위터 공식 블로그 <#KpopTwitter 2021: 진 세계서 78억 트윗 발생+기록 경신! >, https://blog.twitter.com/ko_kj/topics/insights/2021/kpoptwitter-2021
- <그림 4> BBC 뉴스 코리아 <엘비 머스크, 트위터 '파랑새' 대체할 새 로고 'X' 공개 >, <https://www.bbc.com/korean/articles/20221229x>
- <그림 5> SM엔터테인먼트



The Effect of Age, Morphological Structure, and Syllable-Count on the Realization Of Korean Nasal-Liquid Sequences

December 9, 2023

Rok Sim¹, Drew Crosby², D. Eric Holt¹





OVERVIEW:

This presentation investigates variation in the directionality of assimilation in Korean nasal-liquid sequences

1. Data

2. Research Questions

3. Methodology

4. Statistical Findings

5. Formal Analysis (Vestige Theory)

6. Conclusion

Assimilation in Nasal-Liquid sequences in Korean

- The assimilation pattern of the nasal-liquid sequences in Korean is complex (Kang 2002; Sohn 2006; Kwon 2014): [l] in (1) and [nn] in (2).

(1) /nl/ → [l]	(bound morpheme + bound morpheme) a. /kin/ 'diligent' + /lo/ 'work' b. /pan/ 'counter' + /lon/ 'discussion'	[killo] [pallon]	(regressive assimilation) 'labor' 'counterargument'
(2) /nl/ → [nn]	(stem (word) + bound morpheme) a. /nokin/ 'Nokeun' + /li/ 'small town' b. /daehan/ 'Daehan' + /lo/ 'street'	[nokimni] [daehanmo]	(progressive assimilation) 'small town (Li) of Nokeun' 'Daehan Street'

- The key difference in directionality of assimilation: morphological structure

Variation: [nn] ~ [li]

- For some of the data in (2) 'stem (word) + bound morpheme', where [nn] is expected, there is instead show variation in the directionality of the assimilation (Kwon 2014), as in (3).

(3)	a. /nokin-li/ [nokinni] b. /daehan-lo/ [daehanmo]	(i) [nn] (ii) [li]	[nokilli] ʔ[daehallo] 'small town (Li) of Nokeun' 'Daehan Street'
-----	--	-----------------------	--

- Although [nn] is dominant, [li] also surfaces.
- Additionally, we see a discrepancy in allowing for the variation between the words in the same concatenation type.
- For (3a), [nn] and [li] are allowed, but only [nn] is allowed for (3b).

Variable assimilation in Nasal-Liquid sequences in Korean loanwords

Data summary

- Additionally, variation in the directionality of assimilation is observed in English loanword compounds as in (4):

(4) Loanword compounds (/n + l/)

	(i) [nm]	(ii) [nl]
a. /wani/ 'one' + /lumi/ 'room'	[wʌllumi]	[wʌllumi] 'studio apt (lit. one room)'
b. /taun/ 'down' + /loti/ 'load'	[taunnodi]	[taullodi] 'download'

- Canonical assimilation in loanword compounds is illustrated in the (i) forms (Sohn 2006), which shows that the /n + l/ sequence shows progressive assimilation, surfacing as [nm], with regressive assimilation to [l] in the forms of (ii).
- We again observe differential acceptability between (4a) and (4b).

- Type 1: If an /n/ sequence occurs at a boundary between two bound morphemes, [l] is preferred.

Concatenation Type

- Korean : [bound morpheme] + [bound morpheme]

- Type 2: If an /n/ sequence occurs at a word boundary, [nm] is preferred.

- Korean : [stem (word)] + [bound morpheme]
- English loanwords : [word.] + [word.]

KWON (2014)

KIM (2023)

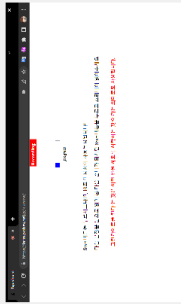
- Kwon (2014) also notes that there is variability in the direction of assimilation in /n/ sequences
- Investigates variation among three 21-year-old Seoul speakers in a factor-controlled study
- Argues that results suggest that
 - words with more syllables tend toward [nm] realizations
 - Hypercorrection is not a factor in assimilation
 - Priming effects are not a factor in assimilation

- Kim (2023) performed an experiment on /n/ assimilation in three-syllable Sino-Korean words among women in their 20s and 30s from the Seoul Capital Area
- Results suggest roles in [nm] or [nl] realization for the following facts
 - Word boundary recognition: [nm] more likely there is a word boundary between /n/ and /l/
 - Subjective lexical familiarity (for three-syllable forms): normative form more common the more familiar a word was (tends toward [nm] when the participants are not familiar with the word).
 - Age: In two-syllable words, women in their 30s realize [nl] more often; they also realize in a task designed to elicit an [nl] reading.

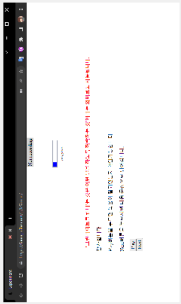
Procedure

Overall data set

- PClibex: <https://www.pclibex.net/> (Zehr and Schwarz 2018)
- Three parts: instruction, practice, and main task
- Main task



Recording screen



Confirmation screen

RESULTS

Raw results: 'bound morpheme + bound morpheme'

- Concatenation type like (1): regressive assimilation

(1) /n/ → [ŋ]

a. /kin/ 'diligent'	+ /o/ 'work'	[kiŋo] 'labor'
b. /pan/ 'counter'	+ /on/ 'discussion'	[paŋon] 'counterargument'

Out of 66 instances, 63 [ŋ] (95.5%) and 3 [nn] (4.5%)

- Variation was found (very marginally present).

**Raw results:
'stem (word) + bound morpheme'**

**Raw results:
'stem (word) + bound morpheme'**

- Concatenation type like (3): expected progressive assimilation

(3)

a. /nokin/ 'Nokseun' + /li/ 'small town' [nokimil] 'small town (li) of Nokseun'

b. /daehan/ 'Daehan' + /lo/ 'street' [daehallo] 'Daehan Street'

Out of 122 instances, 45 [li] (35.2) and 83 [nn] (64.8%)

- Variation was confirmed.

- Effect of frequency (RQ1)**
- Less frequent stem: more [li]
 - Nokseun 152,000 results (Google)
- More frequent stem: more [nn]
 - Daehan 453,000,000 results (Google)
- Words with less frequent stems more likely to be perceived as Type 1 (bound + bound)
- No apparent gender effect

Table 1. The assimilation pattern of the word 'stem (word) + bound morpheme' by frequency

**Statistical results:
Stem (word) + Bound Model**

**Statistical results:
Stem (word) + Bound Model**

- Model: Assimilation Type ~ age + frequency
- Age effect: For every year they were 5.6% less likely to realize [li]
- Frequency effect: High-frequency words are 78.4% less likely to be realized as [li] than low-frequency words
- No significant gender effect

Factor	Estimate	Lower CI	Upper CI	p-value	Significance
Intercept	11.591	2.849	53.602	0.000990	***
Age	0.944	0.913	0.974	0.000504	***
Frequency	0.216	0.0903	0.489	0.000355	***

Note. *** = $p < .0001$, ** = $p < .001$, * = $p < .01$, = $p < .05$

**Raw results:
'word₁ + word₂' (loanwords)**

**Raw results:
'word₁ + word₂' (loanwords)**

Concatenation type like (4): progressive assimilation expected

(4) Loanword compounds (n + V) [m] [l]

a. /wɔn/ 'one' + /lɔm/ 'room' [wɔnmɔm] [wallum] 'studio apt (lit. one room)'
 b. /taun/ 'down' + /lotɔ/ 'load' [taunlotɔ] [taunlotɔ] 'download'

Out of 130 instances, 39 [l] (30.0%) and 91 [m] (70.0%)

- Variation was confirmed.

Assimilation pattern for loanword compounds by initial-morpheme syllable count

Table 2. The assimilation pattern of the word 'word₁ + word₂' by the syllable-count

**Statistical results:
Loanword model**

Factor	Estimate	Lower CI	Upper CI	p-value	Significance
Intercept	1.017,757	32.406	1.64745,4	0.00104	***
Age	0.822	0.707	0.905	0.00136	***
Syllable-count: Two	0.00549	0.000168	0.140	0.00199	***
Gender: Men	0.156	0.000894	8.444	0.409	
Age * Syllable count: Two	1.092	1.005	1.192	0.0371	*
Age * Gender: Men	1.086	0.978	1.265	0.189	

Note. *** = p < .0001, ** = p < .001, * = p < .01, * = p < .05

- Model:** Assimilation Type ~ age + morpheme syllable count + gender + age * morpheme syllable count
- Age effect:** Participants over 35 are 92.0% less likely to realize [l] than participants under 35 (p=0.00136**)
- Syllable-count effect:** Two-syllable initial words are 88.1% less likely to realize [l] than one-syllable words (p=0.00199**)
- Age * Syllable-count interaction:** Much more drastic age effect for one-syllable morphemes than two-syllable morphemes (p=0.371*)
- No significant gender effect

Likelihood of realizing /n/ as [l] by age and initial-morpheme syllable count

Intraspeaker variation (RQ4)

Word (stem) + bound morpheme

% of realization by participant

- Intraspeaker variation** found for both Korean) 'stem + word' (bound morpheme)
- General age pattern visible but:
 - Some older speakers like M63/G produce a fair amount of [l]
 - Many younger speakers like F32C, I are [m] dominant despite the overall pattern of younger speakers being [l] dominant or evenly divided

Intraspeaker variation (RQ4)

Word (word) + bound morpheme (/n + l/)

% of realization by participant

- Intraspeaker variation** found for both Korean) 'word + word' (loanword)
- General age pattern visible but:
 - Some older speakers like M67/G produce a fair amount of [l]
 - Many younger speakers like F31J, I are [m] dominant despite the overall pattern of younger speakers being [l] dominant or evenly divided

DISCUSSION OF VARIATION

- Age effects suggest a change in progress.
- Frequency effects for word (stem) suggest that speakers who may be unfamiliar with the word (stem) for less frequent stems and thus analyze them as one word consisting of three roots rather than a word plus a derivational suffix.
- Loanword compounds with two-syllable initial morphemes favoring [m] realizations over those with one-syllable initial morphemes aligns with Kwon's (2014) suggestion that longer words favor [m] realizations.
- The rapid increase in lateralization of loan compounds with a one-syllable initial word relative to those with a two-syllable initial word suggests that younger speakers may be realizing one-syllable initial words as a Sino-Korean root due to their similarity (Sohn, 1999).
- Frequent interspeaker variation suggest additional demographic factors at play.

Formal analysis of observed variation

Reminder of data:

(3) stem (word) + bound morpheme (/n + l/)
 (3) [l] [m] [l] [l]
 a. /nokin/ 'Nokseun' + /ij/ 'small town' [nokilij] 'small town (L) of Nokseun'
 b. /daehan/ 'Daehan' + /be/ 'street' [daehanbe] ?[daehallo] 'Daehan Street'

(4) loanword compounds (/n + l/)
 a. /wan/ 'one' + /lomy/ 'room' [wanno] [wanno] 'studio apt (lit. one room)'
 b. /taun/ 'down' + /lot/ 'load' [taunlot] ?[taunlot] 'download'

Formal analysis of observed variation, cont.

- The underlying motivation for /nl/ assimilation in Korean is to make the first consonant /n/ as sonorous as the second consonant /l/ in the sequence /nl/.
- Davis and Shin's (1999) OT account employs SyllCon (Syllable Contact Law)
- (6) Syllable Contact Law (Vennemann 1988)
 - When two syllables are in contact, the coda of the first syllable should be of equal or greater sonority than the onset of the second syllable.
 - SyllCon – Avoid rising sonority over a syllable boundary

Other (faithfulness) constraints:

- Max-[lateral]: The feature [lateral] of an input segment is realized in the output
- Max-[nasal]: The feature [nasal] of an input segment is realized in the output

- The markedness constraint SyllCon is ranked higher than the faithfulness constraints Max-[lateral] and Max-[nasal].

(7) /non/ 'discussion' + /li/ 'deal with' [nonli] 'logic'

(8)

/non-li/	SyllCon	Max [lateral]	Max [nasal]
a. non.li	*!		
b. non.ni		*!	
c. *n. nol.li			*!

- Other OT accounts of variation extended analysis based on morpheme type, (Kang 2002) by invoking output-output constraints and extended the analysis to loanwords (Sohn 2006).

Kang (2002)

- Kang (2002) brings up new data that cannot be accounted for by Davis and Shin (1999).
- She identified some data with the /nl/ sequence that surfaces as [nn] contrary to Davis and Shin's (1999) prediction.

(9) /sin.mun/ 'Simmun' + /to/ 'street' [sinmunno] 'Simmun street'

(10) /im.nu/ 'Phonology' + /on/ 'study' [inunnon] 'Phonology'

- Kang (2002) adopts a new Output to Output constraint: Faith-OO – Word nodes in a paradigm are required to be phonologically identical.

- When a word is suffixed with a bound morpheme, that word node (i.e., base) is evaluated by Output-Output correspondence relation.

○ /simmu/ 'Simmun (name)' + /to/ 'street' → [sinmunno] 'Simmun street'

- Kang (2002) ranks Faith-OO higher than Max-IO [lat] and Max-IO [nas].

(13)

/sin.mun + to/	SyllCon	Faith-OO	Max-IO (lat)	Max-IO (nas)
Base: simmun				
a. sin.mun.to	*!			
b. sin.mun.lo		*!		*
c. *n. sin.mun.no			*	

Sohn (2006)

- Sohn (2006) basically extends Kang's (2002) analysis to loanwords.

(16) Loanword compounds (na + L)

a. /na:ni/ 'one'	/na:ni/ 'one'	→ [na:ni:ni]	'studio apt (one room)'
b. /na:ni/ 'down' + /to:si/ 'lead'		→ [na:ni:to:si]	'downlead'

- Note that the liquid is underlyingly specified for laterality in Korean (i.e., /l/), but not specified for laterality in loanwords (i.e., /l/) (Sohn 2006).

More about the laterality in a different lexical strata (Sohn 2006)

- There is just one liquid phoneme in Korean; no distinction between /l/ and /lʰ/ and written as 'ㄹ'
- This 'ㄹ' behaves differently with respect to phonology according to the lexical strata.
- For Korean stratum, the liquid surfaces as [l] word-initially as [l] in intervocalic position, and as [l] word-finally.

(i)

a. /lɔn/ + /ka:ŋ/	[lɔnka:ŋ]	'dispute'	liquid surfacing as [l]	결(論)쟁
b. /ho/ + /lo:n/	[ho:lo:n]	'discussion'	liquid surfacing as [l]	토(론)쟁
c. /mi/	[mi]	'water'	liquid surfacing as [l]	물

- For the loanword stratum in Korean, the liquid surfaces as a flap [ɾ] word-initially or as a lateral [l] in word-finally.

(ii)

a. /lo:pi/	[lo:bi]	'lobby'	liquid surfacing as [ɾ]	로(비)
b. /mi:l/	[mi:l]	'triple'	liquid surfacing as [l]	마(일)

Vestige Theory Kostakis (2010, 2017)

- To formally model this, we employ Kostakis' Vestige Theory approach within Optimality Theory
- Vestige Theory posits that, in accordance with language change, constraints are demoted in the grammar but a **vestige constraint** is left behind that flashes **on/off** according to factors such as social context and word-frequency, thus licensing variation
- Specifically, we propose that the formerly high-ranked constraint Max-(nasal/coda) is demoted to below Max-(lateral)
- This constraint demotion leaves behind an output-sensitive vestige constraint (the shaded column) as in the following tableau:

(17)

	Input	Max-(lateral)	Max-(nasal/coda)	Max-(lateral)	Max-(nasal/coda)
a.	la:mbud	*1			
b.	la:mbud	*1			
c.	la:mbud		*1		
d.	la:mbud			*1	
e.	la:mbud				*1
f.	la:mbud				

- As in loanword adaptation, the input contains unspecified liquid with respect to laterality; I-O correspondence with respect to the features [+sonorant] and [+approximant] needs to be added.
- Sohn (2006) utilizes new constraints such as Max-IO(son) and Max-IO(appr).
- Max-IO(son) – The feature [son] from an input segment is realized in the output.
- Max-IO(appr) – The feature [appr] from an input segment is realized in the output.
- *RR: the flap [ɾ] is less preferred in geminate liquids (Kawahara 2005)
- Max-O(nasal) is a more specified version of Faith-O (cf. Kang 2002)

Conclusion

- This paper investigated a variation of the assimilation pattern in the nasal-liquid sequence in Korean, focusing on what factors induce the variation.
- With results from the production survey, we find that the factors such as 'word frequency', 'syllable count' and 'age' are relevant to the directionality of the assimilation pattern. Furthermore, we found intraspeaker variation in the assimilation pattern.
- To account for the variation, we proposed an analysis within the framework of Vestige Theory.
- We presuppose that a formerly high-ranked constraint (Max-[nasal/coda]) is demoted to a lower position than Max-[lateral] leaving behind a vestige constraint in the original position.
- This vestige constraint flashes **on** and **off** according to both external social factors (e.g., speech community including age-based communities of practice (Eckert and McConnell-Ginet 2007)) and internal linguistic factors (e.g., word frequency and syllable size), which allows for the licensing of the variation.

Input: /nokin.li/	SYLLCON	MAX- (nasal/coda) ^{vestige}	MAX- (lateral)	MAX- (nasal/coda)	MAX- (nasal)
Output: [nm] ~ [ll]	*1	(*1)	*1	*	*
a. nokin.li					
b. nokli.li					
c. Vestige.c. nokin.ni					

This output-sensitive vestige constraint probabilistically activates based on factors such as


- **age:**
 - for older individuals, the probability of the vestige constraint being on is higher, and thus, candidate (c), [nm], is more likely to surface than
 - for younger individuals whose vestige constraint is more likely to be off, and thus to realize candidate (b), [ll]
- **frequency:**
 - for higher-frequency words, the probability of the vestige constraint being on is higher, and thus, [nm] is more likely to surface than [ll]
 - for initial loanwords with more than one syllable the probability of the vestige constraint being on is higher, and thus, the candidate with [nm] is more likely to surface than [ll]

THANK YOU!

- Feedback welcome
- Rok Sim, rsim@email.sc.edu
- Drew Crosby, dmcrosby@korea.ac.kr
- D. Eric Holt, deholt01@mailbox.sc.edu

Directions for future research

- A more in-depth survey with more speakers of various demographic groups
- Coding [nm] or [ll] by acoustic features, not only impressionistically
- Social-indexical meaning of [ll] or [nm] still needs to be explored
- Spelling pronunciations for words comprising 'bound + bound' morphemes reported to exist as spelling pronunciations. (Soo 2022)
- Other weakening of the syllable contact law, e.g., /okilp/ as [toŋilp] or only partial realization examination of spelling pronunciation in future research (Jun 2000)
- --> Additional analysis to account for such pronunciations (e.g., a vestige constraint for SyllCon and possibly multiple SyllCon constraints)
- Use of subjective lexical familiarity (SLF) rather than frequency as a model factor (Song & Dalaja 2019, Kim 2023)



References

- o Benua, Laura. 1997. Transderivational Identity: Phonological Relations between Words. PhD. Dissertation. University of Massachusetts. Amherst.
- o Cucchia-Espinosa, Juan Antonio Cucchias. 2004. Meaningful variability: A sociolinguistically-grounded approach to variation in Optimality Theory. *International Journal of English Studies* 4.2: 165-184.
- o Davis, Stuart and Seung-Hoon Shin. 1999. The Syllable Contact Constraint in Korean: An Optimality-Theoretic Analysis. *Journal of East Asian Linguistics* 8: 285-312.
- o Downing, Laura J., T. Alan Hall, and Renate Raffeliefen. 2005. Paradigms in phonological theory. Oxford University Press.
- o Eckert, P. and S. McConnell-Ginet. 2007. Putting communities of practice in their place. *Gender & Language* 1(1).
- o Fukazawa, Haruka, Mafuyu Kitahara, and Mitsuhiro Ota. 1998. Lexical stratification and ranking invariance in constraint-based grammar. *CLS* 34.2: 47-62.
- o Jun, Jongho. 2000. Preliquid nasalization. *Korean Journal of Linguistics*, 25(2), 191-208.
- o Kang, Hyunsook. 2002. On the optimality-theoretic analysis of Korean nasal-liquid alternations. *Journal of East Asian Linguistics* 11(1): 43-66.
- o Kim, Deulkyeop. 2023. A study on the realization of the /-r- / sequence according to word boundary recognition. Master's Thesis. Korea University.
- o Kostakis, Andrew. 2010. Vestige Theory: Sociolinguistic evidence for output-output constraints. *Lingua* 120: 2476-2496.
- o Kwon, Iksoo. 2014. Arbitrary assimilation of /n/ in Korean. *Language and Linguistics* 64: 1-30.
- o Seo, Yunjeong. 2022. Factors affecting the realization of consonant substitution in Korean. PhD Dissertation. Korea University.
- o Sohn, Ho-Min. 1999. *The Korean Language*. Cambridge University Press.
- o Song, Hyang-Sook. 2006. Morphological basis of sonorant assimilation in Korean. *Studies in Phonetics, Phonology and Morphology* 12(2): 321-338.
- o Sehn, Jyven, and Amanda Dabela. Linguistic entrenchment and the effect of subjective lexical familiarity in Korean. In: *Proceedings of the Linguistic Society of America* 4, no. 1 (2019): 34-1.
- o Venenmann, Theo. 1988. *Preference Laws for Syllable Structure*. Mouton de Gruyter. Berlin.
- o Zähr, Jeremy and Florian Schwarz. 2018. Penn Controller for Internet Based Experiments (IBEX). <https://doi.org/10.17605/OSF.IO/MD832>

STIMULI: BOUND + BOUND

- 근로를 하다가 다치거나 사망하게 되면 산재보상이 도움을 줍니다.
- 잔리를 위하여 몸을 바치는 사람이 되게 하여주십시오.
- 혼란을 제대로 받아야 성공할 수 있다.

STIMULI: STEM (WORD) + BOUND

- 잔흔리를 무대로한 소설이다.
- 거흔리를 거쳐 봉화군까지 달려갔다.
- 여흔리를 통하여 최측으로 가면 설악산이 나온다.
- 대원로를 따라 약 2키로를 직진하면 집이 나온다.
- 음운론을 공부하고 있는 국문과 학생입니다.
- 생선류를 많이 섭취하지 못한다.

STIMULI: LOANWORD COMPOUNDS

- 원품을 여러 개의 공간으로 쪼개서 사용한다.
- 인라인을 이제 막 타기 시작하려는 학생입니다.
- 선라이트를 받고 추천하지 않도록 허락하는 것이라는 의미로 도 사용됩니다.
- '그린라이트를 킨다'는 것은 어떤 일을 하도록 허락하는 것이라는 의미로 사용됩니다.
- 클린랩스를 이용하면서 궁금한 점이 있으면 연락주세요.
- 놀라이언을 가장 존경하는 투수로 꼽았었다.

2023 Linguistic Society of Korea
Winter Conference
Proceedings



한국언어학회
The Linguistic Society of Korea