



한국영어학회(KASELL)-한국코퍼스언어학회(KACL)
2022년 공동 가을 학술대회

Trends and Challenges in Linguistics in Times of New Normal and Big Data

2022.10.22 Sat 9:30~18:00

부산대학교 인문대학 U층 409호, U11호, U12호

공동주회: 한국영어학회(KASELL)

한국코퍼스언어학회(KACL)

주관: 부산대학교 영어영문학과

후원: 연세대 언어정보연구원



한국영어학회(KASELL) · 한국코퍼스언어학회(KACL) 공동 가을학술대회

준비위원회

준비위원장: 장경철(부산대)

위 원: 전혜원(한국해양대) 전혜원(한국해양대) 김태은(방송통신대)

황종배(건국대) 이지연(창원대) 최재호(상명대)

안병길(국립경상대) 육청민(한림대)

조직위원회

공동조직위원장: 장세은(한국해양대) 류병래(충남대)

위 원: 김수정(창원대) 이도경(인제대) 민수정(공주대)

박분주(대구카톨릭대) 우길주(부산교대) 함현종(Univ. of Guam)

이승아(이화여대) 윤영은(이화여대) 김동국(영산대)

박명관(동국대) 엄철주(전남대) 이제영(전주대)

이혜경(아주대) 홍문표(성균관대)

프로그램위원회

공동프로그램위원장: 이석재(연세대) 김한샘(연세대)

위 원: 송상헌(고려대) 최영주(조선대) 홍수민(서울시립대)

김동환(해군사관학교) 백주현(부산교대) 이성용(고신대)

강병규(서강대) 김일환(성신여대) 박선민(KAIST)

오윤정(대만국립정치대학교) 윤영민(연세대) 윤홍옥(제주대)

이은주(이화여대) 조준형(국립경상대) 황규홍(동아대)

황은하(배재대)



한국영어학회(KASELL)-한국코퍼스언어학회(KACL) 2022년 공동 가을 학술대회

Trends and Challenges in Linguistics in Times of New Normal and Big Data

2022.**10.22** Sat **9:30~18:00**

부산대학교 인문대학 4층 409호, 411호, 412호

공동주회: 한국영어학회(KASELL)

한국코퍼스언어학회(KACL)

주관: 부산대학교 영어영문학과

후원: 연세대 언어정보연구원



한국영어학회(KASELL) · 한국코퍼스언어학회(KACL) 공동 가을학술대회

- 주제: Trends and Challenges in Linguistics in Times of New Normal and Big Data

- 일시: 2022년 10월 22일 (토) 09:00-18:00 - 장소: 부산대학교 인문대학 4층 409호, 411호, 412호 (대면)

- 공동주최: 👣 한국영어학회 🔼🖍 한국코퍼스언어학회 - 주관: 🕮 부산대학교 영어영문학과

- 후원: 🔞 연세대학교 언어정보연구원

09:00~09:20 -	구분	코로나 방역 및 등록 Registration/Entrance				
		Opening Ceremony (한국영어학회 장세은 회장/ 한국코퍼스언어학회 류병래 회장)				
09:20~09:30	· 호 호 호 호 호 호 호 호 호 호 호 호 호 호 호 호 호 호 호	전혜원 (한국해양대학교, 한국영어학회 총무)				
	발표	Invited Talk 1 (김종복, 경희대) Pseudogapping in English: A Usage-based Construction Grammar Perspective (K)				
토	론자		최인철 (경북대)			
10:20~10:30			Coffee Break			
		Session A: 영어학 (ROOM 412)	Session B: 코퍼스언어학 (ROOM 411)	Session C: 기타 (ROOM 409)		
조	좌장	이석재 (연세대)	김동국 (영산대)	장경철 (부산대)		
사:	회자	김태은 (한국방송통신대)	전주현 (충남대)	권연진 (부산대)		
10:30~11:00	발표	Aphiwit Liang-Itsara (Mahidol University, Thailand) The Comparison of Motivating Factors for Taking English and Other Foreign Languages among Thai College Students (E)	강남길 (극동대) On <i>Consist of</i> and <i>Made up of</i> . A Corpora-based Analysis (K)	Cuilin Liu and Se-Eun Jhang (Korea Maritime & Ocean University) A corpus-based study of gender-inclusive words in the maritime conventions of "seamen" or "seafarers" (E)		
토	론자	류미림 (한국해양대)	김준기 (동서대)	Mengyu Xu (Korea Maritime & Ocean Univ.)		
11:00~11:30	발표 .	김원빈 (Univ. of Georgia, USA) Application of Artificial Intelligence to Semantic Prosody Research (K)	강아름 (충남대) 부정극어 wh-so-ever의 의미 분석: 허가 조건과 코퍼스 분석 (K)	최예슬 (대덕고), 민수정 (공주대) 우리나라 고등학교 실용영어 교과서 비판적 담화분석 (K)		
토	론자	송상헌 (고려대)	김종복 (경희대)	박혜림 (부경대)		
11:30~12:00	발표	Dong Yu and Yaochen Deng (Dalian University of Foreign Languages, China) Discourse Analysis of the Olympic Reports on U.S. Mainstream Media in Light of Digital Humanities (E)	이혜진 (원광대), 이제영 (전주대) 중등교사 영어 임용시험 수험자를 위한 영 역별 전문어휘 목록 선정 및 지도방안 탐색 (K)	김정렬 (한국교원대) 북한 영어교과서에 나타난 사회주의 교육 테제 (K)		
토	론자	안병길 (국립경상대)	김정렬 (한국교원대)	이제영 (전주대)		
12:00~13:00			점심 (Lunch)			
사:	회자	이선혜 (부경대)	김수태 (신라대)	강형진 (부경대)		
13:00~13:30		전주현 (충남대) 빅데이터시대의 MZ세대를 대상으로 하는 영작문 교육의 딜레마 (K)	김창현 (한국전자통신연구원) 자동요약 학습코퍼스 구축을 위한 가이드라인 (K)	Kevin Parent and Quynh Thuy Bui (Korea Maritime & Ocean Univ.) Thematic Choice in Argumentative Essays by Vietnamese EFL Learners (E)		
토	론자	최재호 (상명대)	이용훈 (충남대)	Guandong Zhang (Korea Maritime & Ocean Univ.)		
13:30~14:00	발표	노강산 (고려대) Would You Like to Exercise or Work Out?: Taking a Collostructional Approach to Korean EFL Learners' Underuse of the Verb-Particle Constructions (K)	김소희(국립국어원) 4차 산업 혁명 대비 국어 빅데이터 사업 (K)	최아녕 (연세대) A Critical Discourse Analysis on the Presidential Discourse: Focusing on Volodymyr Zelenskyy's Linguistic and Discoursal Strategies for Persuasion (K)		
토	론자	황규홍 (동아대)		민수정 (공주대)		
14:00~14:30	46	박슬기, 김희연, 김종복 (경희대) English although-stripping constructions: A corpus-based investigation (E)		안유영 (전남대) Metaphorical Perceptions of Chinese Engineering Students Towards English Learning in EM Context (K)		
	론자	홍수민 (서울시립대)	김한샘 (연세대)	최영주 (조선대)		

14:30~14:40	Coffee Break				
14:40~15:10	사회자	이도경 (인제대)	강아름 (충남대)	정희란 (부경대)	
	발표		이규민, 이정현, 장서현, 송상헌 (고려대) 코퍼스와 기계학습을 통한 사회 현상의 범 학문적 분석: 코로나19 사례를 중심으로 (K)	손지연 (연세대) The Development of Generative Grammar -Focusing on Minimization (E)	
	토론자	장세은 (한국해양대)	김창현 (한국전자통신연구원)	이지연 (창원대)	
15:10~15:40	발표		황동진 (고려대) 한국어 언어 모델의 정보 구조 표상 능력: '은/는'의 관계절 내 형성 제약을 중심으로 (K)	이주원 (전주대) A New Type of Resultative Construction in English (K)	
	토론자	최정인 (부산대)	강남길 (극동대)	장경철 (부산대)	
15:40~16:10	발표	한국영어학회 『영어학』 최우수논문 발표 임수연(한양대)·안현기(서울대) Effects of Parts of Speech on Implicit Prosodic Prominence by Native English Speakers and Korean Learners of English (K)	허진영 (한국교원대) AI 팽톡의 초등영어교육 적용방안 (K)	Suyoung Bae and Myung-Kwan Park (Dongguk University) Resolving some puzzles in Locative Inversions of English (K)	
	토론자	김수정 (창원대)	우길주 (부산교대)	신흥택 (독립연구자)	
16:10~16:20			Coffee Break		
	발표	Invited Talk 2 (류병래, 충남대) Internally-Headed Relative Clauses: Issues, Trends, and Challenges (K)			
	사회자	이용훈 (충남대, 한국코퍼스언어학회 총무)			
	토론자	송상헌 (고려대)			
17:10~17:20			Closing Remarks		
17:20~17:40		한국영어학회 연구윤리 워크숍	한국코퍼스언어학회 연구윤리 워크숍		
17:40~18:00		한국영어학회 총회 (General Meeting of KASELL)	한국코퍼스언어학회 총회 (General Meeting of KACL)		
18:00~20:00		저녁 만찬 (Dinner Reception)			

^{*} 녹색부분은 공동학술대회 주제 관련 발표이고 그 외는 자유 주제 발표임

^{*} 등록비: 학생10,000원, 교수/강사20,000원(프로시딩, 점심, 할인주차권 제공)

목 차

초청강연1. 2 Invited Talk 1 (김종복, 경희대) ------3 Pseudogapping in English: A Usage-based Construction Grammar Perspective Internally-Headed Relative Clauses: Issues, Trends, and Challenges 공동학술대회 주제 관련 Application of Artificial Intelligence to Semantic Prosody Research Discourse Analysis of the Olympic Reports on U.S. Mainstream Media in Light of Digital Humanities 빅데이터시대의 MZ세대를 대상으로 하는 영작문 교육의 딜레마 Would You Like to Exercise or Work Out?: Taking a Collostructional Approach to Korean EFL Learners' Underuse of the Verb-Particle Constructions 자동요약 학습코퍼스 구축을 위한 가이드라인 4차 산업 혁명 대비 국어 빅데이터 사업 코퍼스와 기계학습을 통한 사회 현상의 범학문적 분석: 코로나19 사례를 중심으로 한국어 언어 모델의 정보 구조 표상 능력: '은/는'의 관계절 내 형성 제약을 중심으로 AI 팽톡의 초등영어교육 적용방안

일반 자유주제

1. Aphiwit Liang-Itsara (Mahidol University, Thailand)
The Comparison of Motivating Factors for Taking English and Other Foreign Languages among Thai College Students
2. 박슬기, 김희연, 김종복 (경희대) ····································
English although-stripping constructions: A corpus-based investigation
3. Hyun Jong Hahm (Univ. of Guam, USA)169
Guamanian English and Englishes within
4. 윤인경, 유석훈 (고려대)
A Study on the Readability of American Textbooks and Korean English-Language Textbooks Comparing Conjunction Frequency in Sentence Structures
5. 한국영어학회 『영어학』 최우수논문 발표: 임수연(한양대)·안현기(서울대) ······· 195
Effects of Parts of Speech on Implicit Prosodic Prominence by Native English Speakers and Korean Learners of English
6. 강남길 (극동대)209
On Consist of and Made up of: A Corpora-based Analysis
7. 강아름 (충남대)
8. 이혜진 (원광대), 이제영 (전주대)219
중등교사 영어 임용시험 수험자를 위한 영역별 전문어휘 목록 선정 및 지도방안 탐색
9. Cuilin Liu and Se-Eun Jhang (Korea Maritime & Ocean University)225
A corpus-based study of gender-inclusive words in the maritime conventions of "seamen" or "seafarers"
10 . 최예슬 (대덕고), 민수정 (공주대) ····································
우리나라 고등학교 실용영어 교과서 비판적 담화분석
11. 김정렬 (한국교원대)251
북한 영어교과서에 나타난 사회주의 교육 테제
12. Kevin Parent and Quynh Thuy Bui (Korea Maritime & Ocean Univ.)255
Thematic Choice in Argumentative Essays by Vietnamese EFL Learners
13. 최아녕 (연세대)271
A Critical Discourse Analysis on the Presidential Discourse: Focusing on Volodymyr Zelenskyy's Linguistic and Discoursal Strategies for Persuasion
14. 안유영 (전남대)281
Metaphorical Perceptions of Chinese Engineering Students Towards English Learning in EMI Context
15. 손지연 (연세대)
The Development of Generative Grammar -Focusing on Minimization
16. 이주원 (전주대) ····································
A New Type of Resultative Construction in English
17. Suyoung Bae and Myung-Kwan Park (Dongguk University)
Resolving some puzzles in Locative Inversions of English

한국영어학회(KASELL)-한국코퍼스언어학회(KACL)

2022년 공동 가을 학술대회

초청강연1, 2

Invited Talk 1 (김종복, 경희대)

Pseudogapping in English:

A Usage-based Construction Grammar Perspective

Invited Talk 2 (류병래, 충남대)

Internally-Headed Relative Clauses: Issues, Trends, and Challenges

Pseudogapping in English: A Usage-based Construction Grammar Perspective a

Jong-Bok Kim
Dept. of English Linguistics and Literature
Kyung Hee Univ., Seoul
jongbok@khu.ac.kr

2022 Fall conference of English Linguistics Cohosted by Korean Association for the Study of English Language and Linguistics (KASELL) and Korean Association for Corpus Linguistics (KACL).

Pusan National University Oct 22, 2022

The talk is based on Kim & Runner (2022)

1

Introduction

Basic properties

Pseudogapping and its kin: gapping and VPE

Pseudogapping vs. gapping

Pseudogapping vs. VPE

Previous analyses

Move-and-delete analyses

Nonderivational, base-generation analyses

A Construction-based direct interpretation approach

Fundamentals, inheritance network, and structured discourse

VPE as a micro-construction

Pseudogapping as a micro-construction

Welcome consequences

Conclusion

Introduction	
intibuaction	
Corpus-based vs. corpus-driven approaches	
, , , , , , , , , , , , , , , , , , , ,	
	·

- Corpus-based studies typically use corpus data in order to explore a theory or hypothesis, aiming to validate it, refute it or refine it. Corpus-based research thus relies on established linguistic forms and theory to conduct investigations.
- Corpus-driven studies rely more on corpus data itself in delineating features and the scope of a linguistic investigation. The corpus itself should be the sole source of our hypotheses about language. It is thus claimed that the corpus itself embodies a theory of language (Tognini-Bonelli 2001, Biber 2009).

Usage-based theory

- Usage-based theory takes the view that the cognitive representation of language emerges through, and is shaped by, language use (Langacker 1987, 2000; Barlow & Kemmer, 2000; Bybee 2013; Bybee & Hopper, 2001; Bybee & McClelland, 2005).
- Usage-based approaches reject the strict separation of grammar (or competence)
 and usage (or performance) typical of generative approaches.
- In usage-based approaches, grammar is commonly seen as a vast inventory of symbolic conventions that are extracted from full-fledged utterances through a gradual process of schematization, retaining the syntactic and semantic commonalities across different usage events.

5

Basic properties

Some elliptical constructions

- Pseudogapping additionally differs from gapping and VPE, in that it involves ellipsis of part of a non-finite VP (see, among others, Ross 1967; Levin 1979; Hoeksema 2006; Gengel 2013; Miller 2014; Kubota & Levine 2017; Kim & Runner 2022):
 - (1) a. Kim reads magazines and Lee books. (Gapping)
 - b. Kim reads magazines, and Lee did too. (VPE)
 - c. Kim has read magazines and Lee has books. (Pseudogapping)

7

Two key research tasks

- how to capture its similarities to and differences from other relatives like gapping and VPE,
- how to license the construction

Previous analyses

- Two directions for the first task
 - one is to analyze it as a type of VPE (Jayaseelan 1990; Lasnik 1999; Kubota & Levine 2017)
 - the other is to take pseudogapping as a type of gapping (Agbayani & Zoerner 2004; Johnson 2009).
- two directions for the second task
 - Movement-and-deletion approaches propose that the remnant(s) is moved out of a VP constituent, which is then elided. These approaches vary in the kinds of movements proposed (A-movement vs. A'-movement) and in the direction of movement (leftward vs. rightward) (see Jayaseelan 1990; Lasnik 1999; Agbayani & Zoerner 2004; Thoms 2016).
 - base-generation or direct approaches do not rely on movement; its syntax is licensed by simple syntactic rules while its semantics is resolved by an anaphoric mechanism by inference (Miller 1990; 2014; Kubota & Levine 2017)

a

In this paper

- review key grammatical properties of pseudogapping that share with VPE and gapping as well as differentiate it from these two.
- discuss some possible issues that the previous movement-based as well as surface-oriented analyses wound encounter.
- sketch a construction-based analysis of VPE to lay down a foundation of our analysis for pseudogapping.
- offer a construction-based analysis of pseudogapping that captures its similarities to and differences from VPE.

Pseudogapping and its kin: gapping and VPE

Pseudogapping vs. gapping: similarities and differences

- pseudogapping is quite similar to gapping in that both involve a remnant phrase and a missing main verb. However, they differ in many respects (see, among others, Levin 1979; Miller 1990; Hoeksema 2006; Johnson 2009; Gengel 2013; Kubota & Levine 2017).
- First, gapping shows apparent verbal deletion without any remnant auxiliaries, while pseudogapping must have a finite auxiliary with a remnant, as in (3):
 - (2) a. Pat <u>loves</u> mysteries, and Terry [romances].
 - b. Robin ate beans, and Kim [rice].
 - (3) a. Kim drinks milk more often than he does [water].
 - b. Kim might read the short story, but he won't [the play].

Pseudogapping vs. gapping: similarities and differences

- Second, unlike gapping that in general occurs in the non-initial conjuncts of coordinate structures but not in subordinate structures, pseudogapping is more flexible in subordinate clauses:
 - (4) a. *He will consider your proposal, before he mine.
 - b. He will consider your proposal, before he does mine.
- Third, unlike gapping, pseudogapping is possible in embedded clauses (Levin 1979;
 Johnson 2009):
 - (5) a. *Some had eaten mussels and she clams that others [shrimp].
 - b. Some had eaten mussels and she claims that others had [shrimp].

13

Pseudogapping vs. gapping: similarities and differences

- Fourth, gapping constructions require non-coreferential subjects, but pseudogapping has no such a requirement (Levin 1979; Johnson 2009). Compare gapping in (6) and pseudogapping in (7):
 - (6) a. *Pat; [loves] mysteries, and Pat; romances.
 - b. Pat_i [loves] mysteries, and Lee_i romances.
 - (7) a. Kim; [drinks] milk more often than he; does [water].
 - b. I_i can [shoot] you before you_i can me. (COCA 1997 SPOK)

Pseudogapping vs. VPE: similarities and differences

- The key similarities and differences of the two constructions noted in the previous literature (see, among others, Levin 1979; Miller 1990; Hoeksema 2006; Gengel 2013; Miller 2014; Kubota & Levine 2017).
- First, the literature note that VPE can apply in infinitival clauses, while pseudogapping cannot (Levin 1979; Hoeksema 2006):
 - (8) a. I don't play chess as often as I would like to.
 - b. Kim might not have read the short story, but Lee might have.
 - (9) a. *I don't play chess as often as I would like to [checkers].
 - b. *Kim drank milk more often than he wanted to [water].

15

Pseudogapping vs. VPE: similarities and differences

- Second, it is also noted that VPE readily allows more than one supporting auxiliary, while pseudogapping does not (Levin 1979; Hoeksema 2006):
 - (10) a. Robin has been playing the oboe, and Kim has been too.
 - b. *Robin hasn't been playing the oboe as much as she has been [the bassoon].
- Third, the literature have often noted that VPE can apply backward, but pseudogapping is possible only forward (Levin 1979; Hoeksema 2006):
 - (11) a. Even if Kim could, she wouldn't speak French.
 - b. *Even if Kim could [every Romance language], she wouldn't speak French.

Pseudogapping vs. VPE: similarities and differences

- Fourth, most of the previous literature noted that pseudogapping, unlike VPE, is sensitivity to island constraints (Lasnik 1999; Gengel 2013):
 - (12) a. Robin will fascinate the children, and I believe [the claim [that Kim will too]].
 - b. *Robin won't fascinate the children, but I believe [the claim [that she will the adults]].

17

Pseudogapping vs. VPE: similarities and differences

- However, as noted by Hoeksema (2006) and Miller (2014) pseudogapping allows island violations. Our corpus data also identify island insensitive examples:
 - (13) a. Last night someone put a bullet through his head, just like they did [poor Mr. Felton]. (Left Branching Island, Hoeksema 2006)
 - b. Why don't you put a bullet in his head like you did [that man out there]? (Left Branching Island, Hoeksema 2006)
 - (14) a. Critics say it's not unusual for the police to spend less time investigating the deaths of black citizens than they do [whites]. (CNPC, COCA 1993 SPOK)
 - b. Bring the same kind of carry-ons when traveling by train as you would [by air]. (Adjunct island, COCA 1998 MAG)

Pseudogapping vs. VPE: similarities and differences

- Fifth, as noted earlier, unlike in VP-ellipsis, the elided material in pseudogapping need not be a constituent, as attested by the following (see Miller 2014):
 - (15) a. I would buy one from him before I would [Obama]. (COCA 2012 WEB)
 - b. Those cops spent a whole lot more time looking at Aunt Farrah than they did [me]. (COCA 2008 FIC).
- The antecedent clause, as marked here, contains a putatively elided part that is not a constituent. The putatively elided part can even be discontinuous, as seen from the corpus data:
 - (16) a. The notion probably makes your skin crawl as much as it does mine. (COCA 1993 FIC)
 - b. I just figured the money would do them a lot more good than it would [me]. (COCA 1995 NEWS).

19

Summarized comparison

	Pseudogapping	VPE	Gapping
requires an auxiliary	yes	yes	no
occurs in subordinate clauses	yes	yes	no
occurs in embedded clauses	yes	yes	no
requires a non-coreferential subject	no	no	yes
has a remnant	yes	no	yes
elides a constituent	no	yes	no(?)
occurs only in finite clauses	yes	no	no
may occur in infinitival clauses	no	yes	no(?)
may have more than one auxiliary	no	yes	no
can apply backward	no	yes	no
island sensitivity	yes(?)	no	no(?)

Table 1: Similarities and differences among the three constructions

Previous analyses

Move-and-VP deletion: leftward/rightward, A/A' movement

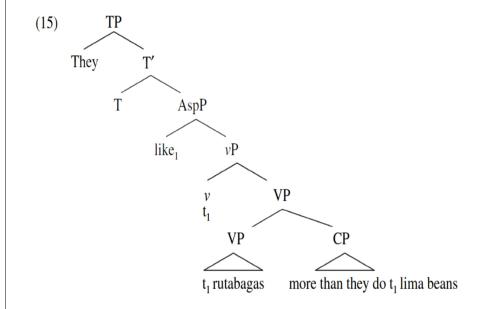
- For instance, Jayaseelan (1990) takes the remnant movement as a heavy NP shift, as shown in the derivation of (17a) in (17b):
 - (17) a. Kim might read the short story, but she won't the play.
 - b. ... she won't $[v_P | \frac{1}{v_P} | \frac{1}{v_P} | \frac{1}{v_P} |$ the play, i.e.
- Lasnik (1999) suggests that pseudogapping involves an overt leftward A-movement with application of a lower VP ellipsis:
 - (18) $\left[_{TP} \text{ she won't } \left[_{AgrP} \text{ the play}_{i} \left[\frac{1}{VP} \text{ read } t_{i} \right] \right] \right]$
- Gengel (2013) and Thoms (2016) suggest that pseudogapping is a rightward A'-focus movement, whose structure is similar to stripping:
 - (19) [Rab has [wine [been drinking]]] and [Tam has [Focp beer [been [drinking]]]].

Move-and-delete: gapping analysis

- takes the movement operation involved in pseudogapping to be a type of gapping (Agbayani & Zoerner 2004, Johnson 2009).
- This direction, positing coordination-like structures for gapping and pseudogapping, suggests that pseudogapping involves ATB (across the board) verb movement.
 - (20) They like _i [_{VP} t_i rutabagas] [_{CP} more than they do t_i lima beans]] Agbayani & Zoerner (2004: (15))

23

A structure from Agbayani & Zoerner (2004)



Interpretive anlaysis

- Non-derivational analyses introduce no covert expressions in the putative clausal source for pseudogapping.
- Miller (1990) introduces ID (immediate dominance) rules like (21) that allow an auxiliary verb to combine with a remnant or remnants.
 - (21) a. $VP \rightarrow H[2]$, NP b. $VP \rightarrow H[7]$, PP
- In addition to these PS-style rules, Miller's analysis introduces an anaphoric interpretation process that refers to a contextual variable corresponding to the elided material.

25

Categorial grammar approach

- Kubota & Levine (2017) provide a non-derivational analysis of pseudogapping within the framework of Hybrid Type-Logical Categorical Grammar.
- The key idea of this Categorial Grammar approach is to assume that pseudogapping involves the ellipsis of a transitive verb and introduces a VP/Pseudogapping operator that behaves like a (di)transitive verb:
 - (22) VP-ellipsis/Pseudogapping operator (Kubota & Levine 2017: (100)): $\lambda \varphi. \varphi; \lambda \mathcal{F}. \mathcal{F}(P); \text{(VP/$)} \uparrow \text{((VP/$)/(VP/$))}$ (where VP/\$ is a metavariable notation for a set of categories (e.g., VP, VP/NP, VP/NP/PP, etc) where any number of arguments and P is a free variable whose value is resolved anaphorically).
- The operator is an empty operator that functions as a (di)transitive verb. It first combines with an auxiliary as its argument and saturates its TV status (since it is a TV/TV)

An example from Kubota & Levine (2017)

(80) Robin has spoken about the war, and Leslie has of similar events. (81)spoken; **(speak-about, speak-of)**; $\frac{\text{VP/PP}_{about} \land \text{VP/PP}_{of}}{\text{spoken; } \mathbf{speak-about; } \text{VP/PP}_{about}} \land^{\mathsf{E}_{l}} \quad \text{about } \circ \text{ the } \circ \text{ war;}}{\mathbf{w}; \text{PP}_{about}}$ robin; $\lambda P.P$; spoken \circ about \circ the \circ war; speak-about(w); VP /E has \circ spoken \circ about \circ the \circ war; **speak-about**(**w**); VP \setminus E NP robin \circ has \circ spoken \circ about \circ the \circ war; speak-about(w)(r); S λφ.φ; has; $\lambda \mathcal{F}.\mathcal{F}(\text{speak-of});$ $\lambda f \lambda x \lambda y. f(x)(y);$ of ○ similar ○ leslie; $(VP/PP_{of}) | ((VP/PP_{of})/(VP/PP_{of})) | (VP/PP_{of})/(VP/PP_{of}) | E$ events; has; $\lambda x \lambda y.speak-of(x)(y)$; VP/PP_{of} l; has \circ of \circ similar \circ events; $\lambda y.speak-of(s-ev)(y); VP$ NP

leslie \circ has \circ of \circ similar \circ events; **speak-of**(**s-ev**)(**l**); S

27

A Construction-based direct interpretation approach

Construction grammar: main features

- The main features of CxG can be summarized as follows (see, among others, Goldberg 2005; 2013, Sag 2012, Kim & Michaelis 2020)
 - All levels of description (including morpheme, word, phrase, and clause) are understood to involve pairings of form with semantic or discourse functions.
 - Constructions vary in size and complexity, and form and function are specified if not readily transparent.
 - Language-specific generalizations across constructions are captured via inheritance networks, reflecting commonalities or differences among constructions.

29

Construction and definition

- Within this view, 'constructions' are thus taken to be the basic units of language and central to all linguistic descriptions and theories of languages.
- A construction consists of a form and a meaning or a function connected with that form
 - (23) Definition of grammatical 'constructions' (Goldberg 2005: 5):

 Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.

Inheritance network

- The constructions identified in a language are related to each other through inheritance hierarchies in which sub-constructions can inherit constructional properties from their super-constructions (see Goldberg 2005, Ginzburg & Sag 2000, Sag 2012, Traugott & Trousdale 2013, Kim & Davies 2020).
 - (24) Constructional schemas: an inheritance hierarchy (Traugott 2007, Traugott & Trousdale 2013)
 - a. Macro-constructions: highly abstract, schematic constructions
 - b. Meso-constructions: a network of related construction types which are still fairly abstract with similar semantics and/or syntax
 - c. Micro-constructions: individual construction types
 - d. Constructs: Instances of micro-constructions, realizations of actual use

31

Elliptical constructions again

- all the elliptical constructions can be defined as the macro-construction, the Ellipsis Construction (*elliptical-cxt*), which licenses ellipsis of one or more words.
 - (25) a. Kim likes wine more than she does beer. (Pseudogapping)
 - b. Kim likes wine, and Lee does, too. (VPE)
 - (26) a. Kim likes wine, and Lee beer. (Gapping)
 - b. Kim likes wine, but not Lee. (Stripping)
- Ellipsis constructions can be classified into at least two meso-constructions with respect to the requirement of an auxiliary verb licensing.

Gapping and stripping: a different mesoconstruction

- Unlike these two auxiliary sensitive constructions, gapping and stripping both typically occur in conjunctions or disjunctions, as in (26) (Johnson 2009, Kubota & Levine 2017).
 - (27) a. I will contact John, if you will. (VPE)
 - b. I'll contact John if you will Mary. (Pseudogapping)
 - (28) a. *I'll contact John if you Mary. (Gapping)
 - b. *I'll contact John if you. (Stripping)

33

Assumed network of constructions

• The four are all macro ellipsis constructions, but they can be classified into different meso-constructions while each functions as its own micro-construction,

elliptical-cxt

aux-ellipsis-cxt nonaux-ellipsis-cxt

pg-cxt vpe-cxt gapping-cxt stripping-cxt

Ellipsis and constructiona constructions

- Following Kehler (2000) and Hardt & Romero (2004), we also assume that the parallelism condition for ellipsis is a condition in discourse structure.
 - (30) Elliptical Construction:

elliptical-cxt
$$\Rightarrow$$
 $\begin{bmatrix} \text{SEM E} \\ \text{FOC nelist} \\ \text{CNXT parallel-rel}(A, E) \end{bmatrix}$

reflects the observed generalization that ellipsis clause (E) has at least one FOC expression, and its meaning E is in a parallel-relation with its antecedent A.

35

DGB (discourse game board)

- One key component of the discourse structure is a Dialogue Game Board (DGB) where the contextual parameters are anchored and where there is a record of who said what to whom, and what/who they were referring to (see Ginzburg 2012).
- As the dialogue progresses, the value of QUD is constantly being updated and the relevant context offers the basis for the interpretation of fragments.

DGB (discourse game board): an illustration

- uttering a question like What do they want? will activate the following information:
 - (31) $\left[\begin{array}{c} \operatorname{FORM} \left\langle \operatorname{What} \operatorname{do} \operatorname{they} \operatorname{want} \operatorname{now?} \right\rangle \\ \operatorname{SYN} \operatorname{S} \\ \operatorname{SEM} \lambda_{x} \left[\operatorname{want} (i,x) \right] \end{array} \right]$
- DGB, as updating the contextual information, also evokes two relevant attributes MAX-QUD (maximal-question-under-discussion) and FEC (focus establishing constituent)

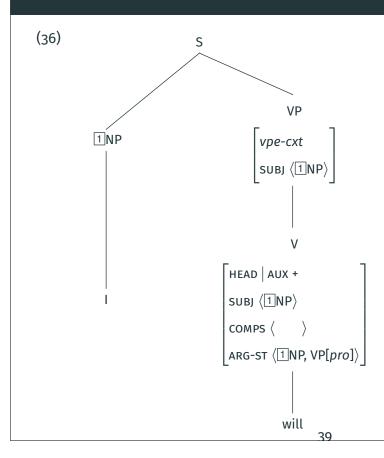
(32)
$$\left[\begin{array}{c} \text{MAX-QUD } \lambda_{X} \Big[want(i, x) \Big] \\ \text{FEC} \left\{ \begin{bmatrix} \text{SYN} \mid \text{CAT NP} \\ \text{SEM } x \end{bmatrix} \right\} \end{array} \right]$$

37

Anaphorical pro

- it has been widely noted that VPE behaves quite like a pronominal expression (Lobeck 1995; López 2000; Kim 2006; Aelbrecht & Harwood 2015; Ginzburg & Miller 2018). For instance, VPE can cross utterance boundaries ((33)), is not sensitive to island constraints ((34)), and is subject to the Backwards Anaphora Constraint ((35)):
 - (33) A: Kim won't leave Seoul soon.
 - B: I don't think Lee will _ either.
 - (34) a. Kim didn't hit a home run, but I know a woman who did . (CNPC)
 - b. That Kim won the batting crown is not surprising, but that Peter didn't know she did is indeed surprising. (SSC)
 - (35) a. *Kim didn't [e] but Lee ate meat.
 - b. Because Kim didn't [e], Lee ate meat.

A licensed structure



Constructional constraints

- The construction, whose immediate superconstruction is *aux-ellipsis-cxt* (indicated by the up-arrow ↑), by inheritance requires its head to be an auxiliary (*aux-vpe-wd*) whose VP argument is unrealized.
 - (37) VPE Construction († aux-ellipsis-cxt):

$$VP\begin{bmatrix} vpe-cxt \\ FOC\ nelist \end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix} aux-vpe-wd \end{bmatrix}$$

Pseudogapping: basic constraints

- As we have seen, pseudogapping shares many of the properties of VPE but behaves more restrictively. The key properties of pseudogapping are given in the following:
 - It requires a finite auxiliary.
 - The remnant functions as a contrastive focus.
- As noted earlier, these two constraints account for the following contrast:
 - (38) a. I played chess as often as I did [checkers].
 - b. *I don't play chess as often as I would like to [checkers].
 - (39) a. I rolled up a newspaper, and Lynn did [a magazine].
 - b. *Kim ate fish, and Lee did [fish] too.

41

Pseudogapping construction

- These constraints are not from general principles but learned as constructional constraints as defined in the following:
 - (40) Pseudogapping Construction (\uparrow aux-ellipsis-cxt, first approximation): $VP \begin{bmatrix} pg\text{-}cxt \\ FOC & \boxed{1} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} aux\text{-}vpe\text{-}wd \\ vFORM & fin \end{bmatrix}, \text{1RP}$
- This simple specification indicates that a pseudogapping VP has a finite auxiliary verb as its head with a focused RP (remnant phrase). Since the finite auxiliary is a type of *v-vpe-wd*, its second argument is not realized at syntax but functions as a pro.

Other constraints

- The remnant and the correlate both need to be in a transitivity situation.
- The remnant is a contrastive focus with respect to the correlate having the same syntactic category.
- The antecedent clause and the pseudogapping clause refer to the same situations with the different referents of the remnant and the correlate.

43

Anaphoric properties

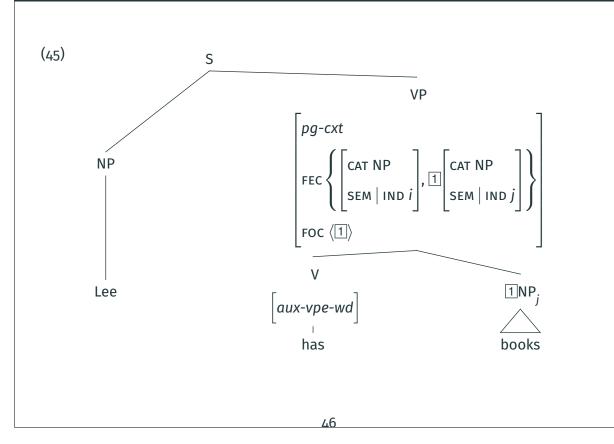
- These constraints are functional constraints on the construction, tied to the
 discourse structure. As noted earlier, VPE and Pseudogapping can take place
 across a discourse, allowing cataphoric reference, but this is not possible in
 gapping (Kehler 2000; Miller 2014):
 - (41) A: Bill will make a statement blasting the media.
 - B: Hilary will, too. (VPE)
 - (42) A: Bill will make a statement blasting the media.
 - B: Hilary will the newspaper reporters. (Pseudogapping)
 - (43) A: Bill will make a statement blasting the media.
 - B: *Hilary the newspaper reporters. (Gapping)

Revised constructional constraints

- To license the anaphoric property of Pseudogapping, we refer to the DGB, as we did for the analysis of VPE, but add additional constructional constraints on pseudogapping, as in (44):
- The members of the FEC can also can refer to syntactic information (Ginzburg & Sag 2000). That is, we assume that the antecedent A of pseudogapping can refer to the syntactic, categorial information (CAT) as well.

45

An example: Kim has read magazines and Lee has books



Resolution process

- The focused remnant has the correlate *magazines* having the same categorial information. The antecedent VP and S would have the following resolution encoded in the DGB(Kim has read magazines and Lee has books.):
 - (46) a. $[antecedent-VP] = \lambda x[read(x, m)]$
 - b. $[antecedent-S] = [read(k, \mathbf{m})]$
- Since the head auxiliary verb *can*, as a VPE introducing one, licenses a *pro* VP as its second argument, it looks for its VP antecedent.
 - (47) a. $[PG-VP] = \lambda x[read(x, b)]$
 - b. [PG-S] = [read(I,**b**)]

7.7

PP remnant with category matching

- Example:
 - (48) a. John speaks to Mary more civilly than he does [Anne].
 - b. You can't count on a stranger, but you can [on a friend]. (Kubota & Levine 2017: (5))
- Resolution
 - (49) a. $[antecedent-VP] = \lambda x[count.on(x, s)]$
 - b. $[antecedent-S] = \neg[count.on(y, s)]$
- DGB information evoked here:

(50)
$$\left[DGB \left[FEC \left\{ \left[CAT \ 2PP[on] \right], \left[CAT \ 2PP[on] \right] \right\} \right] \right]$$

Sprouting with a covert correlate

- The discourse-based analysis implies that the antecedent correlate being in a contrast relation with the RP need not be an overt but can be evoked from the context.
 - (51)We also have to turn our attention as forcefully as we can [to the economy and creating jobs]. (COCA 2012 SPOK)
- The correlate is implicitly provided by the argument structure of the predicate turn (indefinitie/definite null instantiation of an argument, Ruppenhofer & Michaelis 2014):

(52)
$$\begin{bmatrix} \mathsf{FORM} \ \langle \mathsf{turn} \rangle \\ \mathsf{ARG-ST} \left\langle \mathsf{NP}_i, \mathsf{NP}_j, \left(\mathsf{PP} \begin{bmatrix} \mathsf{dn}i \\ \mathsf{PFORM} \ \mathsf{to} \\ \mathsf{IND} \ k \end{bmatrix} \right) \right\rangle \\ \mathsf{CONT} \ \mathsf{turn}(i,j,k) \\ \end{bmatrix} \qquad \mathsf{b.} \left[\mathsf{FEC} \left\{ \begin{bmatrix} \mathsf{SYN} \ | \ \mathsf{CAT} \ | \ \mathsf{PP} \begin{bmatrix} \mathsf{dn}i \\ \mathsf{PFORM} \ \mathsf{to} \end{bmatrix} \right] \right\} \right]$$

b.
$$\left[\text{FEC} \left\{ \begin{bmatrix} \text{SYN} \mid \text{CAT} \mid & \text{PP} \\ \text{PFORM} & \text{to} \end{bmatrix} \right\} \right]$$

Deeply embedded cases

- The present analysis places no restriction on the type of the elided parts. The elided parts need not be a constituent and can be even discontinuous:
 - The notion probably makes your skin crawl as much as it does mine. (53) a.
 - b. Maybe it's easier to get the devil to listen than it is [God].
- Resolution process
 - [antecedent-S] = [make(n, ys, crawl)](54) a.
 - $[PG-S] = [make(n, \mathbf{ms}, crawl)]$

island insensitivity

- Since the present analysis introduces no movement processes, it also allows island insensitivity, as we have noted earlier. Consider one example again:
 - (55) Critics say it's not unusual for the police to spend less time investigating the deaths of black citizens than they do [whites]. (CNPC)
- The remnant NP whites in (55) is within the definite NP in the antecedent clause. The present analysis places no strong conditions on the elided structures: it just refers to the two individuals participating in a contrasting relation.

51

Argument structure mismatch

- mismatch in the argument structure of the antecedent clause and that of the pseudogapping clause:
 - (56) I ask every New Yorker, when you see a police officer today, please offer them condolences as you would [to someone who has lost a family member].
- The antecedent clause here subcategorizes for a ditransitive structure (NP-NP) while the pseudogapping clause for a dative pattern (NP-PP).
 - (57) a. [antecedent-S] = [offer(i,t,c)]
 - b. [PG-S] = [offer(i,s,c)]
- One implication we can have here is that once the syntactic valence frame of the verb *offer* is activated, its semantic argument structure is also evoked. For instance, the verb will activate the argument structure of 'offer(agent, recipient, theme)', which can be syntactically realized either as NP-NP or NP-PP sequences.

Preposition mismatch

- This direction also allows us to account for the preposition mismatch noted by Miller (2014).
 - (58) a. *Kim spoke to Lee more often than he did for Ann.
 - b. Robin has spoken about the war, and Leslie has of similar events. (Kubota & Levine 2017: 19)
- The example in (58a) is also not licensed: since the category of the RP and that of the overt or covert correlate need to be identical. However, examples like (58b) are licensed due to the fact that the verbs *spoke* here share the semantic argument structure
 - (59) a. [antecedent-S] = [speak(k,ce)]
 - b. [PG-S] = [speak(k,s)]
- The second argument here both functions as a goal argument. However, in (58a), the second argument in the antecedent clause is a goal while the one in the pseudogapping is a beneficiary.

53

Voice mismatch

- The analysis, not relying on syntactic identity with the antecedent clause, can also explain voice mismatches, which we noted earlier:
 - (60) a. A whole poached wild striped bass should be taken to the table as you would [a Thanksgiving turkey]...
 - b. These savory waffles are ideal for brunch, served with a salad as you would [a quiche].
- Resolutions:
 - (61) a. [antecedent-VP] = [take.to(x,b,tb)]
 - b. [PG-S] = [take.to(y,tk,tb)]

54

Conclusion	

Conclusion

- Just like other elliptical constructions such as VPE, gapping, and stripping, pseudogapping displays an elliptical construction involving missing expressions with the unusual form-function incongruence.
- The discourse and construction-based analysis, which capitalizes on the inheritance network hierarchy to capture broad similarities and unique differences among these constructions, allows us to account for a much wider range of pseudogapping phenomena.

References

- Aelbrecht, Lobke & William Harwood. 2015. To be or not to be elided: VP ellipsis revisited. *Lingua* 153. 66–97.
- Agbayani, Brian & Ed Zoerner. 2004. Gapping, pseudogappingand sideward movement. Studia Linguistica 58(3). 185–211.
- Gengel, Kristen. 2013. Pseudogapping and ellipsis. Oxford: Oxford University Press.
- Ginzburg, Jonathan. 2012. The interactive stance: Meaning for conversation. Oxford: Oxford University Press.
- Ginzburg, Jonathan & Philip Miller. 2018. Ellipsis in Head-Driven Phrase Structure

 Grammar. In Jeroen van Creanenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*, 75–121. Oxford: Oxford University Press.
- Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations: the form, meaning and use of English interrogatives*. (CSLI Lecture Notes 123). Stanford, CA: CSLI Publications.
- Goldberg, Adele E. 2005. *Constructions at work: constructionist approaches in context.*Oxford: Oxford University Press.
- Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 15–31. Oxford: Oxford University Press.

- Hardt, Daniel & Maribel Romero. 2004. Ellipsis and the structure of discourse. *Journal of Semantics* 21(4). 375–414.
- Hoeksema, Jack. 2006. Pseudogapping: its syntactic analysis and cumulative effects on its acceptability. Research on Language and Computation 4. 335–352.
- Jayaseelan, K. A. 1990. Incomplete VP deletion and gapping. *Linguistic Analysis* 20. 64–81.
- Johnson, Kyle. 2009. Gapping is not (vp)-ellipsis. Linguistic Inquiry 40. 289–328.
- Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23(6). 533–575.
- Kim, Jong-Bok. 2006. Similarities and differences between English VP ellipsis and VP Fronting: An HPSG analysis. *Studies in Generative Grammar* 13(3). 429–460.
- Kim, Jong-Bok & Mark Davies. 2020. English what with absolute constructions: a construction grammar perspective. English Language and Linguistics 24(4). 637–666.
- Kim, Jong-Bok & Laura A. Michaelis. 2020. Syntactic constructions in English. Cambridge: Cambridge University Press.
- Kim, Jong-Bok & Jeffrey T. Runner. 2022. Pseudogapping in English: a direct interpretation approach. *Linguistic Review* 39(3). 457–494.

59

- Kubota, Yusuke & Robert Levine. 2017. Pseudogapping as pseudo-VP ellipsis. *Linguistic Inquiry* 48(2). 213–257.
- Lasnik, Howard. 1999. Pseudogapping puzzles. In ShalomLappin & Elabbas Benmamoun (eds.), Fragments: studies in ellipsis and gapping, 141–174. Oxford University Press.
- Levin, Nancy S. 1979. Main verb ellipsis in spoken English. Ohio State University. (Doctoral dissertation).
- Lobeck, Anne. 1995. Ellipsis: Functional heads, licensing, and identification. Oxford: Oxford University Press.
- López, Luis. 2000. Ellipsis and discourse-linking. *Lingua* 110(3). 183–213.
- Miller, Philip. 1990. Pseudogapping and do so substitution. In Michael Ziolkowski,
 Manuela Noske & Karen Deaton (eds.), Papers from the 26th regional meeting
 ofthe Chicago Linguistic Society, 293–305. Chicago Linguistic Society.
- Miller, Philip. 2014. A corpus study of pseudogapping and its theoretical consequences. In Christopher Pinon (ed.), *Empirical issuesin syntax and semantics* 10, 73–90. Paris: CSSP.
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge, MA: Massachusetts Institute of Technology. (Doctoral dissertation).

- Ruppenhofer, Josef & Laura A. Michaelis. 2014. Frames and the interpretation of omitted arguments in English. In Katz B. Stacey & Lindsy L. Myers (eds.), *Perspectives on linguistic structure and context: Studies in honor of Knud Lambrecht*, 57–86. John Benjamins Publishing.
- 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. Stanford: CSLI Publications.
- Thoms, Gary. 2016. Pseudogapping, parallelism, and the scope of focus. *Syntax* 19(3). 286–307.
- Traugott, Elizabeth Closs. 2007. The concepts of constructional mismatch and type-shifting from the perspective of grammaticalization. *Cognitive linguistics* 18(4). 523–557.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford University Press.

61

-	34	-
---	----	---

한국영어학회(KASELL)·한국코퍼스언어학회(KACL) 가을 공동학술대회 2022년 10월 22일(토), 부산대학교

Internally Headed Relative Clauses in Korean: Issues, Trends, and Challenges

한국어 내핵관계절: 쟁점과 경향 그리고 도전

류병래 (충남대학교 언어학과) <u>http://ling.cnu.ac.kr</u>/ ryu@cnu.ac.kr

1 서론

어느 언어에나 명사 혹은 명사구를 수식하는 절이 존재한다. 한국어에서는 관형절이 여기에 해당한다. 최근 30년 한국어 명사(구) 수식절에 대한 연구는 전통문법의 관형절을 생성문법적 시각에서 관계절로 보는 견해가 주류를 이룬다. 일반적으로 관계절은 다음과 같은 두 가지를 정의적 특성으로 가지고 있다(Grosu (2000, p.85), de Vries (2002, p.14) 참고). 하나는 관계절은 종속절이라는 점이고,다른 하나는 '피봇'(pivot) 구성소를 축으로 주절과 관계절이 통사·의미적으로 연결된다는 점이다.1

관계절 경계를 기준으로 피봇 구성소가 외부에 존재하면 그 관계절을 외핵 관계절(externally headed relative clause, 이하 EHRC)이라고 하고, 그 구성소가 관계절 내부에 존재하면 내핵 관계절(internally headed relative clause, 이하 IHRC)이라 한다 (Keenan (2000,p.141-143), de Vries (2002)). 또한 관계절은 내부에 공백(gap)이 존재한다고 가정하는지 여부에 따라 공백관계절(gapped relative clause)과 무공백관계절(gapless relative clause)로 구분하기도 한다. 이런 이유로

¹ 피봇 구성소란 의미적으로 주절과 관계절에 의해 공유되는 구성소를 말하는데, 학교 문법에서는 이를 '선행사'(antecedent)라고 하기도 하고 학자에 따라서는 '핵어 명사'(head noun)라고 하기도 한다. 이 논문에서는 관계절의 정의적 속성에 수식 (modification)이나 선행사, 혹은 핵어 명사와 같은 개념을 의도적으로 사용하지 않는다. 그 이유는 피봇 구성소의 선형적 순서가 해당 언어의 어순에 따라 달라질 수도 있고 피봇명사구와 수식관계를 형성하지 않는 부가관계절(appositive relatives)도 존재할뿐더러 내핵관계절도 관계절의 일종으로 보는 견해가 이론언어학에서 일반적인 점을 감안하여 논의의 대상을 지나치게 제한하지 않는 것이 합리적이라 보기 때문이다.

외핵관계절은 정의적으로 공백관계절이고 내핵관계절은 무공백관계절이다.

다음 (1)에 제시된 문장은 전형적인 한국어 외핵관계절의 예이다. [RC . . .]로 표기된 관계절은 내부에 공백을 포함하고 있으며 피봇 구성소인 '도둑', '컴퓨터'를 수식하고 있는 구조를 가지고 있다.²

- (1) a. 경찰이 [RC 은행에서 나오는] 도둑을 체포하였다.
 - b. 철수가 [RC 고장난] 컴퓨터를 고쳤다.

한편 한국어에는 문법적 요소 '것'과 선행하는 절이 관형형 어미로 연결된 구문이 존재한다. 예문 (2)가 이 구문의 전형적인 예이다.

- (2) a. 경찰이 [RC 도둑이 은행에서 나오는] 것을 체포하였다.
 - b. 철수가 [RC 컴퓨터가 고장난] 것을 고쳤다.

예문 (2)를 주문장의 술어에 의해 기술된 사건의 대상이 각각 '도둑'과 '컴퓨터'인 것으로 이해하는 데에 모국어 화자들은 어려움을 가지고 있지 않다. 즉, (2)에서 주절 동사의 직접목적어가 마치 [RC . . .] 로 표기된 관계절의 내부에 존재하는 명사(구) 가운데 하나인 것으로 해석할 수 있다는 관찰이 가능하다. 심지어 (2)에 제시된 두 문장이 (1)에 제시된 두 문장과 각각 진리조건적 의미가동일하다고 주장하는 학자들도 있다(Kim, Sells & Yang (2006) 참고). 이 관찰을 기반으로 '것'과 관형형 어미로 연결된 선행절로 구성된 구문이 한국어의 내핵관계절이라는 주장이 1990년대 초반에 제기되었다(Jhang (1991; 1994)). 그 이후 30년 동안이 구문을 내핵관계절로 단정하고 다양한 시각에서 여러 문법이론 틀로 분석하려는 연구가 진행되어 왔으며 그동안 쌓인 문헌의 양도 상당하다(Jung (1995), Y.-B. Kim (1996; 2002),J.-B. Kim (1999), Chung & Kim (2003), M.-J. Kim (2004a; 2004b; 2010), Cha (2005),Kim, Sells & Yang (2006), J. Lee (2006), Song (2021), K.-S. Kim (2022) 등 참조).

외핵 관계절이 여러 언어에서 가장 전형적인 관계절의 유형이지만 내핵관계절도 몇몇 언어에 존재한다. 내핵관계절은 pro-탈락 언어(Culy (1991, p.199)), 관계절이 명사에 선행하는 SOV 어순의 언어(Keenan (1985)), 그리고 의문사가 문두로 이동하지 않아도 되는 언어(Watanabe (1992))에서 주로 발견되는 것으로 알려져 있다. 일본어에서도 (2)에 해당하는 문장이 존재하고 이 문장들을 내핵관

² 내핵관계절은 관계절의 한 유형이고 관계절은 절이기 때문에 경계를 구분할 수 있다. 논의의 편의를 위하여 관계절 경계의 시작과 끝을 명확하게 구분하여 [RC . . .]로 표기하기로 한다. 예문 (2)에 포함되어 있다고 주장하는 내핵관계절은 [RC . . .]로 구분된 종속절로 이해가 된다.

계절 구문으로 간주하고 분석하는 연구가 Kuroda(1976)를 필두로 활발하게 진행되어 왔다(전체적인 일본어 문헌에 대한 조망은 Grosu &Hoshi (2019)를 참고). 한국어와 일본어가 유형론적으로 관련이 깊다는 점을 감안하면, 한국어에도 내핵관계절이 있다는 주장은 그 자체로 이론적으로 흥미롭기도 하지만 크게 이상한 일은 아니다.

그러나 한국어의 내핵관계절에 대한 논의가 활발하게 진행되어 왔음에도 불구하고 아직 그 구성의 본질에 대한 의문에 답을 찾은 것보다는 아직 찾지 못한 것이 더 많은 것이 사실이다. 흥미롭게도 여러 관련 쟁점들 가운데에서 가장 기본적인 쟁점, 즉 (2)에 제시된 '것'과 선행하는 절이 관형형 어미로 연결된 구문을 내핵관계절로 보아야 하는지부터 쟁점이 되고 있다.

언어마다 다른 언어학적 구성을 가질 수 있다는 점을 감안하여 모든 언어의 내핵관계절이 동일한 구성을 보일 필요는 없다는 점을 인정하더라도 한국어의 내핵관계절이라고 주장하는 구문은 다른 언어의 내핵관계절이 보여주는 문법적 양상과 매우 상이하다.3 한국어의 해당 구문은 유형론적 유사성이 강조되는 일본어의 해당 구문과도 상이한 점이 많이 드러난다. 특히, 일본어는 한국어보다해당 구문의 수용도가 더 높아 일본어에서는 가능하지만 한국어에서는 가능하지 않은 문장이 부분적으로 존재한다(연재훈 (2012, p.420)).

(2)에 제시된 예문이 한국어 모국어 화자에게 수용적이라는 점과 그 예문의 구성처럼 '것'과 관형형 어미로 연결된 종속절을 하나로 묶어 한국어의 내핵관계절이라고 주장하는 것은 완전히 다른 문제이다. 그럼에도 불구하고 대부분의 기존 논의에서는 (2)가 수용적이고 (1)처럼 해석할 수 있으며, 피봇명사구에 해당하는 요소가 종속절 내부에 있는 예문을 관찰한 근거로 이 문장을 한국어의 내핵관계절이라고 단정하고 논의를 진행해 왔다.

또한, 아래에서 자세하게 논의하겠지만, 기존 문헌에서는 내핵관계절에 대한 정의나 유사 구문과의 구분법, 나아가서 내핵관계절의 기본 속성으로 제시한 관련 문법 현상에 대한 관찰적 타당성에 대한 성공적인 논의가 없었다. 나아가서 기술의 대상도 대부분 '것'과 관형형 어미로 연결된 종속절이 특정 유형 타동사(예를 들어, '잡다', '먹다', '고치다' 등 대상(theme) 의미역을 하위범주화 속성으로 갖는 전형적인 타동사)의 목적어 자리에 등장하는 경우로 한정해 왔음도 비

³ 조수근 (2014)에서는 한국어 내핵관계절로 알려진 구문에 대한 아동과 성인의 수용성 실험을 통해 한국어 모국어 화자들은 내핵관계절로 알려진 문장을 70% 정도로 제한적으로 수용적인 것으로 판단했다는 보고를 하고 있다. 이 실험 결과를 근거로 그는 한국어에 내핵관계절은 제한적으로 존재한다고 결론을 내리고 있다. 그러나 아래에서 자세하게 논의하듯 문장의 수용성은 한국어에서 내핵관계절의 존재 유무와는 동일시 할 수 없는 두 개의 서로 다른 문제이다.

판적으로 평가된다.

따라서 한국어의 내핵관계절 논의는 정의나 유사 구문과의 구분법, 나아가서 내핵관계절의 기본 속성으로 제시한 관련 문법 현상에 대한 관찰적 타당성에 대 한 근본적인 재검토가 필요하다고 보인다. 저자의 지식 범위 내에서는 (2)와 같 은 예문이 내핵관계절의 정의에 맞는지, 그리고 이 예문이 내핵관계절의 기본 속성으로 알려진 언어학적 현상과 부합하는지 심도있게 논의한 문헌이 존재하 지 않는다. 본 논문은 이 간극을 메우고자 하는 동기로 작성되었다.

본 논문에서 다루고자 하는 쟁점은 크게 두 가지이고 그 두 가지 쟁점에 대한 답을 찾는 것이 이 논문의 목적이다. 하나는 내핵관계절의 정의를 기존 문헌에서 찾아 내핵관계절의 정의적 속성에 한국어의 해당구문이 부합하는지 검토하는 것이다. 다른 하나는 내핵관계절의 기본 속성으로 알려진 무공백 관계절, 무제한의존성, 진리조건적 의미의 동일성, 비제한적 해석 혹은 최대성/전체성 효과를 내핵관계절로 간주된 구문이 과연 보여주는지 밝히는 것이다. 본 논문은 두 가지 쟁점에 모두 부정적인 결론에 도달했음을 내용으로 하는 보고이다.

(2)와 같은 예문에 등장하는 '것'과 관형형 어미로 연결된 선행절로 구성된 구문은 그 자체로는 '명사화된 문장'이 아니라는 점, 매우 제약적인 분포를 보인다는 점, 그리고 피봇명사구가 명시적으로 실현되지 않아도 된다는 점에 비추어한국어에서 내핵관계절로 알려진 구문은 기존 문헌에 제시된 정의에 따르면 내핵관계절이 아니라는 점이 논의를 통해 밝혀질 것이다. 달리 말하면, 한국어에서 대핵관계절로 알려진 구문은 일반적인 관계절과 대핵관계절의 정의에 따르면 대핵관계절로 볼 근거가 없다는 점이 드러날 것이다. 대핵관계절은 관계절이어야 하고, 관계절은 종속절이어야 한다. 그러나 '것'과 관형형 어미로 연결된 선행절로 구성된 구문은 그 자체가 종속절이 아니고 종속절이 아니기에 관계절이라 할 수 없고 관계절이 아니기에 대핵관계절이라고 할 수 없음이 분명해질 것이다.

또한 내핵관계절의 기본 속성으로 알려진 여러 현상 중에서 무공백 관계절, 무제한의존성, 진리조건적 의미의 동일성, 비제한적 해석 혹은 최대성/전체성 효 과에 대한 검증결과도 한국어에서 '것'과 관형형 어미로 연결된 선행절로 구성된 구문을 내핵관계절로 간주하는 견해에 문제가 있음을 보여주고 있다. 정의에 따라 내핵관계절은 무공백관계절이어야 하지만, 내핵관계절로 간주된 구문도 절 내부에 공백을 가정할 수 밖에 없는 문법적 현상을 허용한다는 점, 즉 공백관계 절일 수 있다는 사실을 새로운 자료를 통해 밝히고자 한다. 쟁점이 되고 있는 한국어 구문은 무제한의존성을 보여준다는 점, 내핵관계절로 알려진 구문과 해 당 외핵관계절이 진리조건적 의미의 동일성을 보여주지 않는다는 점, 내핵관계 절로 알려진 구문이 비제한적 해석 혹은 최대성/전체성 효과를 가지고 있다는 기본 속성에 대한 관찰도 경험적 증거로 지지를 받지 못한다는 점을 새로운 자료를 통해 밝히고자 한다. 본 논문은 다음과 같이 구성되어 있다. 제1절 서론에 이어 제2절에서는 내핵 관계절의 정의와 한국어 적용 상의 문제점을 살펴보기로 한다. 2.1 소절에서는 문제가 되고 있는절을 내핵관계절로 보는 학자들이 제시하고 있는 세 가지 정의를 소개한다. 2.2 소절에서는 내핵관계절의 정의적 속성을 명사화된 문장과 생산성에 초점을 맞추어 자세하게 논의하기로 한다. 2.3 소절에서는 Jhang (1994)의 정의에 등장하는 속성인 피봇명사구의 명시성에 대하여 살펴보기로 한다. 중간결론으로 2.4 소절에서 제2절 전체의 내용을 요약하기로 한다. 제3절에서는 문제가 되고 있는 절을 내핵관계절로 분석하는 연구들이 가장 기본적인 것으로 단정하고 있는 기본 속성에 대한 반론을 다룬다.

이 절에서는 크게 네 가지 기본 속성을 쟁점으로 다루고자 한다. 첫째, 문제가 되고 있는 절이 무공백 관계절이어야 한다는 쟁점; 둘째, 문제가 되고 있는 절이 무제한의존성을 보이지 않는다는 속성에 대한 쟁점; 셋째, 외핵관계절과 문제가 되고 있는 절이 진리조건적 의미의 동일성을 갖는지 여부; 그리고 마지막으로 넷째, 문제가 되고 있는 절이 비제한적 관계절이라는 주장과 최대성/전체성효과 를 가지고 있다는 주장의 타당성을 고찰하고자 한다. 마지막으로 제4절에서 이 논문의 전체 논의를 정리하고 남은 문제와 향후 논의의 진행 방향에 대한 제언 을 하고 결론을 맺기로 한다.

2 내핵관계절의 정의와 한국어 적용 상의 문제점

2.1 세 가지 정의

거의 모든 기존 문헌에서는 특별한 정의 없이 (2)와 같은 예문을 제시하고 그예문이 내핵관계절을 포함하고 있다고 가정하고 논의를 시작한다. 내핵관계절은 논란이 되고 있는 한국어에 유일한 관계절이 아니기에 다른 언어에서는 내핵관계절을 어떻게 정의하는지 알아보는 것도 내핵관계절이 무엇인지 이해하는 데에 도움이 될 것이다. 서론에서 언급한 바와 같이 내핵관계절은 정의에 따라 무공백관계절이고 피봇구성소가 그 절의 내부에 존재해야 한다. 우선 Culy (1991, p.199)의 내핵관계절에 대한 정의를 살펴보기로 한다.

(3) "A (restrictive) internally-headed relative clause is a nominalized sentence which modifies a nominal, overt or not, internal to the sentence." (=(13), Culy (1991,p.27))

그의 정의에 따르면 내핵관계절은 명시적으로 실현되어 있든 아니든 문장 내 부에 있는 명사를 수식하는 명사화된 문장(nominalized sentence)이다. 한국어의 내핵관계절의 정의를 시도한 문헌은 저자의 지식 범위 안에서는 Jhang (1994)과 M.-J. Kim (2004a)가 유일하다. Jhang (1994, p.18)은 위에 제시한 Culy (1991, p.27)의 정의와 기본적으로 동일하다. 다른 점이 있다면 한국어의 내핵관계절이 수식하는 핵어 명사가 반드시 명시적으로 실현될 것을 요구하는 제약을 추가한 것이다.

- (4) "Korean IHRCs can be defined as nominalized sentences that modify an overtnominal internal to the sentence." (Jhang (1994, p.18))
- M.-J. Kim (2004a, p.5)의 정의는 (5)에 제시한 바와 같다. 그의 추가 설명에 따르면 정의 (5)에서 내포절(embedded clause)은 공백이 없는 관형절을 의미하고 관계표지(relativeve marker)는 '는' 혹은 'ㄴ'을 지칭한다.
- (5) "The combination of the embedded clause and the relative marker will be referred to as an IHRC." (M.-J. Kim (2004a, p.5))
- M.-J. Kim (2004a, p.5)의 정의에는 명사화된 문장이라는 용어가 등장하지 않는다. 그의 정의에 따르면 내핵관계절은 결과적으로 일반적인 무공백관계절이다. 위의 두 정의와는 달리, 내핵관계절을 무공백관계절과 동일하게 보는 견해는지나치게 포괄적이어서 쟁점이 되고 있는 절과 다른 관계절을 구분하기에 부족하다. 예문 (2)나 (8)에서 확인할 수 있듯이, 한국어에서는 (내핵)관계절 자체가문법기능을 담당하지 못하고 추가로 '것'이라는 요소를 필수적으로 요구한다는점에 비추어 '것'의 필수성을 정의에서 언급하지 않은 점은 정의로서 미흡하다고 평가할 수 있다. 따라서 우리는 내핵관계절을 무공백관계절과 동일하게 보는 M.-J. Kim (2004a, p.5)의 정의를 한국어 내핵관계절의 정의로 간주할 수 없다고보고 더 이상 진지한 논의의 대상으로 삼지 않기로 한다. 대신, Jhang (1994, p.18)의 정의에 등장하는 조건 중 명사화된 문장과 피봇 명사구의 명시적 실현은독립적인 소절로 나누어더 자세하게 아래에서 논의하기로 한다.

2.2 내핵관계절의 정의적 속성

Culy (1991, p.199)에 따르면 내핵관계절은 일반적으로 다음과 같은 세 가지속성을 가지고 있다.

(6) a. 내핵관계절을 갖는 언어는 명사화 정도가 커야 하고, 특히 명사화된 문장이 다른 구문에도 등장해야 한다.

- b. 내핵관계절을 갖는 언어는 SOV가 기본 어순이며 명사구는 적어도 좌분지를 해야 한다.
 - c. 내핵관계절을 갖는 언어는 pro-탈락 언어이다.

Culy (1991, p.199)가 제시한 내핵관계절을 갖는 언어의 속성 세 가지 중에서 마지막 두 가지는 한국어에서는 크게 논란이 되지 않고 있다. 한국어는 기본적으로 SOV 어순을 가지고 있으며 명사구가 좌분지 구조를 가지고 있고 한국어는 pro-탈락 언어의 하나로 간주되고 있기 때문이다.

따라서 우리는 여기에서 첫 번째 속성을 내핵관계절의 정의적 속성이라 부르고, 이 정의적 속성에 제시된 두 가지 하위 속성, 즉 명사화 정도와 명사화된 문장의 분포에 초점을 맞추어 다음 두 소절의 논의를 진행하고자 한다.

2.2.1 명사화된 문장 vs. '것'과 선행절의 결합

한국어에서 내핵관계절이라고 주장된 절은 엄격하게 말하면 '것'을 직접 선행하는 절로서 그 자체로는 명사화된 문장(nominalized sentence)이 아니다. 전형적인 내핵관계절 허용 언어로 알려진 케추아어(Quechua) 예문인 (7)에서 확인할수 있듯이 내핵관계절은 그 자체가 명사화된 문장으로서 문법기능을 담당한다.

(7) [RC nuna bestya-ta ranti-shqa-n] alli bestya-m ka-rqo-n.
man horse-acc buy-perfect-3 good horse-validator be-past-3
'The horse that the man bought was a good horse.' (Ancash Quechua, Cole (1987, p.277))

일반적으로 한국어에서 문장을 명사화하기 위해서는 술어에 접미사 'ㅁ/음'이나 '기'를 붙여야 하고 굳이 말한다면 그렇게 해서 나온 결과물은 명사화된 문장이라고 할 수 있다. 그러나 내핵관계절이라고 알려진 한국어 문장에는 어디에도 접미사 'ㅁ/음'이나 '기'가 붙어 명사화된 예문이 존재하지 않는다. 예문 (8)에 제시한 바와 같이 한국어에서는 관계절 자체가 문법기능을 담당하지 못하고 추가로 '것'이라는 요소를 필수적으로 요구한다.

(8) a. 경찰이 [RC 도둑이 은행에서 나오는] *(것)을 체포하였다. b. 철수가 [RC 컴퓨터가 고장난] *(것)을 고쳤다.

명사화된 문장이라는 내핵관계절의 하위 속성과 관련해서 한국어의 내핵관계 절로 알려진 구문은 관계절만으로는 문법적인 문장을 형성할 수 없고 반드시 '것'이 등장해야 한다는 점을 주목해야 한다. 이 두 가지 측면에서 한국어의 내핵 관계절이라고 주장된 절은 일반적인 내핵관계절의 정의에 부합하지 않는다.4

'것'과 관형형 어미로 연결된 선행절의 연속체가 어떤 구조를 이루며 결합되어 있는지에 관해서도 의견의 일치를 보지 못하고 있다. 선행절이 관계절이라면 수식어-피수식어(modified-modifier) 구조를 갖는 것으로 봐야 하겠지만, 지배적인 견해는 '것'이 핵어이고 선행절이 보충어라고 간주하는 핵어-보충어(head-complement) 구조로 분석하는 것이다(N. Kim (1984)), H. Yoon (1991), Jhang (1994), B.-S. Park (1994), Jo (2002), B.-S.Yang(1993), Chung & Kim (2003)).

또한 핵어-보충어 구조로 분석하는 학자들 사이에서도 '것'의 형태-통사적 지위에 관하여 또 다시 의견이 엇갈려있는 상태이다. '것'은 전통적으로 불완전명사(defective noun) 혹은 의존명사(bound noun)로 간주해왔으나 최근에는 크게네 가지 분석이 대두되어 경합하고 있다. 명사화소(nominalizer)로 보는 견해(예를 들어, N. Kim (1984)), 보문소(complementizer)로 보는 견해(예를 들어, Jhang (1994), I.-S.Yang (1972)), 그리고 명사(nominal)로 보는 견해(예를 들어, B.-S.Yang (1993), Chung & Kim (2003))이 있는가 하면 E-유형 대명사(E-type pronoun)로 보는 견해(예를 들어, M.-J. Kim (2004a))로 구분할 수 있다.5

지금까지 논의가 옳다면 '것'을 직접 선행하는 관형절을 내핵관계절로 볼 것인지 아니면 선행 관형절과 '것'을 통칭하여 내핵관계절로 볼 것인지 딜레마에 빠져있다. 우선 '것'을 직접 선행하는 관형절을 내핵관계절로 보려는 견하는 관형적 자체가 명사화된 절이 아니므로 Culy(1987)의 정의에 따르면 내핵관계절이 아니다. 또한 '것'과 그 요소를 선행하는 관형형 어미로 연결된 절을 하나로 묶어 내핵관계절이라고 하려는 견해는 '-나/는 것'이 하나로 명사화 접미사라는 증거를 제시하지 않는 한 종속절이 아니고 종속절이 아니면 관계절이라고 할 수 없는 모순에 빠져 있다. 따라서 우리는 두 견해 중 어느 견해를 취하든 명명의 오류(misnomer)에 빠져 있다고 결론 내릴 수 있다.

⁴ 한국어 내핵관계절 논의에 지대한 영향을 미친 일본어의 경우 '것'에 해당하는 요소인 '-no'가 나와야 한다. 한국어와 일본어를 제외하고 내핵관계절의 구성 요건에 추가적인 요소가 요구되는 구문 구성 상의 제약은 다른 언어에서는 알려지지 않았다.

⁵ Jhang (1994)에서는 '것'이 명사에서 문법화(grammaticalization) 과정을 거쳐 보문소의 역할을 하게 되었다는 제안을 하고 있다. 내핵관계절의 분석에 'E-형 대명사전략'을 도입할 필요가 있다는 M.-J. Kim (2004a,2007)의 주장은 E-형 대명사가 주어진 문맥에서 현저한 속성을 가진 유일하고(unique) 최대치인(maximal) 개체를 표현한다는 점에 기반을 두고 있다. 이 주장에 대한 경험적 부적절성과 개념적 문제점에 초점을 맞춘 반론은 Grosu (2010)를 참고하기 바람.

2.2.2 생산성 vs. 분포의 제약성

내핵관계절은 정의적으로 핵어 명사가 관계절의 내부에 존재한다는 측면에서 일반적으로 잘 알려진 외핵관계절과 구분이 될 뿐 관계절이라는 속성은 공유한 다. 기본적으로 명사구가 나타날 수 있는 곳에는 핵어 명사를 수식하는 관계절 이 등장할 수 있어야 한다. 핵어명사가 관계절의 외부에 있는 외핵관계절이든 관계절의 내부에 있는 내핵관계절이든 이 속성은 공유해야 한다. 그러나 한국어 의 내핵관계절로 알려진 관계절은 특정한 위치에만 등장한다.

내핵관계절이 등장하는 특정 위치로 J.-B. Kim (1999), Chung & Kim (2003), Kim, Sells & Yang (2006)) 등은 논항위치로 한정하고 있고, 이 논항절로서의 내핵관계절 견해는 주어와 목적어 자리에서만 제한적으로 관찰된다고 본 조수근 (2016)에서도 제시되고 있다. 그러나 Jhang (1991, 1994)에 따르면, 내핵관계절의 분포는 논항자리에만 국한되는 것은 아니다. 그는 심층격의 접근 위계에 의해결정되는 것이 내핵관계절의 분포라고 주장한다(이에 대한 반론은 Y.-B.Kim (2002, p.404) 참고).

- (9) 내핵관계절의 분포는 일종의 심층격 접근 위계에 따라 결정된다.
- a. 주어 내핵관계절: 비대격동사의 주어, 수동문의 주어 그리고 직접 목적어만접근 가능하다.
 - b. 목적어 내핵관계절: 주어와 목적어가 접근 가능하다.
- c. 부가어 내핵관계절: 부가어 내핵관계절은 도구격과 수동문의 행위 자에 국한하다.

내핵관계절이 논항위치에만 국한되어 분포한다는 주장은 (9c)를 통해서 이미 문제점이 지적되었다. 추가로 다음 예문도 그 주장에 대한 반례로 보인다.

(10) a. 철수는 [RC 소년이 물을 끓인] 것으로 커피를 탔다. (Jhang (1994)) b. 나는 [RC 네가 옷감을 준] 것으로 옷을 만들었어. (문숙영 (2012, p.46))

Jhang (1991, 1994)가 제시한 분포의 일반화로 논항 위치로 내핵관계절의 분포가 제한되지 않는다는 것은 분명해졌지만, 그럼에도 불구하고 내핵관계절이라고 주장된 절은 분포적으로 매우 제한적이다. 이러한 분포의 제약성은 내핵관계절로 알려진 구문이 과연 관계절인지 의문을 갖게 만들기에 충분하다.

첫째, 한국어의 내핵관계절이라고 주장된 절은 분포상의 제약이 매우 커서 매우 제한적인 위치에만 등장한다. 즉, 명사화된 문장이 다른 구문에도 등장해야

한다는 내핵관계절의 두 번째 정의적 속성은 극단적인 분포상의 제한으로 한국 어에서 관찰되는 해당 구성의 정의적 속성으로 보기 어렵다. 따라서 (2)에 제시 된 예문이 과연 내핵관계절을 포함하고 있는지 의문을 갖게 만들기에 충분하다.

(2)에서도 확인할 수 있듯이 한국어의 내핵관계절이라고 주장된 절은 주문장의 술어가'체포하다', '고치다', '먹다' 등 전형적으로 대상(theme) 의미역을 하위 범주화하는 타동사의 목적어로 '것'이 등장하고 그 '것'과 선행하는 절이 관형형어미로 연결된 구조를 이루는 경우가 대부분으로 그 분포가 매우 제한적이다.6 관계절은 다른 절과 마찬가지로 생산성을 보이는 것이 일반적이며 내핵관계절도 관계절의 일종이기 때문에 역시 생산성을 보여야할 것으로 기대하지만, (11)에서 볼 수 있듯이 기대가 한국어에서는 충족되지 않는다. 흥미롭게도 케추아어내핵관계절 예문 (7)에 해당하는 한국어 예문은 (11a)에서 볼 수 있듯이 '것'의유무와 상관이 없이 비문법적이다.

(11) a. *[RC 그 사람이 말을 산] (것)은 좋은 말이었다. ((7) 참고)

b. *[RC 도둑이 은행에서 나오는] (것)의 인상착의를 직원들이 기억하지 못한다.(최준수 (2012, p.482) 참고)

둘째, 내핵관계절을 포함한 문장과 그렇지 않은 문장의 구분이 명확하지 않다. 일반적으로 지각동사로 알려져 있는 '보다'가 주절에 쓰인 다음 (12)와 같은 예문은 학자에 따라 내핵관계절 여부의 판단에 이견이 있다. M.-J. Kim (2004a, p.11)은 내핵관계절의 예로 다음을 들고 세 가지로 해석이 가능하여 중의적인 문장으로 보고 있다.

- (12) 진호는 [RC 아이가 국수를 삶고 있는] 것을 보았다.
 - a. 진호는 국수를 삶고 있는 아이를 보았다.
 - b. 진호는 아이가 삶고 있는 국수를 보았다.
 - c. 아이가 국수를 삶고 있었고 진호가 그 사건(= 광경)을 보았다.

그녀에 따르면 예문 (12)에서 진호가 본 것은 아이일 수도 있고, 국수일 수도 있으며 나아가서 아이가 국수를 삶는 사건을 볼 수도 있다는 의견을 제시하고

⁶ 실제로 문헌에 등장한 주문장 동사는 타동사에 국한하는 것을 확인할 수 있으며, 타동사 중에서 다음과 같이 약 10개 정도의 동사가 주문장의 동사에 나오고 소위 말 하는 내핵관계절이 목적어 자리에 나오는 경우에 한정하여 내핵관계절이라고 단정하 고 논의를 진행해 오고 있음을 알 수 있다. 문헌에 나오는 동사는 모두 44개정도로 크 게 보아 50개 미만일 것으로 보인다.

있다. 기본적으로 (12a)와 (12b) 해석이 가능하다는 점에서, 즉 관계절 내부의 명사(구)와 동일한 지시체를 '보다'의 목적어로 해석할 수 있다는 점에서 내핵관계절로 보지 않을 이유가 없다.

그러나 이 동일한 예문을 동일한 저자인 M.-J. Kim (2004b)마저도 다른 문헌에서는 직접지각구문(direct perception construction, DPC)으로 명명하고 내핵관계절이 아니라고 주장하고 있다. Jhang (1994, p.16ff.)도 이 예문을 지각동사의사실적 보충어(factivecomplements of verbs of perception verbs)로 구분하고 대핵관계절로 간주하지 않는 입장을 취하고 있다. 그러나 B.-S. Park (1994, p.41)이나 J.Lee (2006) 그리고 최근 J. Park(2021)에서는 이 예문처럼 주절의 동사가 '보다'인 예문을 대핵관계절 예문으로 간주한다.

2.3 피봇명사구의 명시성

위에서 제시한 바와 같이 Jhang (1994, p.18)은 한국어의 내핵관계절을 '문장 내부의 명시적인 체언을 수식하는 명사화된 문장'으로 정의하고 있다. 그러나 피봇명사구의 명시성 조건은 대상의 자의적 제한이라는 문제점과 한국어에서 과연 내핵관계절로 어떤 구문을 인정하고 그 구문의 기술과 설명에 어떤 기여를 하고 있는지 의문을 갖게 한다. 다음에 그 문제점을 여섯 가지 측면에서 살펴보고자 한다.

첫째, 피봇명사구의 명시성을 내핵관계절일 조건으로 본다면 내포절에 주절술어의 대상이 명시적으로 실현되어 있는 (13a)만이 내핵관계절이고, 비명시적인 대상을 지칭하고 있는 (13b)는 내핵관계절이 아니라고 해야 할 것이다. 그러나 [RC . . .]로 구분된 절은 동일하고 주절의 술어도 대상을 목적어로 갖는 타동사라는 점에서 (13a)와 (13b)를 전혀 다른 구문으로 다루어야 하는 언어학적인이유를 찾기 힘들다.7

(13) a. 교수님은 [RC 학생이 물을 끓인] 것을 문 밖에 버렸다. (명시적: 물)

b. 교수님은 [RC 학생이 물을 끓인] 것을 문 밖에 두었다. (비명시적: 예, 냄비)

둘째, '것'과 선행절이 궁극적으로 정도(degree)나 양(quantity)을 지칭하는 것으로 해석되는 (14b)와 같은 예문은 피봇 명사구의 명시성을 조건으로 한다면 내

⁷ 피봇 명사구가 비명시적인 예문은 주로 주절 술어와 도구(instrument) 혹은 용기 (containment)와 같은의미역 관계를 갖는 경우 발견된다.

핵관계절로 볼수 없다. [RC . . .]로 구분된 절은 동일하고 주어절로서 문법기능도 동일하다는 점에서(14a)만 내핵관계절이라 하고 (14b)는 내핵관계절이 아니라고 주장한다면 이 역시 근거가 확실하지 않은 대상의 배제라는 문제를 드러낸다.

(14) a. [RC 아이가 공을 친] 것이 옆집으로 날아갔다. (명시적: 공) b. [RC 아이가 공을 친] 것이 너무 셌다. (비명시적: 정도)

셋째, Chung & Kim (2003)이 비명시 선행사(implicit antecedent)라 불렀던 명사구도 비명시적이다. 그들은 (15)를 내핵관계절 예문으로 제시하고 있고 철수가 닦은 것은'컵의 물'이 아니라 '넘친 물'이라고 간주하고 있어 피봇명사구의 명시성을 내핵관계절의 조건으로 보는 견해와 모순이다.

(15) 철수는 [RC 컵의 물이 넘친] 것을 닦았다. (비명시적: 넘친 물)

넷째, 일반적으로 둘 이상의 물질이 혼합된 새로운 물질을 지칭하는 명사구는 내핵관계절의 내부에 명시적으로 실현되지 않는다. 이러한 종류의 예문들은 피 봇명사구의 명시성 조건 때문에 해당 내포절을 내핵관계절로 보려는 학자들은 배제하겠지만 이 기준이 언어학적으로 타당한 기준인지 의문을 갖게 만들기에 충분하다.8

- (16) a. 철수가 [RC 영희가 물에 설탕을 탄] 것을 마셨다. (비명시적: 설탕물)
- b. 철수가 [RC 영희가 맥주에 소주를 반반 섞은] 것을 마셨다. (비명 시적: 소맥)

다섯째, 피봇 명사구의 명시성으로 의도한 바는 '것'을 선행하는 절 내부에 주절 술어와 문법적 관계를 맺는 명사(구)가 하나 존재해야 한다는 것이다. M.-J. Kim (2004a)에 따르면, 내핵 관계절 (16)은 (17a)의 의미로도 해석이 가능하고 (17b)의 의미로 해석이 가능하다. 여기에서 주목해 보아야 할 점은 내핵 관계절 (17)이 (17c)의 의미로도 해석이 가능하다는 점이다. 주문장의 술어 '잡다'의 대상으로 '것'을 선행하는 절 내부에 있는 주어와 목적어가 동시에 해석이 가능한 것

⁸ 주절 술어의 상적 특성에 따라 피봇 명사구의 지시 대상이 시간의 추이에 따라 결정되는 경우가 있음을 Y.-B. Kim (2002)와 M.J. Kim (2004a)에서 언급하고 있다. 이러한 종류의 피봇명사구는 각각 결과 대상(resultanttheme)과 점증 대상(incremental theme)으로 명명하고 있다.

은 이다.

- (17) 진호가 [RC 고양이가 쥐를 쫓고 있는] 것을 잡았다. (=(7), M.-J. Kim (2004a))
 - a. 고양이가 쥐를 쫓고 있었고, 진호가 고양이를 잡았다.
 - b. 고양이가 쥐를 쫓고 있었고, 진호가 쥐를 잡았다.
 - c. 고양이가 쥐를 쫓고 있었고, 진호가 고양이와 쥐를 잡았다.

엄격한 의미에서 피봇 명사구의 명시성은 예문 (17)을 (17c)처럼 해석하는 가능성을 배제하는 내용을 조건으로 담고 있지만, M.-J. Kim (2004a, p.8) 주장처럼 내핵관계절의 선행어가 관계절 내부의 지시체를 동시에 복수로 지시할 수 있다면 피봇 명사구의 명시성 조건을 만족시킨다고 보기 어렵다. 예문 (17)을 (17a)와 (16b)로 해석할 경우만 내핵관계절로 분석하고 (17c)로 해석될 경우 내핵관계절로 보지 않는다면 자의적이라는 비판을 면하기 어렵다.9

여섯째, 다음 예문처럼 피봇 명사구가 명시적으로 실현된 내포절과 그렇지 않은 내포절을 전혀 다른 구문으로 — 예를 들어, 하나는 내핵관계절로 다른 하나는 내핵관계절이아닌 절로 — 분석해야 하는 불합리한 경우가 존재한다. 이는 내핵관계절의 포괄 범위를 자의적으로 축소하는 문제와 관련이 되어 있을 뿐더러 포착할 수 있는 일반화를 놓치는 문제를 안고 있다고 봐야 할 것이다.

- (18) a. 경찰이 [RC 도둑이 은행에서 나오는] 것을 체포하였다.
 - b. 경찰이 [RC 은행에서 나오는] 것을 체포하였다.
- (19) a. 철수가 [RC 컴퓨터가 고장난] 것을 고쳤다.
 - b. 철수가 [RC 고장난] 것을 고쳤다.

(18a)와 (19a)에서는 주절 술어의 목적어가 각각 '도둑'과 '컴퓨터'로 해석이 되지만, (18b)와 (19b)에서는 각각 '어떤 사람'과 '어떤 물건'이라는 문맥에 따라 결정될 수 있는 개체를 의미하는 것으로 해석하는 것이 일반적이다. 중요한 것은 두 예문에서 주절 술어의 목적어로 해석하는 개체를 명시적인 경우 내포절에서 찾고 비명시적인 경우 문맥에서 찾는 차이만 존재할 뿐 기본적으로 피봇 명사구를 정하는 방법은 동일한 것으로 보인다는 점이다.

^{9 (17}c)에 예시된 것과 같이 관계절 내부에 있든 외부에 있든 관계절에서 문법 기능을 담당하는 두 개 이상의 명사구가 동시에 피봇으로 기능하는 문장은 일반적으로 관계절로 구성하는 것이 불가능하다. 예측한 바와 같이 (17c)의 해석을 갖는 관계절은 한국어에서 외핵 관계절로도 표현할 수 없다.

지금까지 우리는 피봇 명사구의 명시성을 내핵관계절의 조건으로 제시한 정의의 문제점을 몇 가지 유형으로 나누어 살펴보았다. 그 유형은 피봇 명사구가비명시적이지만 도구나 용기로 해석될 수 있는 경우, 정도(degree)나 양(quantity)을 지칭하는 경우, 동일한 개체를 지시하지 않는 경우, 혼합되어 새로운 개체를 지칭하는 경우, 둘 이상의 피봇명사구와 관련되어 해석되는 경우, 그리고 마지막으로 문맥에 의존하여 결정될 수 있는 경우 등을 포함한다. 어느 경우이든 쟁점이 되고 있는 절을 내핵관계절로 분석하려는 시도가 피봇 명사구가비명시적이라는 이유만으로 배제하기 힘든 예문들이다. 특별한 언어학적 근거없이 지나치게 대상을 제한하는 방법론은 설득력있는 설명을 제공하는 데에 치명적인 약점이라 할 수 있다.

2.4 중간 요약

지금까지 우리는 '것'과 선행하는 절이 관형형 어미로 연결된 구문을 내핵관 계절로 간주하고 분석하는 주장의 출발점인 구문의 정의와 정의적 속성을 일반 이론언어학적인 시각에서 논의하였다. 우선 Culy (1991, p.27)와 Jhang (1994), 그리고 M.-J. Kim (2004a)의 정의를 소개하고, 다른 언어의 내핵관계절과 한국어 에서 문제가 되고 있는 '것'을 선행하는 관형형 어미로 연결된 절은 구성과 분포 에서 상이함을 제시하였다. 또한 정의적 속성으로서 내핵관계절로 알려진 관계 절 그 자체는 명사화된 문장이 아님이 드러났다. 대신 '것'과 선행하는 절이 관형 형 어미로 연결된 구문이 하나의 단위로서 매우 제한된 분포를 보이며 문법적 기능을 하고 있음을 밝혔다. '것'과 선행절 전체를 내핵관계절로 간주하는 경향 적 오류가 있음도 분명하게 밝혔다. 이런 경향적인 단정이 오류인 이유는 내핵 관계절은 관계절이어야 하고 관계절은 종속절이어야 하는데, '것'과 선행절이 결 합한 구성체는 이미 종속절이 아니기 때문이다. 또한 피봇명사구는 명시적으로 '것' 을 선행하는 관형형 어미로 연결된 절 내부에 있을 수도 있고, 명시적으로 나타나 있지 않을 수도 있을 뿐더러 절 내부의 명사(구) 하나만이 아닌 복수의 개체를 지칭할 수도 있다는 증거를 예문을 통해 제시하였다. 결과적으로 피봇명 사구의 명시성도 문제가 되고 있는 구문을 한국어 내핵관계절이라 간주한다 하 더라도 이 구문을 다른 구문과 구분하는 기준으로서는 부적절하다는 점도 드러 났다.

3 내핵관계절의 기본 속성에 대한 반론

M.-J. Kim (2004a, p.5ff.)과 J. Lee (2006, p.12ff.)은 내핵관계절의 기본 속성으로 다음과 같은 다섯 가지를 제시하고 있다. 첫째는 한국어 내핵관계절은 절 내

부에 피봇명사구에 해당하는 공백이 없어야 하기 때문에 무공백관계절이라는 점(J. Lee (2006)), 둘째, 내핵관계절의 의미 핵, 즉 피봇명사구는 무제한의존성 (unbounded dependency)을 보이지 않는다는 점(J. Lee (2006)), 셋째, 한국어 내핵관계절은 비제한적 관계절로서 항상 최대성/전체성 효과(maximality/exhaustivity effects)를 보이고 내핵관계절의 내용은 의미론적 핵어, 즉 피봇명사구가 아닌 주절의 내용을 제한한다는 점(M.-J. Kim (2004a), J. Lee(2006))을 제시하고 있다.

그외 관계절이라고 단정할 경우 당연한 피봇구성소의 이중성을 넷째 속성으로 들고있다(M.-J. Kim (2004a)). 다섯째는 피봇구성소의 지시체는 미결정적이어서 담화 문맥과 주문장 술어의 의미론 등이 복합적으로 작용하는 관련성 조건과 같은 화용론적 요인에 따라 달라질 수 있다는 점(M.-J. Kim (2004a), J. Lee (2006))을 기본 속성으로 제시하고 있다.10

이 절에서는 이 중에서 첫 세 가지 기본 속성의 타당성을 비판적으로 고찰하고자 한다. 마지막에 제시한 두 속성을 여기에서 다루지 않는 이유는 관계절이라고 단정할 경우 정의에서 귀결되는 당연한 속성이거나(넷째 속성), 첫 세 가지기본 속성에 대한 관찰적 타당성이 확보되지 않으면 자연스럽게 논박이 되는 주장이기 때문이다.

3.1 무공백 관계절

M.-J. Kim (2004a, p.5)는 내핵관계절을 내포절과 관계표지의 결합으로 정의하고 내포절은 관계표지 '는'을 선행하는 무공백절이라고 규정하고 있다. '것'을 선행하는 관형절이 무공백절이어야 한다는 입장은 Jhang (1994)과 J. Lee (2006))에서도 유지되고 분명한 언급이 없는 많은 문헌에서도 견지되고 있다. 또한 내핵관계절에서는 외핵관계절과 달리 피봇명사구가 관계절 내부에 존재하기에 정의적으로 무공백관계절이어야 한다. 그러나 한국어의 대핵관계절이라고 주장된 구성에서 '것'을 선행하는 절이 무공백이어야 하는지는 적어도 다음 네 가지 쟁점에 대한 이견이 해소되어야 주장이 가능하다고 보인다.

¹⁰ Kuroda (1976)는 일본어의 내핵관계절인 내포절과 주절이 화용론적으로 상호 관련성이 적어도 다음 세가지 중 하나에 속해야 한다고 주장한다. 첫째, 두 절이 시제적인 배경을 공유해야 하고, 둘째, 두 절은 공간적 배경을 공유해야 하며, 셋째, 그 두절이 의도적으로 혹은 동기적으로 상호 연결되어 있어야 한다는 것이다. Kuroda의 관련성 조건은 종종 외핵관계절과 내핵관계절이 항상 일대 일로 대응하지 아니 한다는 경험적 관찰을 설명하는 하나의 이유로 간주된다. 이 관련성 조건을 수정하고 보완하는 논의가 한국어 대핵관계절 의미핵 결정 논의의 대부분을 차지한다. 이런 맥락의 논의에는 Y.-B.Kim (1996), (2002), M.-J.Kim (2004a) 등이 포함된다.

첫째, 내핵관계절과 외핵관계절이 중첩된 이중관계절이 가능한 이유를 설명할 수 없다. 내핵관계절이 반드시 무공백관계절이어야 한다면 논리적으로 (20)과 같은 이중관계절은 불가능해야 할 것이다. 아니면 내핵관계절로 간주되는 종속절을 내핵관계절이 아닌 것으로 보아야 할 것이다. (20)에서 공백은 문맥상 각각 '도둑'과 '컴퓨터'로 해석이 된다. 이 해석은 서론에 제시한 예문 (2)를 주문장의술어에 의해 기술된 사건의 대상이 각각 '도둑' 과 '컴퓨터' 인 것으로 이해하는 방식과 동일하다.

- (20) a. [RC _i 은행에서 나오는] 것을 현장에서 체포한] 도둑i b. [RC _i 고장난] 것이 틀림없는] 컴퓨터i
- (2)를 내핵관계절이라고 한다면 (20)에 제시한 예문의 [RC . . .] 로 구분된 절도 내핵관계절이라고 해야 할 것이다. 그러나 (20)의 [RC . . .] 로 구분된 절은 공백을 포함하고 있으므로, 내핵관계절은 무공백관계절이라는 주장에 대한 반례가 된다.

둘째, (21c)와 같은 다중대격표지구문의 가능성도 무공백 조건과 모순된다(다 중격표지구문에 대한 자세한 논의는 류병래 (2014) 참고). 다음 예문 (21a)는 관계절의 주어 자리에 주격 명사구인 '도둑이'는 가능하지만 대격 명사구인 '도둑을'은 불가능함을 보여준다. 대격 명사구인 '도둑을'이 관계절 밖에 위치한 (21b)는 다중격표지구문을 형성하고 있는데, 이 경우 관계절의 주어는 관계절 내부에서 실현될 수 없고 (21c)처럼 공백으로 남아있어야 한다.

- (21) a. 경찰이 [RC {*도둑을/도둑이} 은행에서 나오는] 것을 현장에서 체포하였다.
- b. 경찰이 도둑을 [RC (*도둑이) 은행에서 나오는] 것을 현장에서 체포하였다.
 - c. 경찰이 도둑을 [RC 은행에서 나오는] 것을 현장에서 체포하였다.

문제가 되고 있는 '것'과 선행절의 연속체를 내핵관계절이라고 보는 견해의 핵심은, 간략하게 말해서, '것'이 관계절 내부에 있는 명사구의 지시체와 동일한 것으로 해석할 수 있다는 것이다. (21c)에서 내포된 관계절의 주어는 '도둑'으로 해석된다. 즉, (21c)의 해석 방식은 내핵관계절이라고 주장된 절에서와 기본적으로 동일하다. 따라서 (21c)에 제시한 바와 같이 주어 자리가 비어있어야 하는 다중대격표지구문은 내핵관계절이 무공백관계절이어야 한다는 주장에 대한 반례의 하나이다.

셋째, 한국어에서 피봇명사구가 미명시적이면 그 자리는 공백으로 나타나는

데, 그런 공백을 포함한 예문도 일반적으로 문법적이다. 예를 들어, 예문 (22)는 '누군가 은행에서 나오는 사람을 체포하였다'는 의미와 '무언가 고장난 물건을 고쳤다'는 의미로 자연스럽게 해석이 가능하다.

(22) a. 경찰이 [RC 은행에서 나오는] 것을 체포하였다. (= (18b)) b. 철수가 [RC 고장난] 것을 고쳤다. (= (19b))

(22)에 내포된 관계절을 무공백관계절이 아니라는 이유로 내핵관계절이 아니라고 주장하는 것은 타당성을 현저하게 잃었다고 봐야 한다. 왜냐하면 관계절의 피봇명사구, 즉 의미핵의 결정 방식은 공백이 있는 경우나 무공백인 경우나 동일하기 때문이다. 따라서 (22)와 같은 예문을 무공백이 아니라는 이유로 내핵관계절이 아니라고 하거나 내핵관계절은 무공백관계절이어야 한다는 점을 기본속성으로 주장하는 것은 자의적이고 타당성을 인정받기 어렵다고 보인다.

이미 언급한 바와 같이 내핵관계절을 갖는 다른 언어에서도 핵어 명사는 명시적으로 관계절 내부에 실현되어 나타날 수도 있고 실현되지 않을 수도 있다 (Culy (1991)). 예를 들어 다음 예는 하다차어(Hidatsa)의 내핵관계절인데 핵어 명사가 실현되지 않은 것이다(Boyle (2016, p.259)). 핵어 명사가 실현되어 있지 않은 경우에도 우리는 내핵관계절이 아니라고 하지 않으며 이 경우 what, something, it을 의미하는 것으로 미명세적으로 의미해석을 한다.

(23) Mary e aku-híri-š wa-rúci-c (=(10), Boyle (2016, p.259))

Mary e REL.S-make-DET.D 1A-buy-DECL 'I bought (what/something/it) Mary made.'

이 논의가 맞다면 내핵관계절의 정의적 속성으로 관형절에 공백이 나타나지 않아야 한다고 보는 것은 잘못된 관찰이라고 결론 내릴 수 있다.

넷째, 어떤 경우에는 정반대로 무공백관계절이라는 조건을 만족시키기 위하여 논항을 채우면 해당 관형절의 수용성이 현저하게 떨어지기도 한다 (최준수 (2012) 참고).

(24) a. 철수는 디저트로 [RC (??중국집 주인이) 감을 얼린] 것을 먹었다.

b. 북어는 [RC (?? 어부가) 명태를 말린] 것이고, 동태는 [RC (?? 누군가) 명태를 얼린] 것이고, 코다리는 [RC (??업자가) 명태를 코를 꿰어 꾸덕꾸덕하게 말린] 것이다.

외핵관계절과는 달리 내핵관계절은 피봇 명사구가 관계절 내부에 존재하기에 무공백인 경우가 대부분이다. 이점에서 내핵관계절은 외핵관계절과 극명하게 대비가 된다. 그러나 지금까지 논의가 맞다면 한국어에서 내핵관계절이라고 주장된 구성에서 '것'을 선행하는 절이 무공백이어야 한다는 점은 주장하기 힘들 것으로 보인다.

이 문제 제기는 한국어 내핵관계절은 공백을 가지고 있다고 주장하고 있는 K.-S. Kim(2022)에서도 확인되고 있다. 다만 내핵관계절 내부에 공백이 있으면 내핵관계절은 무공백 관계절이어야 하는 정의적 속성과 모순된다는 지적 또한 피할 수 없다.

3.2 무제한의존성

영어의 경우 관계절은 Wh-의문문 및 주제화구문과 더불어 가장 전형적인 무제한의존구문가운데 하나이다. 문제가 되고 있는 예문 (2)를 내핵관계절이라고 가정하고 분석하는연구에서는 외핵관계절인 (1)과는 달리 의미 핵, 즉 피봇명사구가 절 경계를 넘어 관계를맺을 수 없다고 주장한다. 즉, 내핵관계절은 무제한 의존성을 드러내지 않는다는 것이다.이 주장은 일본어의 해당 구문의 특징으로 주장한 Watanabe (2004)를 한국어에 적용한것으로, 이를 수용한 J. Lee (2006, p.19)에서는 그 증거로 다음 예문을 들고 있다.

- (25) a. *경찰은 [RC 사람들이 [RC 도둑이 은행에서 나온다고] 말한] 것을 잡았다.
 - b. *철수는 [RC 사람들이 [RC컴퓨터가 고장났다고] 말한] 것을 고쳤다.

그러나 내핵관계절이 반드시 무공백관계절이어야 한다는 주장에 대한 문제점을 지적하면서 제시한 예문 (20)을 포함하고 있는 다음 예문은 이 주장에 문제가 있음을 보여준다. 아래에 가독성을 위하여 다시 적는다.

(20')a. [RC 경찰이 [RC _i 은행에서 나오는] 것을 현장에서 체포한] 도둑i 이 묵비권을 행사하고 있다.

b. [RC 철수가 [RC _i 고장난] 것이 틀림없는] 컴퓨터i를 고쳤다.

(20'a)에서 '도둑'은 이중으로 내포된 절의 주어로 해석이 되는데, 이는 피봇명 사구인'도둑'이 절 경계를 두 번이나 넘어 문법적 관계를 형성하고 있다는 말이다. (20'b)의'컴퓨터'도 마찬가지로 절 경계를 두 번이나 넘어 문법적 관계를 형성하고 있다.

(26) a. [RC 경찰이 [RC 도둑이 은행에서 나오는] 것을 현장에서 체포한] 것을 시민들이 에워쌌다.

b. [RC 철수가 [RC 컴퓨터가 고장난] 것을 고친] 것을 영희가 또 고장 냈다.

피봇명사구가 절 경계를 넘어 문법적 관계를 형성할 수 있음은 해당 관계절이 무공백절로서 가장 깊이 내포되어 있는 경우에도 가능하다. 무제한의존성을 보여주지 않는 것처럼 보이는 예문 (25)를 J. Lee (2006, p.19)에서 비문법적으로 본이유는 첫 내포절의 술어인 '말하다'가 소위 말하는 내핵관계절을 허용하지 않는 술어이기 때문인 것으로 보는 것이 합리적이다. 예문 (26)이 가능한 것으로 보아, 예문 (25)를 근거로 내핵관계절이라고주장하는 절 자체가 무제한의존성을 가지지 않는다고 주장하기 어렵다고 보인다. 따라서문제가 되고 있는 구문을 내핵관계절로 가정하고 내핵관계절은 무제한의존성을 보여주지 않는다고 하는 주장은 경험적 타당성을 얻기 힘들다고 결론 내릴 수 있다.

3.3 진리조건적 의미의 동일성

한국어 내핵관계절 논의의 출발점은 외핵관계절로 표현된 문장 (1)과 (2)가 각각 동일한 의미를 가진다는 관찰이다. Kim (1999), Chung & Kim (2003), Kim, Sells & Yang (2006)등은 외핵관계절과 내핵관계절에 명확한 진리조건적 차이가 존재하지 않는다고까지 주장한다. 그러나 이 주장은 적어도 다음 두 가지 반례에 의해 그 타당성이 의문시된다. 첫째, 내핵관계절의 핵어명사가 명사구로 명확하게 드러나지 않는 표현들은 두 관계절 유형의 진리조건적 의미동일성에 문제를 제기한다. 예문 (27a)에서 철수가 산 것은 '아이스크림 케익'이라고 해석하는 것이 가장 일반적이다. 그러나 (9b)에 제시된 심층격접근 위계에 따르면 예문 (27a)의 피봇명사구는 주어인 '체리'가 될 것이다. 그러나 체리가 피봇명사구로나와 있는 외핵관계절 (27b)와는 그 의미가 다르다.

- (27) a. 철수는 [RC 체리가 아이스크림 위에 얹혀 있는] 것을 사왔다.
 - b. 철수는 [RC 아이스크림 위에 얹혀 있는] 체리를 사왔다. (≠ (27a))
- c. 철수는 [RC 체리가 위에 얹혀 있는] 아이스크림을 사왔다. (≠ (27a))
- d. 철수는 [RC 체리가 (위에) 얹혀 있는] 아이스크림 케익을 사왔다. (=(27a))

(27a)와 동일한 의미를 가진 외핵관계절은 (27d)일테지만, 피봇명사구인 '아이스크림 케익'은 (27a)에서 종속절이나 주절 어디에도 등장하지 않는다. 철수가산 것은 (27b)에서는 피봇명사구인 '체리'이고 (27c)에서는 '아이스크림'이다. 즉, 예문 (27a)는 외핵관계절 예문 (27b)와도 (27c)와도 동일한 의미를 가지고 있지않다. 나아가서 '아이스크림 케익'은 주어로도 목적어로도 종속절에 명시적으로 등장하지 않기 때문에 내핵관계절을 가정하는 기존의 분석법으로는 설명이 불가능하다.

둘째, 외핵관계절은 복수의 지시체를 피봇 명사구로 가질 수 없으나 내핵관계절이라고 주장하는 예문에서는 가능하다. 예문 (28)에 제시된 바와 같이, 심지어 M.-J. Kim (2004a,p.8)에서는 내핵 관계절의 선행어가 관계절 내부의 지시체를 동시에 복수로 지시할 수 있다고 주장한다.

- (28) 진호가 [RC 고양이가 쥐를 쫓고 있는] 것을 잡았다. (=(7), Kim(2004a: 8))
 - a. 고양이가 쥐를 쫓고 있었고, 진호가 고양이를 잡았다.
 - b. 고양이가 쥐를 쫓고 있었고, 진호가 쥐를 잡았다.
 - c. 고양이가 쥐를 쫓고 있었고, 진호가 고양이와 쥐를 잡았다.

그에 따르면 내핵 관계절 (28a)은 외핵 관계절 (28b)의 의미로 해석이 가능하기도하고, (28c)의 의미로도 해석이 가능하다. 주목할 만한 점은 (28c)의 경우 해당 외핵관계절이 아예 존재하지 않는다는 점이다.

셋째, J.Park (2021)에 지적하는 바와 같이 소위 말하는 '없어진 종이 퍼즐'도 내핵관계절과 외핵관계절의 진리조건적 의미동일성에 문제가 있음을 보여준다. (29a)에서 본것은 '종이가 타 버리고 남은 재'이지만, (29b)에서 내가 본 것은 '종이'다. 물론 이문장이 화용론적으로 적절하지 않지만 적절하다고 가정했을 경우에 그렇다.

- (29) a. 나는 그 집에 들어가자 마자, [RC 종이가 완전히 타 버린] 것을 보았다. (=(24), J.Park (2021))
- b. ??나는 그 집에 들어가자 마자, [RC 완전히 타 버린] 종이를 보았다.

따라서 우리는 내핵관계절이 포함된 문장의 진리치와 외핵관계절이 포함된 문장의 진리치가 같다는 주장은 수용하기 힘들다고 결론을 내릴 수 있다. 두 관 계절 유형의 진리조건적 의미 동일성에서 출발한 내핵관계절 기본 속성에 관한 관찰은 전면 재검토가 필요한 이유이기도 하다.

3.4 비제한적 해석 혹은 최대성/전체성 효과

영어의 경우 비제한적 관계절은 피봇명사구와 쉼표로 구분되어야 하고 관계 대명사로[+WH] 형식이 나와야 하며 생략되거나(φ) that이 나올 수 없다. de Vris (2006)의 예문을 우선 살펴보기로 한다.

- (30) a. (I spoke to) the lecturers that failed the test on didactics. (제한 적 관계절)
- b. (I spoke to) the lecturers, who failed the test on didactics. (비 제한적 관계절)
 - c. (I spilled) the coffee that there was in the pot. (최대성 관계절)

(30a)에서는 시험에 떨어진 강사 그룹에게만 말하고 시험에 합격한 강사는 언급하지 않았다. (30b)에서는 담화 영역의 모든 강사와 대화를 나누었는데, 강사들은 모두 시험에 떨어졌다는 의미로 해석된다, 그러나 (30c)에서는, 다른 커피와 대비는 없지만, 주전자 안에 있는 커피의 양 최대치, 즉 전부가 쏟아졌다는 의미로 해석이 된다. 즉 양이나 정도의 상대적인 최대성을 표현하고 있으며, 이최대성의 표현 방법이 비제한적 관계절, 그 중에서도 최대성 관계절 (maximazing relative clause)이다. 이 세 번째 유형의 관계절은 Carlson (1977)과 Grosu & Landman (2017)에서 자세하게 논의되었다.

외핵관계절과 내핵관계절이 통사적으로도 다른 제약을 받고 의미적으로도 다른 해석을 받는다는 주장이 일본어 자료를 근거로 제기되어 왔다 (Hoshi (1995), Shimoyama(1999), M.-J. Kim 2005, Watanabe (2004) 참고). 이 견해에 따르면 외핵관계절은 제약적 관계절 해석을 받지만, 내핵관계절은 비제한적 해석을 받는다고 본다. 이 주장은 한국어의 대핵관계절 문헌에서도 일부 받아들여져, 한국어외핵관계절에서는 제한적인 해석만이 가능하지만 대핵관계절에서는 비제한적관계절에서와 같이 최대성/전체성 효과(maximality/exhaustivity effects)를 보인다는 주장으로 이어진다(M.-J. Kim (2004a),J.Lee (2006) 참고).11

¹¹ de Vries (2018)가 소개하고 있듯이 내핵관계절은 유형론적으로 두 가지 유형이 있다고 알려져 있다. 일부 언어(예: 라호타, 모하비)에서는 내핵관계절이 제한적 해석 만을 허용하지만 다른 언어(예: 케추아어, 일본어, 한국어)에서는 의무적으로 최대성 해석, 즉 비제한적 해석만을 허용한다는 것이다. 제한적 내핵관계절은 비한정적 해석을 허용해서 존재 양화사와 공기할 수 있지만, 비제한적 내핵관계절은 그런 해석을 허용하지 않으며 중첩(stacking)도 허용하지 않는 것으로 알려져 있다. 또한 비제한적 내핵관계절에서는 관계절 내부의 피봇명사구가 강한정사(예컨대 정관사 및 전칭 양화사)

이 주장에 따르면, (31a)는 강도가 세 명이 넘을 수 있는 상황에서 경찰이 체 포한 강도는 세 명이라는 해석을 갖고, (31b)는 강도가 세 명이 전부이고 그 세 명을 경찰이 모두 체포했다는 해석을 갖는다.

(31) a. 경찰이 [RC 은행에서 나오는] 강도 세 명을 체포하였다. b. 경찰이 [RC 강도 세 명이 은행에서 나오는] 것을 체포하였다.

그러나 이 주장 혹은 관찰이 타당한지 검토하기 위해서 우리는 적어도 다음 몇 가지 쟁점을 해소해야 한다.

첫째, 예문 (31b)의 의미로부터 강도가 세 명이 전부라는 해석은 도출되지 않고,다만 강도 세 명이 동시에 은행에서 나오다 경찰에 체포되었다는 의미로 해석이 된다고 보는 것이 보편적이다. 즉, 예문 (31a)와 (31b)는 '동시성'에서 해석의 차이가 드러나지 최대성/전체성 효과가 (31b)에서만 존재한다고 보기 어렵다.

한국어에는 제한적 관계절과 비제한적 관계절 혹은 최대화 관계절을 구분하는 형태 통사적 장치가 없다. 따라서 외핵관계절은 제한적 해석을 갖는 데에 반하여, 여기에서 쟁점이 되고 있는 내핵관계절은 비제한적 해석을 갖는다는 주장은 한국어에서는 직관 외에 다른 문법적 증거를 통해 증명되지 않은 한 그 주장의 타당성이 의심받을 수 밖에 없고 공감대를 얻기 어렵다.

둘째, 비제한적 관계절로서 내핵관계절이라고 주장하는 절이 최대성/전체성 효과를 보인다는 주장에 대한 반례가 많은 것이 문제이다. (31b)가 수(number)에 관한 최대성/전체성 효과에 대한 반례라면 다음 예문은 양(quantity)에 관한 최대성/전체성 효과가없다는 것을 보여주는 반례이다.

(32) a. 철수는 [RC 탁자에 넘친] 컵의 물을 닦았다. b. 철수는 [RC 컵의 물이 탁자에 넘친] 것을 닦았다.

(32b)가 내핵관계절로서 최대성/전체성 효과를 보인다는 주장이 타당하려면 컵의 물이 모두 탁자에 넘쳐 흘렀다는 해석이 가능해야 하지만, 실제로는 이 해 석이 불가능하다. (32b)는 컵의 물 일부만이 넘쳐 탁자로 흐른 것으로 보는 것이 가장 자연스러운 해석이다. 셋째, 영어의 경우 제한적 관계절과는 달리 비제한적 관계절은 특정 명사구, 예를 들어 고유명사와 공기할 수 있다는 사실이 잘 알려 져 있는데, 한국어에서는 제한적 관계절로 간주되는 외핵관계절 (33a) 뿐만 아니

와 결합할 수 있지만, 제한적 내핵관계절에서는 외부지정어만이 강한정사일 수 있는 차이가 있다. 이런 이유로 제한적 관계절은 NP의 부가어로, 비제한적 관계절은 DP의 부가어로 통사구조를 분석하려는 시도가 있었다(Demirdash (1991)).

라 내핵관계절이라고 주장된 예문 (33b)에서도 고유명사는 피봇명사구로서 등장할 수 있다.

- (33) a. 철수는 [RC 고함치는] 영희를 나무랐다.
 - b. 철수는 [RC 영희가 고함치는] 것을 나무랐다.

따라서 우리는 외핵관계절은 제한적 관계절이고 내핵관계절은 비제한적 관계절이라는 이분법적 분석이 타당하지 않음을 예문 (33)을 통해서 증명되었다고 말할 수 있다.12

넷째, 영어의 경우 비제한적 관계절은 특정 양화사, 예를 들어 존재양화사의 하나인 any와는 공기하지 않는다. 그러나 한국어의 내핵관계절이라고 주장된 예 문에서는 이러한 제약이 적용되지 않는다.

- (34) a. *Any lion, which eats small animals, is cowardly.
 - b. Any lion that eats small animals is cowardly.
- (35) a. 아이들은 [RC 송사리를 잡아먹는] 어떤 물고기든 잡았다.
 - b. 아이들은 [RC 어떤 물고기든 송사리를 잡아먹는] 것을 잡았다.

이 소절에서는 관계절의 해석의 차이를 기반으로 외핵관계절과 내핵관계절을 구분하여 한국어 내핵관계절에서는 비제한적 관계절에서와 같이 최대성/전체성 효과를 보인다는 주장을 네 가지 반례를 들어 비판적으로 고찰하여 보았다. 지금까지의 논의가 맞다면, 한국어의 '것'과 이를 선행하는 절을 내핵관계절로 간주하고 그 속성으로 최대성/전체성효과를 보인다는 주장은 받아들여지기 힘든 것으로 보인다.

지금까지 우리는 '것'과 이를 선행하는 절을 한국어의 내핵관계절로 분석하는 문헌에서 주장하는 그 구문의 기본 속성을 비판적으로 고찰하여 보았다. 그 중 에서 이 절에서 우리가 타당성을 검증해 본 속성은 네 가지였다. 첫째는 한국어 의 내핵관계절은 무공백관계절이라는 주장, 둘째는 한국어의 내핵관계절은 무제 한 의존성을 보이지 않는다는 주장, 셋째는 내핵관계절과 외핵관계절은 동일한

¹² 의미적으로 제한적 관계절은 제한적으로 해석되어 두 개체의 집합의 교집합으로 해석이 되는 것이 일반적이다. 그러나 최준수 (2012)에 따르면 내핵관계절은 핵어명사의 지시체의 집합을 제한하는 것으로 해석되지 않는다고 주장한다. 그러면서 이러한 비교집합적 해석은 아마도 내핵관계절이 고유명사를 핵어명사로 취하지 못하는 현상과 관계가 있을 것으로 추측을 하고 있다.

진리조건적 의미를 가지고 있다는 주장, 그리고 마지막으로 넷째는 내핵관계절 은 비제한적 관계절로서 최대성/전체성 효과를 보여준다는 주장이었다.

이 절의 논의를 통하여 우리는 한국어에서 '것'과 이를 선행하는 절이 위에 논 의한 네 가지 기본 속성을 가지고 있는 내핵관계절이라는 주장은 전면적으로 재 검토가필요하다는 결론을 내릴 수 있다.

4 결론

한국어에 내핵관계절이 있는가 하는 질문은 매우 오래 전부터 제기되어 왔다. 이 질문은 특정 술어가 등장하는 문장의 특정 위치, 예를 들어 '잡다', '먹다' 등 대상 의미역을 하위범주화하는 완전 타동사의 목적어 자리나 비대격동사 등 일부 자동사의 주어 자리에 나온 '것'과 그 직접 선행절을 아울러서 한국어의 내핵관계절이라고 주장하기 시작한 1990년대 초반으로 거슬러 올라간다. 주류를 형성하고 있는 견해는 '것'과 그 직접 선행절을 한국어의 내핵관계절이라고 간주하는 것이다. 그러나 한국어의 내핵관계절에 대한 논의가 활발하게 진행되어 왔음에도 불구하고 아직 그 구성의 본질에 대한 의문에 답을 찾은 것보다는 아직 찾지 못한 것이 더 많은 것이 사실이다.

본 논문의 대상은 '것'과 관형형 어미로 연결과 그의 직접 선행절의 결합체이었다. 이 논문에서 우리는 이 결합체와 관련된 여러 질문들 가운데에서 가장 기본적이라고 할 수 있는 질문, 즉, '것'과 선행하는 절이 관형형 어미로 연결된 구문을 한국어의 내핵관계절로 볼 수 있는지 여부에 초점을 맞추어 논의를 진행했다.

본 논문에서 다룬 쟁점은 크게 두 가지이고 그 두 가지 쟁점에 대한 답을 찾는 것이 이 논문의 목적이었다. 하나는, 내핵관계절의 정의를 기존 문헌에서 찾아 대핵관계절의 정의적 속성에 한국어의 해당 구문이 부합하는지 검토하는 것이었다. 다른 하나는, 내핵관계절의 기본 속성으로 알려진 무공백 관계절, 무제한 의존성, 진리조건적 의미의 동일성, 비제한적 해석 혹은 최대성/전체성 효과를 내핵관계절로 간주된 구문이 과연 보여주는지 밝히는 것이었다. 본 논문에서 다룬 첫번째 쟁점과 관련해서 우리는 다음과 같은 결론에 도달했다.

일반적인 관계절과 내핵관계절의 정의에 따르면 한국어에서 내핵관계절로 알려진 구문은 내핵관계절로 볼 근거가 없다는 점이 드러났다. 내핵관계절은 관계절이어야 하고, 관계절은 종속절이어야 한다. 그러나 '것'과 관형형 어미로 연결된 선행절로 구성된 구문은 그 자체가 종속절이 아니고 종속절이 아니기에 관계절이라 할 수 없고 관계절이 아니기에 내핵관계절이라고 할 수 없다. '것'과 관형형 어미로 연결된 선행절로 구성된 구문은 그 자체로는 '명사화된 문장'이 아니라는 점, 매우 제약적인 분포를 보인다는 점, 그리고 피봇명사구가 명시적으로

실현되지 않아도 된다는 점에 비추어 한국어에서 내핵관계절로 알려진 구문은 기존 문헌에 제시된 내핵관계절의 정의에 부합하지 않음을 우리는 밝혔다. 본 논문의 두 번째 쟁점과 관련해서 우리는 내핵관계절의 기본 속성으로 알려진 여 러 현상 중에서 무공백 관계절, 무제한의존성, 진리조건적 의미의 동일성, 비제 한적 해석 혹은 최대성/전체성 효과를 문제가 되고 있는 내핵관계절로 알려진 구문이 보여주는지 새로운 자료들을 통해 검증하였다. 우선, 정의에 따라 내핵관 계절은 무공백관계절이어야 하지만, 내핵관계절로 간주된 구문도 절 내부에 공 백을 가정할 수 밖에 없는 문법적 현상을 허용한다는 점을 우리는 확인할 수 있 었다. 즉, 내핵관계절이라고 주장된 구문도 공백관계절일 수 있다는 사실을 새로 운 자료를 통해 밝혔다. 나아가서, 쟁점이 되고있는 한국어 구문은 무제한의존성 을 보여주지 않는다는 점, 내핵관계절로 알려진 구문과 해당 외핵관계절이 진리 조건적 의미의 동일성을 보여주지 않는다는 주장, 내핵관계절로 알려진 구문이 비제한적 해석 혹은 최대성/전체성 효과를 가지고 있다는 주장 모두 경험적 뒷 받침을 받지 못하고 있음이 명백해졌다. 두 번째 쟁점에 대한 논의 결과도 한국 어에서 '것'과 관형형 어미로 연결된 선행절로 구성된 구문을 내핵관계절로 간주 하는 견해는 재고를 할 필요성이 있다.

결론적으로 본 논문의 논의가 맞다면 한국어에서 '것'과 관형형 어미로 연결된 선행절로 구성된 결합체는 내핵관계절의 정의에도 부합하지 않고 기본 속성이라고 알려진 문법적 속성도 공유하고 있지 않다. 따라서 우리는 지금까지 알려진 한국어의 내핵관계절은 그릇된 명칭(misnomer)이라고 결론을 내릴 수 있으며, 이 구문에 관한 본격적인 분석은 원점에서 다시 시작해야 할 것으로 보인다. 기존의 분석은 문제가 되고 있는 구문이 내핵관계절이라는 전제 하에 제안된 것으로 이에 대한 대안의 제시는 당연히 이루어져야 하지만 이는 본 논문의목적이 아니고 지면의 제약이 있기에 후일의 연구로 남겨 둔다.

참고문헌

Boyle, John P.(2016), The Syntax and Semantics of Internally Headed Relative Clauses in Hidatsa, Catherine Rudin & Bryan J. Gordon (eds.), Advances in the Study of Siouan Languages and Linguistics, volume 10, 255 - 287, Berlin: Language Science Press.

Carlson, Greg N.(1977), Amount Relatives, Language 53(3), 520 - 542.

Cha, Jong-Yul(2005), Constraints on Clausal Complex Noun Phrases in Korean withFocus on the Gapless Relatice Clause Construction, Ph.D. thesis, University of Illinoisat Urbana-Champaign.

Chung, Chan & Jong-Bok Kim (2003), Differences between Externally and Internally Headed Relative Clause Constructions, On-line Proceedings of HPSG 2002, 3 - 25.

- Cole, Peter(1987), The Structure of Internally Headed Relative Clauses, Natural Language & Linguistic Theory 5(2), 277 302.
- Culy, Christopher Douglas(1991), The Syntax and Semantics of Internally Headed Relative Clauses., Ph.D. thesis, Stanford University.
- Demirdash, Hamida (1991), Resumptive Chains in Restrictive Relatives, Appositives, and Dislocation Structures, Ph.D. thesis, Massachusetts Institute of Technology.
- Grosu, Alexander(2000), Type-Resolution in Relative Constructions: Featural Marking and Dependency Encoding, The Syntax of Relative Clauses 32, 83.
- Grosu, Alexander(2010), The Status of the Internally-Headed Relatives of Japanese/Korean within the Typology of Definite Relatives, Journal of East Asian Linguistics19(3), 231 274.
- Grosu, Alexander & Koji Hoshi(2019), Japanese Internally-headed and Doubly-headed Relative Constructions, and a Comparison of Two Approaches, Glossa: A Journal of General Linguistics 32(1), 1 31.
- Grosu, Alexander & Fred Landman (2017), Amount Relatives, The Companion to Syntax1, 190 245.
- Hoshi, Koji(1995), Structural and Interpretive Aspects of Head-internal and Head-external Relative Clauses, University of Rochester.
- Jhang, Sea-Eun (1991), Internally Headed Relative Clause in Korean, S. Kuno, I.H. Lee,
- J. Whitman, Park S.-Y, Y.-S. Kang & Y.-J. Kim (eds.), Harvard Studies in Korean Linguistics, volume IV, 269 280, Seoul: Hanshin.
- Jhang, Sea-Eun (1994), Headed Nominalizations in Korean: Relative Clauses, Clefts, and Comparatives, Ph.D. thesis, Simon Frazer University.
- Jo, Mi-Jeung (2002), The Structure of Relative Clauses in Korean, Studies in Generative Grammar 12(1), 107 137.
- Jung, Yunsun (1995), Internally Headed Relative Clauses in Korean, Harvard Studies in Korean Linguistics VI, 235 248, Department of Linguistics, Harvard University.
- Keenan, Ed (1985), Relative Clauses, Tim Shopen (ed.), Language Typology and Syntactic Description. Volume II. Complex Constructions, 141 170, Cambridge: CambridgeUniversity Press.
- Kim, Jong-Bok (1999), Grammatical Interfaces in Korean Internally Headed Relative Clause Construction, Language Research 17, 257 275.
- Kim, Jong-Bok, Peter Sells & Jaehyung Yang (2006), A Computational Implementation of Internally Headed Relative Clause Constructions, Advances in Natural Language Processing, 24 31, Springer.
- Kim, Kwang-sup (2022), Korean IHRC has a Gap, Korean Journal of Linguistics 47(2), 261 288.
- Kim, Min-Joo (2004a), Event-Structure and the Internally-Headed Relative Clause Construction in Korean and Japanese, Ph.D. thesis, University of Massachusetts at Amherst.
- Kim, Min-Joo (2004b), Three Types of kes-Nominalization in Korean, Harvard

- Studies in Korean Linguistics 10, 479 492.
- Kim, Min-Joo (2007), Formal Linking in Internally Headed Relatives, Natural Language Semantics 15(4), 279 315.
- Kim, Nam-Kil(1984), The Grammar of Korean Complementation, volume 11, University of Hawaii Center for Korean.
- Kim, Yong-Beom (1996), Internally Headed Relative Clause Constructions in Korean, Proceedings of the 11th Pacific Asia Conference on Language, Information and Com putation, 403 413.
- Kim, Yong-Beom (2002), Relevancy in Internally Headed Relative Clauses in Korean, Lingua 112, 541 559.
- Kuroda, S-Y (1976), Headless Relative Clauses in Modern Japanese and the Relevancy Condition, Annual Meeting of the Berkeley Linguistics Society, volume 2, 269 279.
- Lee, Jeongrae (2006), The Korean Internally-Headed Relative Clause Construction: Its Morphological, Syntactic and Semantic Aspects, Ph.D. thesis, The University of Arizona.
- Park, Byung-Soo (1994), Modification vs. Complementation: the so-called Internally Headed Relative Clauses Reconsidered, Proceedings of the 1994 Kyoto Conference: A Festschrift for Professor Akira Ikeya, 41 48.
- Park, Chongwon (2021), Korean Relative Clauses: Metonymy, Zone Activation, and Reference Point, Studia Linguistica 1 40.
- Ryu, Byong-Rae (2014), Semantic Constraints on Multiple Case Marking in Korean, Anja Latrouite Doris Gerland, Christian Horn & Albert Ortmann (eds.), Meaning and Grammar of Nouns and Verbs, 77 - 112, Düsseldorf: düsseldorf university press.
- Shimoyama, Junko (1999), Internally Headed Relative Clauses in Japanese and E-type Anaphora, Journal of East Asian Linguistics 8(2), 147 182.
- de Vries, Mark (2002), The Syntax of Relativization, Ph.D. thesis, University of Amster dam.
- de Vries, Mark (2006), The Syntax of Appositive Relativization: On Specifying Coordination, False Free Relatives, and Promotion, Linguistic Inquiry 37(2), 229 270.
- de Vries, Mark (2018), Relative Clauses in Syntax, Oxford Research Encyclopedia of Linguistics, Oxford University Press.
- Watanabe, Akira (1992), Subjacency and S-Structure Movement of Wh-in-situ, Journalof East Asian linguistics 1(3), 255 291.
- Watanabe, Akira (2004), Parametrization of Quantificational Determiners and Head internal Relatives, Language and Linguistics 5(1), 59 97.
- Yang, Byung-Sun (1993), Clause and Information Structure of Korean Relative Clauses in Role and Reference grammar, M. Bernstein (ed.), Proceedings of the 1992 Eastern States Conference on Linguistics, volume 8, 282 293, Ithaca: Cornell University Press.
- Yang, In-Seok (1972), Korean Syntax: Case Markers, Delimiters, Complementation, and Relativization, University of Hawai'i at Manoa. Yoon, Hang-Jin (1991),

Functional Categories and Complementation in English, Koreanand Turkish Linguistics, Ph.D. thesis, University of Wisconsin-Madison.

김민국 (2010), 핵어 명사의 관계화 제약에 대한 연구, 한국어학 47, 131 - 162.

김종복, PeterSells, 양재형, 강우순(2010), 말뭉치 자료 조사를 통한 한국어 '것' 구문에 대한 통사·의미구조 분석 및 전산학적 구현, 한국연구재단(NRF) 연구보고서.

문숙영 (2012), 유형론적 관점에서 본 한국어 관계절의 몇 문제, 개신어문연구 35, 31 - 68.

연재훈 (2012), 유형론적 관점의 한국어 관계절 연구, 국어학 63, 413 - 457.

조수근 (2014), 한국어 내핵관계절의 존재 유무에 대한 연구, 언어정보 19, 189 - 205. 조수근 (2016), 구어 말뭉치 분석을 통한 한국어 내핵관계절 연구, 언어정보 23, 77 - 94.

최준수 (2012), 우리말 관형절 갈래와 생김새, 생성문법연구 22(2), 459 - 495.

한국영어학회(KASELL)-한국코퍼스언어학회(KACL)

2022년 공동 가을 학술대회

공동학술대회 주제 관련

- 1. 김원빈 (Univ. of Georgia, USA)
 - Application of Artificial Intelligence to Semantic Prosody Research
- 2. Dong Yu and Yaochen Deng (Dalian University of Foreign Languages, China)

Discourse Analysis of the Olympic Reports on U.S. Mainstream Media in Light of Digital Humanities

3. **전주현** (충남대)

빅데이터시대의 MZ세대를 대상으로 하는 영작문 교육의 딜레마

4. 노강산 (고려대)

Would You Like to Exercise or Work Out?: Taking a Collostructional Approach to Korean EFL Learners' Underuse of the Verb-Particle Constructions

5. 김창현 (한국전자통신연구원)

자동요약 학습코퍼스 구축을 위한 가이드라인

6. 김소희 (국립국어원)

4차 산업 혁명 대비 국어 빅데이터 사업

7.이규민, 이정현, 장서현, 송상헌 (고려대)

코퍼스와 기계학습을 통한 사회 현상의 범학문적 분석: 코로나19 사례를 중심으로

8. **황동진** (고려대)

한국어 언어 모델의 정보 구조 표상 능력: '은/는'의 관계절 내 형성 제약을 중심으로

9. **허진영** (한국교원대)

AI 팽톡의 초등영어교육 적용방안

Application of artificial intelligence to semantic prosody research

Wonbin KIM

University of Georgia, United States

ABSTRACT

This study aims to present a new method to analyze semantic prosody in an efficient way. Specifically, the semantic prosody of the Korean neologism \mathcal{H} is examined. The neologism \mathcal{H} - is slang and came from the existing prefix \mathcal{H} -. It is attached to adjectives or verbs which have negative meaning (e.g., 개짜증나다). However, interestingly, the combinations of the neologism and a positive word have been observed today (e.g., パ예쁘다). To check whether there is change in its semantic prosody, I investigated adjectives and verbs where the neologism *게*- is attached. For data collection, I used snscrape in Python to scrape about 35,000 tweets including ${\mathcal H}$ as a keyword per month from 2010 to 2019. For data preprocessing, I only extracted adjectives and verbs where \mathcal{H} is attached from tweets by means of KoNLPy. To sort those extracted adjectives and verbs into positive and negative, I employed Long short-term memory (LSTM). As training data to the LSTM model, I used KNU Korean Sentiment Lexicon, which is a Korean sentiment dictionary including a list of domainindependent sentiment words. After training the LSTM model, I utilized the model to classify the adjectives and verbs into positive and negative. To track change in the semantic prosody, for each year, I counted the token frequencies of the sorted positive and negative adjectives and verbs and added them up depending on the category. The result showed that the proportion which positive words account for in the token frequency has gradually increased over the ten years. This finding demonstrates that the semantic prosody of the neologism *게*is shifting from negative toward positive. This study is significant in that it has shown the possibility that artificial intelligence can be used as an efficient method to analyze the semantic prosody of a word for the first time.





Application of Artificial Intelligence to Semantic Prosody Research

Oct.22.2022

Wonbin Kim wk32657@uga.edu

Department of Linguistics University of Georgia

1

Motivation

[the existing prefix kay-]

- attached to nouns with negative meaning
- used to amplify the degree of the negative meaning from nouns
- 'severe' or 'heavy'

[the neologism kay-]

- attached to <u>adjectives or verbs with</u> <u>negative meaning</u>, <u>functioning as an</u> <u>adverb which means 'really' or 'very'</u>

- The combinations of the neologism and a positive word have been observed today.

2

Research Objective

This phenomenon suggests that

the semantic prosody of the neologism kay- is shifting from negative toward positive.

> **Semantic prosody**: a consistent aura of meaning with which a form is imbued by its collocates (Louw 1993)

e.g. largely + negative words/expressions broadly + positive words/expressions

My research aims to track

how the semantic prosody of the neologism kay- has changed over the past ten years from 2010 to 2019 using artificial intelligence.

3

Data Collection

Why Twitter data?



How collect the data?



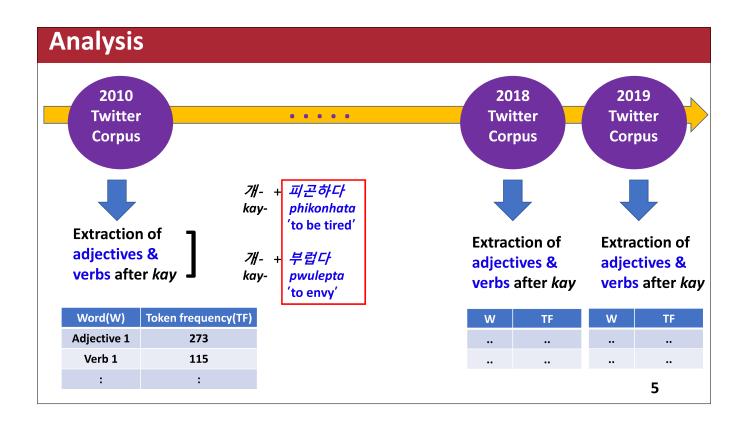
- used in informal settings.
- 2020).
- short & less refined sentences
- ungrammatical expressions & slang

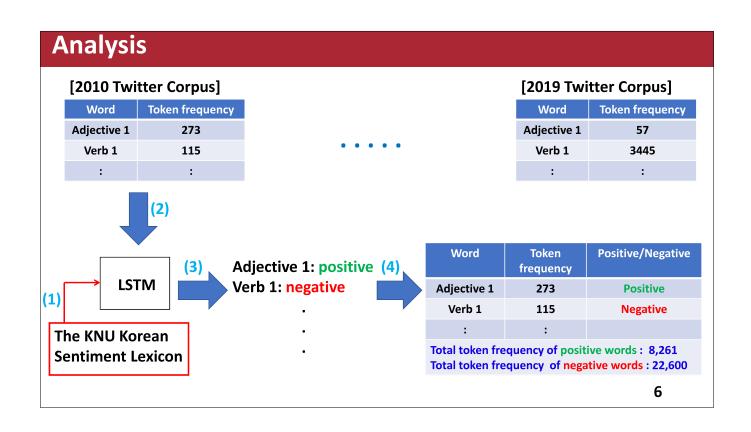
- The neologism kay- is slang and snscrape: a scraper that scrapes posts, hashtags, and so on from social networking services
- The Korean language used in With the aim of making a corpus consisting of Twitter is close to spoken Korean about 5 million morphemes per year, about rather than written Korean (Shi, 35,000 Korean tweets including kay as a keyword were scraped per month from 2010 to 2019 using snscrape.

Ten Twitter Corpora

2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
TC									

(TC: Twitter corpus)





KoNLPy

Konley: a Python package for natural language processing of the Korean language, including various Korean morpheme analyzer tools (e.g. Okt, Kkma, Komoran, Mecab, etc.)

>>> from konlpy.tag import Okt
>>> okt = Okt()

>>> print(okt.morphs(u'단독입찰보다 복수입찰의 경우'))
['단독', '입찰', '보다', '복수', '입찰', '의', '경우']

>>> print(okt.nouns(u'유일하게 항공기 체계 종합개발 경험을 갖고 있는 KAI는'))
['항공기', '체계', '종합', '개발', '경험']

>>> print(okt.phrases(u'날카로운 분석과 신뢰감 있는 진행으로'))
['날카로운 분석', '날카로운 분석과 신뢰감', '날카로운 분석과 신뢰감 있는 진행', '분석', '신뢰', '진행']

>>> print(okt.pos(u'이것도 되나욬ㅋㅋ', norm=True, stem=True))

Example code source: https://konlpy.org/ko/latest/api/konlpy.tag/#okt-class

7

Data Preprocessing_ Twitter Data

[('이', 'Determiner'), ('것', 'Noun'), ('도', 'Josa'), ('되다', 'Verb'), ('ㅋㅋ', 'KoreanParticle')]

Step	Data preprocessing
1	To reduce duplicate tweets to one tweet
2	To remove other things such as punctuation marks, numbers, and other foreign languages from each tweet
3	To reduce duplicate tweets to one tweet once again to filter the retweets which were not filtered in the first step
4	To split sentences in each tweet into eojuls* and only extract eojuls including kay from each tweet -> e.g. 저v사람v <mark>개나쁘네</mark> (eojul: a content word itself or the morphosyntactic combination of a content word and thing(s) in charge of grammatical function (e.g. particles and endings) Eojuls compose a Korean sentence and they are separated by spacing.)
5	To use KoNLPy* to convert adjectives and verbs after kay into their stems and extract them from the eojuls -> e.g. 개나쁘네 -> 나쁘다

8

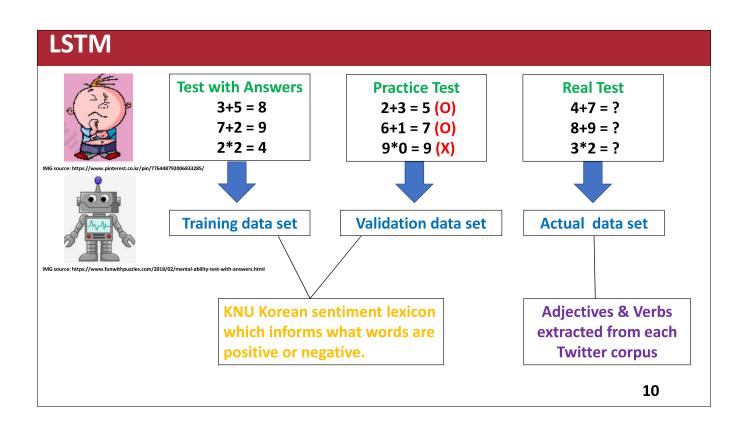
Data Preprocessing_KNU Korean Sentiment Lexicon

► KNU Korean Sentiment Lexicon: a Korean sentiment dictionary including a list of domainindependent sentiment words constructed by Park, Na, Choi, Lee, and On (2018)

All the words and phrases in the KNU sentiment lexicon are on <u>a five-level scale</u>: -2 (very negative), -1 (negative), 0 (neutral), 1 (positive), and 2 (very positive).

Step	Data preprocessing
1	To reduce those words and phrases to binary levels (i.e. negative and positive), excluding the neutral words and phrases (0: negative words & phrases / 1: positive words & phrases)
2	To divide every word and phrase into morphemes and make them stems using KoNLPy
3	To convert those consisting of stem morphemes to the sequences of integers based on frequency using Keras* (e.g. ['거래', '가', '금지', 되다'] → [11, 107, 5, 457]) (*Keras: a Python library for artificial neural networks)
4	To adjust the lengths of all of the sequences of integers so that every sequence can have the same length

9



Training LSTM KNU Korean sentiment lexicon (imbalanced) • # of Negative words & phrases: 9,795 VS # of Positive words & phrases: 4,839 • the ratio of negative (N) to positive (P) is about 2:1. (Test with Answers) (Practice Test) (Real Test) Actual data set **Training data set** Validation data set **KNU Korean sentiment lexicon Twitter data Training data set Validation data set** N:P = 2:1N:P = 2:111

rainin	g LSTM					
		Stratified k-fold cross-validation 1) Validation Accura 2) Confusion Matrix 3) Precision-Recall (onfusion Matrix	
Fold 1	Training data set N:P = 2:1	Training data set N:P = 2:1	Training data set N:P = 2:1	Training data N:P = 2:1		Validation data set N:P = 2:1
Fold 2	Training data set N:P = 2:1	Training data set N:P = 2:1	Training data set N:P = 2:1	Validation da N:P = 2:1		Training data set N:P = 2:1
Fold 3	Training data set N:P = 2:1	Training data set N:P = 2:1	Validation data set N:P = 2:1	Training data N:P = 2:1		Training data set N:P = 2:1
Fold 4	Training data set N:P = 2:1	Validation data set N:P = 2:1	Training data set N:P = 2:1	Training data N:P = 2:1		Training data set N:P = 2:1
Fold 5	Validation data set N:P = 2:1	Training data set N:P = 2:1	Training data set N:P = 2:1	Training data N:P = 2:1		Training data set N:P = 2:1
						12

Evaluation of Performance_Validation Accuracy

Fold	Validation accuracy
F1	0.9351
F2	0.9870
F3	0.9935
F4	0.9969
F5	0.9969
Averaged validation	0.98188
accuracy	

- Validation accuracy: how correctly the model can predict the validation data sets
- The averaged value means that the model predicts the validation data sets in about 98% accuracy on average.

Confusion Matrix

- Confusion Matrix: a specific table layout which represents the prediction results of classification

[Confusion matrix with two class labels]

		Predicted Values		
		Positive	Negative	
	Positive	True Positive	False Negative	
\ctual 'alues		(TP)	(FN)	
Act Val	Negative	False Positive	True Negative	
		(FP)	(TN)	

14

Evaluation of Performance_ Confusion Matrix

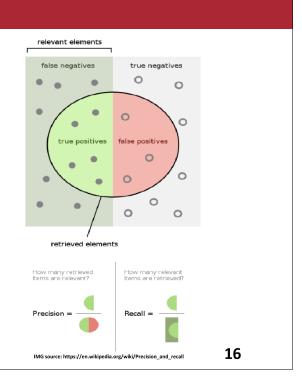


- Top left: the instances where the model predicted negative words as 'negative'
- Bottom right: the instances where the model predicted positive words as 'positive'.
- For the prediction of negative words, the probability of answering correctly is about 99% and for the prediction of positive words, the probability of answering correctly is about 96%.

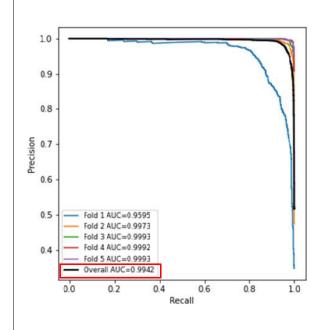
15

Precision-Recall Curve

- Precision-Recall Curve: a plot of the yaxis of precision against the x-axis of recall
- ▶ It shows the trade-off between precision and recall for a predictive model.
- Area under the precision-recall curve (AUC-PR): an indicator showing a model's performance
- ► The higher the AUC-PR value is, the higher both precision and recall are.



Evaluation of Performance_ AUC-PR



AUC-PR (Area Under the Precision-Recall Curve)

- Recall: how many positive words the model can detect out of the actual positive words
- Precision: how accurately the model can predict positive words
- The model can detect almost every positive word out of the actual positive words and predict positive words in almost 100% accuracy concerning the validation data sets.

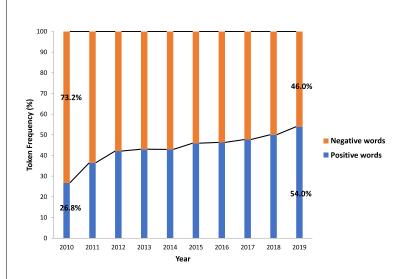
17

Results_Total Token Frequencies

Year	Total Token Frequency of Negative words	Total Token Frequency of Positive words
2010	22,600	8,261
2011	28,815	16,565
2012	32,513	23,704
2013	30,022	22,674
2014	29,508	22,227
2015	41,714	35,500
2016	44,399	38,297
2017	53,922	49,242
2018	60,741	60,899
2019	66,245	77,642

18

Results_Change in Semantic Prosody



The proportion which positive words account for has gradually increased over the last ten years.



The semantic prosody of the neologism *kay*- is <u>shifting from negative toward positive</u>.

19

Conculusion & Significance

- ► This study is the first study using Al and Korean social media data to track change in the semantic prosody of a Korean neologism over time.
- ▶ This study has confirmed the possibility that AI can be used to analyze the semantic prosody of a word efficiently.
- ▶ The results of this study support usage-based linguistics by demonstrating that the meaning of a lexical item is determined by how people use the lexical item.
- ▶ The new method proposed in this study will be able to be applied to the field of language learning as well as linguistics.

20



Topic Model Analysis of the Olympic Reports on U.S. Mainstream Media

Deng, Yaochen; Yu, Dong (DUFL)

2022.10.3

ABSTRACT: Taking the Associated Press (AP) as an example, this study aims to uncover the implicit topics, the diachronic distributional features of the topic model and the hidden stances and attitudes in the Olympic reports on mainstream U.S. media. Combining the methods of topic modeling (TM) and critical discourse analysis (CDA), the study trained and analyzed topic models based on a corpus of AP's Olympic reports from 2012 to 2022, and produced some useful findings. First, the topics clustered into two major categories according to the diachronic distributional structure of the model. Second, the topic model reflected significant attitudinal differences among different news coverage object. The results of the study offered a promising direction of combining the qualitative and quantitative perspectives in the analysis of media discourse.





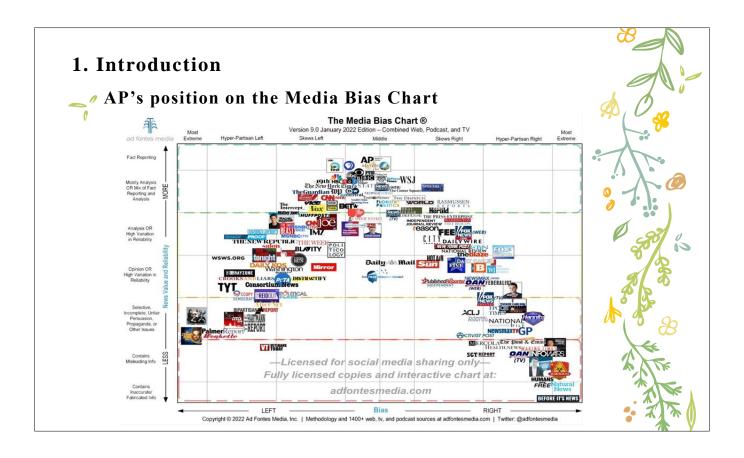
- 1. Introduction
- 2. Methodology
- 3. Results and Discussion
- 4. Conclusion





- The importance and representativeness of Olympic reports as a kind of public discourse
- + The Associated Press as the representative agency of U.S. mainstream media
- + Research object: Olympic report discourse on AP (2012-2022)



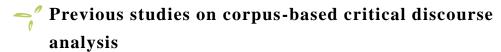




Previous studies on the Olympic reports

- + Since 2008
- + Framework of media reports on the Olympic Games
- + The construction of the national image
- + Lack of attention





- + Qualitative analytical approach in CDA
- + The number of texts selected for critical discourse analysis studies is too small and unrepresentative, and the interpretation of discourse lacks objectivity and systematicity (Fowler 1996)
- + (Krishnamurthy 1996; Baker 2008; Grundmann & Krishnamurthy 2010; Jaworska & Krishnamurthy 2012...)
- + TM as a new approach in combining the qualitative and quantitative perspectives in the analysis of media



2. Methodology



Research objectives:

- + Objective 1: topic mining
- + Objective 2: uncovering the diachronic distributional features of the topic model.
- + Objective 3: uncovering the implicit stances and attitudes in AP's Olympic report discourse.



2. Methodology



Methods:

- + LDA and Topic Modeling
- + MALLET(NLP toolkit) as the TM instrument
- + The three-dimensional analytical framework of Discourse-historical approach (Wodak 2009)



2. Methodology



Data set:

- + Data source: Factiva, a Dow Jones news database
- + Time span: 2012-2022 (6 Olympic events included)
- + Standard of collection: The top 100 reports (in order of relevance) of each Olympic Games
- + Corpus size: 180,000 words

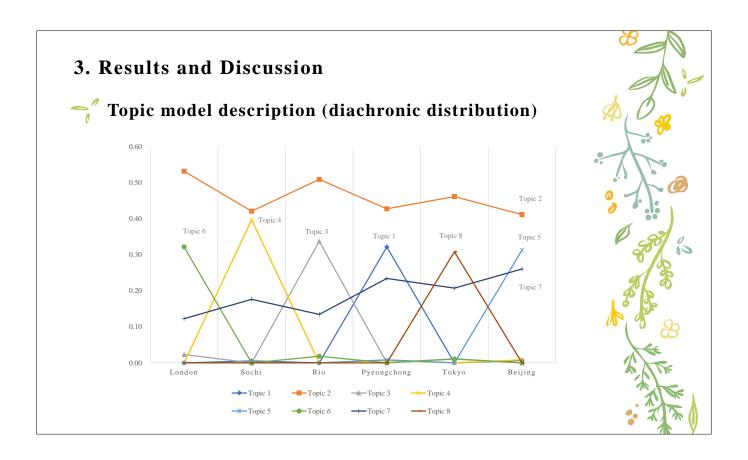


2. Methodology

Research procedures:

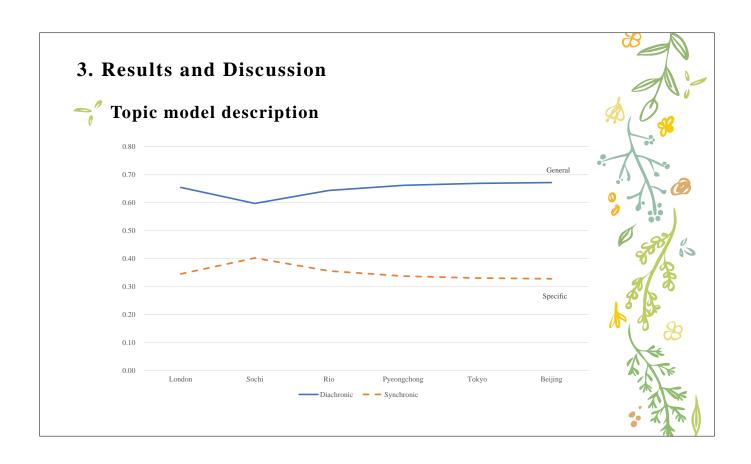
- + Phase 1: text cleaning and data type conversion
- + Phase 2: topic model training (10 trials with the variation of topic number 6-15)
- + Phase 3: finalizing the model (8) by the comparison of the readability and clustering tendency of each trial
- + Phase 4: interpreting the model (keywords and diachronic distribution)
- + Phase 5: topic analysis based on Discourse-historical approach (DHA)

3. Results and Discussion Topic model description (keyword clusters) Table 1 Top 20 words assigned to each topic in AP's Olympic reports Weight← Key Words⊖ 0.15715€ Pyeongchang north korea korean south kim ski sochi winter slalom luge skiing russian cross-country alpine halfpipe runs run bobsled hirscher← 6.49057€ olympic gold games medal olympics world won team women u.s. win final medals men time bronze day years back silver€ 0.194666 rio bolt brazil lochte meters police brazilian neymar water track janeiro field swimmers meter phelps match 0.16192€ sochi russia games hockey russian canada winter security putin goal bach ice finland dutch game skating vancouver sweden alpine slalom← 0.124356 beijing valieva ice russian winter china skating doping figure humphries bobsled jackson hockey meyers skaters skate skiing skater taylor ski← 0.14401€ london britain british stadium boxing beijing meter china australia bolt phelps park chinese field badminton pool rogge swimming organizers volleyball€ 3.091886 olympics event joc sports athletes team russia time sport lot days athlete compete competition events canada. year-old seconds end podium← 0.05463€ tokyo japan biles japanese pandemic camara rio tsimanouskaya covid belarus dressel health mental china pool rugby djokovic sevens osaka match



Topic model description + Category 1: General Topics + Category 2: Specific Topics

3. Results and Discussion



3. Results and Discussion



Topic model description

- + General topics:
- + Topic 2—Team USA
- + Topic 7—Competition
- + Outstanding performance and achievements
- + Women's roles and gender equality



3. Results and Discussion



Topic model description

- + Specific topics:
- + Topic 1—Pyeongchang 2018
- + Topic 3—Rio 2016
- + Topic 4—Sochi 2014
- + Topic 5—Beijing 2022
- + Topic 6—London 2012
- + Topic 8—Tokyo 2020



3. Results and Discussion

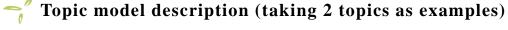


Topic model description

- + Specific topics in chronological order:
- + Topic 6—London 2012
- + Topic 4—Sochi 2014
- + Topic 3—Rio 2016
- + Topic 1—Pyeongchang 2018
- + Topic 8—Tokyo 2020
- + Topic 5—Beijing 2022



3. Results and Discussion



- + Topic 6—London 2012
- + "bolt" and "phelps"; "beijing" and "chinese"
- + Topic 4—Sochi 2014
- + "putin"; "Putin Games"

4. Conclusion

+ First, the topics clustered into two major categories (GTs and STs) according to the diachronic distributional structure of the model.

- + Second, the topic model reflected significant attitudinal differences among different news coverage object.
- + Third, the U.S. was portrayed as a sports powerhouse with a developed sports industry and high-level athletes in AP's discourse.



Pedagogical Approaches Based on Analysis of EFL College Students' English Academic Writing Assisted by AI-based Translating Tools

- Pedagogical Dilemmas of Teaching Academic English Writing to Generation MZ in Big Data Era -

> Presented by Joo Hyun Chun Instructor, Chungnam National University

> > **Contents**

Contents

- 1. Research questions
- 2. Literature review
- 3. Methodology
- 4. Findings
- 5. Conclusions

Research Motivation

Research questions; dilemmas

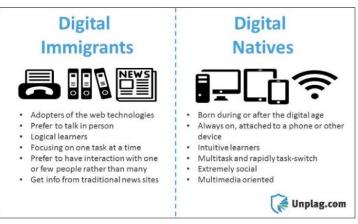
- 1. Who do we teach?
- 2. What era do we live?
- 3. How have AI translators evolved?
- 4. Let them use them or not?
- 5. Possible to acquire authentic academic English writing?
- 6. Unfortunately, students don't know what's wrong with their writing...

Research Motivation

Who do we teach? Generation MZ? No! Gen Z



Digital Immigrant vs Digital Native



Difference between 'Digital Native' and 'Digital Immigrant'

Source: Closing The Gap Between 'Digital Native' and 'Digital Immigrant' | by Mia Carina Friska | Medium

Literature review

Who do we teach? Generation MZ, No! Gen Z

Information accessibility



Generation Z consists of people born between 1997 and 2012.

Generation - Wikipedia

Source: http://www.koreanlii.or.kr/w/index.php/File:MZers_bigdata.png

Literature review

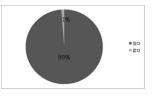
What era do we live? Post Covid19

- Digital transformation in educational environment
- "We cannot return to the world as it was before." by UNESCO
- The 2030 Agenda for Sustainable Development (UNESCO in 2019) the International Commission on the Futures of Education presenting nine(9) ideas for concrete actions today that will advance future education
- Challenges and opportunities of English education

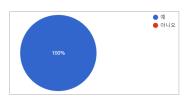
Literature review

What era do we live? Post Covid19

- Survey results from H. University by Chun(2022)
- 104 students of College English course leaning academic English writing
- Question: do you use a translator when writing your English essays?







〈그림 3〉 자동번역기 사용유무

(그림 4) 사용한탁기 사용복색(다음립) Lim (2017)

Chun (2022)

What era do we live? Post Covid19

Literature review

- Kim(2021) suggests a new paradigm of teaching foreign languages for improving digital literacy.
- 5 Steps for Teaching Principles (Kim, 2021)
- 1. Providing various language inputs using digital-based foreign language learning materials and programs for learners to become familiar with real English
- 2. Educating search ability, critical evaluation, and comprehension strategies to select and analyze digital language data for students' autonomous learning
- 3. Improving linguistic and cultural interaction skills by providing various online English communication opportunities
- 4. Developing utilizing skills and production skills to apply technology and web resources to learn English
- 5. Improving students' comprehensive language skills by enabling them to perform technology-based assignments, presentations, and creative problem-solving projects in class

 Kim (2021, p.96)

Methodology

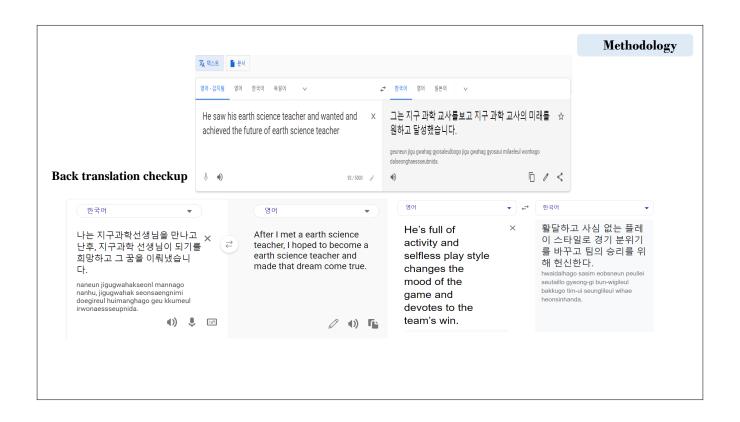
Corpus

- A type of corpus learner corpora, a collection of the writing acquiring English as foreign language
- Cross sectional or longitudinal analysis
- Longman Learners' Corpus 10 million words
- The Cambridge Learner Corpus (CLC) 20 million words
- HKUST Corpus of Learner English -10 million words
- How to design the learner corpus from College English students' essays
- DIY corpora, which is specialized corpus to cover a very small size of data

Methodology

Examples of students' writing using translators

- She told "Didn't give up, then continued to try" on many shows.
- Make an effort on their talent is more difficult than have a talent.
- After acne appeared on the skin, the ingredients of cosmetics were determined.
- Becoming researcher is hard.
- Becoming an architect is my dream.
- My career path is computer engineer.
- The successful person in my opinion is Lee Kun-hee.
- I think that BTS is a successful person. He comforted many fans by making the word I purple you.



Findings & results

How have AI translators evolved?

- Google translator
- a. 어제 시장을 봤다.
- b. I saw the market yesterday. (Papago; Google; SYSTRAN)
- Naver Papago
- c. 어제 시장 보러 갔다.
- Kakao i translator
- d. I went to the market yesterday. (Papago; Google; SYSTRAN)
- Bing translator
- e. 어제 장보러 갔다.
- Willi.ai translator
- f. I went grocery shopping yesterday. (Papago)
- SYSTRAN Translate g. I went shopping yesterday. (Google; SYSTRAN)

Findings & results

Would allow them to use translator or not?

- a. 나는 농구는 모른다. 하지만 조던은 안다.
- b. I don't know basketball. But Jordan knows.
- c. 방탄소년단은 성공한 사람이라고 생각한다. 많은 팬을 위로했다.
- d. I think that BTS is a successful person. He comforted many fans.
- e. 나는 바가지를 이용해서 목욕했다.
- f. I took a bath using a rip-off.

Findings & results

Semantic and syntactic problems

- <u>He saw</u> his earth science teacher and <u>wanted and achieved</u> the future of earth science teacher. (No subject)
- I think that BTS is a successful person. <u>He</u> comforted many fans by making the word I purple you. (Pronoun)
- He's full of activity and <u>selfless play style changes</u> the mood of the game and devotes to the team's win.

Findings & results

Common pitfalls of machine translators

- Structural ambiguity Woman without her, man is nothing. Black taxi drivers
- Context awareness
- Cannot overcome the difference btw language structure
- World English and new words
- Dialects
- Acceptable essay organizations depending on language, people's logic, and culture difference

Conclusions

Not yet!

"The difference between the right word and the almost right word is really a large matter — it's the difference between lightning and a lightning bug."

- Mark Twain -

References

- Allen, J. (2003). Post-editing. In H. Somers (Ed.), Computers and translation: A translator's guide (pp. 297-317). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kim, Haeyoung (2021). The future direction and challenges of English education in the post-COVID-19 era. English Teaching, English Teaching, vol. 76 S1 (2021), p. 87-105
- ImHeeJoo 2017 The university students' perceptions or attitudes on the use of the English automatic translation in a general English class Based on English writing lessons
- Jeong, Namsuk (2021). A study on the effects of machine translators on college students' writing proficiency and affective attitude. Multimedia Assisted Language Learning, 24.1, 134-157.
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge? Contemporary Issues in Technology and Teacher Education, 9(1), 60-70.
- Yi Liu and 김양희. (2019). College English Writing Instruction in the Big Data Era From the Process Approach Perspective. 언어학 연구, 53, 161-180.
- "[문해력 리포트] ② 한국 청소년 '디지털 문해력'마저…OECD 바닥권." 연합뉴스. 2021년 12월 18일 수정, 2022년 1월 17일 접속, https://www.yna.co.kr/view/AKR20211209134600501
- "인공신경망 VS 통계기반 번역, 뭐가 다를까?"지디넷코리아. 2016년 12월 25일 수정, 2022년 1월 10일 접속, https://zdnet.co.kr/view/?no=20161223190944. Statista. (2022, January 26). Retrieved February 7, 2022, from https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/
- 과과고 https://papago.naver.com/?tk=en
- 구글번역 <u>https://translate.google.co.kr/?hl=k</u>
- UNESCO Education in a post-COVID world: https://en.unesco.org/sites/default/files/education_in_a_post-covid_world-nine_ideas_for_public_action.pdf

발표자 정보

Presentation Language: Korean

Presenter (Korean / English): 전주현 / Joo Hyun Chun

Affiliation: Chungnam National University, Republic of Korea

Presentation Title: Instructor

Email address: joochun@cnu.ac.kr

- Mn	-
------	---



Would You Like to
Exercise or Work Out?
Taking a Collostructional
Approach to Korean EFL
Learners' Underuse of
the Verb-Particle
Constructions

Noh Kang San Korea University

Contents

- 1. Introduction
- 2. Background
- 3. Methodology
- 4. Results
- 5. Implications

The present study aims to detect Korean EFL learners' underuse of the English verb-particle construction, which is also known as English phrasal verbs.

- (1) a. Do not **disappoint** me.
 - b. Do not let me down.

Based on the theoretical grounds of *Construction Grammar*, this study takes a constructionist approach to language and its acquisition.

1. Introduction

The primary purpose of this thesis is to observe Korean EFL learners' underuse of the English verb-particle construction by using L2 learner corpora.

That is, will Korean EFL learners avoid the English verb-particle construction and instead prefer single-word verbs?

Then follows the two questions.

One concerns whether Korean EFL learners prefer the order of [V-Prt-OBJ] construction over that of [V-OBJ-Prt].

The other concerns whether Korean EFL learners prefer the verb-particle constructions with literal meanings over those with figurative meanings.

2. Background

The English verb-particle constructions, often referred to as *phrasal verbs*, have been defined as one of the multi-word verb constructions along with prepositional verbs, phrasal-prepositional verbs, and other multi-word verb constructions (Biber et al., 2002).

Phrasal verbs consist of a verb followed by an adverbial particle.

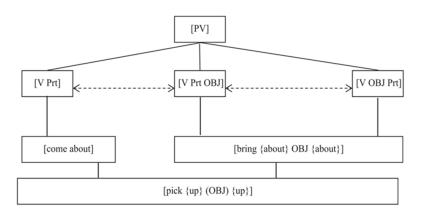
(Biber et al., 2002)

2. Background

Based on the construction network of phrasal verbs (from Gilquin 2015: 57), this study classifies the verb-particle constructions into three levels:

- [1] The Super-construction of [PV]
- [2] Three Sub-constructions: [V Prt], [V Prt OBJ], [V OBJ Prt]
- [3] The Concrete Instantiations of [2]

2. Background



3. Methodology

O Distinctive Collexeme Analysis

Distinctive collexeme analysis is performed to measure the preference of a lexeme to a specific construction over a semantically or functionally similar counterpart (Gries and Stefanowitsch 2004).

That is, it is to identify a lexeme that shows a strong attraction to one member of the pair as opposed to the other, thereby identifying subtle distributional differences between the two.

 \rightarrow Thus, it can be applied to a specific linguistic phenomenon known as grammatical alternation.

3. Methodology

◎ The alternation between [V-OBJ-Prt] and [V-Prt-OBJ]

The verb-particle constructions can undergo the process of *alternation* by changing the position of an object.

Distinctive collexeme analysis is applied in such cases of alternation, as in (2):

(2) a. John picked up the book. (Gries and Stefanowitsch 2004: 110)

b. John **picked** the book **up**.

3. Methodology

O Performing Distinctive Collexeme Analysis

The two-by-two table is required to perform distinctive collexme analysis.

The p-value of *make up* is 4.20E-12.

expected frequency

	make up	make up ☐ make up				
[V – Prt – (OBJ)]	652 (603.47)	2,668	3,320			
[V - OBJ - Prt]	18 (66.53)	348	366			
Column Totals	670	3,016	3,686			

3. Methodology

As for learner corpora, the two Korean EFL learner corpora were used: the Gachon Learner Corpus (GLC) and the Yonsei English Learner Corpus (YELC).

The GLC is composed of 2,507,899 words and the YELC is composed of 1,082,295 words, respectively.

The list of verb-particle constructions is based on the top 100 phrasal verb lemmas in BNC by Gardner and Davies (2007).

4. Results

Q1: Will Korean EFL learners avoid the English verb-particle construction and instead prefer single-word verbs?

A1: From the selected 90 verb-particle constructions, 73 of them were underused compared to their single-word counterparts. Only 17 verb-particle constructions were more used than single-word counterparts.

GLC + YELC						
Verb-Particle Constructions	Freq	Single-Word Verbs	Freq			
go on	564	continue	399			
pick up	169	lift	70			
go out	469	exit	72			
find out	122	discover	91			
make up	670	fabricate	2			
go down	107	descend	10			
give up	327	abandon	62			
get out	177	exit	72			
come down	42	descend	10			
break down	134	collapse	24			
take back	19	repossess	0			
break up	45	disperse	10			
break out	52	erupt	5			
put in	74	insert	26			
put off	20	postpone	3			
come through	16	penetrate	14			
come off	4	detach	3			

<The 17 Verb-Particle Constructions More Used than Single-Word Counterparts>

4. Results

Q2: Do Korean EFL learners prefer the order of [V-Prt-OBJ] construction over that of [V-OBJ-Prt]?

A2: From the selected 53 transitive verb-particle constructions, 38 of them turned out to be distinctive for the [V-Prt-OBJ] construction. The remaining 15 were distinctive for the [V-OBJ-Prt] construction.

→ Korean EFL learners preferred the [V-Prt-OBJ] construction.

V - Prt - (OBJ)	p.value	V - Prt - (OBJ)	p.value	
make up	make up 4.20E-12		0.1961	
take out	7.53E-06	go over	0.3186	
give up	1.35E-05	pick out	0.3186	
go through	2.26E-05	turn down	0.3336	
break down	9.03E-05	carry out	0.3436	
find out	0.0001918	carry on	0.4157	
put on	0.0008229	bring in	0.4575	
get on	0.003829	hold up	0.4575	
bring about	0.004728	turn over	0.4575	
break up	0.02502	pick up	0.464	
bring up	0.0281	set out	0.5064	
take on	0.05177	hold out	0.5064	
set up	0.06304	break off	0.5064	
look up	0.1027	bring down	0.6386	
get through	0.1027	take up	0.7269	
put off	0.1365	set about	0.7398	
give in	0.1468	take over	0.787	
come through	0.1832	give out	0.787	
put in	0.1886	turn off	0.9223	

The 38 Verb-Particle Constructions Distinctive for the [V-Prt-OBJ] Construction

V - OBJ - Prt	p.value
give back	p-value < 2.2e-16
put back	1.67E-09
get back	1.82E-08
take down	0.0001794
make out	0.0005689
bring back	0.001569
put up	0.002993
take off	0.1381
put down	0.1536
hold back	0.1751
put out	0.1913
take back	0.1913
set off	0.3133
take in	0.4773
bring out	0.9305

The 15 Verb-Particle Constructions Distinctive for the [V-OBJ-Prt] Construction

4. Results

Q3: Do Korean EFL learners prefer the verb-particle constructions with literal meanings over those with figurative meanings?

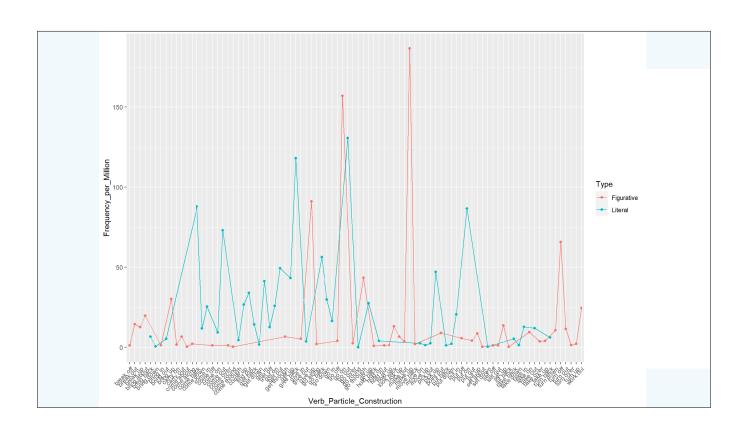
[1] Literal: pick up, go down, etc.

[2] Figurative: carry out, bring about, etc.

[3] Completive: break down, take off, etc.

There are 41 literal, 47 figurative, and 2 completive verb-particle constructions in total.

A3: It turned out that Korean EFL learners generally prefer literal verb-particle constructions over figurative ones.



5. Implications

The analysis of the two Korean EFL learner corpora shows that Korean EFL learners generally prefer single-word verbs over verb-particle constructions.

In addition, the corpus data shows the following two results: (1) they prefer [V-Prt-OBJ] construction over [V-OBJ-Prt] construction and (2) literal verb-particle constructions over figurative ones.

The current study revealed how Korean EFL learners use the English verb-particle constructions by taking a constructionist approach and applying distinctive collexeme analysis.

5. Implications

The current study also shows the need to apply learner corpus data into the research of L2 learners' over/underuse behavior.

The application of learner corpora makes it possible for L2 instructors to provide the learner-specific instruction geared towards a certain group or level of L2 learners.

L2 instructors might make use of L1 corpus data (e.g., BNC or COCA) to show L2 learners how to put certain constructions or structures into use like native speakers.

Reference

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). London: Longman.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL quarterly*, 41(2), 339-359.

Gilquin, G. (2015). The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory*, 11(1), 51-88.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective onalternations'. *International journal of corpus linguistics*, *9*(1), 97-129.

Thank you for listening!

자동요약 학습코퍼스 구축을 위한 가이드라인

김창현(한국전자통신연구원, chkim@etri.re.kr)

자동요약은 많은 내용을 포함하는 문서/대화 컨텐츠를 핵심내용 위주로 짧게 다시 작성하는 기술로써, 디지털 컨텐츠의 생산 속도가 너무나 빨라 소비가 따라갈 수 없는 요즘 상황에서 그 필요성이 더욱 커지고 있는 기술이다. 최근 딥러닝 기술을 기반으로 자동 요약 기술의 품질이 많은 향상을 보이고 있으나, 이러한 딥러닝을 이용하기 위해서는 많은 학습 데이터가 필요하다. 영어권에서는 최근 자동요약을 위한 학습용 데이터가 활발히 구축되고 있으며, 이를 기반으로 한 실험결과들이 보고되고 있다. 자동 요약용 학습데이터 구축은 난이도가 높은 작업이며, 시간과 비용 또한 많이 든다. 또한 구축 작업자별, 관점의 차이로 인해 요약 결과도 상당히 다른 결과를 보인다. 한국어를 위한 자동요약 학습용 데이터는 이제 막 구축 노력을 기울이는 초기 단계이며, 이러한 측면에서, 자동요약 학습용 데이터 구축을 위한 최소한의 일정한 가이드라인 필요하다. 본 논문에서는 자동요약을 위한 최소한의 가이드라인을 제시함으로써 향후 이루어질 자동요약 데이터 구축의 활성화 및 토대를 마련하고자 한다.

국립국어워 4차 산업혁명 대비 국어 빅데이터 구축 사업의 현황과 전맛1)

김소희(국립국어위)

국립국어원은 4차 산업혁명 시대 국가 공공재로서 국어 말뭉치 구축이 필요함에 따라 2018년부터 '4차 산업혁명 대비 국어 빅데이터(말뭉치) 구축'사업을 추진하고 있다. 이 사업의 주요 내용은 인공 지능이 최신의 언어 자원을 학습할 수 있도록 신문 기사, 일상 대화, 메신저 대화 등 다양한 언어 환경을 반영하는 언어 자료를 매년 수집하여 원시 말뭉치로 구축하고 고도화된 인공 지능 언어처리 기술 개발에 대응하여 다양한 분석 정보를 부착하여 국가 공공재로서 민간에 공개하고 활용을 장려함으로써 한국어 인공 지능 기술 발전의기반을 마련하는 것이다.

국립국어원은 "모두의 말뭉치(https://corpus.korean.go.kr)" 개통과 함께 2020년 8월 18억어절 규모의 말뭉치를 공개한 이래, 신규로 구축한 말뭉치를 매년 분기별로 공개하고 있으며, 2022년 9월 현재 34종 21억 어절을 공개하고 있다. 국립국어원이 구축한 말뭉치는 대기업, 중소기업, 새싹기업 등 산업계에서 언어 모델 개발, 챗봇 개발, 띄어쓰기 교정 등 다양한 기술을 개발하는 데 주로 활용되고 있으며, 말뭉치를 이용한 언어학 연구 자료로도 이용되고 있다.

한편, 국립국어원은 말뭉치가 인공 지능 기술 개발과 언어 연구의 발전을 견인할 수 있도록 다양한 활용 방안을 모색하고 있다. 2021년에 2009년부터 매년 개최해 온 '국어정보처리시스템 경진대회'를 한국어 인공 지능 기술 환경에 맞추어 '인공 지능 언어 능력 평가대회'로 개편하고 공통의 과제와 평가 지표를 활용하여 성능을 평가할 수 있는 순위표(리더 보드) 체계를 도입하였다. 올해부터는 말뭉치를 이용한 언어 연구에 관심 있는 이용자들을 위한 온라인 강좌 및 특별 강연 등을 차례로 공개하고 있다.

국립국어원은 앞으로 변화하는 기술 환경과 산업 수요에 부합하는 인공 지능 학습 및 성능 평가용 말뭉치 구축을 확대하는 동시에 한국어-외국어 병렬 말뭉치, 글쓰기 자동 채점시스템 개발을 위한 언어 자료 등 어문 연구 및 정책 수행 기관으로서 국어원의 전문성을살릴 수 있는 언어 자원을 연구하고 구축할 계획이다. 또한 '인공 지능 언어 능력 평가 체계'를 상시로 운영하고 한국어 인공 지능 분야의 평가 과제들을 누적하여 공개함으로써 연구 개발자들이 인공 지능의 한국어 이해 및 생성 능력을 다양한 방면에서 검증하고 성과를 공유하는 장을 마련할 계획이다.

¹⁾ 이 발표 논문은 국립국어원이 추진 중인 '4차 산업혁명 대비 국어 빅데이터 구축' 사업에 현황과 전망에 관한 내용으로, 관련 정책 보고서와 연구 보고서의 내용을 상당 부분 포함하였다.

1. 4차 산업혁명 대비 국어 빅데이터 구축 사업

국립국어원은 4차 산업혁명 시대 국가 공공재로서 국어 말뭉치 구축의 필요에 따라 2018년부터 '4차 산업혁명 대비 국어 빅데이터(말뭉치) 구축'사업을 추진하고 있다. 이사업의 주요 내용은 인공 지능이 최신의 언어 자원을 학습할 수 있도록 신문 기사, 일상대화, 메신저 대화 등 다양한 언어 환경을 반영하는 언어 자료를 매년 수집하여 원시 말뭉치로 구축하고 고도화된 인공 지능 언어처리 기술 개발에 대응하여 어휘 의미 분석, 함의 분석 등 다양한 분석 정보를 부착하여 국가 공공재로서 민간에 공개하고 활용을 장려함으로써 한국어 인공 지능 기술 발전의 기반을 마련하는 것이다.

'21세기 세종계획'(1998년~2007년)을 통해 2억 어절가량의 말뭉치가 구축되었으나, 인공 지능 언어 처리 시스템 개발을 위한 기초 자원으로 활용되기에는 매우 적은 규모였으며, 분석의 층위도 다양하지 않았다. 국립국어원은 '4차 산업혁명 대비 국어 빅데이터 구축' 사업을 통해 4차 산업혁명 시대의 도래와 함께 급속도로 발전하고 있는 인공 지능 언어 처리 기술의 기본이 되는 데이터(말뭉치)를 연구하고 구축하며 민간에서 다양한목적으로 활용할 수 있도록 지원하고 있다. 이 논문에서는 국립국어원에서 이 사업을 통해 구축한 말뭉치들을 소개하고, 이 말뭉치들의 사용 방법과 활용 사례를 살펴본다. 특히현재 국립국어원에서 추진 중인 인공 지능 언어 능력 평가 체계에 대해 간략히 알아볼것이다. 마지막으로 '4차 산업혁명 대비 국어 빅데이터 구축' 사업의 향후 방향에 대해언급하는 것으로 이 글을 마치고자 한다.

2. 사업 추진 현황

가. 말뭉치 구축 현황

1) 원시 말뭉치 구축

국립국어원은 신문 기사 및 일상 대화 언어 자료를 매년 수집하여 원시 말뭉치로 구축하고 있다. 원시 말뭉치는 언어 자료 원문을 수집한 후 컴퓨터가 읽어 들일 수 있는 형태로 처리하여 구축된다. 그러나 단순히 자료를 입력하는 것으로 원시 말뭉치를 구축하는 것은 아니며, 메타 정보를 부착하고, 민간에서 활용할 수 있도록 개인 정보를 삭제하고 다른 정보로 대체하는 등의 처리 작업이 필요하다. 그리고 무엇보다도 국립국어원이 구축한 말뭉치를 언어 연구 또는 인공 지능 기술 개발 및 상용화를 위한 목적으로 활용할 수 있도록 저작권을 해결하는 과정이 필요하다. 이렇게 구축한 대량의 원시 말뭉치는 인공지능 언어 모델의 입력으로 활용된다.

국립국어원의 신문 기사 수집 목표 어절 수는 매년 1억 어절 규모이며, 월별 수집 규모, 분야별 수집 규모, 매체 선정 등에 대하여 사전에 검토하고 결정한다. 이러한 결정 사항에 따라 자료를 수집하고 원문을 수집할 때에는 저작권자와 이용 허락 계약을 체결한다. 이는 원문의 저작권을 국어원에 모두 양도한다는 의미는 아니며, 특정 목적으로만 활용할 수 있도록 저작권자에게 허락을 받는 절차이다. 이렇게 자료가 수집되고 나면, 미리정의한 정제와 분할 수준에 따라 참조할 수 있는 기본 정보를 부착하고 자료를 분할하여컴퓨터가 처리할 수 있도록 원시 말뭉치를 구축한다. 특히 말뭉치의 활용성을 높이기 위해 기본적으로 문단 분할을 수행한다. 한편, 신문 기사를 원시 말뭉치로 구축할 때에는 중복 기사나 유사 기사를 제거하는 과정과 작업자들이 직접 기사들을 읽어 가면서 문장으

로 볼 수 없는 기사 내용과 본문에 불필요한 내용을 삭제하는 작업과 저작권 이용 허락범위에 해당하지 않는 기사들을 제거하는 작업들도 필요하다.

국립국어원 일상 대화 말뭉치는 화자들의 대화 상황을 녹음하고 녹음 내용을 전사하는 과정을 통해 구축되는 구어 말뭉치의 한 종류이다. 대화 자료는 수집의 난도가 높아 학계와 산업계에서 요구가 더 높다. 국립국어원은 민간의 수요를 고려하여 매년 일정 시간 이상의 일상 대화를 수집하여 말뭉치를 구축하고 있다. 일상 대화 말뭉치의 경우도 신문과 마찬가지로 화자의 수, 화자의 유형, 대화의 주제, 전사의 방법, 품질 관리 방법 등 말뭉치 구축의 다양한 요소를 반영한 구축 기준을 사전에 정의하고 이에 따라 음성 대화 자료수집을 시작한다. 특히 국립국어원의 일상 대화 자료는 일반적인 언어 변화를 반영하는데 목적을 두고 있으므로 대화의 주제를 결정하고 녹음을 희망하는 화자를 모집하는 과정이 상대적으로 까다롭다. 또한 일정한 환경에서의 음성 녹음을 가정하므로 참여자뿐 아니라 전문적인 녹음 장비 및 녹음 전문 인력이 필요하다. 일상 대화 자료를 수집할 때에도 신문 기사와 마찬가지로 저작권자와 이용 허락 계약을 맺는다. 이러한 과정을 통해 구축한 원시 말뭉치는 최종적으로 제이슨(json) 구조로 변환된다. [그림 1]은 신문 말뭉치의 예시이다.

```
"id": "NIRW1900000001"
    "metadata": {
"title": "국립국어원 신문 말뭉치 NIRW1900000001",
         "creator": "국립국어원",
"distributor": "국립국어원",
         "vear": "2019".
         "category": "신문 > 인터넷 기반 신문",
         "annotation_level": [
         "sampling": "부분 추출 - 임의 추출"
    "document": [
             "id": "NIRW1900000001.1",
             "metadata": {
    "title": "오마이뉴스 2009년 기사",
                 "author": "선대식",
"publisher": "오마이뉴스",
                 "date": "20090101",
"topic": "사회",
                 "original_topic": "경제"
             "paragraph": J
                      "id": "NIRW190000001.1.1".
                      "form": "\"대통령, 시장 방문만 하지 말고 실천해달라\""
                     "id": "NIRW190000001.1.2".
"form": "2008년의 마지막 새벽, 언론의 카메라는 서울 여의도를 향했다. 방송법 등 주
요쟁정 법안이 상정될 국회 본회의장을 두고 여야 의원들의 전쟁을 기다리고 있었던 것."
```

[그림 1] 배포용 신문 말뭉치 예시

이 예시에서 'id' 신문 말뭉치의 고유 id를 말한다. 'metadata'에는 이 말뭉치의 명칭, 구축자, 배포자, 구축 연도, 주석 수준 등의 정보가 포함되어 있다. 'document'에는 신문 기사별 고유 아이디, 메타 정보, 기사 본문이 포함되어 있다. 특히 기사 본문이 앞서 언급한 대로 문단별로 구분되어 있음을 확인할 수 있다.

```
"sex": "여성",
"birthplace": "서울",
"id": "SDRW2000000002"
"metadata": {
    "title": "국립국어원 구어 말뭉치 SDRW2000000002",
                                                                                        "principal_residence": "서울",
                                                                                       "current_residence": "서울",
"education": "대재"
    tute : 국립국어된 구여 달
"creator": "국립국어원",
"distributor": "국립국어원",
                                                                                  }
    "year": "2020",
                                                                               "setting": {
     'category": "구어 > 사적 대화 > 일상대화".
                                                                                   "relation": "기타"
    "annotation_level": [
                                                                              }
         "워지"
                                                                           utterance": [
     "sampling": "본문 전체"
                                                                                   "id": "SDRW2000000002.1.1.1"
"document": [
                                                                                   "form": "반려동물을 키우고 계신가요?".
                                                                                   "original_form": "반려동물을 키우고 계신가요?",
                                                                                   "speaker_id": "SD2000003",
         "id": "SDRW2000000002.1".
         "metadata": {
"title": "2인 일상 대화",
"author": "개인 발화자"
                                                                                   "start": 2.78903.
                                                                                   "end": 4.92608,
"note": ""
              "publisher": "개인 발화 녹음",
              "date": "20200602",
"topic": "반려동물 > 보험, 유튜버, 동물학대, 작명",
                                                                                   "id": "SDRW2000000002.1.1.2",
                                                                                   "form": "혹시 안 키우고 계시다면은".
"original_form": "혹시 안 키우고 계시다면은".
              "speaker": [
                                                                                    speaker_id": "SD2000003",
                       "id": "SD2000003",
                                                                                   "etart": 1 93608
                       "age": "30대",
                                                                                   "end": 6.70908,
                        "occupation": "전문가 및 관련 종사자".
                                                                                   "note":
```

[그림 2] 배포용 일상 대화 말뭉치 예시

[그림 2]는 일상 대화 말뭉치의 예시이다. 일상 대화 말뭉치의 경우 화자의 성별, 거주지, 연령 등 정보를 포함하고 있으며, 화자가 녹음한 내용은 억양구 단위로 분할되어 있음을 확인할 수 있다. 또한 음성 자료와 연계하기 위해 발화 시작점과 끝점의 위치를 표시하고 있다.

이밖에도 국립국어원은 온라인 대화 말뭉치, 웹 말뭉치 등 다양한 매체의 언어 자료를 매년 수집하여 말뭉치로 구축하고 있다. 이처럼 언어 변화 현실을 반영하는 원시 말뭉치를 지속해서 구축하여 공급하면 인공 지능 언어 처리 기술이 최신의 언어 자료를 학습할 수 있으므로 현실에 맞는 예측을 잘 수행할 수 있다.

2) 분석 말뭉치 구축

국립국어원은 수집한 원문 자료에 한국어의 특성을 반영한 정보를 부착한 말뭉치를 구축한다. 이는 주석 말뭉치, 레이블링 데이터 세트 등 다양한 명칭으로 불릴 수 있으나, 국립국어원에서는 원시 말뭉치를 언어학적 측면에서 분석한 정보를 부착한다는 의미로 분석 말뭉치라는 용어를 사용한다. 2019년 형태, 어휘의미, 구문 분석, 무형대용어 복원, 상호참조 해결, 의미역 등 다양한 층위의 언어 분석 말뭉치를 구축한 이래, 최근에는 인공 지능이 언어와 관련된 현실의 문제를 더 잘 이해하고 해결할 수 있도록 형태 분석, 어휘 의미 분석, 구문 분석 등 과거에 구축한 유형 외에도 다양한 분석 말뭉치를 구축하여 공개하고 있다.

한국어는 어순이 자유롭고 문장의 구성성분 생략이 빈번하다. 무형대용어 복원 말뭉치는 한국어의 사용 맥락을 정확하게 이해하는 데 도움을 준다. [표 1]과 [그림 3]은 무형대용어 복원 말뭉치의 예이다. 두 번째 문장에서 밑줄 친 '개발하다'는 술어(predicate)로, 주어(subject)와 목적어(object)가 생략되어 있으며 첫 번째 문장에서 해당 구성성분을 찾을 수 있다. 무형대용어 복원 말뭉치는 [표 1]의 가장 아래 줄과 같이 주격무형대용어와 목적격무형대용어를 논항으로 가지는 술어를 찾아 주어와 술어를 복원한 모양으로 구성된다.

구분	내용
	(#1) 창원시는 기존 누비자의 단점을 보완, 내구성·안전성·편의성 등을 획기적으로 개선한 '명품 누비자'를 이달 말 보급한다고 26일 밝혔다. (#2) 삼천리자전거와 공동 개발한 '명품 누비자'는 내구성 강화를 위해 프레임을 고급 알투 미늄 합금 소재로 제작했고, 프레임 직경을 기존 55mm에서 57mm으로 키웠다.
무형대용어 복원	(#2) 삼천리자전거와 공동 <u>개발한</u> '명품누비자'는 내구성 강화를 위해 프레임을 고급 알루미늄 합금 소재로 제작했고, 프레임 직경을 기존 55mm에서 57mm으로 키웠다. ※생략어 복원: [주어창원시가] [목적어누비자를] 개발하다

[표 1] 무형대용어 복원 사례



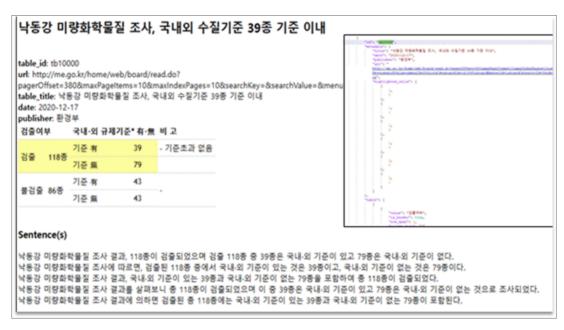
[그림 3] 무형 대용어 복원 말뭉치 예시

언어 사용 결과로부터 언어 사용자의 감정을 추출하는 작업은 언어의 사용 현실을 보여주기도 하지만 상품에 대한 소비자의 감정을 이해하고자 하는 목적에서 산업적으로도 관심이 높다. [표 2]는 속성 기반 감성 분석 말뭉치의 예이다. 속성 기반 말뭉치는 문장에 나타나는 개체와 속성 범주를 식별하고, 이 범주에 대한 후기 작성자의 감성 극성(궁정또는 부정)을 부착한 말뭉치로 의미 평가 국제 워크숍(International Workshop on Semantic Evaluation, SemEval)의 과제를 참조하여 구축하였다. [표 2]에 따르면 후기 작성자는 본품(제품 본체)의 품질에 대해 부정적인 감성 극성을 표현하고 있으며, 품질의 속성을 가진 본품을 가리키는 개체는 '기어'이다.

구분	내용
식별 부호	"id": "nikluge-sa-2022-train-00001"
ㅁ기. 처리	sentence_form": "둘쨋날은 미친듯이 밟아봤더니 기어가 헛돌면서 틱틱 소리
문장 형태	가 나서 경악."
주석 내용	"annotation": [["본품#품질", ["기어", 16, 18], "negative"]]}

[표 2] 속성 기반 감성 분석 말뭉치 예시

한편, 국립국어원에서는 위에서 소개한 두 유형의 말뭉치와 같은 자연어 이해와 관련된 분석 말뭉치를 구축하기도 하지만, 자연어 생성과 관련된 분석 말뭉치를 구축하기도 한다. [그림 4]는 국립국어원에서 구축 중인 표 기반 유사 문장 말뭉치의 예시이다. 이 말뭉치는 구글이 발표한 표 정보를 텍스트로 생성하는 인공 지능 학습 데이터 세트 Controlled Table-Text Generation, ToTTo)를 참고하여 구축하였다. 여기서 '낙동강 미량화학물질 조사 결과, 118종이 검출되었으며 검출 118종 중 39종은 국내·외 기준이 있고 79종은 국내·외 기준이 없다.'는 표의 제목과 헤더, 표의 음영 부분을 한국어로 기술한것이다. 이 말뭉치는 연내 공개 예정이다.

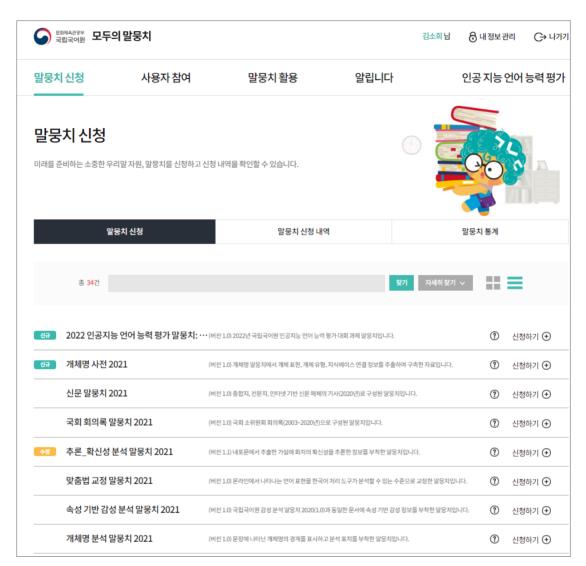


[그림 4] 국립국어원 표 기반 유사 문장 말뭉치 예시

분석 말뭉치를 구축할 때에는 전문가의 국어학적 전문성을 살리되 일관성을 갖춘 지침을 연구하여 적용하는 것이 중요하다. 4차 산업혁명 대비 국어 빅데이터 구축 사업을 통해 구축하는 말뭉치는 언어 연구 자원으로써 활용되기도 하지만 인공 지능의 언어 처리성능 향상이라는 실용적인 목표를 달성하는 데 초점을 두고 있어 전문가의 국어학적 지식을 충분히 활용하되 일반 언어 사용자의 상식에 맞춘 분석 지침 마련이 필요할 때도 있다. 이렇게 구축한 분석 말뭉치는 인공 지능의 자연어 이해 또는 자연어 생성 성능을 향상하기 위한 언어 자원으로 주로 활용된다.

나. 말뭉치 배포 및 활용 현황

국립국어원은 "모두의 말뭉치(https://corpus.korean.go.kr)" 플랫폼 개통과 함께 2020년 8월 18억 어절 규모의 말뭉치를 공개한 이래, 신규로 구축한 말뭉치 또는 수정·보완한 말뭉치를 매년 분기별로 공개한다. 2022년 9월 현재 공개 중인 말뭉치는 34종이며 "모두의 말뭉치"신청 화면에서 확인할 수 있다. "모두의 말뭉치"의 신청 화면은 다음과 같다. 각 말뭉치 설명에서 '?'를 누르면 말뭉치 설명서를 확인할 수 있다.



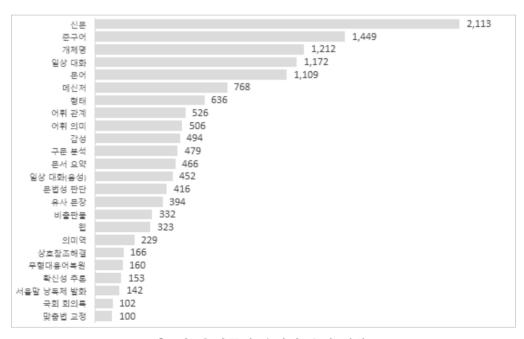
[그림 5] 모두의 말뭉치 신청 화면

국립국어원이 공개한 말뭉치는 "모두의 말뭉치" 누리집에서 회원 가입 후 신청 및 약정서 작성 절차를 거쳐 이용할 수 있다. 말뭉치를 신청할 때에는 다음의 사항에 주의할 필요가 있다. 말뭉치 이용 주체는 '개인'을 원칙으로 하며 말뭉치 이용 결과물에 말뭉치 원문 또는 원문 일부가 포함되어서는 안 된다(단, 연구 논문 등에 인용하는 것은 가능하다). 또한 개별 말뭉치의 이용 허락 범위 내에서 말뭉치를 활용하여야 하며, 제삼자에게 양도, 대여 공유할 수 없다. 이는 말뭉치 이용자들이 원문 저작권자와의 이용 허락 계약 내용을 준수하여 말뭉치를 이용할 수 있도록 하기 위한 것이다. 따라서 국립국어원은 이용자들이 말뭉치를 신청할 때 연구 목적 및 연구 기간 등을 명확히 밝히도록 안내하고 있으며, 관리자는 이용자들이 국어원의 정책에 맞게 말뭉치를 이용하는지 확인한 후 승인한다. 한편 말뭉치 이용자들이 말뭉치를 활용한 결과물을 외부에 공개하고자 할 때에는

"모두의 말뭉치" 플랫폼상의 "말뭉치 활용/결과물 공개 신청"메뉴를 이용하여 검토 요청을 하여야 한다. 이와 같은 절차는 다소 까다로워 보이기는 하지만 이용자가 내려받은 말뭉치를 약정 내용을 준수하여 이용하였는지를 확인하기 위해 필요한 절차인 동시에, 국립국어원이 이용자들이 말뭉치를 어떤 목적으로 이용하고 있는지 파악하여 더 나은 서비스를 제공하기 위한 절차이기도 하다.

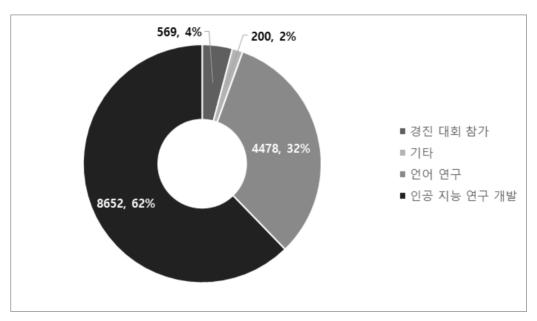


[그림 6] 결과물 공개 신청 화면



[그림 7] 말뭉치 유형별 승인 현황

[그림 7]은 2022년 9월 현재 말뭉치 유형별 승인 현황이다. 말뭉치 유형별 규모와 공개시기가 다르기는 하지만 이용자들이 선호하는 말뭉치 유형을 확인할 수 있다. 신문, 일상대화, 메신저 대화 등 원시 말뭉치의 이용량이 높은 것을 확인할 수 있다. 분석 말뭉치중에는 개체명 말뭉치를 가장 많이 이용하고 있다. 개체명 말뭉치는 수요에 맞게 2019년부터 2022년까지 4년간 구축하고 있으며 분석 말뭉치 중 가장 규모가 크다, 또한 형태, 어휘 의미, 구문 분석 등 전통적인 분석 말뭉치들의 이용량도 많다. 마지막으로 감성 분석, 문법성 판단 말뭉치 등 자연어 이해 성능 향상을 위해 활용되는 말뭉치들도 많이 이용되고 있다. 한편, 말뭉치 이용 목적별 승인 현황은 [그림 8]과 같다. 이 사업의 취지에맞게 인공 지능 연구 개발 목적으로 말뭉치를 신청하는 비중이 높으며, 언어 연구 목적으로도 활용되고 있음을 알 수 있다.



[그림 8] 말뭉치 이용 목적별 승인 현황

이밖에도 국어원은 빅데이터의 중요성을 알리고 모두의 말뭉치의 활용도를 높이기 위한 노력도 지속하고 있다. 올해 9월 23일에는 '우리 삶을 변화시킨 한국어 인공 지능'이라는 주제로 챗봇, 기계 번역, 음성 인식 등의 기술을 활용한 서비스 사례를 소개하는 특별 강연회를 개최하였다. 또한, 말뭉치를 활용하는 데 어려움을 느끼는 언어 연구자들을 위해 2022년 하반기 총 6개의 동영상 강좌를 제작하여 공개할 계획이다. 이 동영상 강좌에는 회원 가입, 말뭉치 신청 방법, 제이슨 구조를 이해하고 파이선 언어(python)를 이용하여 처리하는 방법, 특정 표현의 용례와 빈도를 추출하는 방법 등의 내용이 담길 예정이다. 국어원은 앞으로도 "모두의 말뭉치" 플랫폼을 중심으로 구축한 말뭉치를 공개하는 말뭉치 관련 짧은 강좌 등 학습 자료를 제공함으로써 이용자들이 말뭉치를 더 쉽게 이용할 수 있도록 지원하고자 한다.

다. 인공 지능 언어 능력 평가

인공 지능 기술의 급격한 발전과 함께, 최근에는 사람과 유사하게 언어를 이해하고 추론하는 시스템을 개발하고 성능을 경쟁하면서 기술을 발전시키는 환경이 전 세계적으로 활성화되어 있다. 예를 들어 인공 지능 언어 처리 연구자들은 미국의 글루(GLUE), 슈퍼글루(SuperGLUE) 등이 제공하는 공통된 과제를 성능 평가의 기준으로 삼아 인공 지능 언어처리 기술들(BERT, GPT-3)을 발전시켜 왔다. 국립국어원은 이와 같은 체계가 국어를 기반으로도 운영되어야 한다는 민간의 요구를 반영하여 2009년부터 연간 개최한 '국어정보처리 시스템 경진대회'를 2021년부터 '인공 지능의 언어 능력 평가 대회'로 개편하여자연어 처리 기술의 성능을 평가하기 위한 평가 체계를 운영하고 있다.

대부분의 자연어 처리 기술 개발이나 평가 환경에서 데이터 세트는 훈련용(Train), 개발·검증용(Dev.), 시험용(Test) 데이터 세트로 분할된다. 훈련용 데이터 세트는 인공 지능 언어 모델이 학습하기 위한 입력으로 사용되며, 사람이 처리한 정답 주석이 포함되어 있다. 개발 또는 검증용 데이터 세트는 개발 중인 시스템의 적합한 파라미터 값을 찾는 등성능을 향상시키기 위해 훈련용에서 따로 떼어 둔 데이터 세트이다. 시험용 데이터 세트는 개발이 완료된 시스템을 종합 평가하기 위한 데이터 세트로, 정답은 공개하지 않으며, 비공개 정답과 비교하여 산출한 점수만 순위표에 게시한다.

다음은 '2022 국립국어원 인공 지능 언어 능력 평가'경진 대회의 '속성 기반 감성 분석'과제의 훈련용, 시험용 데이터 세트의 예시이다. 훈련용 데이터 세트는 데이터 아이디(id), 분석 대상 문장(sentence form), 정답 주석(annotation)으로 구성되어 있으며, 시험용 데이터 세트에는 정답 주석이 빠져 있다. 참고로 속성 기반 감성 분석이란 자연어이해 과제 중 하나로 분석 대상 문장에서 나타나는 개체와 속성 범주를 식별하고 이 범주에 대한 후기 작성자의 감성을 추론하는 과제이다. 훈련용 데이터 예시에 따르면 후기 작성자는 제품 본품(전자 기기의 본체 등)의 품질에 대해 부정적인 감성 극성을 표현하고 있다. 경진대회 참가자들은 시험용 데이터의 정답 부분을 ["제품 전체#일반", "positive"]와 같이 채워 넣어 국립국어원 모두의 말뭉치(https://corpus.korean.go.kr) 내 '인공 지능 언어 능력 평가'에서 제출하며, 국립국어원에서는 비공개 정답과 비교한 점수를 산출하여 순위표에 게시한다. 참가자들은 답안을 시스템에 제출하여 나의 순위를 확인함으로써 내가 만든 모델의 성능을 가늠할 수 있다.

항목	내용
훈련용	{"id": "nikluge-sa-2022-train-00001", "sentence_form": "둘쨋날은 미친 듯이 밟아봤더니 기어가 헛돌면서 틱틱 소리가 나서 경악.", "annotation": [["본품♯품질", ["기어", 16, 18], "negative"]]}
시험용	{"id": "nikluge-sa-2022-train-00003", "sentence_form": "간사하게도 그 이후에는 라이딩이 아주 즐거워져서 만족스럽게 탔다.", "annotation": []}

[표 3] 2022 국립국어원 인공 지능 언어 능력 평가 대회 과제(속성 기반 감성 분석)의 훈련용·시험용 데이터 예시

인공 지능 언어 능력 평가 체계를 구성할 때에는 대상 과제를 해결하기 위해 일반적으로 활용되고 있는 모델과 모델의 성능을 함께 제공한다. 이를 베이스라인 모델(baseline model)이라고 하며, 참가자들은 이 모델을 활용하여 직접 개발한 모델의 성능을 개선할 수 있다. 올해 과제에는 해당하지 않지만 어떤 과제에서는 사람 수행 점수(Human

Performance Score)를 함께 제시하기도 하는데, 이 점수는 인공 지능 언어 처리 기술이 도달해야 할 목표 점수로서 의미가 있다. '2022 국립국어원 인공 지능 언어 능력 평가대회'의 접수 기간은 10월 4일부터 11월 4일까지이며 '모두의 말뭉치 (https://corpus.korean.go.kr)' 플랫폼에서 접수한다. 답안은 하루 한 번 제출할 수 있으며 채점 결과는 다음 날 순위표에 게시된다.

3. 향후 계획

국립국어원은 2018년 '4차 산업혁명 대비 국어 빅데이터 구축' 사업을 시작한 이래 변화하는 기술 환경과 산업 수요에 맞추어 사업을 추진해 왔다. 사업 초기에는 신문, 일 상 대화, 준구어, 웹, 메신저 등 다양한 매체의 언어 자료를 대규모로 수집하여 원시 말뭉 치를 구축하고 형태 분석, 어휘 의미 분석, 구문 분석, 개체명 분석, 무형대용어 복원, 의 미역 분석, 상호 참조 해결 등 언어학적 분석 층위를 고려한 분석 말뭉치를 구축하였다. 현재는 인공 지능이 최신의 언어 자료를 학습할 수 있도록 신문 기사와 일상 대화 원시 말뭉치는 매년 구축하고 인공 지능 언어 처리 기술 개발에 대응한 다양한 분석 말뭉치를 구축하는 것으로 구축 방향을 일부 수정하였으며 이에 따라 인공 지능의 자연어 이해 능 력과 생성 능력을 평가할 수 있는 말뭉치를 구축하게 되었다.

국립국어원은 앞으로 변화하는 기술 환경과 산업 수요에 부합하는 인공 지능 학습 및 성능 평가용 말뭉치 구축을 지속하되, 한국어의 특성과 사용 현실을 고려한 말뭉치에 중점을 두어 말뭉치를 연구하고 구축할 계획이다. 또한 어문 연구 및 정책 수행 기관으로서 국어원의 전문성을 살리고, 국민의 언어생활 편의를 향상할 수 있도록 다양한 말뭉치들을 구축할 계획이다. 현재 한국어-외국어 병렬 말뭉치, 국민의 글쓰기 평가(채점, 첨삭 등)말 뭉치 구축 등을 계획하고 있다. 또한 '인공 지능 언어 능력 평가 체계'를 상시로 운영하고 한국어 인공 지능 분야의 평가 과제들을 누적하여 공개함으로써 연구 개발자들이 인공지능의 한국어 이해 및 생성 능력을 다양한 방면에서 검증하고 성과를 공유하는 장을 마련할 계획이다.

참고문헌

문화체육관광부(2022) 국어 발전과 보전에 관한 시책 및 시행 결과 보고서 2021, 문화체육관광부.

국립국어원(2021) 2021년 신문 기사 원문 수집 및 정제 보고서. 국립국어원.

국립국어원(2021) 2021년 일상 대화 말뭉치 구축 보고서. 국립국어원.

국립국어원(2021) 2021년 유사 문장 생성 말뭉치 연구 분석 보고서. 국립국어원

국립국어원(2020) 구문 및 무형대용어 복원 말뭉치 연구 분석 보고서. 국립국어원.

국립국어원 모두의 말뭉치 누리집 URL: https://corpus.korean.go.kr

국립국어원(2022) 국립국어원 신문 말뭉치 2021(버전 1.0).

국립국어원(2022) 국립국어원 인공 지능 언어 능력 평가 대회 말뭉치-속성 기반 감성 분석(버전 1.0).

국립국어원(2021) 국립국어원 무형 대용어 복원 말뭉치 2020(버전 1.0).

국립국어원(2021) 국립국어원 일상 대화 말뭉치 2020(버전 1.2).

코퍼스와 기계학습을 통한 사회 현상의 범학문적 분석 : 코로나19 사례를 중심으로

이규민*, 이정현**, 장서현***, 송상헌****†

본 연구는 신종 코로나바이러스 감염증(코로나19)로 인해 발생한 사회현상을 대용량 의 비정제 텍스트 데이터에 기반하여 분석하는데 목표를 둔다. 코로나19는 2020년 1 월 국내에서 첫 번째 감염자가 확인된 이후 2년 이상의 시간 동안 한국 뿐 만 아니라 전 세계 사람들의 일상 생활과 사회에 여러 변화를 가져왔다. 이러한 사회적 변화를 들여다보기 위하여 본 연구진은 코로나19와 관련된 영어 언론 기사로 구성한 코퍼스 인 BYU Coronavirus Corpus에 대응할 수 있는 한국어 코퍼스인 Korean Coronavirus Corpus를 구축하였다. Korean Coronavirus Corpus는 2020년 1월 1일부터 코로나 관련 한국 언론 매체 기사를 수집한 데이터로 180만개 이상의 기사. 4억여 어절, 10억여 형태소로 이루어져 있다. 해당 코퍼스를 활용한 본 연구진의 기 존 연구의 연장으로, 본 발표에서는 코로나19 대유행 기간 동안 국내에서 이루어진 두 번의 중앙 선거를 방역의 관점에서 비교한다. 또한, BYU Coronavirus Corpus 중 미국을 기반으로 둔 매체의 기사만을 활용하여 코로나19 대유행 속에서 등장한 '백신 관광(Vaccine Tourism)'의 사례를 탐색한다. 마지막으로, 미국과 한국 두 나라 가 모두 대유행 속에서 대선을 치른 만큼 방역이슈를 중심으로 두 나라의 대선을 비 교한다. 이러한 분석에는 빈도 및 용례와 같은 전통적인 코퍼스 분석 방법에서 기계 학습을 활용한 텍스트 분석 까지 다방면의 전산언어학 기술을 활용하였으며, 이러한 기술로 도출한 결과를 사회학적으로 분석함으로써 범학문적 연구를 꾀하였다. 본 연 구는 코로나19 현상을 직접적으로 바라볼 수 있는 대규모의 한국어 코퍼스 데이터의 구축 뿐 만 아니라 이에 대한 전산언어학적 기술의 적용 및 범분야적 지식을 융합하 여 코로나19 대유행 속에서 국내외 비정제 텍스트 데이터를 통해 사회 현상을 살펴보 아 단순한 수치로 드러나지 않는 사회의 단면을 분석한다는 점에서 의의가 있다.

^{*} 고려대학교 언어정보연구소

^{**} 고려대학교 언어학과

^{***} 고려대학교 사회학과

^{****} 고려대학교 언어학과

[†] 교신저자

서울특별시 성북구 안암로 145 고려대학교 문과대학, sanghoun@korea.ac.kr, 010-6750-0311

한국어 언어 모델의 정보 구조 표상 능력: '은/는'의 관계절 내 형성 제약을 중심으로

황동진(고려대)

초록: 본 연구는 실험을 통해 관계절 내부 주어짐성 화제(aboutness topic) 실현 제약에 대해 살펴보고, 이러한 제약을 통해 한국어 언어 모델의 정보 구조 표상 능력에 대해 논의하고자 한다. 한국어의 정보 구조는 일반적으로 조사를 통해 형태론적인 층위에 표시된다(Choi, 1996). 본 연구에서는 정보구조적 기능을 가지는 조사들 중, 주어짐성 화제를 표시하는 '은/는'에 초점을 맞추어 논의를 진행한다. 한국어에서 주어짐 화제를 표시하는 '은/는'은 일반적인 경우 관계절 내부에 나타나지 않는다(Song, 2014a).

1) 이씨가 김씨{가/??는} 읽은 책을 샀다. (Song, 2014b)

(1)에서와 같이, 언어에서 관계절 내부에 화제가 실현될 수 없는 이유는 관계절의 머리명사가 이미 화제로 실현되었기 때문이라고 지적되었다(Kuno, 1976). 즉, 화제 표지인 '은/는'의 관계절 내부 형성 제약을 언어 모델이 충실히 표상한다면, 이는 언어 모델이 문장 내부의 정보 구조와 관련한 특징들을 잘 학습한 결과라고 볼 수 있는 것이다.

이를 살펴보기 위해 실험은 크게 두 가지가 진행되었다. 먼저, 한국어 화자를 대상으로 한 수용성 판단 실험을 통해 한국어에서의 관계절 내부 주어짐성 화제 실현 제약을 밝혔다. 이어, 한국어 화자의 수용성을 조사한 같은 예문을 바탕으로 한국어 딥러닝 언어 모델인 KR-BERT(Lee et al., 2020)를 통해 관계절 내부에 자리하는 조사들의 놀라움(surprisal)을 측정했다.

실험 결과, 한국어 화자들은 관계절 내부에 화제 표지 '는'이 위치하는 문장들에 대해 낮은 수용성을 보였다. 마찬가지로 딥러닝 언어 모델들 또한 관계절 내부에 '는'이 위치할 경우 다른 조사들의 놀라움보다 더 높은 수준의 놀라움 분포를 보여주었다. 이러한 결과는 한국어 화자들의 직관에 한국어 화제표지에 대한 관계절 내부 형성 제약이 존재함을 보여주며, 동시에 이러한 정보 구조와 관련된 언어 사용자들의 직관을 딥러닝 언어 모델이 잘 포착하고 있음을 증명한다. 이러한 결과를 바탕으로 본 연구는, 한국어 언어 모델이 형태적층위를 통해 실현된 정보 구조를 표상할 수 있다고 주장한다.

1. 서론

본 연구는 한국어 화자를 대상으로 한 수용성 판단 실험을 통해 관계절 내부 주어 짐성 화제(aboutness topic) 실현 제약에 대해 살펴보고, 언어 모델이 이러한 제약을 어느 정도로 근사하게 표상하고 있는지를 확인하고자 한다. 이를 통해 궁극적으로 한국어 언어모델의 정보 구조 표상 가능성에 대해 논의할 것이다. 한국어의 정보 구조는 일반적으로 조사를 통해 형태론적인 층위에 표시된다(Choi, 1996). 본 연구에서는 정보구조적 기능을 가지는 조사들 중, 주어짐성 화제를 표시하는 '-은/는'에 초점을 맞추어 논의를 진행한다. 한국어에서 주어짐 화제를 표시하는 '-은/는'은 일반적인 경우 관계절 내부에 나타나지 않는다(임동훈, 2012; Song, 2014a; 2014b).

(1) 이씨가 김씨{가/??는} 읽은 책을 샀다. (Song, 2014b)

(1)에서와 같이, 언어에서 관계절 내부에 화제가 실현될 수 없는 이유는 관계절의 머리명사가 이미 화제로 실현되었기 때문이라고 지적되었다(Kuno, 1976). 즉, 화제 표지인 '-은/는'의 관계절 내부 형성 제약을 언어 모델이 충실히 표상한다면, 이는 언어 모델이 문장 내부의 정보 구조와 관련한 특징들을 잘 학습한 결과라고 볼 수 있는 것이다. 때문에 이 논문에서는 형태통사적 요소를 통해 드러내는 문장 내 정보구조의 경우에는 언어 모델이 이미 조사의 분포적 특징을 통해 표상을 능력을 가지고 있음을 주장한다.

2. 배경

2.1 한국어 관계절 내부의 화제 표시 '-은/는'의 형성 제약

관계절 형성 시, 관계절의 수식을 받는 머리명사나, 관계사 등에 화제가 부여된다는 주장은 한국어뿐만 아니라 다른 언어에서도 있어 왔다(Kuno, 1976; Bresnan and Mchombo, 1987; Bjerre, 2011). 특히, Song(2014a; 2014b)은 한국어 관계절 내부에도 화제 표지의 형성이 자유롭지 않다고 지적한다. 임동훈(2012) 또한 "-은/는"이 실현하는 화제의 실현에 존재하는 몇 가지 제약을 제시한다. 첫째, 주어는 참조의 주어짐성(givenness)에 따라 의미 있는 제약을 갖는다. 이때, 화제는 기본적으로 제한적이고 개념적으로 일반적이어야 하며 그렇지않으면 구체적이거나 최소한 대조적이어야 한다. 두 번째는, 위치 제약 조건이다. 학자들의의견이 완전히 일치하지는 않지만, 문장의 시작 부분에 주어가 위치해야 한다는 주장도 있다. 마지막 제약 조건은 내포절 제약 조건이다. 기본적으로 주어를 표시하는 '은/는'은 내포절에 표시할 수 없다.

다만, 임동훈(2012)과 Song(2014a) 모두 한국어 조사 '-은/는'이 대조 초점을 나타내기 위해서 사용되는 경우에는 관계절 내부에 실현이 가능함을 인정한다.

(2) [향기는 좋은] 꽃이 핀다. (임동훈, 2012)

(2)의 경우에서와 같이 꽃은 다른 특성과 '향기'라는 특성이 대조 맥락을 이루고 있는 경우, '향기'는 정보 구조상 대조 초점을 이루게 되고, 이러한 경우에는 '-은/는'의 실현이 관계절 내부에 가능함을 보여준다. 다만, 이 때 조사 '-은/는'의 경우는 정보 구조에서는 화제가 아닌 대조 초점의 역할을 하고 있다. 즉, '-은/는'이 주어짐성 화제를 나타내기 위하여 관계절 내부에 형성되는 것은 여전히 불가능한 것으로 보인다. 그러나, 기존의 연구들은 대부분의 경우 언어학자의 직관에 의존한 연구들이다. 코퍼스를 통해 이러한 제약을 살펴보거나 피험자를 대상으로 한 수용성 판단 연구는 진행된 적이 없는 것으로 보인다. 때문에, 본 연구에서는 선행연구들에서 제시한 언어학자들의 직관을 성인 한국어 화자들을 대상으로 한 수용성 판단 연구를 통해 확인하고자 한다.

2.2 한국어 보조사의 특성과 정보구조

기존에 관계절과 정보구조 사이의 관계들에 대한 연구는 한국어가 아닌 다른 언어들을 중심으로 이루어졌다(Bresnan and Mchombo, 1987; Kuno, 1976.)이번 절에서는 정보구조를 표시하는 '한국어' 조사들의 특징에 대해 살펴보고, 이러한 선행 연구들이 본 연구에어떤 맥락을 제공하는지 살펴보도록 하겠다.

장석진(2000)은 담화 기능으로서의 화제와 초점의 실현은 음운, 어휘, 통사 간의 상호 접촉과 제약을 통해 실현된다고 주장한다. 본 논문은 한국어 정보구조의 실현에 있어 조사의 역할을 직접적으로 고려하지는 않지만, 한국어 정보구조가 한 층위의 언어적 실현에만 의존하지 않는다는 것을 보여준다.

다음으로, 서민정(2007)은 한국어 후치사 중 '-이/가', '-을/를', '-은/는', '-도'를 동일한 범주로 결합할 수 있는 분포적 특성에 대해 이야기한다. 논문에 따르면 기존 조사의구별은 '격조사'와 '보조사''의 통사적 관계를 나타내는 조사와 의미적, 화용적 기능을 나타내는 조사로 구분되었다. 그러나 본 논문은 국어 조사 '-이/가', '-은/는'의 분포 특성을 들어예외적인 조합을 보이는 경우의 특성을 통해 동일한 범주의 조사로 분류하는 것이 가능하다고 주장한다. 즉, 화용적 층위에서 구분되는 정보구조의 특징이, 한국어의 형태통사적 층위의 특징 중 하나인 조사를 통해 실현되며, 이러한 보조사의 분포적 특징을 통해 그러한 화용론적 층위에서 정보구조의 구분을 정당화하고 있는 것이다. 이러한 주장은, 보조사의 분포적 특징이 정보구조와 긴밀한 관계를 가지고 있음을 보여주며, 완전한 의미의 '화용적' 정보구조의 표상은 아닐지라도, 형태통사적 층위에서 분포적으로 실현될 수 있는 문장의 정보구조적 특징의 경우 언어 모델에 표상되는 것이 가능할 것임을 예상할 수 있게 한다.

다음으로 김일규와 지윤주(2020)는 실험을 통해 한국어 관계의 제약조건을 살펴본다. 본 연구는 조사의 사용 또는 정보구조와 조사의 사용과의 관계를 명시적으로 밝히지 않는다. 그러나 본 논문에서는 한국어 정보구조의 해석이 관계절의 수용에 영향을 미친다는 주장을 펼친다. 이러한 주장은, 한국어의 관계절 제약이 순수하게 통사적인 것이 아님을 보여준다. 특히 관계절의 수용성에 영향을 미치는 대조 관계의 경우, 강한 대조 맥락에서만 문장의 수용성이 유의하게 높다. 본 연구는 정보구조적 특징이 관계절과 관계절 머리명사의

관계에서도 드러나며, 이는 관계절의 수용성에 직접적인 영향을 미칠 수 있다는 것이다. 즉, 이러한 두 현상 사이의 긴밀한 관계는, 모델의 관계절 표상을 통해 모델이 문장의 정보 구조를 어느 정도로 파악하고 있는지에 대한 근사가 가능함을 보여준다.

3. 실험 방법

이를 살펴보기 위해 실험은 크게 두 가지가 진행되었다. 먼저, 한국어 화자 44명¹⁾을 대상으로 한 수용성 판단 실험을 통해 한국어에서의 관계절 내부 주어짐성 화제 실현 제약을 밝히고자 했다. 또 다른 실험의 경우 언어 모델을 통해 실험 항목의 '놀라움(surprisal)'을 측정하는 방식으로 이루어졌다.. 데이터의 경우 초기에 243문장이 구성되었고 파일럿 실험을 통해 수용성이 높은 60문장이 실험 대상으로 선택되었다. 처음에는 기본 템플릿이 사용되었다.

(3) A는 B가 C하는 D를 E했다.

토큰화의 효율성을 극대화하기 위해 각 A에서 E까지 KR-BERT 사전에서 3개의 단어를 선택했다. 또한 섹션별로 한국어 빈도가 비슷한 단어를 선정했다. 그 빈도는 김한샘(2005)에 따른 것이다.

	Vocab	Rank	Freq.
	그녀	72위	3767회
Α	ュ	13위	14058회
	교수	2위	1312회
	우리	22위	11213회
В	후보	759위	415회
	정부	116위	2652회
	요구하다	590위	608회
C	생각하다	58위	4570회
	제공하다	659위	530회
	지원	592위	606회
D	활동	230위	1647회
	관심	286위	1324회
	설명하다	464위	809회
Е	지적하다	797위	376회
	기록하다	816위	362회

¹⁾ 현재 추가적인 피험자들을 대상으로 실험이 진행중이며, 실험이 완료되면 전체적인 데이터와 내용을 수정 보완할 계획이다. 현재 피험자의 수가 40명을 넘었기에 분석을 진행하여 발표 논문에 싣게 되었다.

표1. 예문 구성에 사용된 어휘들의 한국어 사용 빈도(김한샘, 2005)

(3)의 템플릿과 표1.의 단어를 사용하여 예문들을 구성했다. 이렇게 구성된 예문들 중 파일 럿 실험을 통해 만장일치로 수용 가능한 것으로 평가된 60개의 문장을 선정했다. 파일럿 실험에서는 관계절 주어의 조사로 모두 동일한 '-가'를 사용했다.

이렇게 선정된 문장들을 총 4 세트를 구성하는데 사용되었다. 한 세트에 포함된 실험 예문은 총 60개로, 15개씩 총 4 부분로 나누어, 각각 부분의 관계절 주어 뒤에 '-도', '-만', '-가', '-는' 각각의 조사를 첨가하여 수용성 판단 실험을 위한 예문을 구성하였다. 이과정에서, 각 문장이 서로 다른 세트에 속할 경우, 관계절 내부에 서로 다른 조사가 부여되어 들어갔다. 이런 식으로 구성된 예문들은 라틴 스퀘어(Latin Square) 방식으로 피험자들에게 제시되었다. 이후, 이 예문들을 가지고 44명의 성인 한국어 화자들에게 수용성 설문조사를 진행했다. 조사를 위해 7점 리커트 척도가 사용되었으며, 1은 그 수용성이 가장 낮은 것 7은 그 수용성이 가장 높은 것을 고르게 하였다. 진행 과정에서 90개의 필러 예문을 사용했다. 90개의 필러 예문은 수용성이 확실하게 좋을 것으로 기대되는 24개와, 반대로 수용성이나쁠 것으로 기대되는 예문 24개, 그리고 수용성이 이 두 예문의 사이 정도의 수준을 보일 것으로 예상되는 예문 42개로 구성되었다.

모델 실험을 위해서는 파일럿 실험을 통해 만장일치로 수용 가능한 것으로 평가된 60개의 문장의 내포절 주어 자리에 (2)와 같은 형태로 [MASK] 토큰을 사용했다.

(4) 그는 정부[MASK] 요구하는 활동을 설명했다. 가 는 도 만

이후, 이 60개의 문장으로 내포된 절을 추출하여, 원본 문장에서 관계절을 형성했던 문장을 통해 주절을 만들었다. 즉, 관계절 정보 구조에서 주제를 할당받을 것으로 예상되는 머리명 사를 주절의 목적어 자리에 위치하게 만든 것이다.

(5) 정부[MASK] 활동을 요구했다. 가 는 도 만

그렇게 함으로써, 주어의 조사가 내포절의 주어에 후행하는지, 모절의 주어에 후행하는지가 요인 설계로 사용되었다.

모델의 경우 KR-BERT(Lee, 2020)가 사용된다. 이 모델은 BERT를 기반으로 하지만 한국 데이터로 사전학습을 시행했다. 이 모델은 한국어 데이터를 기반으로 모델이기 때문에 베이스라인으로 선택되며, 다른 한국 BERT 모델과 비교해도 성능 면에서 비교우위가 있다 (박권식 외, 2021). 모델은 [MASK] 토큰을 각 조사, '가', '는', '도', '만'으로 대체하고 각각 조사의 놀라움을 측정한다(Hale, 2001; Levy, 2008). 놀라움은 주어진 문맥에서 단어의 음의로그 확률로 이해 어려움을 표현함으로써 인간의 인지 부하를 반영한다(Hale, 2001; Levy, 2008). 원래 목적은 인지 부하를 모방하기 위해 제안되었지만, 확률론적 모델을 활용하기 때

문에 언어 모델이 얼마나 '놀라운'지를 나타내는 척도로 머신러닝/딥러닝 영역에서도 자주 사용된다(Wilcox et al., 2019; Da Costa and Chaves, 2020). 즉, 언어 모델이 계산한 놀라움의 정도가 높을 경우, 주어진 맥락에서 해당 토큰이 기대되는 정도가 낮으며, 따라서 수용성 또한 상대적으로 낮을 것임을 예상할 수 있고, 반대의 경우에는 수용이 상대적으로 높을 것임을 기대할 수 있다.

4. 결과

4.1 수용성 판단 실험 결과

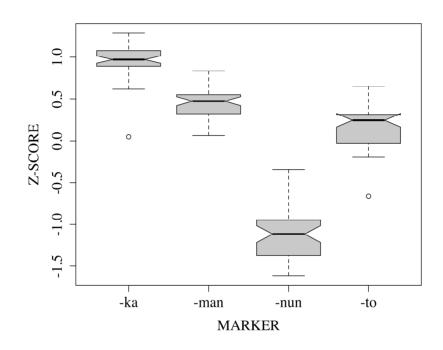


그림 1. 관계절 내부 조사 사용에 따른 한국어 화자들의 수용성 분포

그림1은 44명의 한국어 화자들이 관계절 내부 주어 자리에 각각 다른 조사가 사용된 문장들에 대해 보인 수용성을 보여준다. 일원분산분석에 따르면, 이 네 종류의 조사가 사용된 예문들의 경우 그 수용성에서 유의미한 차이를 보이는 것을 확인할 수 있다. (F(3, 176) = 548.9, p <.0001). 특히 조사 '-가'가 사용된 문장들의 경우, 다른 조사들이 사용된 문장들에 비해 유의하게 높은 정도로 높은 수용성을 보여주었다. '-만'과 '-도'의 경우, '-가'에 비해 상대적으로 낮은 정도의 수용성을 보이긴 했지만, 여전히 Z점수 0이상의 정문에 가까운 수용성을 보여주었다. 반면, '-는'의 경우 네 종류의 조사 중 가장 낮은 Z점수로 가장 낮은 수용성을 보여주었다. 즉, 결과를 요약하면, 주격조사 '-가', '-도', '-만'의 경우는 관계절 내부에 사용되어도 문장을 수용성을 크게 낮추지 않았다. 반면, 한국어의 주어짐성 주제 표지인 '-는'의 경우에는 관계절 내부에 나타날 경우 문장 전체의 수용성을 크게 낮추었다. 이

러한 결과는 관계절의 경우에는 머리명사가 관계사에 주제가 할당되기 때문에 별도의 주제가 관계절 내부에 형성될 수 없다는 기존의 분석과 일관된 결과라고 할 수 있다.

4.2 딥러닝 모델의 한국어 조사의 놀라움 측정 결과

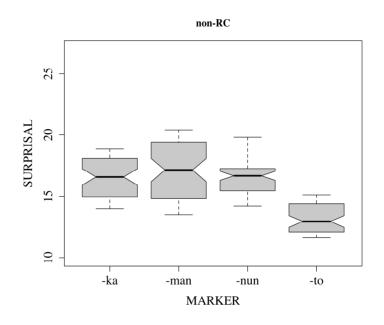


그림 2. 모절 주어 위치 사용된 조사들의 놀라움 분포

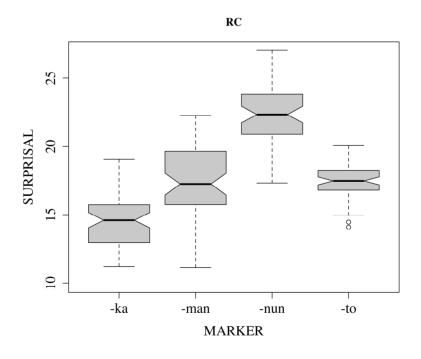


그림 3. 관계절 주어 위치에 사용된 조사들의 놀라움 분포

그림 3은 non-RC절, 즉 모절에서 조사를 화제 표지로 사용한 경우 KR-BERT가 보여준 각조사들의 놀라움 분포를 보여준다. 이 경우, 조사 사이의 차이가 유의하게 존재하기는 하지만(F(2,236) = 73.854, p <.0001), 이 유의성은 "-도'에만 국한된다. 나머지 조사의 경우 그차이가 유의하게 나타나지 않는다. 이러한 결과는, 주어진 실험 예문 중, 모절 예문의 경우, 경우 문두의 주어 뒤에 올 수 있는 모든 종류의 조사가 한국어 언어 모델에 의해 비슷한 수준으로 기대되며, 그 중 '-도'가 다른 조사들에 비해 유의하게 높은 수준으로 기대된다는 사실을 추론할 수 있게 한다. 이는 모절 실험 예문의 수용성이 사용되는 조사의 특징에 따라크게 달라지지 않을 타당한 예문임을 보여주는 것이다.

반면, 그림 4는 화제 표지가 관계절에서 사용될 때 조사의 놀라움의 분포를 보여준다. 이 경우에도 조사들 사이에 유의한 차이(F(3,236) = 153.94, p <.0001)가 존재하며, 화제 표지들이 모절에 사용된 경우와는 다르게 '-만'과 '-도'를 제외한 모든 조사들 사이의 차이가 유의하다. 이 때, 그림 1나.에서 '-가'는 모든 조사들 중 가장 낮은 수준의 놀라움 분포를 나타내고, '-는'은 가장 높은 수준의 놀라움 분포를 보여준다. 이 결과는 앞서 언급된 연구자들이 지적한 한국어 화제 표지의 관계절 내 화제 표지 형성 제약과 일관된 결과라고 볼 수있다.

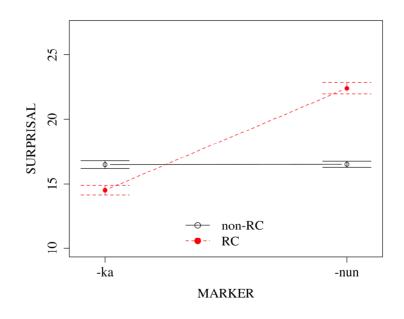


그림 4. 조사의 실현 위치와, 조사의 종류 사이의 상호작용 플롯

그림 5는 조사 '-가'와 '-는'이 사용된 경우의 상호작용을 나타내고 있다. 위 그림은 실험 예문에 사용된 '조사' 요인에 따른 효과와, 실험 예문이 모절이 경우와 관계절인 경우에 따른 효과가 모두 유의함을 보여준다. 또한, 두 요인 사이의 상호작용효과 또한 존재함을 확인할 수 있다. 그 안의 경향을 살펴보면, '-는'의 경우 관계절 내부에서 사용되었을 때 그

놀라움의 분포가 모절에 사용된 경우의 분포보다 높은 것을 확인할 수 있다. 반면 '-가'의 경우, 오히려 관계절 내부에 사용된 경우 그 놀라움의 분포가 낮게 나타나고 있다. 이러한 결과는 다시 한 번, 한국어에서는 일부 대조 초점의 경우를 제외하고는 화제 표지 '는'이 관계절 내부에 형성될 수 없다는 제약을 가진다는 선행 연구들(임동훈, 2012; Song 2014a)의 주장을 뒷받침한다.

5. 논의 및 결론

먼저, 실험 결과를 통해 한국어 화자들은 관계절 내부에 화제 표지 '-는'이 위치하 는 문장들에 대해 낮은 수용성을 보인다는 것을 알 수 있다. 앞서 결과에서 언급했듯이, 한 국어 성인 화자들은 실험에 사용된 '-가', '-도', '-만', '-는' 네 종류의 조사들 중, '-는'에서 가장 낮은 수용성을 보였고, '-는'에서는 모든 예문들이 0 이하의 Z점수를 보였다. 이 논문 에서 제시된 결과는 관계절의 경우에는 '핵어'에 주제가 부여되고, 그러한 이유로 관계절 내 부에 주제 표지 형성이 불가능하다는 주장이 완전하게 설명하는 근거로 작용하지는 못한다 는 한계를 지닌다. 왜냐하면, 첫 번째 한국어의 주제 표지 책략 중 하나인 강세와 같은 운율 적인 요소가 실험에서 배제되었다. 두 번째로, '-은/는'이 주어짐성 화제 표지가 아닌 대조 초점을 표시하기 위해 사용되는 경우에는 관계절 내부에도 사용이 가능하다는 주장과 관련 한 예문들은 함께 살펴보지 않았다. 즉, 관계절 내부 '-은/는'의 형성 제약에 있어 전체적인 현상을 살펴보도록 예문이 설계되지 못했다. 그러나, 이번 실험 결과가 관계절의 핵어와 관 련된 정보 구조의 성질을 밝히기에는 일부 부족함이 존재하지만, 관계절 내부에 한국어의 화제 표지 형성에 일정한 제약이 있다는 사실에 대해서는 확실한 근거를 제시한다. 전영철 (2009)은 정보구조에서 화용적이거나 의미론적 충돌이 발생할 경우 형태론적 차원에서 특정 조합이 발생하는 것을 방지한다고 주장한다. 즉, 실험 결과는 정보 구조와 같은 화용적, 의 미적 요소가 언어의 형태적 층위에 제약을 가하는 것을 보여주는 하나의 예시가 된다. 만약, 한국어 관계절에서는 머리명사에 주제가 할당되며, 한국어 관계절에서 화제 표시 '-은/는'이 사용되는 경우, 조사를 통해 표시된 화제가 정보 구조 내에서 이미 화제인 머리명사와 충돌 한다는 가설을 받아들인다면, 이러한 현상은 형태통사론적 층위 의미화용적 층위 사이의 상 호작용을 보여주는 현상으로 이해할 수 있다. 이러한 관점에서, 만약 언어 모델이 관계절 내 부의 화제 표지 형성 제약을 충분하게 표상해낸다면, 형태통사적 층위와 의미화용적 층위 사이의 상호작용을, 국소적인 현상이기는 하지만 표상하는 능력이 있다고 주장할 수 있을 것이다.

이어서, 한국어 언어 모델인 KR-BERT는 보조사를 통해 실현되는 문장 내적인 정보 구조와 관련된 언어 사용자들의 직관을 딥러닝 언어 모델이 잘 포착하고 있음을 보여준다. 이번 연구에 사용된 언어 모델에서의 놀라움의 분포는 인간을 대상으로 수용성 판단 실험과 상당히 유사한 경향을 보이고 있다. 놀라움의 경우 인간이 가지고 있는 수용성과 완전히 같 은 지표로 이해되지는 못하지만, 상당한 상관관계를 가지고 있는 것으로 이해된다(Wilcox et al., 2019; Da Costa and Chaves, 2020). 여기에 더해, 모델의 놀라움 분포에서 조사 '-가'가 가장 낮은 놀라움의 정도를 보이고, '-도'와 '-만'은 그 중간 정도, '-는'은 가장 높은 놀라움 정도를 보여주는 것은 한국어 화자가 네 종류의 조사를 사용한 문장에 보인 수용성의 분포 경향과 완전히 일치한다. 즉, 이러한 결과는 두 가지 함의점을 가지고 있다. 먼저, 놀라움이라는 지표가 한국어 화자의 수용성과 상당히 유사한 분포를 보인다는 것이다. 이러한 결과는 심리언어학적 관점에서의 언어 처리의 부하를 놀라움 지표를 통해 포착할 수 있다는 주장과 일관되다. 두 번째로 한국어 언어 모델이 관계절 내부의 조사 형성과 관련한 제약을 상당 부분 학습하고 있다는 사실을 보여준다. 이러한 결과는 앞서 관계절 내부 화제 표지형성 제약의 형태통사-의미화용적 함의점에 비추어 생각해 볼 때, 언어 모델이 형태통사적인 요소들의 분포를 통해 더 추상적인 의미화용적 구조인 '정보구조'를 표상하는 능력까지 갖출 수 있음을 의미한다. 이러한 맥락에서 본 연구는, 한국어 언어 모델이 형태통사적 층위를 통해 실현된 정보 구조를 표상할 수 있다고 주장한다.

참고 문헌

- 김일규, & 지윤주. (2020). 한국어 관계절 제약의 본질에 대한 실험연구. 언어, 45(4), 777-796.
- 김한샘. (2005). 현대 한국어 사용 빈도 조사2. 국립국어원
- 박권식, 김성태, 송상헌. (2021). 최소대립 문장쌍을 활용한 한국어 사전학습모델의 통사 연구 활용 가능성 검증. 언어와 정보, 25(3), 1-21
- 서민정. (2007). 명사토의 특수한 현상에 대한 일반론적 해석. 우리말글, 40, 77-96.
- 임동훈. (2012). '은/는'과 종횡의 의미 관계. 국어학(國語學) 64, 217.
- 장석진. (2000). 통합문법: 주제,초점과 시점. 언어와 정보사회, 1, 145-175.
- 전영철. (2009). '이/가' 주제설에 대하여. 담화와인지, 16, 217-238.
- Choi, H. W. (1996). Optimizing structure in context: scrambling and information structure. Stanford University. ERIC.
- Da Costa, J. K., & Chaves, R. P. (2020). Assessing the ability of transformer-based neural models to represent structurally unbounded dependencies. *Proceedings of the Society for Computation in Linguistics*, 3(1), 189-198.
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second Meeting* of the North American Chapter of the Association for Computational Linguistics. 1-8.
- Lee, S. Jang, H. Baik, Y. Park, S. & Shin, H. (2020). Kr-bert: A small-scale Korean-specific language model. *arXiv preprint arXiv*:2008.03979.
- Levy, R. 2008. Expectation-based syntactic comprehension. Cognition 106.3, 1126-1177.
- Song, S. (2014a). A grammar library for information structure. University of Washington. ERIC.
- Song, S. (2014b). Information structure of relative clauses in English: a flexible and computationally tractable model. *Language and Information*, 18(2), 1-29.
- Kuno, S. (1976). Subject, theme, and the speaker's empathy: a reexamination of relativization phenomena. In C. N. Li (Ed.), Subject and Topic (pp. 417-444). American Academic Press.
- Wilcox, E., Levy, R., & Futrell, R. (2019). What syntactic structures block dependencies in RNN language models? *arXiv preprint arXiv*:1905.10431.

- 134	-
-------	---

AI펭톡의 초등영어교육 적용 방안

허진영

I. 배경 및 목적

코로나19로 불가피하게 등교수업이 어려워질 때, 플립러닝(flipped learning)1), 블렌디드 러닝 (blended learning)2) 등으로 오프라인 수업뿐만 아니라 온라인 혼합 수업이 각광 받고 있다. 현재 수업받는 학생들인 MZ³⁾세대(MZ Generation)에게는 휴대폰, 컴퓨터 등의 디지털 환경이 익숙하고, 이를 통해 편리함과 재미를 추구한다. 이러한 상황에서 교사는 적절한 교육매체를 활용하여 효과적인 수업을 해야한다. 2022개정 교육과정 총론 시안 발표에서는 디지털 기초소양 함양 교육을 강화하였다. 뿐만 아니라, 정부는 2022년부터 2026년까지 5년간 디지털 인재 100만명을 양성하는 '디지털 인재양성 종합방안'을 발표하였다.

이러한 현실을 반영하여, 초등 영어 교과에서 AI펭톡(교육매체)를 활용하여 학생들에게 디지털 기초소양을 함양시키며, 개별 맞춤 학습을 할 수 있는 환경을 조성하고자 AI펭톡을 활용한 수업 적용 방안을 모색하였다.

AI펭톡	교육부와 한국방송공사(EBS)가 기획한 인공지능 기반 개인 맞춤형 영어 말하기 학습 서비스
교육매체 사용 이유	-무료로 사용할 수 있어 누구나 동일한 영어 학습이 가능함다른 기기가 필요하지 않고, 휴대폰과 PC로 구글 플레이 스토어/앱 스토어를 통해서 AI펭톡 앱(APP)을 설치하면 쉽게 사용할 수 있음학교 영어 출판사를 미리 설정하여, 교육과정과 연계한 학습 실현 -해당 주제에 다양한 활동들과 여러 단계가 있어서, 학생들이 자기주도적으로 자기 수준에 맞는 영어 학습 가능함펭수라는 학생들이 좋아하는 캐릭터 사용으로 영어 학습에 대한 자신감과 흥미를 부여함학생들에게 좀 더 전문적으로 개별 피드백이 가능(억양, 발음, 강세, 발전 향상, 원어민 비교 등)
의사소통 역량 향상	-매주마다 랭킹(별점, 마일리지)를 학생들이 확인할 수 있어 선의의 경쟁 을 하며 학습 동기를 부여함. -AI와 친구가 되어 함께 의사소통할 수 있음. -자신의 목소리를 녹음하고, 듣는 과정에서 부족한 부분을 확인하고 연습 할 수 있음.

¹⁾ 온라인을 통한 선행학습 이후 오프라인 강의를 통해 교수와 토론식 강의를 진행하는 역진 행 수업 방식

²⁾ 온, 오프라인 학습을 결합한 학습 방법

^{3) 1980}년대 초~2000년대 초 출생한 Millennials세대와 1990년대 중반~2000년대 초반 출생한 Z세대

Ⅱ. 학교적용 방법

1) 수업 전









- -가정통신문 「AI펭톡 안내 및 서비스 이용 동의서」를 배부함.
- -학생 목소리 녹음, 학생 개인정보 수집 등으로 AI 펭톡 서비스 이용 동의서 회신받음.
- -인공지능 기반 AI 개인 맞춤형 영어 말하기 학습 서비스의 효과도 함께 안내하여 학생들이 스스로 활용할 수 있도록 자세히 안내함.
- -AI펭톡 앱(APP)을 설치하는 방법을 명시하여 가정에서도 학습할 수 있도록 함.

② LCMS 관리자로 회원 코드 배부 및 학생 숭인

				-44.4.4.4.4				▶ 검색건	선수 : 3	48건	
학교영문값	3학년 출판사	4학년 출판사	5학년 출판사	6학년 출판사				No		학생코드	상세정보
CLD	VDM/7I\	VDM/71\	rila	da	#ITIM	AZG 14	110	1		21CLR00001	상세정보
CLR	YBM(김)	YBM(김)	대교	대교	이신성		348	2		21CLR00002	상세정보
V			H7	1:1-1건수:1건				3		21CLR00003	상세정보
<u> </u>				1.1-107.10				4		21CLR00004	상세정보
각	각 학년 영어 출판사 설정				선생님&학생 매핑		학생 회원 코드 부여				

2) 수업 중

①AI펭톡 활용 방법

소개



〈AI펭톡 주요 기능 설명〉

-토픽월드: 교육과정 기반한 수준별 학습

-스피킹: 실용 영어 말하기

-렛츠톡: 펭수와 자유롭게 대화하기

-스캔잇: 사진을 찍어 사물을 인식한 영어 단어 학습

-스쿨톡: 반 친구들과 모둠별로 자유롭게 토론

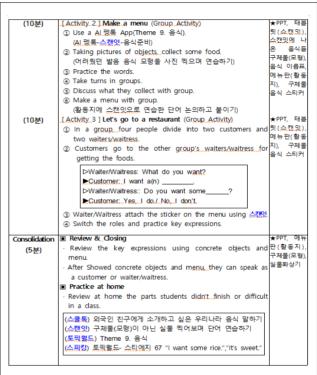
③ 교육과정에 적용한 AI펭톡 활용 우수 사례



-대상: 4학년 2반 학생들 〈학습목표〉

토픽월드와 스피킹을 통해 자기 수준에 맞게 표현을 연습하고, 스캔잇으로 자신에게 필요한 음식 표현을 묻고 답할 수 있다.

할습과정	교수 학습 활동	Materials (★) & Teaching Aids (*)
Introduction (5분)	■ Warm up & Motivation ○ Using the △是馬, checked the homework. · Thems, Which Korean food do you want to introduce to a foreign friend? ○ Show EBSe, "Who am I?" · we can talk about robot waitress which can connect AI思惠 and restaurant. · EBSe, contents; Drawing 68. He's, putting milk in the coffee. ■ Clarify Learning point · Students guess what we learn about. We can talk about what food I want with AI思惠.	★PPT, 슸쿨 톺Feedback EBSe콘텐츠 (Who am I?), 미러링
Development	[Activity1.] Food(WholeActivity)	★PPT, 태블
(10분)	O Introduce on words and expressions. Learnwords and expressions practically using concrete objects. O Dialogue practice with AI 정통 -Practice the dialogue with AI 정통 They can Listen and speak. Also They can hear their recording voice through AI. ③(Topic world)-Theme 9. 음식 ②(Speaking) 토핑물도-Stage 67 "I want some rice.","It's sweet."	混)



3) 수업 후



- -EBSe에서 학생 개인별 학습 시간, 활동 내용 등 활용 방법을 안내하여 학생 개별 피드백을 할 수 있도록 함.
- -매주 랭킹 (마일리지, 별점)을 학생 스스로 확인하고, 더 열심히 학습 할 수 있도록 함.
- -EBS에서 매달 성과 이벤트 결과를 안내하고, 상품을 수여하여 더 열심히 영어 학습할 수 있도록 계기를 부여함.
- -비밀번호 초기화 등 학생들이 AI펭톡을 하면서 어려운 부분에 대한 의견을 수렴하고 방법을 안내함.

Ⅲ. 결과



AI펭톡을 활용한 개별맞춤학습으로 학생들의 기초적인 영어 의사소통 능력을 향상 시키고자 AI펭톡을 활용한 영어 수업 적용 방안을 모색 하였다. 학생들에게 친숙한 기기를 활용하여 영어 노출 시간을 늘릴 수 있는 환경을 조성하였다. 또한, Affective filter(감정적 여과장치)를 낮춤으로써 자신감과 흥미를 갖고 즐겁게 영어 학습에 참여하는 모습을 확인할 수 있었다. 본 연구를 통하여 다음의 결론을 얻을 수 있다.

AI 펭톡 카테고리	적용 대상	적용내용	적용결과
스쿨톡	초4	■ Introduction -사전에 스쿨톡으로 "Which Korean food do you want to introduce to a foreign friend?" 주제를 EBSe 영상과 함께 올리고 그룹별로 학생들이 자유롭게 답변할수 있도록 함. -전체적인 피드백을 교실에서 하고, 개별 피드백을 가정에서 확인하도록 함.	-3학년 교육과정과 연계한 체계적인 지도를 위하여 3학년에서 배운 내용을 적용하여 스쿨톡에서 복습하도록 함스쿨톡에 올라온 학생들의 답변으로 호기심유발함개별 피드백으로 가정에서 지속적으로 확인하며 영어 학습할 수 있도록 함.
토픽월드	초4	 ■ Activity1 - "Theme 9. 음식" 활용함. -주요 표현을 AI펭톡으로 전체 연습한 후, 학생 태블릿으로 개별 연습함. 	-학생들이 자기주도적으로 수준에 맞는 표현 활동을 선택하여 연습함자신의 목소리를 녹음하며 들으며 어려운부분을 스스로 파악함.
스피킹	초4	 ■ Activity1 - "Stage 67 I want some rice./ It's sweet." 활용함. -주요 표현을 AI펭톡으로 전체 연습한 후, 학생 태블릿으로 개별 연습함. 	-자기주도적으로 활동을 선택하여 좀 더 영어에 흥미를 갖고 수업에 참여할 수 있음수업 시간 내에 마치지 못한 내용과 어려웠던 부분들을 가정에서 연결하여 학습하도록 함.
스캔잇	초4	■ Activity2 & Activity3 -스캔잇 "Theme 9. 음식" 활용함. -구체물을 스캔잇으로 사진을 찍어 음식을 수집하고, 모둠 친구들과 의사소통을 통해 메뉴판을 만듦. -스캔잇으로 만든 메뉴판을 활용하여 실생활 (식당)에 적용하여 역할 놀이하며 필요한 음식을 문고 답하기 연습을 함.	-구체물(모형)을 스캔잇으로 듣고 말하기 연습하며 수 업에 대한 재미와 호기심을 유발함 -학생이 어려웠던 부분이나 관심 가지고 있는 단어들을 자기주도적으로 학습함구체물을 스캔잇하며 해당 영어를 학습하는 활동 자체에 신기해하며, 적극적인 태도로 수업에 참여함궁금했던 내용을 AI가 응답해주면서 개인수준에 맞게 효과적인 학습을 할 수 있음.

영어 공부에 대한 의욕은 있으나 가정형평상 사교육 접근이 어려운 학생들 그리고 영어 흥미도가 낮은 학생들에게 영어에 대한 흥미 증진 활동들은 더욱 필요해 보였다. AI펭톡을 활용한 수업은 자기주도적 학습 및 개별화 학습 그리고 학생 수준별 학습의 새로운 대안이라고 생각된다. 그러므로 초등영어 교육에서 AI펭톡을 활용하고 적용할 수 있는 다양한 방법들이 앞으로도 모색되어야 할 것이다. 이와 더불어, 교사들에게 AI펭톡 활용 수업 연수를 제공하여 학교 현장에서의 AI펭톡 수업 활용도를 높이는 지원이 수반된다면 우리가 꿈꾸는 행복한 영어 학습이 실현될 것이다.

한국영어학회(KASELL)-한국코퍼스언어학회(KACL)

2022년 공동 가을 학술대회

일반 자유주제

1. Aphiwit Liang-Itsara (Mahidol University, Thailand)

The Comparison of Motivating Factors for Taking English and Other Foreign Languages among Thai College Students

2. **박슬기**. **김희연**. **김종복** (경희대)

English although-stripping constructions: A corpus-based investigation

3. Hyun Jong Hahm (Univ. of Guam, USA)

Guamanian English and Englishes within

4. 윤인경, 유석훈 (고려대)

A Study on the Readability of American Textbooks and Korean English-Language Textbooks Comparing Conjunction Frequency in Sentence Structures

5. 한국영어학회 『영어학』 최우수논문 발표: 임수연(한양대)·안현기(서울대)

Effects of Parts of Speech on Implicit Prosodic Prominence by Native English Speakers and Korean Learners of English

6. **강남길** (극동대)

On Consist of and Made up of: A Corpora-based Analysis

7. **강아름** (충남대)

부정극어 wh-so-ever의 의미 분석: 허가 조건과 코퍼스 분석 (K)

8. **이혜진** (원광대), **이제영** (전주대)

중등교사 영어 임용시험 수험자를 위한 영역별 전문어휘 목록 선정 및 지도방안 탐색

9. Cuilin Liu and Se-Eun Jhang (Korea Maritime & Ocean University)

A corpus-based study of gender-inclusive words in the maritime conventions of "seamen" or "seafarers"

10. 최예슬 (대덕고), **민수정** (공주대)

우리나라 고등학교 실용영어 교과서 비판적 담화분석

11. 김정렬 (한국교원대)

북한 영어교과서에 나타난 사회주의 교육 테제

12. Kevin Parent and Quynh Thuy Bui (Korea Maritime & Ocean Univ.)

Thematic Choice in Argumentative Essays by Vietnamese EFL Learners

13. **최아녕** (연세대)

A Critical Discourse Analysis on the Presidential Discourse: Focusing on Volodymyr Zelenskyy's Linguistic and Discoursal Strategies for Persuasion

14. 안유영 (전남대)

Metaphorical Perceptions of Chinese Engineering Students Towards English Learning in EMI Context

15. **소지연** (연세대)

The Development of Generative Grammar -Focusing on Minimization

16. **이주원** (전주대)

A New Type of Resultative Construction in English

17. Suyoung Bae and Myung-Kwan Park (Dongguk University)

Resolving some puzzles in Locative Inversions of English



The Comparison of Motivating Factors for Taking English and Other Foreign Languages among Thai College Students

Aphiwit Liang-Itsara

The Graduate Program in Applied Linguistics
Faculty of Liberal Arts, Mahidol University, Thailand
Email: aphiwit.lia@Mahidol.edu



Introduction

- Motivation is an essential factors that affect both processes and results in acquiring a foreign language (Dörnyei, 2005), and has known to be on of the best predictors in conducting effective teaching and learning.
- Research on second or foreign language motivation has been based primarily on ESL or EFL, leaving large number of other languages from the field of study (Boo, Dörnyei & Ryan, 2015; Dörnyei & Al-Hoorie 2017; Ushioda & Dörnyei, 2017; Ushioda, 2017; Yang & Chanyoo, 2022).



Introduction

- Using the Second Language Motivation
 System (L2MSS) as a theoretical framework,
 studies on motivation in learning languages
 revealed different results.
 - Idea L2 Self as the strongest predictors for English and German in Hungary (Csizér & Lukács, 2010) and English and Mandarin in Hong Kong (Dörnyei & Chan, 2013)
 - Ought-to L2 Self for L3 Japanese and EFL in most countries
 - L2 Learning Experience for languages other than English (LOTEs; Ushioda & Dörnyei, 2017; Yang & Chanyoo, 2022).



Study Objectives and RQs

- 1.To examine motivational factors for college students in learning English;
- 2.To examine motivational factors for college students in learning English and LOTEs; and
- 3. To identify the differences (if any) of the motivational factors for college students in learning English and LOTEs.

RQ1: What are the motivational factors for college students in learning English?

RQ2: What are the motivational factors for college students in learning LOTEs?

RQ3: Are there any differences in the L2MSS factors for college students learning English compared to LOTEs?



Conceptual Framework

Motivation in learning English

Ideal L2 Self Ought-to Self Past L2 Experience



Motivation in learning LOTEs

Ideal L2 Self Ought-to Self Past L2 Experience



Methodology

Design: mixed method: questionnaire and interview **Participants:** 173 undergraduate students taking one of the five languages, including English, French, Chinese, Japanese, and Korean.

	No. of	No. of
Language	participants	interviewees
English	31	5
French	24	5
Chinese	42	5
Japanese	36	5
Korean	40	5
Total	173	25



Research Instruments

1. Questionnaire: The questionnaire was based on the L2MSS (Dörnyei, 2005; You & Dörnyei, 2016; You, Dörnyei, & Csizér, 2016), consisting of three parts:

Part 1: consisted of forty-two 6-rating scale questions asking about L2MSS. Among all question items, 15 asked about the Ideal L2 Self, 20 about the Ought-to L2 Self, and 7 about the Past L2 Learning Experience.

Part 2: five demographic questions asked about their personal information such as gender, language taken, age, and experience in living in a target language environment or studying with a native speaker to the target language.

2. A semi-structured interview. The interview lasted for ten minutes for each session3. Validity and reliability of the instruments



Validity and Reliability

- Three experts were asked to validate the question items adapted from the original version.
 - The index of item-objective congruence (IOC) was calculated for each item. With minor suggestions on clarification of language use, all question items were achieved 1.0 on the IOC score.
- The interview questions were validated and approved by three experts in the field. The experts advised no major revision.
 - Internal consistency using the Corrected Item Total Correlation (CITC) method was employed to test for reliability of the questionnaire and achieved a Cronbach's alpha score of .937 for the set.



Data Collection

- After getting approval from the Institutional Review Board and done with the instrumentation process, the researcher posted an invitation of participation on an online social platform and asked assistance from teachers in respective foreign language classes to spread the call for participation.
- All participants agreed to participate voluntarily. The participants completed an online version of the questionnaire. Those who agreed to join the interview were asked questions via a voice call on a social media application. All participants were given a small token of appreciation worth about US\$6.



Data Analysis

- The quantitative data were analyzed by frequency, percentage, mean, and standard deviation for L2MSS factors within each language group. A one-way and two-way Analysis of Variance (ANOVA) was used to identify differences in the L2MSS factors across languages.
- Content analysis was used to analyze interview data. A graduate student in applied linguistics was asked to verify the interview transcription to assure the reliability of the coding.



Results: RQ1

- Thai English learners reported the Ought-to L2 self as the strongest motivational factor for them in taking English course (M = 5.12; SD = .55), followed by L2 Learning Experience (M = 4.97; SD = .73), and Ideal L2 Self (M = 4.23; SD = .63) respectively.
- There was a statistically significant difference among sources of motivation, as determined by one-way ANOVA (F(2, 90) = 16.923, p = .000). A Dunnett T3 post hoc analysis revealed that the average motivational scores of the Ideal L2 Self were significantly lower than the Past L2 Learning Experience and the Ought-to L2 Self at ps<.05. No statistical significance was found between the Ought-to L2 Self and the Past L2 Learning Experience (p = .768).



Results: RQ2

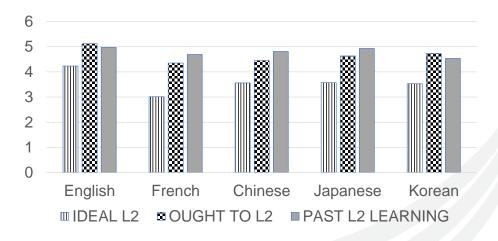
MEAN SCORES OF MOTIVATIONAL FACTORS FOR STUDENTS IN TAKING COLLEGE LOTE COURSES

		Ideal L2-	Ought-to L2	Past L2
Language	n	Self (SD)	Self (SD)	Learning
				Experience
				(SD)
French	24	3.01 (.78)	4.35 (.58)	4.69 (.81)
Chinese	42	3.56 (.98)	4.45 (.69)	4.80 (.75)
Japanese	36	3.57 (.64)	4.63(.44)	4.93 (.79)
Korean	40	3.53 (.71)	4.73 (.62)	4.53 (.97)
Overall	142	3.60 (.84)	4.66 (.63)	4.78 (.85)



Results: RQ3

Overall motivational factor for Thai undergraduate students to take foreign language courses was Past L2 Learning Experience (M = 4.78; SD = .85), followed by Ought-to L2 Self (M = 4.66; SD = .63), and Ideal L2 Self (M = 3.60; SD = .84), respectively.





Results: RQ3

- The LSD pairwise comparisons were performed to identify different levels of overall motivational factors across five target languages. The highest motivational score in taking college foreign language courses was English (M = 4.78; SD = .74), followed by Japanese (M = 4.38; SD = .86), Chinese (M = 4.27: SD = .96), Korean (M = 4.26; SD = .97), and French (M = 4.01; SD = 1.03), respectively.
- Motivational scores of undergraduate students who took Chinese, Japanese and Korean were significantly higher than those who took French courses, *p*s < .05.



Discussion: RQ1

- the Ought-to L2 Self as the significant motivational factor for the participants to study English, followed by the L2 Learning Experience and the Ideal L2 Self.
- They needed to achieve a certain level of proficiency to graduate, continue their studies, or get a well-paid job. Moreover, few participants reported that their parents and significant people expected them to be good at English, so they need to perform well. The findings are in line with those of Huang et al. (2015), Sugita McEown et al. (2017), Kong et al. (2018), and Mendoza & Phung (2018) in the Asian context.
- Asian students did not have a solid Ideal L2 Self in studying a foreign language. They were not sure why they needed to motivate themselves in studying languages (Sugita McEown et al., 2017).



Discussion: RQ2

- The Past L2 learning experience was the most potent motivating factor for Thai students to study the LOTEs, followed by the Ought-to L2 self, and the Ideal L2 Self (except for Korean students).
- Previous studies in the EFL context also revealed the importance of enjoyment in students' courses.
- For example, Saito, Deweale, Abe, and In'nami (2018) emphasized joyful learning experiences that influence students' motivation. They also furthered that anxiety negatively affected motivation, placed on the other end of enjoyment.



Discussion: RQ3

- The Past L2 Learning Experience became the primary motivational factor for learners to keep studying a foreign language, followed by the Ought-to L2 Self, and the Ideal L2 Self.
- It is not surprising that English language learners reported the Ought-to L2 self because the English language assumed its position as a mandatory language subject for Thai students.
- The current findings clearly showed that the Past L2 Learning Experience was a primary motivating source for students who had taken or were taking LOTEs.
- Past L2 Learning→ Ideal L2 Self = proficiency



Teaching Implications

- "Positive emotions and attitudes can make language learning more effective and enjoyable."
- Building a supportive learning environment and psychologically secured atmosphere in the classroom may reduce students' negative feelings and boost positive learning experience the students.
- The practice aligns with Stephen Krashen's affective filter (1986) that the learners would engage more in the language learning classes when their anxiety level was low.
- Ni (2012) also mentioned that success in language learning occurs in the atmosphere in which learners' values and manners are supported, learners take a confident and enjoyable stance towards language. They would thus utilize the target language comfortably. That will eventually contribute to their confidence to use the TL.



Recommendations for future studies

- Future studies should expand to investigate motivational components for students in taking other foreign languages such as neighboring and other modern languages offered in schools or universities. Understanding sources of motivation will benefit foreign language teachers to prepare their classes more effectively.
- The variable of the length of time for taking a foreign language class should be considered a significant motivational factor for students to take language classes.
- Future studies should investigate whether the findings from this study hold or there may be some other factors provided by foreign language learners.



Conclusion

- The current mixed-method study investigated different motivating factors of Thai undergraduate students who took foreign language courses at a Thai university in the Greater Bangkok area.
- The findings revealed different motivating sources for the students who had taken or were taking foreign language courses. The positive <u>L2 Learning</u> <u>Experience</u> is directly related to the strong perception of <u>the Ideal L2 Self</u>.
- Therefore, the study suggested that teachers provide an adequate and supportive learning experience to foreign language students, leading to higher motivation and engagement in foreign language classrooms.

English although-stripping constructions

A corpus-based investigation

Seulkee Park, Hee-Yeon Kim, and Jong-Bok Kim (seulkeepark@khu.ac.kr & heeyeonkim@khu.ac.kr & jongbok@khu.ac.kr) Kyung Hee University

2022 Fall KASELL-KACL Joint Conference October 22, 2022 Pusan University

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
●0000000	0	00000000000	000000	000	0	00

Stripping

- Stripping (or 'Bare Argument Ellipsis') occurs when one element remains in a clauses and other components are not expressed, where it is in the coordinated structure with its antecedent clause (Hankamer and Sag 1976).
 - (1) a. Lauren can play the guitar, and Mike, too.
 - b. Lauren can play the guitar, **but not Mike**. (Merchant 2013: 20)
- Stripping generally occurs in the coordinate structure, and a coordinating conjunction 'and' with a parallel meaning or a connective with a contrast meaning such as 'but' must appear in the stripped clause:
 - (2) a. Parallel: A and B too
 - b. Contrast: A but not

Park & Kim (KHU)

English although-stripping constructions

•

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
0000000	0	00000000000	000000	000	0	00

Concessive (al)though-stripping constructions

- However, a subordinating conjunction *although* can also have a fragmental remnant as in the coordinating structure with two sub-patterns:
 - (3) **Typical** although-stripping
 - a. That's possible to do, although tricky.
 - b. That's possible to do, although that's tricky to do.
 - (4) **Negative** although-stripping
 - a. Abby speaks passable Dutch, although not Ben. (Merchant 2003: (38))
 - b. Abby speaks passable Dutch, although Abby does not speak passable Dutch.
- Ellipsis in *although*-stripping is not obligatory, which allows its non-elided counterpart to be grammatical.

Park & Kim (KHU)

English although-stripping constructions

2

References

Embedded context

- As in the coordinated stripping, although-stripping also does not occur in the embedded context.
 - (5) *John was eating natto, but I think (that) **not sushi**. [Stripping] (Yoshida et al. 2015: 326)
 - (6) *John was eating natto, but I think (that) although not sushi.
- When the antecedent clause and the stripped clause are coordinated, they can be embedded together (Merchant 2003):
 - (7) Jack regretted that Abby got to go,
 - a. and Ben, too. (= and Ben got to go, too)
 - b. but not Ben. (= but Ben didn't get to go) (Merchant 2003: (22)-(23))
- The distribution also applies to subordinate although-stripping:
 - (8) All argue **that although difficult**, it might be possible to create a supranational monetary system.

(= although it is difficult to create a supranational monetary system)

Park & Kim (KHU)

English although-stripping constructions

Sentential approach of subordinate if-stripping

- Myers and Yoshida (2018) argue that subordinate conjunction *if* can have a stripping construction, as in *although*-stripping:
 - · if is a complementizer in a stripped clause
 - Focus movement and clausal ellipsis:
 The remnant moves to specifier of FocusP in a SplitCP model and if-stripping undergoes a clausal ellipsis of lower CP projection (FinP), rather than TP.
 - (9) John likes to drink whiskey. **If scotch**, I'll pour him an Islay. [ForceP If FocusP scotch Finn Try John ForceP If FocusP scotch] ...]]]]

Park & Kim (KHU)

Subordinate stripping

English although-stripping constructions

Data discussion

Theoretical implications

7

References

0000•000 0 0000000000 000000

Arguments for sentential approaches

Research questions

- The connectivity effects support the argument that *if* (*not*)-stripping has the identical syntactic structure with the putative source clause.
- Binding Condition A (Myers and Yoshida 2018:2)
 - (10) a. A: John; is criticizing someone. B: **If himself**; it is unfair.
 - b. A: Johni's friends are criticizing someone. B: *If himselfi, it is unfair.
- Binding Condition C (Myers and Yoshida 2018:3)
 - (11) a. She; was selling some pictures. *If of Mary;, her mother will be upset.
 - b. Her_i sister was selling some pictures. If of Mary_i, her mother will be upset.

Park & Kim (KHU)

English although-stripping constructions

Ę

Connectivity effects of although-stripping

- The remnants in *although*-stripping show grammatical dependencies on missing linguistic material based on their putative sources.
- Case-matching
 - (12) Some, **although not** [NP[nom] I], may be drawn to the implicit neo-political message in its subtitle, "The Value of Values." (2008 FIC)
- · Subcategorization requirement
 - (13) a. The lack of logical or moral consistency in their own positions is vexing, although not surprising. (2012 BLOG) VFORM[ing]
 - b. The weight of managing all the needs of all the kids in the orphanage has been relieved, **although not forgotten**. (1991 NEWS) VFORM[en]
- · Preposition stranding/identity
 - (14) a. **Although not a Quaker,** he is listed as an attendant at the Green Street Meeting House in Philadelphia. (2005 MAG)
 - b. Schmidt eventually moved back to the Volga Basin, although not to Saratov. (1993 NEWS) PFORM[to]

Park & Kim (KHU)

English although-stripping constructions

6

Subordinate stripping OOOOO●O

Research questions

Corpus investigation

Data discussion

Theoretical implications

Summar

References

Non-negated although-stripping

- Merchant (2003) suggests that the Neg head in stripping has a lexical specification checked by 'but, if, C[Q,wh], or' and this selects for an FP with an E feature.
- The distribution with the negator also can be applied with a licensor '(al)though' in stripping.
 - (15) a. *Abby speaks passable Dutch, (al)though Ben too.
 - b. *Abby speaks passable Dutch, (al)though Ben.
 - c. Abby speaks passable Dutch, (al)though not Ben.

(Merchant 2003: (38))

- Merchant (2003) argues that stripping structure with (al)though is only allowed with a negator 'not'.
- He also states that the positive (al)though-stripping example (19b) has narrower distribution than its negative counterpart (17c)¹.

¹As observed later in this research, the remnant NP has significantly lower frequency in positive *although*-stripping than in negative one from the extracted corpus data.

Polarity reversal and parallelism

- Wurmbrand (2017) assumes that although is a coordinator like 'but' rather than a subordinator, which requires a contrast between the two conjuncts and shows polarity differences.
- She also suggests that the remnant XP combining with *although* in FocP counts as a phase where its complement allows stripping under the domain;
- The stripping with *although* is only possible where the context allows both polarity reversal and parallelism.
 - (16) a. [FocP Abby [TP[-neg] tabby speaks passable Dutch]] although

 *[FocP Ben [TP[-neg] taben speaks passable Dutch]] *polarity

 *[FocP Ben [TP[-neg] taben doesn't speak passable Dutch]] *parallelism

 b. [FocP Abby [TP[-neg] tabby speaks passable Dutch]] although

 [FocP not Ben [TP[-neg] taben speaks passable Dutch]] ✓ polarity, parallelism

 (Wurmbrand 2017: (38))

Park & Kim (KHU)

English although-stripping constructions

Q

Subordinate stripping Research questions Corpus investigation Data discussion Theoretical implications Summary References 00000000 000 000 000 000 000

Research questions

- To look into the grammatical properties and distributional environment of concessive although-stripping constructions
- To identify how ellipsis is licensed in the stripped clause
- · To account for although-stripping based on the authentic data findings

Park & Kim (KHU)

English although-stripping constructions

ç

Methods

Data and tokens

- Observation with a web-based corpora COCA (Corpus of Contemporary American English)
- Collected approximately 350 tokens of randomly selected examples from spoken and written registers (e.g. fiction, news, academic, etc.)

Searching strings

- Subordinating conjunction although with two sub-patterns:
 - (17) a. although [remnant XP] (PUNC) [145 tokens]
 - b. although not [remnant XP] (PUNC) [205 tokens]
- Other concessive subordinate conjunctions such as though, while, or if were not used.
- · Considerations when selecting the remnant:
 - Definiteness and various forms of the phrases
 - Length of phrases for the grammatical weight
- Additional adverbial expressions are only allowed in the remnant.

Park & Kim (KHU)

English although-stripping constructions

1

Data variables

- · Variables are tagged on each example based on the form, function, and usage.
 - Remnant type based on the syntactic categories and functions (e.g., NP[nom], V[aux], VP[pass], quantifier, ...)
 - correlate type to demonstrate the relation with the remnant (e.g., Correlate with NP[nom] & its remnant with NP[acc]: case mismatch)
- **Reconstruction**: Recovering the elided part in *if* (*not*)-stripping from the explicitly or implicitly suggested context to identify the syntactic or semantic identity
 - Syntactic match/mismatch (tense, category, case, voice, polarity, finiteness, modality, and verb forms)
 - Syntactic (non-)connectivity (binding condition, island constraint, preposition identity, etc.)

· Overtness of the correlate

- Overt: The correlate of a remnant is overtly expressed in the antecedent.
- · Covert: The correlate can be covertly represented.
- Exophoric: With no correlate, the remnant can be reconstructed with an implicit argument or with the surrounding context.

Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping Research questions Corpus investigation Data discussion Theoretical implications Summary References

Remnant types (cont'd)

• A remnant in *although* (*not*)-stripping is restricted to phrasal AP, NP, VP[*en*], VP[*ing*], PP, or AdvP, or lexical adjective or quantifier.

- (18) a. **Although [**_{AP} **bright]**, Wes was never really challenged at the Bronx schools he attended. (2001 MAG)
 - b. We grew up across the street from each other same community and comparable income levels, although [_{NP} different schools]. (2012 BLOG)
 - c. The words were in English, although [vp[en]] heavily accented]. (2013 FIC)
 - d. Things are still a bit shaky, although [vPlina] getting better]. (2012 BLOG)
 - e. The blood was stopped, although [$_{\mathbf{pp}}$ with difficulty]. (2012 WEB)
 - f. Although [_AdvP frankly], the guy could have run away. (2010 SPOK)
- (19) a. I eventually produced a decent, **although** [Adj tentative], paper. (1993 MAG)
 - b. Inimical to Opposition Earth and many Elementals, although not [Quantifier all]. (2019 FIC)

Park & Kim (KHU)

English although-stripping constructions

1

Remnant types

- To look into the internal and external relations with antecedent, syntactic category was tagged on each remnant.
 - (20) a. They did have a chapel, although [_Ap small]. AP (2006 MAG)
 - b. The Collins-Hochul race happened also in the region,
 although not [_{NP}the city]. NP (2012 WEB)
 - c. That was just a rumour, **although** [VP[en] well-founded]. VP[EN] (2016 MOV)

Table 1: The distributions of syntactic categories of remnants of although-stripping (raw frequencies (%))

		Syntactic category of remnants						
	AP	NP	VP[en]	VP[ing]	AdvP			
although	66 (45.5%)	5 (3.4%)	15 (10.3%)	7 (4.8%)	39 (26.9%)			
although not	53 (26.3%)	47 (22.9%)	16 (7.8%)	2 (1%)	54 (26.3%)			
	PP	proform	quantifier	polarity particle	TOTAL			
although	7 (4.8%)	2 (1.4%)	2 (1.4%)	2 (1.4%)	145 (100%)			
although not	23 (11.2%)	o (o%)	10 (4.9%)	o (o%)	205 (100%)			

Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
00000000	0	000000000000	000000	000	0	00

Overt correlate

- A remnant in *although*-stripping can have its overt correlate in the antecedent, which can only be in the contrastive relation.
 - (21) **Although bright**, he was insecure, melancholy, and introverted. (1994 MAG)
- The remnant in negative *although*-stripping seems to be contrastive with or repeated from its overt correlate with broader spectrum.
 - (22) a. I think it's a disgrace that you've taken this stance, although not a surprise. (1997 SPOK)
 - b. She needs to be careful, although not this careful. (1993 FIC)
- However, the repeated and coreferential correlates of negative *although*-stripping are conclusively in a contrastive relation due to the negator 'not'.
- Therefore, the remnants both in positive and negative *although*-stripping is only allowed to be contrastive with its correlate.

Park & Kim (KHU)

English although-stripping constructions

1

Subordinate stripping Research questions Corpus investigation Data discussion Theoretical implications Summary References

Correlate types

- Syntactically (mis)matching relations can be identified by remnant's corresponding correlate type, when it is overtly represented.
 - (23) a. Their miniature portraits, **although rare**, are also of great quality.

 AP-PP (2009 FIC)
 - b. It will be difficult for him to lead, although not impossible. AP-AP (1998 SPOK)

Table 2: The distributions of syntactic categories of correlates in *although*-stripping (raw frequencies (%))

		Syntactic category of correlates							
	AP	NP	VP[en]	VP[ing]	AdvP				
although	18 (12.4%)	4 (2.8%)	1 (0.7%)	1 (0.7%)	2 (1.4%)				
although not	46 (22.4%)	32 (15.6%)	11 (5.4%)	3 (1.5%)	12 (5.9%)				
	PP	S	quantifier	Overt correlates	TOTAL				
although	2 (1.4%)	2 (1.4%)	4 (2.8%)	34 (23.4%)	145 (100%)				
although not	11 (5.4%)	3 (1.5%)	7 (3.4%)	125 (61.0%)	205 (100%)				

Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping Research questions Corpus investigation Data discussion Theoretical implications Summary References

Match / Mismatch variable

- We also looked into any possible mismatches between antecedents and ellipsis sites.
 - (24) a. This is largely true, although not completely. MATCH (2006 MAG)
 - b. I was a cyclist for many years, **although not any longer**. MISMATCH TENSE (2012 WEB)

Table 3: The distribution of match and mismatch between a remnant and its overt correlate (raw frequencies (%))

		although	although not
Match		70 (82.4%)	155 (82.5%)
	category	12 (14.1%)	32 (16.9%)
	finiteness	2 (2.4%)	0 (0.0%)
Mismatch	voice	1 (1.2%)	0 (0.0%)
	tense	0 (0.0%)	1 (0.5%)
	agreement	0 (0.0%)	1 (0.5%)
Total		85 (100%)	189 (100%)

Park & Kim (KHU)

English although-stripping constructions

16

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
0000000	0	000000000000	000000	000	0	00

Covert correlate

- · However, the remnant can also have a covert one in the antecedent clause.
 - (25) a. They did have a chapel, **although small**. (2006 MAG) (= although they did have a small chapel)
 - b. I realized it too, although later. (1995 FIC)(= although I realized it later)
 - (26) a. The word absconded contains five letters of the alphabet, A-B-C-D and E, in order, **although not consecutive**. (1996 SPOK)

 (= although the word absconded does not contain five letters of the
 - (= although the word absconded does not contain five letters of the alphabet in consecutive order)
 - b. The paternity tests, the DNA tests, may be revealed, although not publicly. (2007 SPOK)

(= although the paternity tests may not be revealed publicly)

Park & Kim (KHU)

English although-stripping constructions

Antecedent types

- However, there are cases when, even if the antecedent is suggested in the context, only noun phrase can be source of ellipsis with a copula verb.
 - (27) The air was dead calm and [the waves], although small, were beautifully formed and breaking at regular intervals. (2000 MAG)(= although the waves are small)
- Even a discontinuous clause can be the antecedent, especially when the stripped clause is within the relative clause.
 - (28) Holden was considered [a rebellious, ungrateful, disrespectful teenager] that, **although rare**, [is a worldwide epidemic]. (2010 NEWS)
- Sometimes the antecedent can be non-linguistically given as the situational antecedent.
 - (29) A: We do not get the money into a heap.
 B: It Is your problem. Plunder although the bank. (2009 MOV)
 (= although there is a bank / the bank can keep your money)

Park & Kim (KHU)

Subordinate stripping

English although-stripping constructions

1

Overtness of the correlates

Research questions

Corpus investigation

Data discussion Theoretical implications

Summary

References

- Based on the variables and reconstructions, the structural relations between a remnant and its antecedent can be found with the distribution:
 - (30) a. **Although not a citizen**, Membrero was in the U.S. legally. **OVERT**
 - b. They did have a chapel, although small. COVERT
 - c. **Although promising**, neither of [the new methods] has been proved.

 DEICTIC (2002 NEWS) (= although the new methods are promising)
 - d. Maybe I'll go with you, although the shooting. SITUATIONAL

Table 4: The distribution of overtness of the correlates and antecedents (raw frequencies (%))

Total		145 (100%)	205 (100%)
Situational	N/A	2 (1.4%)	o (o%)
Deictic argument	N/A	60 (41.4%)	16 (7.8%)
Covert	Mismatch	1 (0.7%)	o (o.6%)
Covert	Match	48 (33.1%)	64 (31.2%)
Overt	Mismatch	12 (8.3%)	34 (16.6%)
Overt	Match	22 (15.2%)	91 (44.4%)
		although	although not

Park & Kim (KHU)

English although-stripping constructions

Grammatical functions of the remnant

- Based on the putative source, the grammatical function of the remnant in *although*-stripping can be conjectured.
 - (31) a. Most of those seized have been released, although not the Japanese. (2004 NEWS) (Subject)
 - b. There has been similar interest in Teach For America, which recruits new college graduates, although not career-switchers. (2009 NEWS) (Object)
 - c. The partial anarchy of the international system has only been mitigated, although not removed. (VP Complement)
 - d. Both loads are highly effective, although excessively powerful. (2003 MAG) $\langle Predicative\ complement \rangle$
 - e. Sometimes, although rarely, these visitors are very big rocks asteroids or comets. (1992 MAG) \langle Adjunct \rangle

Park & Kim (KHU)

Subordinate stripping

English although-stripping constructions

Data discussion

000000

20

00000000 0 00000000000

Corpus investigation

Theoretical implications

Summar

References

Grammatical function types

Research questions

- The remnants show various grammatical functions from the putative sources.
 - (32) a. Some, **although not I** SUBJ , may be drawn to the implicit neo-political message in its subtitle, "The Value of Values." (2012 WEB)
 - b. The Trojan Horse is unique and improbable, **although not impossible** PRED . (2007 MAG)
 - c. They still plan to do that, although not tonight ADJUNCT . (1992 SPOK)
 - d. The Story with training is also good, **although not uniformly so PROP** . (2003 ACAD)

Table 5: The distribution of the grammatical function types of remnants (raw frequencies (%))

	although	although not
Subject	2 (1.4%)	4 (2.0%)
Predicate	87 (60.0%)	119 (58.0%)
Adjunct	55 (37.9%)	81 (39.5%)
Propositional anaphor	1 (0.7%)	1 (0.5%)
Total	145 (100%)	205 (100%)

Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
0000000	0	00000000000	●00000	000	0	00

Adjacency

- Generally, the remnant occurs linearly adjacent to its contrastive correlate.
 - (33) a. The size of the federal debt does have a very important, although indirect, effect on interest rates. (1995 NEWS)
 - b. The Latin translations were all that was subsequently available to European (although not to Islamic) culture from the decline of Rome until the Renaissance. (2012 WEB)
- Different from phrasal remnants, certain remnants such as an attributive adjective or a quantifier can be a lexical one.
 - (34) a. I eventually produced a decent, **although tentative**, paper. (1993 MAG)
 - b. Inimical to Opposition Earth and many Elementals, although not all. (2019 FIC)

Park & Kim (KHU)

English although-stripping constructions

2

Adjacency: Lexical remnants

- However, such remnants are allowed not to be adjacent to their contrastive correlate, even though they are in the same scope of the phrase.
 - (35) a. The Burj, **although new**, is an easy way to frame Dubai as a travel destination. (2012 BLOG)
 - b. The democrats expect to win <u>some</u> seats in the legislature China will impose on Hong Kong, **although not many**. (1997 NEWS)

Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping Research questions Corpus investigation Data discussion Theoretical implications Summary References

Adjacency: Lexical remnants in coordination

- The different distribution of the lexical remnants can be found from appositives examples from Griffiths (2015: 81)
 - (36) a. The chief, **or main**, problem is the military. b. *The chief problem is the military, **or main**.
- Griffiths (2015) analyzed appositives with clausal coordination structure to account for the unacceptable reordering of the sub-apposition.
- Also, this can be applied to the stripping with an attributive adjective remnant.
 - *Our most urgent problem is the military, **and main**, too. (from Griffiths 2015: (62a))
- Therefore, the acceptable cases in although-stripping may be due to the subordinate structure, which is distinct from coordinate one in stripping or apposition.

Park & Kim (KHU)

English although-stripping constructions

24

Copula clauses with a deictic argument

- Some remnants with no corresponding correlate can be reconstructed with a
 definite implicit argument which refers to its antecedent and copula verb within
 the context.
- Such remnants in although-stripping are restricted to occur as AP or VP[en].
 - (38) a. [The new class of 1998]_i, **although** (it_i is) **small**, could subtly alter that chemistry. (1998 NEWS)
 - b. Frankly, [the Federal Reserve]_i, **although** (it_i is) **well-intentioned**, is not doing the right job. (1992 SPOK)
- Interestingly, in negative *although*-stripping, this configuration was shown only from the remnant AP or NP, not from VP[*en*].
 - (39) a. Although (he_i was) not a philosopher, [the manic comedian Jerry Lewis]_i captured this willing surrender to sensation ... (2015 ACAD)
 - [Family, friends, career, fame, social duties, etc.] are all insurance policies, although (they; are) not unimportant. (2012 BLOG)

Park & Kim (KHU)

English although-stripping constructions

Remnant restrictions

- As mentioned earlier, allowing syntactic categories of remnants in although-stripping is more restricted than those in concessive if-stripping where any lexical and phrasal remnant is allowed.
- Especially, when the remnant functions as a (part of) predicate or in the putative source, it can be reconstructed as a copula clause with a pronominal deictic subject.
 - (40) a. His skin is sallow, although freshly shaven. (2000 FIC)
 - b. although-stripping = although his skin is freshly shaven
 - c. although-stripping = although it is freshly shaven
 - (41) a. Both loads are highly effective, although excessively powerful. (2003 MAG)
 - b. although-stripping = although both loads are excessively powerful
 - c. although-stripping = although they are excessively powerful

Park & Kim (KHU)

English although-stripping constructions

2

Remnant restrictions (cont'd)

- As in the predicative remnant, a lexical attributive remnant can also be reconstructed as a copula clause.
 - (42) a. They become about the same size, although not exactly even. (2000 FIC)
 - b. although-stripping = although they do not become about even size
 - c. although-stripping = although it (=size) is not even
- However, the other lexical remnant type 'quantifier' cannot be understood as a copula clause.
 - (43) a. Not Mikey, and not Margalo, and there were a few others, too, although not many. (2002 FIC)
 - b. although-stripping = although there were not many others
 - c. although-stripping \neq although they (=others) are many

Park & Kim (KHU)

English although-stripping constructions

Subordinate st	tripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
0000000	00	0	00000000000	000000	•00	0	00

Polarity reversals

- As stated in the semantic properties, although-stripping presupposes an implicitly
 opposite meaning of assumption evoked from the antecedent's content as in the
 following repeated example.
 - (44) **Although bright**, he was insecure, melancholy, and introverted.
 - Dr. Curt Freed at Colorado said the risk of the surgery appears to be low, and the limited available data so far suggest most patients are helped, although not cured. (1991 NEWS)
 - (46) Montelena is Napa Green certified, although not organic.
- Likewise, the utterance evoked by *although* in stripping denotes to the negated proposition of antecedent, and it semantically asserts the reversed polarity of the proposition.
 - (47) a. Semantic content of although-stripping: ASSERT $\neg(p)$
 - b. Semantic content of although not-stripping: ASSERT $\neg(\neg p)$

Park & Kim (KHU)

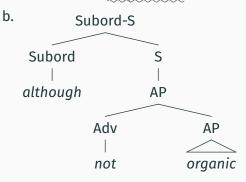
English although-stripping constructions

28

Subordinate strippingResearch questionsCorpus investigationData discussionTheoretical implicationsSummaryReferences0000000000000000000

Sentential negation

- According to Kim (2020), the fragmental negated remnant accompanying the negator 'not' may license two possible structures.
 - (48) a. A: Who could do such a thing?
 - b. B: $[_{NP}$ Not $[_{NP}$ Ella]]. \langle Constituent negation \rangle
 - c. B: $[_{\rm S}$ Not $[_{\rm S}$ $[_{\rm NP}$ Ella]]]. \langle Sentential negation \rangle (2010 FIC)
- Although not-stripping can take the structured meaning from the negator 'not' as a sentential operator which combines with the NSU projected from the remnant.
 - (49) a. Montelena is Napa Green certified, although not organic.



Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
0000000	0	00000000000	000000	00•	0	00

Predicational remnant

- Mikkelsen (2008) points out that in a copula question and answer pair, only predicational copula clause is permitted to be the answer to QUD, while a copula clause with the specificational predicate cannot.
 - Q: Who/What is Mary?
 A1: Mary is the graduate advisor. (Predicational)
 A2: #The graduate advisor is Mary. (Specificational) (Mikkelsen 2008: (13))
- The remnant in *although*-stripping can be interpreted or be alternatively paraphrased as a predicational copula clause.
 - (51) **Although not a philosopher**, [the manic comedian Jerry Lewis captured this willing surrender to sensation] ... (2015 ACAD)
 - (52) Although (he; was) not a philosopher, [the manic comedian Jerry Lewis]; captured this willing surrender to sensation ...

Park & Kim (KHU)

English although-stripping constructions

30

Subordinate stripping	Research questions	Corpus investigation	Data discussion	Theoretical implications	Summary	References
00000000	0	00000000000	000000	000	•	00

Summary

- As in coordinate stripping, subordinating conjunction *although* can also license stripping.
- Although-stripping shares some similarities with typical stripping constructions in many syntactic aspects.
- Connectivity effects such as binding condition or preposition identity support sentential approaches from previous analyses.
- However, there are some evidence that contradicts polarity reversal and parallelism from previous studies.
- Our authentic data findings provide the base-generation account of although-stripping.

Park & Kim (KHU)

English although-stripping constructions

Subordinate stripping Research questions Corpus investigation Data discussion Theoretical implications Summary References

Selected references

- Depiante, Marcela Andrea. 2000. The syntax of deep and surface anaphora: A study of null complement anaphora and stripping/bare argument ellipsis. PhD Dissertation. University of Connecticut.
- Griffiths, James. 2015. Reformulative appositions and clausal ellipsis. *Lingua* 165: 70-91.
- Hankamer, Jorge. and Ivan A Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3): 391-426.
- Hardt, Daniel, and Maribel Romero. 2004. Ellipsis and the structure of discourse. Journal of Semantics 21(4): 375-414.
- Izutsu, Mitsuko Narita. 2008. Contrast, concessive, and corrective: Toward a comprehensive study of opposition relations. *Journal of Pragmatics* 40(4): 646-675.

Park & Kim (KHU)

English although-stripping constructions

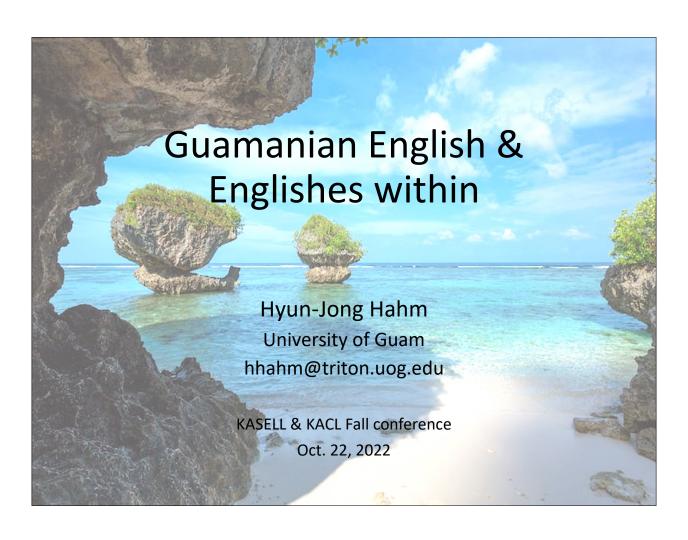
3

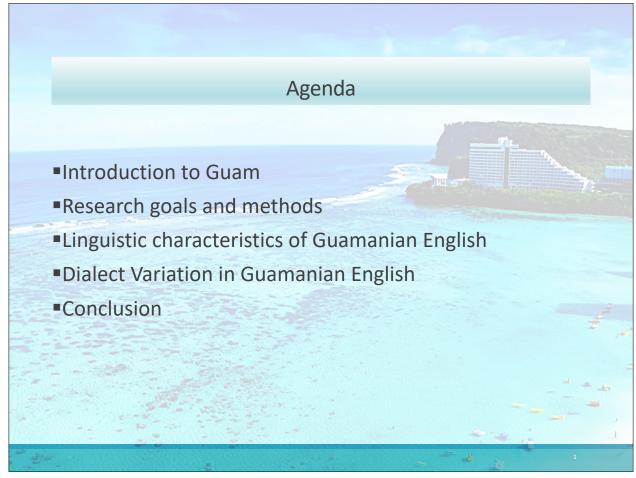
Selected references (cont'd)

- Konietzko, Andreas. 2016. Bare argument ellipsis and focus. Amsterdam: John Benjamins.
- Merchant, Jason. 2003. Remarks on stripping. Ms. University of Chicago.
- Myers, Ethan and Masaya Yoshida. 2018. What can if-stripping tell us about ellipsis? A handout from The 49th North East Linguistic Society. Ithaca, Cornell University, NY. October, 2018
- Reinhart, Tanya. 1991. Elliptic conjunctions non-quantificational LF. In Asa Kasher (ed.), *The Chomskyan turn*, 360-384. Oxford: Blackwell.
- Wurmbrand, Susi. 2017. Stripping and topless complements. *Linguistic Inquiry* 48(2): 341–366.

Park & Kim (KHU)

English although-stripping constructions





Dialect

- ■Dialect: a variety of a language shared by a group of speakers
- Speaking a language is to speak a dialect of that language
- Every dialect exhibits variation in all facets of language. Not every speech participant has same social factors.
- •Dialect diversity continues to exist in mass communication, increased mobility, and seeming cultural homogenization due to many reasons such as:
 - Geographical reason;
 - Group and individual identity;
 - Social interaction and practices
- ■There has been a lot of thorough research in dialects of the states of US, but not about Guam, US territory.

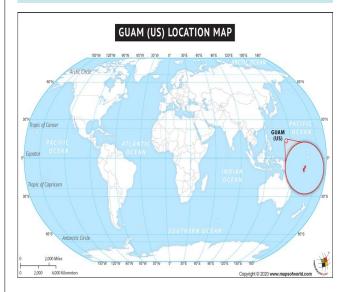
Guam is a Fiesta!

- America is known as a melting pot, but multiculturalism uses a salad bowl to describe US, where different cultures from different immigrants with diverse culture and language come about and make one big culture together. Still, they all make salads.
- •Guam is not just a salad bowl, but a fiesta. At a fiesta, people bring different dishes, and everyone appreciates the uniqueness in each and everything that makes a great fiesta.
- •Guam has so much diversity in people, language, food, culture, and any other social aspects (sexual orientation, etc.) compared to the size of the geographic location, and each group is unique and makes Guam as a whole.
- If I ask someone what is Guamanian English like, their answer is "it depends". Depending on who you talk to, it will be quite different.

Research goal and methods

- •Guam is a tiny but multi-ethnic multi-cultural island. There have been immigrants coming to Guam for over 400 years, and new people continue to move to Guam via marriage or work. Also, Guamanians have constant contact with foreigners due to tourism being the main source of economy.
- This is a preliminary descriptive research on the dialect of English spoken on Guam. The goal is to find linguistic characteristics of Guamanian English and its varieties based on personal data collection and interviews.
- Consultants who participated in this data collection process are students and colleagues at the University of Guam as well as local friends on Guam. To examine the dialectal features accurately, the participants are all born and raised on Guam, in the age of 20s to 50s in both genders.

Geographical info on Guam, where America's Day begins (UTC +10)



https://www.mapsofworld.com/usa/guam/location-map.html

- An unincorporated US territory (since 1899): People born on Guam are US citizens but have no voting rights in the US presidential elections while living there.
- Westernmost location among US states and US territories (Guam, CNMI, American Samoa, US Virgin Islands, and Puerto Rico)
- A tiny island of 561 km² (217 sq. mile)

Historical background

- Pre-colonization era:
 - Indigenous people of the island: Chamorro, who settled in about 2000-1500 BC.
- ■Spanish era (1521-1898):
 - 1521: Portuguese explorer, Magellan, in service of Spain's short visit, the first visit by a European
 - 1565: Official claim by Spanish (colonization of the Philippines begins; some migration from the Philippines happens)
 - 1668: Official colonization by Spain
- ■American era (1898-1941)
 - 1898: the capture of Guam by the US from the Spanish without a fight
 - 1917: "No Chamorro" policy and education in English; punishment for speaking Chamorro

Historical background

- ■Japanese era (1941-1944)
 - Japanese occupies Guam during World War II
 - Lexical borrowing (e.g., jori)
- ■Post-war era (1944-present)
 - 1944: the US recaptures Guam from Japan
 - 1950: the Organic Act Guamanians get US citizenship
- **1970**:
 - Chamorro as one of the official languages of Guam
 - Mandatory to teach Chamorro in school
- ■1980-1990: Chamorro Activist group formation
- ■2000- Chamorro revitalization

Language contact

- •During all colonization periods, the first task by the colonizers was to suppress the use of the indigenous language.
 - During Spanish era, a huge amount of lexical borrowing from Spanish took place, including last names (*Blas, Cruz, Duenas, Perez*, etc.) and basic vocabulary, such as the days of the week, *Lunes* 'Monday' and *adios 'bye'*; Christianity (esp. Catholicism) was adopted.
 - Chamorro language was discouraged or banned for about 450 years of colonization by different countries; and most people started to stop using Chamorro and speak only English.

- Not only language contact with the colonizers' languages, Spanish, English, and Japanese, there has been constant Immigration from other countries:
 - A huge number of Immigrants from the Philippines as construction workers after the war and typhoons (1945-)
 - There are more than 120 languages spoken in the Philippines:
 Filipino languages are heard everywhere on Guam (Tagalog as well as other Philippine languages such as Ilocano, Pangasinan, Bisayan/Ilongo, Cebuano, Sambal, etc.)
 - Chamorro and Philippine languages are closely related under the same language family more so than other Micronesian languages spoken nearby.
 - Migration of Micronesians (from Federated States of Micronesia) since 1986, when they became under the Compact of Free Association with the US, which lets them to live and work in the US. Due to the proximity to Guam, many of them considers Guam as their first step in living in America.
 - Other Asians, such as Korean and Chinese, also continue to immigrate to Guam.

Tourism

- There are more than a million tourists from different countries visiting Guam every year (Guam Business Bureau)
- The largest number of the visitors are from Korea and Japan. In the past, we also had many tourists from Russia and China. This depends on whether there is a direct flight to Guam.
- The ethnicity of the visitors affects what languages Guamanians are interested in learning since tourism is the biggest source of Guam economy.
- Many students are interested in learning Japanese and Korean.
- The language contact also influence Guamanian English.

10

Guam Demographic Profile in 2010

(US Census Bureau.)

- ■Total population: 159,381
- Ethnicities:
 - Chamorros 37.3% (About 55% of Chamorros speak only English at home)
 - Filipinos: 26.3% (About 24% of Filipinos speak only English at home)
 - Asians except Filipinos (Koreans (3,437), Chinese, and Japanese) 5.9%
 - Micronesians (Chuukese alone 7%, Pohnpeians, Palauans, etc.) 10%
 - Whites 7.1%
 - Two or more ethnic origins: 9.4%
- Language spoken at home in total population:
 - English only: 43.6%

Other than English: 56.4% (No English 0.8%, More than English 37.9%, Equally often as English 35.7%, Less often than English 26%)

- Filipino 21.2%
- Chamorro 17.8%
- other Pacific Island languages 10%
- Asian languages 6.3%
- other 1.1%

Lexicon unique to Guamanian English



(At the Guam airport)

- ■Hafa Adai: 'Hello,' greetings, or 'Are you serious?/What?!'
- ■Balutan, N or V '(take) the leftover food after a party or fiesta'
 - Take more balutan! Please, balutan the food!
- Cherry 'good/nice'
 - "How was the party?" "It was cherry!"
- ■Nai: 'didn't I say so?' 'right?'
- ∘ I told you to come here, nai!
- ■Aircon: 'AC'
- Taotaomona 'ancestral spirits, ghost (literally, people of before)'
 - She is the last taotaomona.
- *Kelaguen*: local Chamorro dish of raw meat cooked in lemon (like ceviche)
- Adobo: Filipino dish, whose name comes from Spanish

12

Lexicon Unique to Guamanian English

Chamorro expressions:

- · Laña: expressing a mild surprise, 'what?!'
 - Laña, what did he do?!
- Ma'goddai: getting strong feeling when you see cute chubby babies and wanting to pinch or squeeze them
 - Aww, she makes me feel ma'goddai!

Filipino expressions:

- Ai na ko! or Susmaryosep! 'Oh my goodness, Oh no! OMG!' (An expression of surprise or disappointment)
- Gigil 'feeling of excitement when seeing a very cute baby or puppy, or people'
 - I just feel really gigil.

Lexicon Unique to Guamanian English

- ■The following expressions are used by **younger generation** very commonly:
- Lech / Leche / Lechay / Laputa: an expression showing different feelings such as endearment, anger, frustration, surprise, etc. Lech(e), what's wrong with you?!
- Par or Prim 'bro', 'friend'
- Fokai: 'beat someone up' from a popular sports company name I am going to fokai you.

14

Guamanian English variation in Lexicon

- •Guamanian English:
 - Result:
 - Running shoes 40.4%
 - Rubber shoes 30.8%
 - Tennis shoes 23.1%
 - Others (gym shoes, workout shoes, sneakers) about 7%
 - Observations:
 - The most favored word now is 'running shoes'.
 - Consultants who use 'rubber shoes' were solely Filipinos; Filipinos who do not use 'rubber shoes' are all younger generation in 20s.
 Thus, I expect that the number of speakers using 'rubber shoes' will decrease.
 - There were some people who use either 'running shoes' or 'tennis shoes'. Some said that they used the latter when young, but now they use the former, which attests dialect change.
- FYI. Dialect Survey (http://dialect.redlog.net/)
 - In the US, sneakers 45.5%, tennis shoes 41.3%, gymshoes 5.6%, running shoes 1.42%
 - E.g., NY: sneakers 93% of participants in NY uses 'sneakers', while in TX, 69% of participants chose 'tennis shoes' and 20%, sneakers.

Phonological characteristics: Palatalization

- Palatalization: Pre-vocalic alveolar plosives change to post-alveolar affricates when preceding a retroflex.
 - Train, trip, trick, Trump, etc. [tʃɹ...] instead of [tʰɹ...]
 - o Dream, dragon, drawer, Drake, etc. [dʒɹ...] instead of [dɹ...]
- ■Even when following another consonant in the cluster: [stʃɹ...] (not [stɹ...] or [ʃtʃɹ...] having another palatalization)
 ∘ Stress [stʃɹɛs], straight, structure, etc.
- ■Not just by precedence, but when together in the onset position:
 - He was no**t r**aised on Guam.
 - The boy jumped an**d r**ang the bell.

16

Phonological characteristics: Palatalization

- ■Some, if not many, American English dialects have this phenomenon. Some people mention that their <t, d> before <r> are similar but not the same as [ʧ, ʤ], in the discussion forums on languages online.
- No literature in phonology or dialect studies discussing this in my knowledge, although many analyze the palatalization before [j].
- In Guamanian English, this rule is very systematic and there seems no exception to any word in that phonetic environment.
- •Guamanian English do recognize that they pronounce them with affricates exactly same as other affricates in church or judge.

Phonological characteristics: Cot-caught merger

- Cot-caught merger:
 - A major unconditioned merger of low back vowels in North America
 - The complete merger is dominant in Canada, the West, Eastern New English and western Pennsylvania. Merger took place before a nasal first (*Don* vs. *dawn*) (Labov 2006)
- ■In Guamanian English, also this merger took place.
 - ∘ Water [warəɹ], instead of using [ɔ]
 - Caught and cot, both with [a]
 - Wonder and wander, both with [a]

18

Guamanian English variation in Cot-caught merger

- Cot-caught merger is a very common phenomenon that Guamanian speakers went through.
- I still found the variation in pronunciation of words that could have gone through the merger.
- Examples:

awesome: 89% [a]; 11% [ɔ]
caught: 89% [a]; 11% [ɔ]
bought: 78% [a]; 22% [ɔ]
taught: 67% [a]; 33% [ɔ]

- Christopher: 30% [α]; 70% [ɔ] (This is a name, so the pronunciation varies.)
- The speakers in the same gender, age, and ethnicity, who were born and raised on Guam, showed different variation. The reason behind this variation requires further research.

Phonological characteristics: Final Obstruent Devoicing

- ■Final Obstruent Devoicing
 - It is well-known that English has assimilation in voicing for plural and past morpheme.
 - However, in Guamanian English, the obstruent is devoiced at the end of the words.
 - The final consonants in the following words are all neutralized to [s] or [t].

Houses; bags, drags/dragged, tags/tagged, ...

- •Many speakers apply this to roots of the words as well:
 - ∘ Hug [hʌk], above [əbʌf], ...
- ■Both Chamorro and Tagalog, the most spoken languages other than English, have no voiced fricatives in their sound inventories.
- Also, devoicing of the final obstruent is observed in Tagalog as well.
- •Final obstruent devoicing in Guamanian English is the result of language contact.

20

Morphological Characteristics

Analogy:

- Mass Ns to Count Ns: *luggage* and *baggage* as a count noun *I have luggages to carry.*
- Gradual disappearance of past participle –en forms (e.g., driven, given, etc.) (Many verbs have the homophonous morpheme for past tense and past participle.)

I have drove around. (cp. I have driven around.)

The suffix -er attaches to a noun, not only a verb (cp. sing/swim/drive-(e)r)

I just hired a windshield wiper-er (a windshield wiper-er 'a person who cleans the windshield'

2:

Morphological Characteristics

■Comparative doubling (more + -er morpheme) and *-er* morpheme for even long adjectives or a new adjective from a noun via conversion like 'fun' (a fun thing to do):

```
(More) beautiful-er(Cp. More beautiful)(More) intelligent-er(Cp. More intelligent)(Cp. More fun)
```

- Not all Guamanian English speakers use comparative doubling, but many do so, and everyone is aware of it.
- Conversion:
 - Worth: A to V 'Does it worth it?' (cp. Is it worth it?)
 - ∘ Joke N>>V 'I joke you'

22

Syntactic variation

- Subject verb inversion in embedded questions:
 - o I don't know what is it.

This construction appears in a few dialects of English as well. (Yale University 2022)

- ■Prodrop after main verbs:
 - Do you have some coins? I have. (instead of I do.)
 - Do you like that? Yeah, I like.

Different factors affecting the dialect variation

■Media:

 American English variety influence younger generation from movies, music, etc. Many slangs that are used by young people in the states are also used by Guamanian younger generation as well. (chad, dope, lit, etc.)

■Multi-ethnic background:

- The vast of languages spoken in this small island.
- A lot of Austronesian language descendants reside (mainly Chamorro, Tagalog and other languages of the Philippines, and Chuukese and other Micronesian languages)
- Lots of people speak their heritage language at home.

Social interaction and language contact:

- People interact with different social groups at work or school although Guam is culturally family-oriented.
- Many locals interact with millions of tourists due to the main tourism industry here.

■Style shift:

- Many informants, especially females, told me they switch their styles and word choice as well as pronunciation depending on speech participation.
- Their subconscious minds already understand dialect variations as well as different linguistic features.

24

Conclusion

- •Many systematic linguistic characteristics exist in Guamanian English.
- •Also, dialect variation is apparent due to Guam's unique diversity in language and culture.
- •More detailed research is necessary in Guamanian English dialects.
- •Guamanian English continue to undergo changes as any other languages and dialects.
- •Recording and examining contemporary Guamanian English dialects is potent to see the direction of those changes.

Future research

- •Extending the research to examine the variants affecting their dialectal variation in the spectrum of more standard vs. vernacular patterns, by gathering more detailed information on individuals' social networks (the groups of people they socialize and what kinds of activities they commonly do, etc.) as well as linguistic and demographic background.
- ■As for phonological features, performing the production and perception test to Guamanian English speakers with apparent (on-going) merger cases in the states to examine which region Guamanian English dialect shares the most with.

26

Selected references

- Campbell, Bruce L. 1987. *The Filipino Community of Guam (1945-1975)*. MA thesis, University of Hawai'i.
- •Foster, Sofie and Dirk Anthony Ballendorf. "Guam." *Encyclopedia Britannica*. https://www.britannica.com/place/Guam.
- •George Mason University. 2021. The Speech Accent Archaive. http://accent.gmu.edu/.
- Labov, William, Sharon Ash, and Charles Boberg. 2006. Atlas of North American English: Phonetics, Phonology and Sound Change. Berlin, New York: Mouton de Gruyter.
- •Wolfram, Walt and Natalie Schilling. 2016. American English: Dialects and Variation, 3rd ed. Wiley Blackwell.
- •Ygdp.yale.edu. 2022. *Inversion in embedded questions* | *Yale Grammatical Diversity Project: English in North America*. https://ygdp.yale.edu/phenomena/inversion-embedded-questions.

A Study on the Readability of American Textbooks and Korean English-Language Textbooks Comparing Syntactic Complexity and Conjunction Frequency in Sentence Structures

윤인경

(고려대학교 언어학과 석사과정)

유석훈

(고려대학교 언어학과)

The purpose of this study is to investigate the differences in readability as defined by Kim (2002) between textbooks used in grades 1~5 in American textbooks and 9~11 in Korean Englishlanguage textbooks. Based on Park (2015)'s observations that the aforementioned textbook groupings have similar sentence numbers and lengths, it has been hypothesized that—of the factors that affect textbook readability—the number of conjunctions used in complex sentences differentiate readability when it comes to the two groups. Using 17 textbook samples from three widely used American textbook publishers ranging from grades 1~5 and 39 samples from eight official Korean English-language textbooks ranging from grades 9~11—for a total of over 200,000 words—the study used the L2 Syntactic Complexity Analyzer (L2SCA), the Flesch-Kincaid Grade Level (FKGL), and the Flesch Reading Ease (FRE) scale, finding that the American textbooks had far more frequent instances of conjunctions used in between complex sentences, leading to decreased readability—as supported by Lee (2010), who found the readability of Korean Englishlanguage textbooks to be higher than corresponding American English textbooks. Contrary to the belief that Korean-made textbooks catering to Korean students may include more conjunctions due to the characteristics and frequency of run-on sentences in the Korean language, it was found that textbooks exhibited the opposite results, showing relative coordination with their American counterparts. The study was limited in the scope of its sampled data and further studies should include datasets from more publishers and sources in order to corroborate the present findings.

Keywords: readability, Syntactic Complexity Formula, conjunctions, textbook analysis, English-language textbooks.

Reading is a complex process involving a number of factors and the development of fluent reading abilities is a challenging task for EFL learners (Song & Kim, 2021), due to the fact that

textbooks are a medium by which English education policies, school classes, teachers, and learners are directly connected (Hong & Sohng, 2017). A number of studies surrounding the development of Korean English textbooks have focused upon vocabulary, writing, and listening (Kim & Chun, 2008; Lee & Hyeon, 2013; Jeon, 2011). It is, however, difficult to find studies connected to the development of English textbooks for countries in which English is a native language. Since Korean English textbooks aim to cultivate learners who are similar in English levels to those who speak English natively, there is a need for research surrounding the education methods and development of English textbooks in their respective countries. Although a direct copy of the pedagogical methods is not to be expected and inaccurate due to the difference in grading structure and systems between different countries (Park, 2015), an investigation into the reading items in textbooks and their syntactic complexity could lead to insights for the more appropriate development of Korean English textbooks.

This study aims to investigate how coordinate conjunctions—and, but, or—are presented in American textbooks from grade 1 to 5 and Korean English textbooks from grade 9 to 12. The findings are expected to help textbook developers, English instructors and educators, and teachers to develop and adapt learning materials for appropriate grade levels by researching how important connectives such as conjunctions make reading materials easier or more difficult, based on the English textbooks from two countries: American and South Korea.

2. Measurement and Readability

Readability refers to how easily text can be read and understood—in other words, text with high readability can easily be understood, while text with low readability cannot easily be read for specific reasons. Scholars have held a number of differing definitions for readability. Cho (1982) found that text with high readability include the following factors: content, coherence, vocabulary, and syntax. Kim (2002) divided the definitions for readability into three categories: first, readability refers to the surface-level, physical aspects of text (such as the shape of the text, their arrangement on a screen, and so on) which affect the understanding of said text. The second factor involves the effects of readers' levels of interest and background knowledge regarding the text. Lastly, the lexical aspects including word elements, word length, and

syntactic complexity have been defined as affecting readability. However, with regards to syntactic complexity, it is not always the case that simpler syntactic structures define text with higher readability. Klare (2000) asserted that an emphasis should not be on making sentences as short as possible, or even on making all sentences short—such focus does not necessarily decide the readability of text. For example, the sentence, "To be, or not to be: that is the question" can be considered as high in readability when it comes to syntactic complexity. However, in terms of its underlying meaning and contents, it is quite complex, complicated, and low in readability (Botel & Granowsky, 1972).

2.2. Conjunctions and Readability

Conjunctions are essential parts of the English grammar system since they help to connect words, phrases, clauses, and sentences. Moreover, conjunctions are cohesive devices which link clauses of sections of a text to demonstrate meaningful patterns (Bloor & Bloor, 1995). Geva (1986) highlighted the importance of the role of conjunctions in L2 literature, referring to Cohen and Fine (1978), who found that non-native adult speakers of English have insufficient exploit cohesive textual links, therefore they fail to adequately comprehend expository texts. With regards to readability, many studies have shown that the knowledge of conjunctions has a close relationship to reading comprehension (Cain, 2003: Geva, 1992; Robertson, 1968). Consider the following two examples below:

- (1) Nick survived the party. Kimi enjoyed it.
- (2) Nick survived the party, and Kimi enjoyed it.

As mentioned in Cho (1982), while sentences such as (2) add a coordinating conjunction, there is no apparent difference in readability between (1) and (2). However, as conjunctions are also able to link meanings between sentences, it must be understood that the removal of conjunctions has the possibility for increasing the difficulty of readability for sentences.

2.3. Syntactic Complexity and Readability

Syntactic measures in readability formulas have generally been used to control the reading

difficulty of elementary grade materials (such as in the 1948 Dale-Chall formula and the 1953 Spache formula). However, syntactic measures including those of sentence length offer little indication of the grammatical markup and complexity of a sentence. In other words, the complexity or readability of a sentence should not be judged from a word count of the sentence. For example, Shakespeare's "To be or not to be: that is the question" would be rated as having primary level difficulty in terms of sentence length (and vocabulary frequency), despite the fact that it is semantically and pragmatically complex in underlying meaning (Botel & Granowsky, 1972; Cho, 1982).

Based on the aforementioned issues in and applications for connecting syntactic complexity and readability, this study aims to investigate how differences in sentence complexity may affect the readability of American textbooks (from Grades 1 to 5) and Korean English textbooks (Grades 9 to 12).

3. Methods

To investigate a relationship between conjunctions and readability, this study chose American textbooks from three textbook publishers, McGraw-Hill 'Treasures', Houghton Milfflin Harcourt 'Journeys', and Scott Foresman 'Reading Street'. Based on Grades 1 to 5 for American textbooks (denoted as G1 ~ G5) and Grades 9 to 11 for Korean English textbooks (denoted as M3, E, E1, and E2¹), this study aims to measure the complexity of sentence structures based on the number of conjunctions used. The research methods were largely based off Lu (2010)'s study, which involved the analysis of syntactic complexity through a Syntactic Complexity Analyzer, in addition to a study focusing upon syntactic complexity for Korean middle and high school English textbooks conducted by Song and Kim (2021). The Mean Length of Sentence (MLS) and Mean Length of Clause (MLC) metrics were used to determine syntactic complexity within the samples from the textbooks, further filtered by sentences containing the conjunctions to be examined, 'and' and 'but.' In addition, to the sentence

¹ Because American and Korean academic systems and grade (year) systems are not compatible and two of the Korean textbooks differ in pacing from the rest, the Korean grade system was marked as follows: M3 is the third year of middle school, E is the first year of high school, and E1 and E2 are—respectively—the first and second halves of the second year of high school.

complexity ratios relating to amounts of coordination (C/S), the Flesch-Kincaid Grade Level (FKGL) was used in coordination with the Flesch Reading Ease (FRE) scale in order to determine the difficulty of the text. Lastly, SPSS 26 and JMP 16 were used to aggregate and analyze the data for statistical significance and mean differences.

4. Results

The following tables outline the comparative results derived from the five American textbooks and four Korean English textbooks, divided by MLC and MLS, in addition to their C/S ratios, FKGL, and Flesch Reading Ease levels.

Table 1. Comparative results based on the MLC and MLS data from sentences containing coordinating conjunctions in American textbooks (Grade 1~5) and Korean English textbooks (Grade 9~12)

Category	Source	n	Mean ²	p
	G1	122	5.23	< .001
	G2	145	7.02	< .001
	G3	185	7.64	< .001
	G4	199	7.86	< .001
MLC	G5	206	8.60	< .001
	M3	183	7.42	< .001
	Е	241	8.48	< .001
	E1	204	9.01	< .001
	E2	204	9.75	< .001
	G1	122	11.43	< .001
	G2	145	15.09	< .001
MLS	G3	185	16.3	< .001
	G4	199	16.75	< .001
	G5	206	18.22	<.001

² The mean in all tables refer to the mean word frequency per sentences containing coordinating conjunctions.

M3	183	15.91	< .001
Е	241	17.98	< .001
E1	204	19.11	< .001
E2	204	20.59	< .001

For both the MLC and MLS metrics, it can be seen that the number of sentences containing coordinating conjunctions gradually rise in frequency from the earliest grade of 'G1' in the American textbooks (n = 122) to the latest of 'G5' (n = 206), while they exhibit the highest frequency in the 'E' level for the Korean English textbooks (n = 241), followed by both 'E1' and 'E2' (n = 204), and then 'M3' (n = 183).

For both the American and Korean English textbooks and in both the MLC and MLS categories, their mean frequency of words per sentence rise through the grade levels.

Table 2. Comparative results based on the syntactic complexity of sentences containing conjunctions in American textbooks (Grade 1~5) and Korean English textbooks (Grade 9~12)

Category	Source	n	Mean	p
	G1	122	1.36	< .001
	G2	145	2.12	< .001
	G3	185	2.46	< .001
	G4	199	2.35	< .001
C/S	G5	206	2.5	< .001
	M3	183	2.26	< .001
	Е	241	2.21	< .001
	E1	204	2.5	< .001
	E2	204	2.72	< .001

The C/S scale measures the amounts of coordination between clauses, which have been found, in Table 2, to increase from the lowest grades to the highest in both American and Korean English textbooks. The mean number of coordination used in the American textbooks rose

rapidly (1.36~2.5), whereas it had a more gradual incline in the Korean English textbooks (2.26~2.72).

Table 3. Comparative results based on the readability of sentences containing conjunctions in American textbooks (Grade 1~5) and Korean English textbooks (Grade 9~12)

Category	Source	n	Mean	p
	G1	122	2.16	< .001
	G2	145	4.29	< .001
	G3	185	5.29	< .001
	G4	199	5.4	< .001
FKGL	G5	206	6.33	< .001
	M3	183	5.93	< .001
	Е	241	7.15	< .001
	E1	204	8.74	< .001
	E2	204	9.05	< .001
	G1	122	95.97	< .001
FRE	G2	145	86.69	< .001
	G3	185	83.06	< .001
	G4	199	83.57	< .001
	G5	206	77.81	< .001
	M3	183	78.79	< .001
	Е	241	72.32	< .001
	E1	204	65.51	< .001
	E2	204	65.89	< .001

The FKGL scale, which shows the recommended grade level for texts ranging from 1 to 14, shows a number of significant findings. First is that the average mean grade level for the American textbook data demonstrates a recommendation of one to two grades higher than the grade levels for which the books were written (e.g. 'G1' has an FKGL mean of 2.16, while 'G2'

has a higher differential in showing '4.29'). In contrast, the FKGL mean for the Korean English textbooks demonstrate an initially lower recommended rating, with the 'M3' mean of 5.93 corresponding roughly to the 'G4' mean of 5.4, followed by ratings that more closely match their American textbook grade equivalents (e.g. 'E2' at 9.05).

For the American textbooks, the FRE readability scale initially starts high for 'G1' (n = 95.97) and drops steadily to its lowest point at 'G5' (n = 77.81). In near coordination with their American counterparts, the Korean English textbooks' FRE readability starts similarly to 'G5' with 'M3' at 78.79, dropping to its lowest at 'E2' at 65.89.

5. Discussion

The data produced from the five American textbooks and four Korean English textbooks provided a number of findings corresponding to the aims of this study. First, the number of sentences containing coordinating conjunctions were seen to gradually rise through both types of textbooks, with their mean frequency of words rising in both scales for clauses and sentences, offering evidence that higher-grade textbooks from both sources provide longer clauses and sentences in order to match the learners' respective levels of English. Second, the C/S scale data also found that more conjunctions are used on average as grade levels go up for both sources, providing evidence that conjunctions are also a measure by which difficulty is measured. Third, the FKGL scale's recommendations for appropriate grade levels showed that the American textbook sources contained more complex and difficult clauses and sentences than intended, while the Korean English textbook sources started off lower in difficulty but inclined up to American textbook levels in later grades. Lastly, the FRE readability levels were nearly linear and predictive of increasing difficulty between the two sources.

These findings indicate a positive relation between the frequency of conjunctions and the difficulty of texts for both sources, in addition to the frequency of words on both clausal and sentential levels, as originally posited by Park (2015). This is followed by a steady decrease in readability as grades rise, which is seen as generally appropriate for the respective levels, based on the FKGL data and Lee (2010)'s findings. The results lead to the possibility for further studies investigating the effects of conjunctions on text complexity and readability for a larger

dataset and more independent sources. The primary limitation of this study included that textbooks were randomly sampled per grade and not all textbooks per publisher or grade were equally represented. Future studies including more datasets and correlations will lead to more significant findings which will aid in the analysis and development of textbooks which use the most effective sample and test items for English learning.

References

- 김국종. (2002). 영어 이독성 분석 연구. 석사학위논문. 한국교원대학교
- 박정은. (2015). 미국 교과서와 한국 영어교과서의 가독성 비교. 한국영미어문학회 학술대회 발표논문집, 2015(3), 170-180.
- 이정민, & 현태덕. (2013). 중학교 3 학년 영어 교과서 읽기 자료 비교 분석. *현대영어교육. 14*(4), 263-286.
- 이재은. (2010). Comparison of readabilities of textbooks: With special reference to textbooks for 1-6 in the U.S.A and English textbooks for 8-10 in Korea. 석사학위논문. 경북대학교 교육대학원.
- 전문기. (2011). Coh-Metrix 를 이용한 중학교 1 학년과 2 학년 개정 영어교과서 읽기 자료의 코퍼스 언어학적 연계성 분석. *언어과학연구, 56*, 201-218.
- 조석주. (1982). Readability와 syntactic complexity. 어학연구, 18(2), 325-336.
- Baumann, J. F. (2014). *Houghton Mifflin Harcourt Journeys Grade 1, 2, 4, 5*. Florida: Houghton Mifflin Harcourt.
- Bear, D. R., & Dole, J. A. (2007). *MacMillan McGraw-Hill Treasures Grade 1, 2, 3, 4*. New York: McGraw-Hill.
- Bloor, T. & Bloor, M. (1995). The Functional Analysis of English London. New York: Arnold.
- Botel, M., & Granowsky, A. (1972). A formula for measuring syntactic complexity: A

- directional effort. Elementary English, 49(4), 513-516.
- Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's functional narratives. *British Journal of Developmental Psychology*, 21, 335-351.
- Cohen, A. D., & Fine, J. (1978). Reading history in English: Discourse analysis and the experience of native and non-native readers. *Working Papers on Bilingualism Toronto*, 16, 55-74.
- Foresman, S. (2011). Reading Street Grade 1, 2, 3, 4, 5. Illinois: Scott Foresman.
- Geva, E. (1986). Reading comprehension in a second language: The role of conjunctions. *TESL Canada Journal, Special Issue 1*.
- Geva, E. (1992). The role of conjunctions in L2 text comprehension. *TESOL Quarterly*, 26(4), 731-747.
- Hong, S. W., & Sohng, H. S. (2017). Developing students' moral virtues in high school English class: Focused on the analysis of high school English textbooks. *Journal of Curriculum Evaluation*, 20(3), 25-49.
- Houghton Mifflin Harcourt. (2017). Journeys: Research Foundations. *Houghton Mifflin Harcourt*.
- Kim, J. R., & Chun, Y. (2008). A study on the continuity of elementary and middle school English textbooks through a corpus-based analysis: The 6th grade elementary school English textbook and the 1st grade middle school English textbook. *English Teaching*, 63(2), 361-394.
- Klare, G. R. (2000). The measurement of readability: useful information for communicators. *ACM Journal of Computer Documentation (JCD)*, 24(3), 107-121.
- Roberston, J. E. (1968). Pupil understanding of connectives in reading. *Reading Research Quarterly*, *3*, 387-417.
- Song, J. H. & Kim, J. H. (2021). An analysis of the reading passages represented in the textbooks of middle/high school and national achievement English tests through syntactic complexity and readability. *Secondary English Education (SEE)*, 14(2), 48-66.

Presentation Language: Korean

Presenter (Korean): 윤인경

Affiliation: 고려대 언어학과 석사과정, 대한민국

Presentation Title: A Study on the Readability of American Textbooks and Korean English-

Language Textbooks Comparing Syntactic Complexity and Conjunction Frequency in

Sentence Structures

E-mail address: <u>kimiyo@korea.ac.kr</u>

-	194	-
---	-----	---

Effects of Parts of Speech on Implicit Prosodic Prominence by Native English Speakers and Korean Learners of English

Suyeon Im(Hanyang Univ.)

Hyunkee Ahn (Seoul National Univ.)

It has been proposed that content words are more likely to be prosodically prominent than function words. Such a binary distinction in word classes (content words versus function words) has been considered insufficient to predict the occurrence of prosodic prominence in more recent research. This study investigates the effects of (1) word classes and (2) parts of speech on prosodic prominence by native English speakers and Korean learners of English in silent reading of public speech. Results showed that Korean learners of English were more likely than native English speakers to assign prosodic prominence on content words as well as function words. The difference was greatest for verb for content words and determiner for function words. Also, Korean learners of English tended to show coarser mapping between prosodic prominence and parts of speech than native English speakers. In particular, native English speakers favored pronoun, among function words, which may convey co-referential information in discourse context, but this part of speech was treated equally with other function words by Korean learners of English. Based on the results, we propose a hierarchy of (sentence-level) stressability. These results are not identical to those from the previous studies on perception of prosodic prominence, suggesting a weak perceptionproduction link of prosodic prominence. Overall, the present study expands our understanding of (a) the relationship between prosodic prominence and word classes/parts of speech and (b) the perception-production link of the suprasegmental feature in L2.

한국영어학회-한국코퍼스언어학회 가을 학술대회 2022/10/22

Effects of Parts of Speech on Implicit Prosodic Prominence by Native English Speakers and Korean Learners of English

Suyeon Im (Hanyang Univ.)

Hyunkee Ahn (Seoul National Univ.)

Introduction

- Prosodic prominence (or sentence-level stress) highlights words relative to other neighboring words in an utterance. If a word is prominent, the word tends to be produced with enhanced acoustic cues (e.g., higher pitch, longer duration, or greater intensity), (Beckman 1986, among many others).
- Prosodic prominence conveys the information status of a word (Pierrehumbert & Hirschberg 1990) and can be subject to various linguistic or paralinguistic factors, including rhythm, speech style, speech mode, etc. (Hirschberg 1993, among many others).

Introduction

- In English, the occurrence of prosodic prominence appeared to be most predicted by parts of speech (Hirschberg 1993).
- The function word tends to be phonetically reduced and cliticized to the adjacent content word in an utterance.
- Not every content word was accented by the speakers (Hirschberg 1993). Prosodic prominence on all the content words would sound unnatural in English.

Damon fried an omelet.

3

Introduction

- More recent studies argue that prosodic prominence cannot be fully addressed by the dichotomous distinction in word classes (content *vs* function).
- Each part of speech may have an inherently different strength of prosodic prominence (stressability).

Class	Anttila (2015)	Shih (2018)
1	weak pronoun	conjunction, determiner, preposition to
2	strong pronoun, finite auxiliaries	preposition, pronoun, wh-adverb, wh-determiner
3	modal, <i>wh</i> -words	particle, modal auxiliary, some adverbs, past tense verb, present tense verb, wh-pronoun
4	adjective, adverb, noun, verb	cardinal number, <i>wh</i> -pronoun, adjective, noun, proper noun, present tense verb, some adverbs

Introduction

- Korean is an edge-prominence language (Jun 2005). Prosodic prominence is realized at the edge of a prosodic unit, which may consist of content words or function words.
- · Content words might not vary with stressability.
- If the final syllable is an affix, it may become acoustically enhanced. This clearly differs from function words in English, which tend to be non-prominent and acoustically reduced.

아버지가 방에 들어가신다.

5

Introduction

 The differences in prosodic system between English and Korean may raise difficulties in learning English prosody for Korean learners of English.

Production

(Yoo 2014)

- Group difference in function words, but not in content words
- KLE may fail to produce a reduced form of function words.

Introduction

 The differences in prosodic system between English and Korean may raise difficulties in learning English prosody for Korean learners of English.

Perception (Im 2019)

- Group difference in content words, but not in function words
- KLE perceive content words, esp. N & V, as prominent although those are not always acoustically salient.

Production (Yoo 2014)

- Group difference in function words, but not in content words
- KLE may fail to produce a reduced form of function words.

Introduction

 The differences in prosodic system between English and Korean may raise difficulties in learning English prosody for Korean learners of English.

Perception (Im 2019)

- Group difference in content words, but not in function words
- KLE perceive content words, esp. N & V, as prominent although those are not always acoustically salient.

Implicit (Silent) Production

???

Production (Yoo 2014)

- Group difference in function words, but not in content words
- KLE may fail to produce a reduced form of function words.

This Study

- To investigate how native English speakers and Korean learners of English assign prosodic prominence in relation to parts of speech in silent reading of public speech in English:
 - 1) How do native English speakers assign prosodic prominence in relation to (a) word classes and (b) parts of speech? What is their hierarchy of stressability?
 - 2) How do Korean learners of English assign prosodic prominence as a function of (a) word classes and (b) parts of speech? What is the difference in the hierarchy of stressability between native English speakers and Korean learners of English, if any?

9

Experiment

- Participants
 - 35 native English speakers
 - 30 advanced Korean learners of English (965/990 in TOEIC) for few difficulties in the semantic or pragmatic meaning of a word
- Rapid Prosody Transcription (Cole & Shattuck-Hufnagel 2016)
- Speech material
 - A TED Talk (361 words) with a non-technical topic and a plain vocabulary

Experiment

Imagine that you would deliver a speech to an audience in public and select words that you would highlight for listeners on a transcript of a speech.

A few years ago, I felt like I was stuck in a rut, so I decided to follow in the footsteps of the great American philosopher, Morgan Spurlock, and try something new for 30 days. The idea is actually pretty simple. Think about something you've always wanted to add to your life and try it for the next 30 days. (...)

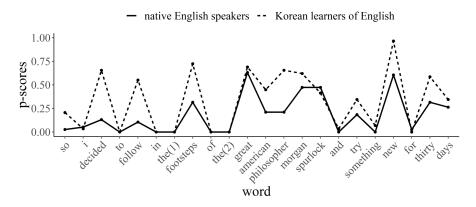
(https://www.ted.com/talks/matt_ cutts_try_something_new_for_30_days)

11

Analyses

- Annotation of part of speech
 - The Penn Treebank P.O.S. Tags (Taylor et al. 2003), used for previous research (Anttila 2015, Shih 2018)
 - 11 POS (adjective, adverb, noun, verb, conjunction, determiner, interjection, modal, number, preposition, pronoun)
- Statistical analyses
 - Two generalized linear mixed-effects models
 - Packages: Ime4 (Bates et al. 2015), Ismeans (Lenth 2016) in R (R Core Team 2019)

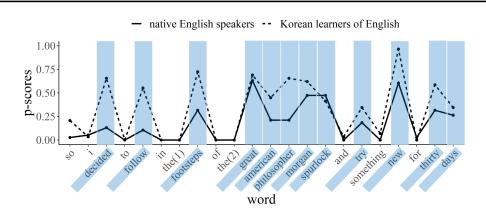
Prominence Scores



• The overall patterns of prosodic prominence rated by native English speakers and Korean learners of English

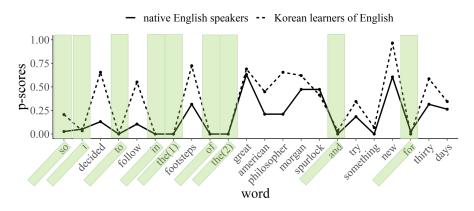
13

Prominence Scores



• For (content) words, the prominence scores tend to be higher for Korean learners of English than for native English speakers.

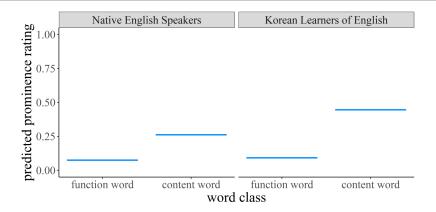
Prominence Scores



• For most function words, except "so", prominence scores tend to be similar between native English speakers and Korean learners of English.

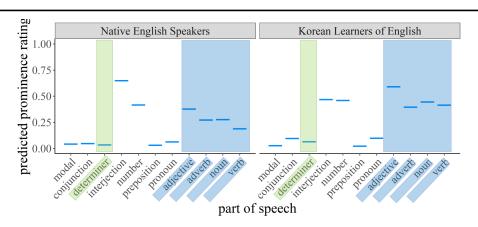
15

Effects of Word Class by L1 Group



• For content words and function words, Korean learners of English showed higher estimates than native English speakers.

Effects of POS by L1 Group



- For all the content words and few function words, Korean learners of English showed higher estimates than native English speakers.
- The difference between the groups increased in the following order: adverb < determiner < noun < adjective < verb.

Discussion

• Prosodic prominence is probabilistically mapped with the parts of speech within the identical word classes (Shih 2018).

Class	Native English Speakers	Korean Learners of English
1	preposition, determiner, modal, conjunction	preposition
2	pronoun	modal, determiner, conjunction, pronoun
3	verb	adverb, verb, noun, number, interjection
4	adverb, noun	adjective
5	adjective, number	_
6	interjection	_

18

Discussion

• The differences in prosodic system between English and Korean raise difficulties in learning English prosody for Korean learners of English.

Perception (Im 2019)

- Group difference in content words, but not in function words
- KLE perceive content words, esp. N & V, as prominent although those are not always acoustically salient.

Implicit (Silent) Production

- Group difference in content words and function words
- KLE would produce prominence on content and function words.

Production (Yoo 2014)

- Group difference in function words, but not in content words
- KLE may fail to do not produce a reduced form of function words.

19

Discussion

- Previous instruction guidelines of English rhythm (Kim 2000, Yoon 1999, among many others) have introduced mapping between prosodic prominence and word classes.
- We believe that the instruction of simple mapping between prosodic prominence and word classes is important and efficient for Korean learners of English.
- In addition to this, we propose that the gradual and probabilistic relationship between prosodic prominence and parts of speech (i.e., hierarchy of stressability) needs to be supplemented in the instruction of English rhythm.

Take-home messages

- This study provides evidence of probabilistic mapping between prosodic prominence and parts of speech (Anttila 2015, Shih 2018). We present the parts of speech along the (sentence-level) stressability obtained from native English speakers and Korean learners of English.
- Also, we suggest that there is a weak perception-production link for suprasegmental feature (prosodic prominence), in alignment with segmental feature (Levy & Law 2000) in second language acquisition.

21

References

- Anttila, A. 2017. Stress, phrasing, and auxiliary contraction in English. In V. Gribanova and S. Shih, eds., The Morphosyntax-Phonology Connection: Locality and Directionality at the Interface, 143-170. Oxford, UK: Oxford University Press.
- Bates, D., M. Mächler, B. Bolker, and S. Walker, 2015. Fitting linear mixed-effects models using Ime4. *Journal of Statistical Software* 67(1), 1-48.
- Beckman, M. 1986. Stress and Non-Stress Accent. Dordrecht, Netherlands: Fortis.
- · Cole, J., and S. Shattuck-Hufnagel. 2016. New methods for prosodic transcription: Capturing variability as a source of information. Laboratory Phonology 7, 1-29
- Hirschberg, J. 1993. Pitch accent in context predicting intonational prominence from text. Artificial Intelligence 63(1-2), 305-340.
- Im, S. 2019. Word class information in perception of prosodic prominence by Korean learners of English. Phonetics and Speech Sciences 11(4), 1-8.
- Jun, S-A. 2005. Korean intonational phonology and prosodic transcription. In S-A. Jun, ed., Prosodic Typology: The Phonology of Intonation and Phrasing, 201-229. New York: Oxford University
- Kim, C-H. 2000. Teaching the rhythm of English. *Journal of Linguistic Studies* 5, 45-59.
- Lenth, R. V. 2016. Least-Squares Means: The R Package Ismeans. Journal of Statistical Software 69(1), 1-33.
- Levy, E. S., and F. F. Law. 2010. Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. The Journal of the Acoustical Society of America 128(3), 1290-1305.
- Pierrehumbert, J., and J., B. Hirschberg, 1990. The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, eds., Intentions in Communication, 271-311. Cambridge: MIT Press.
- R Core Team. 2019. R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/
- Shattuck-Hufnagel, S. 1995. The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": Comments on Grabe and Warren, and Vogel, Bunnell, and Hoskins. In B. Connell and A. Arvaniti, eds., Papers in Laboratory Phonology IV, 128-140. Cambridge, UK: Cambridge University Press.
- Shattuck-Hufnagel, S. 1995. The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": Comments on Grabe and Warren, and Vogel, Bunnell, and Hoskins. In B. Connell and A. Arvaniti, eds., Papers in Laboratory Phonology IV, 128-140. Cambridge, UK: Cambridge University Press.
- Taylor, A., M. Marcus, and B. Santorini. 2003. The Penn treebank: an overview. In A. Abeillé, ed., Treebanks: Building and using parsed corpora, 5-22. Dordrecht, Netherlands: Springer
- Yoo, H-B. 2014. Phonetic reduction of English function words in the passage reading by Korean EFL learners. English Language and Linguistics 20(1), 179-202
- Yoon, Y-B. 1999. Teaching pronunciation of suprasegments in English. Korean Journal of Elementary Education 10(1), 365-392.

Thank you

(suyeonim@hanyang.ac.kr)

23

Appendix: Statistical Analyses

• Two generalized linear mixed-effects models

	Model 1		Model 2	
DV		prominence rating (0 non-prominent, 1 prominent)	prominence rating (0 non-prominent, 1 prominent)	
	IV1	L1 group (NAE, KLE)	L1 group (NAE, KLE)	
IVs	IV2	Word class (content, function)	Part of speech (adj., adv., conj., det., etc.)	
	IV3	Interaction between IV1 and IV2	Interaction between IV1 and IV2	

On Consist of and Made up of: A Corporabased Analysis¹

Namkil Kang
Far East University
South Korea

Abstract-The ultimate goal of this article is to show that consist of and made up of are somewhat different in the British National Corpus (20, April 2022. Online https://corpus.byu.edu/bnc). With respect to the BNC (20, April 2022. Online https://corpus.byu.edu/bnc), it is interesting to note is that the British prefer using made up of rather than using consist of. A further point to note is that consist of is 57.14% the same as made up of. Talking about consist of and made up of in the Hansard Corpus (20, April 2022. Online https://english-corpora.org/hansard/), the former is preferable to the latter in the British parliament. Quite interestingly, consist of was the most preferable one (942 tokens) for British politicians in 1960, but it was the least preferred one (49 tokens) in 1800. On the other hand, made up of reached a peak (1,020 tokens) in the British parliament in 1990. More importantly, the BNC (20, April 2022. Online https://corpus.byu.edu/bnc) shows that consist of people is the most frequently used one in the UK, whereas made up of people and made up of representatives are the most widely used ones. Finally, it is worthwhile noting that 6.66% of forty five nouns are the collocations of both consist of and made up of.

Index Terms- British National Corpus, Hansard Corpus, consist of, made up of,

I. INTRODUCTION

As Murphy (2016, 2019) points out, consist of and made up of are used synonymously. As argued by Kang (2022), the British National Corpus (20, April 2022. Online https://corpus.byu.edu/bnc) is a database that is heavily used in English learning. The main purpose of this paper is to compare consist of with made up of in the British National Corpus (20, April 2022. Online https://corpus.byu.edu/bnc) and the Hansard Corpus (20, April 2022. Online https://englishcorpora.org /hansard/). First, we consider the similarity between consist of and made up of by examining seven genres. Also, we measure the distance between *consist of* and *made up of* in seven genres. By doing so, we can see how similar consist of and made up of are in seven genres. Second, we consider which type is the preferred one in the British parliament. Also, we consider when consist of and made up of were the most frequently used in the British parliament. Finally, we consider the collocations of consist of and made up of in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). By using the software package NetMiner, we can see how similar consist of and made up of are. This paper is organized as follows. In section 2, we argue that the British prefer using made up of to using consist of. We further argue that consist of is 57.14% the same as made up of. We also contend that consist of and made up of show the highest similarity in the non-academic genre. In section 3, we maintain that consist of was preferable to made up of in the British parliament. More interestingly, consist of and made up of reached a peak in 1960 and 1990, respectively. In section 4, we argue that consist of people is the most frequently used one (4 tokens) in the UK, whereas made up of people and made up of representatives are the most widely used ones (19 tokens vs. 19 tokens). Finally, we show that 6.66% of forty five nouns are the collocations of both consist of and made up of.

II. CONSIST OF AND MADE UP OF IN THE BNC

In what follows, we consider the similarity between *consist of* and *made up of* by examining seven genres. Table 1 shows the frequency of *consist of* and *made up of* in seven genres:

¹ This work was supported by the 2022 Far East University Research Grant (FEU2022S02).

TABLE 1 FREQUENCY OF CONSIST OF AND MADE UP OF IN SEVEN GENRES (BNC)

GENRE	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON- ACAD	ACADMIC	MISC
Consist of	993	23	22	67	25	230	318	308
Made up of	1,129	53	49	102	89	273	272	291

An immediate question is "Which type is the preferred one in the UK?" Table 1 clearly shows that *made up of* is preferable to *consist of*. To be more specific, the frequency of *consist of* is 993 tokens, whereas that of *made up of* is 1,129 tokens. This amounts to saying that the British prefer using *made up of* rather than using *consist of*.

Interestingly, consist of ranks first (318 tokens) in the academic genre, whereas made up of ranks first (291 tokens) in the miscellaneous genre. Quite interestingly, there is no similarity between consist of and made up of in rank-one. With respect to the academic genre, it is worth noting that the frequency of consist of (318 tokens) is somewhat higher than that of made up of (272 tokens). This in turn implies that the British prefer using consist of rather than using made up of in the academic field. With respect to the miscellaneous genre, on the other hand, it is interesting to note that consist of (308 tokens) is preferable to made up of (291 tokens).

Noteworthy is that *consist of* ranks second (308 tokens) in the miscellaneous genre, whereas *made up of* ranks second (273 tokens) in the non-academic genre. Again, there is no similarity between *consist of* and *made up of* in rank-two. More interestingly, in the non-academic genre, the frequency of *made up of* (273 tokens) is much higher than that of *consist of* (230 tokens). This amounts to saying that the British prefer using *made up of* in the non-academic genre.

It is worth pointing out that *consist of* ranks third (230 tokens) in the non-academic genre, whereas *made up of* ranks third (272 tokens) in the academic genre. Quite interestingly, *consist of* does not show the same property as *made up of* in rank-three, hence showing no similarity.

It is worthwhile mentioning that *consist of* and *made up of* are the fourth most frequently used ones (67 tokens vs. 102 tokens) in the magazine genre. Interestingly, there is a similarity between *consist of* and *made up of* in the magazine genre.

It is worthwhile noting that *consist of* and *made up of* are the fifth most frequently used ones (25 tokens vs. 89 tokens) in the newspaper genre. Quite interestingly, there is a close similarity between *consist of* and *made up of* in the newspaper genre. It should be pointed out, however, that the frequency of *made up of* (89 tokens) is three times higher than that of *consist of* (25 tokens) in the newspaper genre. We take this as implying that British journalists prefer to use *made up of* rather than use *consist of* in their newspapers.

It is interesting to note that *consist of* (23 tokens) and *made up of* (53 tokens) are the sixth most frequently used ones in the spoken genre. Again, there is a similarity between *consist of* and *made up of* in the spoken genre, hence the same pattern in rank-six. Interestingly, the frequency of *made up of* (53 tokens) is higher than that of *consist of* (23 tokens) in the spoken genre. This might imply that the British prefer to use *made up of* rather than use *consist of* in the spoken genre.

Finally, it is worthwhile pointing out that *consist of* and *made up of* are the seventh most frequently used ones (22 tokens vs. 49 tokens) in the fiction genre. More interestingly, there is a similarity between *consist of* and *made up of* in the fiction genre, thus showing a high similarity in rank-seven. It must be noted, however, that *made up of* is favored over *consist of* in the fiction genre. We infer this from the fact that the frequency of *made up of* (49 tokens) is even higher than that of *consist of* (22 tokens). To sum up, *consist of* and *made up of* show the same pattern in four genres, but they show a different pattern in three genres. From all of this, we may conclude that *consist of* is 57.14% the same as *made up of*.

Now let us turn to the percentage of *consist of* and *made up of* in seven genres:

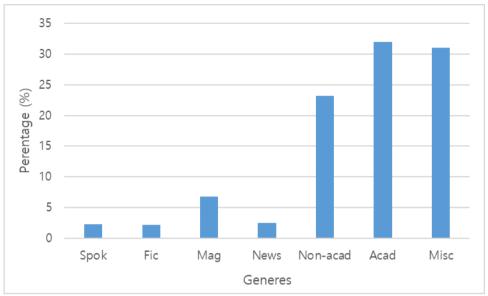


FIGURE 1 PERCENTAGE OF CONSIST OF IN SEVEN GENRES

As can be seen from Figure 1, the academic field is the most influenced by *consist of*, followed in descending order by the miscellaneous genre, and the non-academic genre.

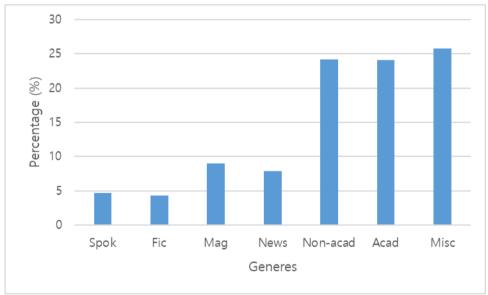


FIGURE 2 PERCENTAGE OF MADE UP OF IN SEVEN GENRES

As shown in Figure 2, the miscellaneous field is the most influenced by *made up of*, followed by the non-academic genre, and the academic genre.

Now attention is paid to the distance between *consist of* and *made up of* in seven genres. Here the Euclidean distance provides an indication of how similar *consist of* and *made up of* are in seven genres. We define the Euclidean distance in the following way:

(1) The Euclidean distance:

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

The following table shows the degree of the similarity between *consist of* and *made up of* in seven genres:

TABLE 2 EUCLIDEAN DISTANCE BETWEEN CONSIST OF AND MADE UP OF IN THE BNC

Genre	Spok	Fic	Mag	News	Non-acad	Acad	Misc
Percentage of consist of	2.31	2.21	6.74	2.51	23.16	32.02	31.01
Percentage of made up of	4.69	4.34	9.03	7.88	24.18	24.09	25.77
Euclidean distance	2.38	2.13	2.29	5.37	1.02	7.93	5.24

More interestingly, consist of is the furthest from made up of in the academic genre. The Euclidean distance between consist of and made up of in the academic genre is 7.93. This shows that consist of and made up of show low similarity. Interestingly, in the non-academic field, consist of is the nearest to made up of. Their distance is 1.02, which is the lowest. This shows that consist of and made up of show high similarity. We thus conclude that in the non-academic genre, consist of is the nearest to made up of.

III. CONSIST OF AND MADE UP OF IN IN THE HANSARD CORPUS

In this section, we aim to consider the diachronic aspects of *consist of* and *made up of* from 1800 to 2000. Table 3 shows the use of *consist of* and *made up of* in the British parliament (1800-2000):

TABLE 3
FREQUENCY OF CONSIST OF AND MADE UP OF IN THE HANSARD CORPUS

Period	Consist of	Made up of
1800	49	9
1810	63	16
1820	74	25
1830	256	67
1840	206	52
1850	507	64
1860	397	94
1870	524	94
1880	598	176
1890	534	142
1900	629	212
1910	795	227
1920	604	266
1930	801	291
1940	820	348
1950	825	356
1960	942	525
1970	919	669
1980	918	851
1990	837	1,020
2000	453	591

An important question is "Which type was the preferable one for British politicians?" Table 3 clearly indicates that *consist of* was the preferred one in the British parliament from 1800 to 2000. That is to say, the frequency of *consist of* is 11,751 tokens, but that of *made up of* is 6,095 tokens. This shows that *consist of* was preferable to *made up of* in the British parliament. Note that as observed earlier, the British prefer to use *made up of* rather than use *consist of*. The following graph shows the diachronic use of *consist of* and *made up of* from 1800 to 2000:

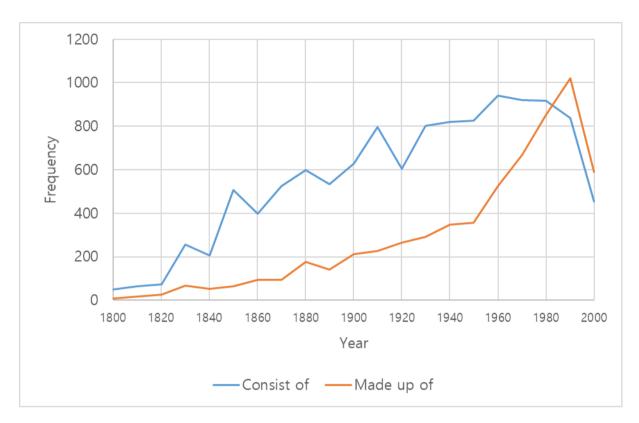


FIGURE 3 USE OF CONSIST OF AND MADE UP OF IN THE BRITISH PARLIAMENT FROM 1800 TO 2000

It is interesting to note that the frequency of *consist of* increased from 1800 to 1850. Note that there was a decline (a decline of 110 tokens) in 1860. With respect to the frequency of *consist of*, it is interesting to point out that there was a gradual increase (an increase of 201 tokens) from 1860 to 1880. Notice, however, that there was a sudden fall (a decline of 64 tokens) in the frequency of *consist of* in 1890. Interestingly, the frequency of *consist of* increased steadily (a rise of 408 tokens) from 1890 to 1960 except 1920. However, the figure of *consist of* decreased suddenly (a decline of 489 tokens) from 1960 to 2000.

Interestingly, the figure of *made up of* increased gradually (a rise of 58 tokens) from 1800 to 1830. Interestingly, the figure decreased suddenly (a decrease of 15 tokens) in 1840. After this period, there was a gradual rise (an increase of 124 tokens) in the frequency of *made up of* from 1840 to 1880. With respect to the frequency of *made up of*, it is worth noting that there was a sudden decrease (a decline of 34 tokens) in 1890. Quite interestingly, the figure of made up of increased dramatically (a rise of 878 tokens) from 1890 to 1990. Note that there was a decrease of 429 tokens in 2000. Most interestingly, *made up of* had the highest frequency (1,020 tokens) in 1990. Most importantly, *consist of* was always preferable to *made up of* from 1800 to 1980. Note, however, that *made up of* was favored over *consist of* in the British parliament in 1990 and 2000.

IV. COLLOCAIONS OF CONSIST OF AND MADE UP OF IN THE BNC

In the following, we aim to provide a collocation analysis of *consist of* and *made up of* in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). This list was cut off in the top 24:

TABLE 4 COLLOCATION OF CONSIST OF

Number	Collocation	Frequency
1	Consist of people	4
2	Consist of fish	2
3	Consist of cash	2
4	Consist of money	2
5	Consist of molecules	2
6	Consist of modules	2
7	Consist of mixtures	2
8	Consist of hundreds	2

9	Consist of interviews	2
10	Consist of periods	2
11	Consist of persons	2
12	Consist of representatives	2
13	Consist of sheets	2
14	Consist of volunteers	2
15	Consist of water	2
16	Consist of words	2
17	Consist of obsidian	1
18	Consist of objects	1
19	Consist of oats	1
20	Consist of noun	1
21	Consist of non-lawyers	1
22	Consist of neutrons	1
23	Consist of music	1
24	Consist of mortgage	1

Now which collocation is the most preferred by the British? Table 4 indicates that the collocation *consist of people* is the most commonly used one (4 tokens). This in turn indicates that *consist of people* is the most preferable one (4 tokens) for the British. It is interesting to note, on the other hand, that *consist of mixtures* ranks second (2 tokens) in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). It is also interesting to point out that *consist of representatives* ranks second (2 tokens) in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). Finally, it is worthwhile mentioning that *consist of water* ranks second (2 tokens) in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). It can thus be concluded that the collocation *consist of people* is the most preferred one (4 tokens).

Now consider Table 5:

TABLE 5 COLLOCATION OF MADE UP OF

Number	Collocation of made up of	Frequency
1	Made up of people	19
2	Made up of representatives	19
3	Made up of members	10
4	Made up of parts	7
5	Made up of individuals	6
6	Made up of players	4
7	Made up of staff	4
8	Made up of street	4
9	Made up of thousands	4
10	Made up of units	4
11	Made up of waves	4
12	Made up of hundreds	3
13	Made up of electrons	3
14	Made up of couples	3
15	Made up of cells	3
16	Made up of amino	3
17	Made up of quarks	3
18	Made up of women	3
19	Made up of components	2
20	Made up of attitudes	2
21	Made up of atoms	2
22	Made up of employees	2
23	Made up of elements	2
24	Made up of experts	2

Table 5 clearly indicates that *made up of people* and *made up of representatives* are the most widely used ones (19 tokens vs. 19 tokens) in the UK. This in turn suggests that *made up of people* and *made up of representatives* are the most preferred ones (19 tokens) in the UK. As illustrated in Table 5, *made up of people* and *made up of representatives*

are the most preferred (19 tokens vs. 19 tokens) by the British, followed by made up of members, made up of parts, made up of individuals, and made up of players, in that order. Quite interestingly, consist of people and made up of people rank first (4 tokens vs. 19 tokens) in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). On the other hand, consist of representatives and made up of representatives rank second (2 tokens vs. 19 tokens) in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). It is worth observing that made up of members ranks third (10 tokens) in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc). We thus conclude that made up of people and made up of representatives are the most preferred (19 tokens vs. 19 tokens) by the British.

Now let us turn to the visualization of the collocations of *consist of* and *made up of* in the top 24:

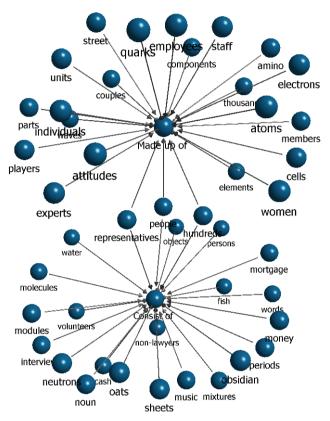


FIGURE 4 VISUALIZATION OF THE COLLOCATIONS OF CONSIST OF AND MADE UP OF IN THE BNC

As exemplified in Figure 4, twenty four nouns are linked to *consist of* and *made up of*, respectively. Most importantly, three nouns are linked to both *consist of* and *made up of*, which indicates that they are the collocations of both *consist of* and *made up of*. The three nouns linked to *consist of* and *made up of* are *people*, *representatives*, and *hundreds*. Most importantly, 6.66% of forty five nouns are the collocations of *consist of* and *made up of*. It can thus be concluded that the types *consist of* and *made up of* are somewhat different.

V. CONCLUSION

To sum up, we have compared *consist of* with *made up of* in the British National Corpus (20, April 2022. Online https://corpus.byu.edu/bnc) and the Hansard Corpus (20, April 2022. Online https://english-corpora.org /hansard/). In section 2, we have argued that the British prefer using *made up of* rather than using *consist of*. We have further argued that *consist of* is 57.14% the same as *made up of*. Also, we have shown that *consist of* is the nearest to *made up of* in the non-academic genre. In section 3, we have maintained that *consist of* was preferable to *made up of* in the British parliament. Quite interestingly, *consist of* was the most preferable one (942 tokens) for British politicians in 1960, whereas it was the least preferred one (49 tokens) in 1800. On the other hand, *made up of* reached a peak (1,020 tokens) in the British parliament in 1990. In section 4, we have contended that *consist of people* is the most frequently used one (4 tokens) in the UK, whereas *made up of people* and *made up of representatives* are the most widely used ones (19 tokens vs. 19 tokens). Finally, we have shown that 6.66% of forty five nouns are the collocations of both *consist of* and *made up of*. This in turn suggests that *consist of* and *made up of* are low similarity synonyms in the BNC (20, April 2022. Online https://corpus.byu.edu/bnc).

REFERENCES

- British National Corpus (BNC). 20, April 2022. Online https://corpus.byu.edu/bnc
 Hansard Corpus (HC). 20, April 2022. Online https://english-corpora.org/hansard/
 Kang, N (2022). A Comparative Analysis of Impressed by and Impressed with in Two Corpora. *Theory and Practice in Language* Studies 11 (11): 819-827.
- [4] Murphy, R. (2016). *Grammar in Use*. Cambridge University Press. Singapore. [5] Murphy, R. (2019). *English Grammar in Use*. Cambridge University Press. New York.

부정극어 wh-so-ever의 의미 분석: 허가 조건과 코퍼스 분석

강아름(충남대학교)

본고의 목표는 영어 wh-so-ever의 허가 조건을 코퍼스를 통해 실증적으로 분석하고 의미를 분석하는데 목적이 있다. 영어 사전에 등재된 whatsoever의 의미는 다음과 같다:

- (1) Used after a negative phrase to add emphasis to the idea that is being expressed:
 - a. He has no respect for authority whatsoever.
 - b. "Did you have any idea what was happening at the time?" "None whatsoever."

위 정의에 따르면 whatsoever는 주로 부정구문 뒤에 사용되어 강조를 표현한다고 되어 있다. (1a)에서는 부정사 no 이후에 발생하여 '그는 authority에 대한 어떠한 존경도 '전혀' 없음'을 강조하고 있고 (1b)에서는 none 이후에 발생하여 '화자는 어떠한 아이디어도 '전혀' 없음'을 강조한다. 하지만 COCA(The Corpus of Contemporary American English)에서 추출한 많은 예시들에서 whatsoever는 명시적 부정구문에 의해 허가되지 않는 듯 하다. (2a)는 조건문에 의해, (2b)는 n-word 'nothing'에 의해, (2c)는 명사 absence에 의해, (2d)는 의문문에 의해 발생하는 듯 하다.

- (2) a. It will blow up in your face if any stress whatsoever is placed on it.
 - b. In particular, the authenticating server learns nothing whatsoever about the identity of the specific user that now entered the system!
 - c. There was a total absence of any parental authority, whatsoever.
 - d. Do they respect any law whatsoever?

통사 이론적 측면에서 부정 극성(negative polarity)은 반형태나 반부가 부정어의 유무에 제약을 받는 강부정극어(Strong NPI)를 의미하는데(Ladsaw 1980, Zwarts 1998 등) (2)의 예문들은 강부정극어에 적용되지 않는 듯 하다. 따라서 약부정어(weak NPI)를 허가하는 하향함의나 비진언적 연산자와같은 의미적 극어 환경(Giannakidou 1999 등), 혹은 척도적 극성이 유발되는 적절한 맥락 하에서 극어가 척도상의 끝점(extreme)을 지시하는 화용적 극어 환경, 특히 수사적 특성이 고려되어야 할 것이다(Israel 2001 등).

이와 같은 현상에 기반하여 본고에서 재조명하고자 하는 wh-so-ever의 극성의 의미는 다음 두 가지로 진행할 것이다. 첫째, Word2vec을 사용한 워드 임베딩을 이용하여 COCA에서 제공되는 약 2만 여개의 문장 데이터에서 whatsoever의 관련 어형의 관계를 코사인 유사도를 통해 살핀다. 둘째, 컴퓨터 결과에 대한 언어학적 해석 및 설명을 시도한다. 본고의 의의는 다음과 같다. 첫째, 이론적 분석에 그쳤던 기존의 논의에 더해 본고는 코퍼스를 활용한 의미론적 연구로 wh-so-ever의 분석 영역을 확장한다. 둘째, 기존 Word2vec이 어휘 간의 의미 관계(유의어, 반의어, 상하의어, 어휘 의미변화 등)를 설명하기 위한 분석도구로서의 사용에 그친데 반해 본고는 극어 whatsoever의 허가 환경 조건에 대한 의미화용적 규칙(rule)을 도출해내는 분석도구로 그 역할을 확장한다.

참고문헌

- Giannakidou, Anastasia. 2011. "Negative and positive polarity items," Semantics (HSK 33.2), 1660-1712. Israel, Michael. 2001. "Minimizers, maximizers and the rhetoric of scalar reasoning," Journal of Semantics, 18-4, 297-331.
- Ladusaw, William. 1980. "On the notion "affective" in the analysis of negative polarity items, "Journal of Linguistic Research 1, 1-23.
- Zwarts, Frans. 1998. "Three Types of Polarity," In F. Hamm and E. Hinrichs (eds.) Plurality and Quantification, 177-238.

중등교사 영어 임용시험 수험자를 위한 영역별 전문어휘 목록 선정 및 지도방안 탐색¹⁾

이혜진(원광대학교) 이제영(전주대학교)

I. 서론

교육부와 한국직업능력개발원이 실시한 2019년 진로교육 현황조사에 따르면 중고등학생들의 희망직업 1위는 교사로 지난 2009년과 2015년에 행해진 조사결과와 동일하게 학생들의 교사에 대한 직업 선호도는 최상위권을 유지하였다. 하지만 중등교사 양성과정인 사범대학의 경우 졸업자 대비 임용률이 낮아 경쟁이 치열한며, 외국어 전공의 경우 임용시험이 해당외국어로 치러지기 때문에 학습 부담이 배가 될 수밖에 없다.

근래의 코퍼스 기반 연구의 비약적인 발전과 더불어 중등영어 임용시험을 대상으로 한 코퍼스 분석은 기존의 선행 연구에서도 여러모로 진행되고 있지만, 대부분 특정 전공영역에 한정되어 있거나, 임용시험에 출제되었던 평가영역의 구분 없이 전체적인 접근에서의 분석이 대다수였다(예: 김민경, 신창원, 2019; 이혜진, 2020). 이에 본 연구는 분석 대상의 대표성 (representativeness) 및 균형성(balancedness)을 확보하고자 중등영어 임용시험에 출제되는 각 하위 영역(영어교육학, 영어학, 영문학, 일반영어)을 특수 코퍼스로 구축하여 각 영역별 전문어휘 목록을 구축하고자 한다. 단순히 어휘목록을 제시하는 것만으로는 원하는 목표를 달성하기란 요원하므로 선정된 어휘목록을 바탕으로 대면/비대면 수업에서 활용할 수 있는 게임형식의 교육적 방안들도 추가로 제안하고자 한다.

Ⅱ. 연구 방법

1. 분석 대상

본 연구의 분석 대상은 한국교육과정평가원 홈페이지(www.kice.re.kr)의 자료마당에 공개된 2002~2021학년도 중등 영어과 임용시험 기출문제를 기반으로 구축한 코퍼스이다. 전공분야별(영어교육학, 영어학, 영미문학, 일반영어) 전문어휘 추출을 위해 2인의 연구자가 2002~2021학년도 중등 영어과 임용시험의 각 문항의 전공 분야를 상의하여 분류하였다. 그결과 다음 〈표 1〉과 같이 총 679문항을 토대로 13만 단어 수준의 코퍼스가 구축되었다. 분야별로는 영어교육학 분야의 문항이 252개로 가장 많았고 총 5만 단어 수준의 서브 코퍼스가 구축되었으며, 다음으로 일반영어(184문항), 영어학(163문항), 영미문학(80문항) 순의로 서브 코퍼스의 규모가 컸다. TTR(Type-Token Ratio)의 경우 전체 코퍼스가 6.62였고, 다음으로 영어교육학(7.44), 영어학(11.88), 일반영어(14.04), 영미문학(17.11)의 순이었다.

¹⁾ 본 발표는 이혜진, 이제영(2021)의 내용을 보완 및 재구성한 것임.

〈표 1〉 중등 영어과 임용시험 코퍼스의 규모

10000	1 1 0 1 1 0 10	- 1 11		
분야	문항 수	토큰	타입	TTR
영어교육학	252	49,957	3,719	7.44
영어학	163	26,908	3,198	11.88
영미문학	80	21,326	3,649	17.11
일반영어	184	32,496	4,562	14.04
전체	679	130,687	8,647	6.62

2. 분석 도구 및 자료 처리 방법

중등 영어과 임용시험 코퍼스의 전문어휘 추출은 AntWord Profiler 1.4.1을 활용해 진행하였다. 이 프로그램은 General Service List(이하 GSL; West, 1953), Academic Word List(이하 AWL; Coxhead, 2000) 등 기존 학계에서 개발된 어휘 목록, 또는 사용자가 특정 목적을 위해 작성한 어휘 목록에 포함된 단어의 활용 빈도 및 분포를 분석하는 기능을 제공한다.

전문어휘의 선정 기준으로는 빈도(frequency)와 분포(range)를 활용하였다. 전문어휘는 중 등 영어 임용시험 코퍼스에서 일정 빈도와 일정 분포 이상을 기록한 내용어 중 GSL과 AWL에 포함되지 않은 어휘를 선정하는 방식으로 결정하였다. 먼저 빈도는 기존 선행연구(이제영, 김정렬, 2013; 이제영, 이혜진, 2019; Coxhead, 2000)의 기준을 고려하여 서브 코퍼스의 규모가 5만 단어 수준인 영어교육학 분야의 경우 최소 5회, 서브 코퍼스가 2~3만 단어 수준인 영어학, 영미문학, 일반영어 부분의 경우는 최소 3회 이상의 빈도를 기록한 경우 전문어휘로 판단하였다. 그리고 1문항에서만 특정 단어가 반복하여 등장한 경우를 제외하기 위해 최소 2개 이상의 문항에서 등장한 어휘를 전문어휘로 선정하였다.

Ⅲ. 연구 결과

1. 기본 어휘와 학술 어휘의 분포

약 13만 단어 수준의 중등 영어과 임용시험 코퍼스에 포함된 기본 어휘(GSL)와 학술 어휘 (AWL)의 분포를 임용시험 전체와 전공 분야별로 살펴보았고 그 결과는 다음 <표 2>와 같다.

〈표 2〉기본 어휘(GSL)와 일반 어휘(AWL)의 분포

,				
분야	GSL 1000	GSL 2000	AWL	합계
영어교육학	39,262	2,904	4,166	46,332
	(78.59%)	(5.81%)	(8.34%)	(92.74%)
영어학	20,488	1,548	1,694	23,730
	(76.14%)	(5.75%)	(6.30%)	(88.19%)
영미문학	17,417	1,322	478	19,217
	(81.67%)	(6.20%)	(2.24%)	(90.11%)
일반영어	25,575	1,850	2,171	29,596
	(78.70%)	(5.69%)	(6.68%)	(91.08%)
전체	102,742	7,624	8,509	118,875
	(78.62%)	(5.83%)	(6.51%)	(90.96%)

2. 영어교육학 분야의 전문 어휘

앞서 연구방법에서 밝힌 것처럼 영어교육학 분야의 서브 코퍼스에서 최소 5회 이상, 2개이상의 문항에서 등장한 어휘 중 GSL과 AWL에 포함되지 않은 어휘를 영어교육학 분야 전문 어휘로 선정하였다. 그 결과 〈표 3〉에 제시한 것과 같이 총 113개의 영어교육학 분야 전문어휘가 선정되었다. 113개의 어휘는 지도의 편의를 위해 빈도 순으로 목록당 50개씩 어휘를 제시하였다.

〈표 3〉영어:	교육학 분야의 전문 어휘
하위 그룹	단어
List 1	classroom, feedback, vocabulary, score, textbook, grammatical, syllabus, movie, comprehension, authentic, dialogue, phrase, linguistic, video, essay, online, utterance, semester, negotiation, lexical, multimedia, oral, proficiency, blank, clip, comprehensible, clue, peer, discourse, worksheet, engage, web, diet, handout, postcard, repetition, checklist, competence, corrective, fluency, interview, pronunciation, homework, underline, cloze, exam, excerpt, reference, session, verbal
List 2	activate, curriculum, interactional, interlocutor, referenced, script, sematic, software, television, transcript, blog, cue, dialog, holder, instructional, pace, rubric, schema, accomplish, appeal, clothe, collocations, comprehend, declarative, incidental, interlanguage, jacket, magazine, metalinguistic, morphological, picnic, portfolio, routine, scaffolding, segment, vacation, workshop, adjective, audio, auxiliary, bike, brainstorm, chat, cognitive, collocation, column, convey, delete, drill, elicit
List 3	embarrass, graphic, hobby, interpersonal, intonation, kid, museum, outdoor, piano, predetermined, questionnaire, rewrite, usage

3. 영어학 분야의 전문 어휘

영어학 분야의 전문 어휘는 해당 분야의 서브 코퍼스에서 최소 3회 이상, 2개 이상의 문항에서 등장한 어휘 중 GSL과 AWL에 포함되지 않은 어휘를 선정하였다. 위와 같은 기준을 적용한 결과 아래 〈표 4〉에 제시한 것과 같이 총 149개의 영어학 분야 전문어휘가 선정되었다. 이 부분 역시 빈도 순으로 목록당 50개씩 어휘를 제시하였다.

〈표 4〉영어형	학 분야의 전문 어휘
하위 그룹	단어
List 1	syllable, consonant, phrase, adjective, suffix, pronoun, grammatical, utterance, modifier, ungrammatical, underline, preposition, adjunct, affix, assimilation, alveolar, lax, maxim, articulation, nasal, syntactic, adverbials, antecedent, coda, phonological, reflexive, intonation, segment, ungrammaticality, utter, convey, exemplify, infinitive, predicate, diphthong, pitch, prefix, pronunciation, literal, morphemes, peak, adverb, conversational, illocutionary, inflectional, linguistic, matrix, obligatory, phonetic, reference
List 2	rhyme, cluster, conjunction, delete, deletion, derivational, interlocutor, inversion, morpheme, phonemes, phrasal, rhythm, chocolate, embed, flap, glottal, node, onset, paraphrase, specifier, switch, syllabic, usage, accent, addressee, adverbial, anaphor, assimilate, attributive, barn, competence, complementizer, corpus, counterpart, homonymy, inflection, jacket, lateral, lexical, penultimate, phonemic, phonetically, placement, prepositional, quantifier, semantic, singular, velar, affricate, auxiliary
List 3	bilabial, casual, conjoin, discourse, fricative, generalization, guy, implicatures, innocent, intransitive, movie, nucleus, oral, overt, participle, professor, proposition, realizations, rental, tonic, velarized, allophones, alveolars, articulate, asterisk, clausal, dialect, dialogue, generic, glide, gradable, grammaticality, homework, interview, locutionary, monosyllabic, museum, negation, opponent, palatalization, phoneme, pie, preposed, repetition, salient, segmental, semantically, sociolinguistic, vocabulary

4. 영미문학 분야의 전문 어휘

영미문학 분야의 전문 어휘 역시 영어학 분야와 동일하게 서브 코퍼스에서 최소 3회 이상, 2개 이상의 문항에서 등장한 어휘 중 GSL과 AWL에 포함되지 않은 어휘를 선정하였다. 그 결과 아래 〈표 5〉에 제시한 것처럼 총 49개의 영어문학 분야 전문어휘가 추출되었다.

〈표 5〉 영미문학 분야의 전문 어휘

하위 그룹	단어
List 1	biff, felicity, bun, stare, hound, narrator, monster, thou, tone, genius, glance, literally, melancholy, mistress, personality, talent, tiny, awful, cheek, clothe, instinct, metaphor, scream, simile, solitude, stair, terror, thy, vague, bead, boot, cash, confront, desolate, doorway, evoke, gasp, graduate, haunt, horrible, mansion, mood, murmur, nervous, nutmeg, obscure, passion, personification, politely

5. 일반영어 분야의 전문 어휘

일반영어 분야의 전문 어휘 역시 영어학 및 영미문학 분야와 동일한 기준을 적용해 선정 작업이 이루어졌다. 그 결과 아래 〈표 6〉에 제시한 것과 같이 총 116개의 일반영어 분야 전 문어휘가 선정되었다. 이 부분 역시 빈도 순으로 목록당 50개씩 어휘를 제시하였다.

〈표 6〉일반영어 분야의 전문 어휘

하위 그룹	단어
List 1	emotion, professor, gene, museum, interview, cell, classroom, disorder, appeal, biological, emotional, genetic, molecular, species, bug, drug, intellectual, kindergarten, organism, alcohol, atmosphere, cognitive, engage, inanimate, peer, personality, stimulate, bacteria, kid, linguistic, mainstream, portrait, recruit, socialization, symptom, alzheimer, anthropologists, biology, canvas, categorize, cigarette, divine, electronic, embrace, executive, graphic, grocery, meter, multimedia, pitch
List 2	protein, recall, render, scholar, strive, television, therapy, traffic, video, viewer, web, workplace, accomplish, arithmetic, authoritarian, boyfriend, breed, capture, cling, coach, connotation, contend, depict, developmental, diet, dye, endeavors, feed, feminism, genius, ginger, grasp, hobby, hormone, impression, infection, mount, movie, mutation, neurological, personnel, physicist, physiological, pill, prescription, prey, primitive, protest, rattle
List 3	ritual, seasonal, semester, sensor, spectrum, stimulus, thrive, tiny, trash, unhealthy, urgent, utter, verbal, versa, vice, victim, virus

참고문헌

- 김민경, 신창원. (2019). 코퍼스 기반 중등 영어교원 임용시험 출제 경향 분석: 2009-2019학 년도 영어교육학 문항을 중심으로. *영어영문학, 24*(3), 219-243.
- 이제영, 김정렬. (2013). 학술 전문 코퍼스를 활용한 영어교육 전문어휘 추출. Foreign Languages Education, 20(1), 99-122.
- 이제영, 이혜진. (2019). 음성학/음운론 분야 학술논문에서의 학술어휘 및 전문어휘의 분포: 코퍼스 기반 연구. *언어과학*, *26*(1), 95-113.
- 이혜진. (2020). 중등영어임용시험 핵심어휘목록 구축 및 활용방안. 영어교과교육, 19(2),

181-199.

이혜진, 이제영. (2021). 중등영어 임용시험 수험자를 위한 각 영역별 전문어휘목록 선정 및 지도방안 연구. *영어교과교육*, 20(1), 161-178.

Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34(2), 213-238.

이혜진

주소: (54538) 전북 익산시 익산대로 460 원광대학교 영어교육과

이메일: lee.hyejin0419@gmail.com

이제영

주소: (55069) 전북 전주시 천잠로 303 전주대학교 영어교육과

이메일: jylee@jj.ac.kr



A Corpus-Based Study of Gender-Inclusive Words in the Maritime Conventions of "Seamen" or "Seafarers"

October 22, 2022

LIU, Cuilin & JHANG, Se-Eun (Korea Maritime & Ocean University)

The 2022 KASELL and KACL Joint Fall Conference Pusan National University, Busan, South Korea

1

Abstract and Key Words

Gender-inclusive words are interesting in the study of the relations between gender and language. The present study aims to diachronically investigate the trends and context of a pair of gender-related profession terms such as seamen/seafarers and manned/crewed in maritime conventions (self-compiled Corpus of Maritime Conventions of Semen or Seafarers (COMC), compared with that used in general English (The Corpus of Historical American English (COHA)) from a perspective of corpus linguistics. In this presentation, we focus on a pair of seamen and seafarers only. Specifically, based on a self-compiled corpus of maritime conventions relating to seamen or seafarers over time from 1920s to 2010s, we will present some findings on the **trends** in use of these two words in maritime conventions and general English texts. The context of the gender-inclusive terms, seamen vs. seafarers is also our concern because some implications of the developing trends in both specifically purposed English texts of some maritime conventions and generally purposed English texts can be found. The word *seamen* showed a reduction trend in both corpora and it has disappeared in the maritime conventions except for a crew ranking title such as able seamen. In contrast, seafarers showed an increasing trend typically not in COHA but in maritime conventions. But there was an outstandingly high frequency in use of seafarers in both corpora in the same periods of 1950s and 2000s. Seamen seems to have been replaced as a gender-inclusive term seafarers in maritime conventions as well. With newly-added concerns found by analyzing the collocations, some implications were put forward for potential future seafarers and maritime industry in order for us to try to answer 'a big question' such as why we should use **no sexist language** such as gender-inclusive words in a linguistic perspective view.

Key Words: corpus, gender biased words, gender-inclusive words, maritime conventions, seamen, seafarers, trend analysis, maritime conventions, COHA, diachronic approach



CONTENTS

- 1 Introduction
- **2** Previous Studies
- 3 Data and Methodology
- 4 Results and Discussion
- 5 Conclusion

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Conventions of "Seamen" or "Seafarers" —



1. Introduction

1.1 Background of this study: women in the global workforce including maritime sectors

Seafaring is one of the traditionally male-dominated occupations for a long time. In 1992, the International Maritime Organization (IMO) made an estimation of women's representation in seafaring, just 1 to 2 % of the world's 1.25 million seafarers (Belcher et al., 2003, p. 9). The latest large workforce survey reveals that women seafarers in deck and engine departments represent only 1% of all seafarers.(BIMCO& ICS, 2016)

What about women workers in the global workforce?

- Women account for **47.7%** of the global workforce in 2022. (https://teamstage.io/women-in-the-workforce-statistics/#statistics-on-women-in-the-workplace)



Movement for non-sexist language in different/similar periods?

Gender bias in language → gender-neutral term movements in different periods

General English gender-inclusive forms

(e.g., firemen \rightarrow firefighters: late 1970s, Yoder and Aniakudo, 1995, 1996)

- **Family or friends** (boyfriend or girlfriend \rightarrow partner, husband or wife \rightarrow spouse, mother or father \rightarrow parent, sister or brother \rightarrow sibling, son or daughter \rightarrow child, aunt or uncle \rightarrow pibling (parent's sibling))
- <u>Male-dominated professions</u> in the workplace (*mailman* → *mail carrier* or *mail worker*, *chairman* or *chairwoman* → *chairperson*, *policeman* → *police officer* (Holdaway and Parker 1998), and others.

Definition of gender-inclusive words

It has been defined as "speaking and writing in a way that does not discriminate against a particular sex, social gender or gender identity, and does not perpetuate gender stereotypes" (United Nations, 2020)

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Conventions of "Seamen" or "Seafarers" —



Background

What about Maritime English in the movement for non-sexist language?

The language used in maritime sectors can also illustrate such a dominance of males. A number of documents on the basic assumption that seafarers are men were described, and the expression of 'he', 'his' and 'him' can be observed in the literature. In fact, the word 'seafarers' is gender-neutral and has been replaced by the masculine forms, such as 'seaman'. However, there are still some arguments about which word is appropriate. In the 294th session of the International Labor Organization (ILO, 2005), ILO officially proposed the use of the word 'seafarer' instead of 'seaman' in Maritime Labor Convention (2006).

Hence, the idea underlying the present study comes from this historical issue. We intend to check out how gender-inclusive terms such as 'seafarers' instead of 'seamen' and 'crewed' instead of 'manned' have changed over time in both maritime English and general English.



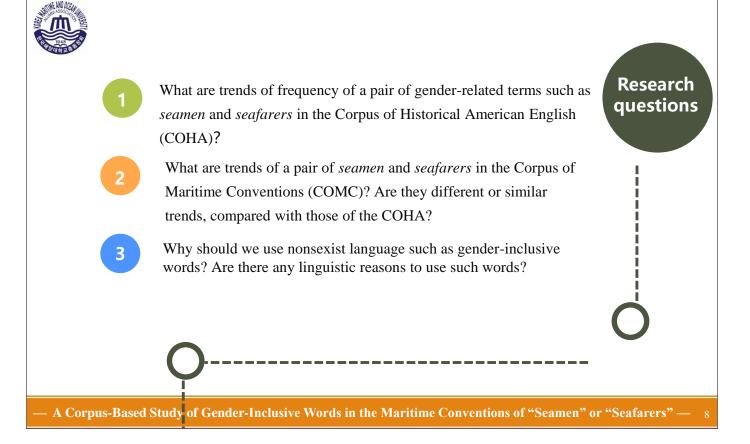
The purpose of this paper

The present study aims to diachronically investigate the **trends and context** of a pair of gender-related profession terms such as *seamen/seafarers* and *manned/crewed* in maritime conventions (self-compiled Corpus of Maritime Conventions of Semen or Seafarers (COMC), compared with that used in general English (The Corpus of Historical American English (COHA)) from a perspective of corpus linguistics. In this presentation, we focus on a pair of seamen and seafarers only.

Specifically, based on a self-compiled corpus of maritime conventions relating to *seamen* or *seafarers* over time from 1920s to 2010s, we will present some findings on the trends in use of these two words in maritime conventions and general English texts.

The context of the gender-inclusive terms, *seamen* vs. *seafarers* is also our concern because some implications of the developing trends in both specifically purposed English texts of some maritime conventions and generally purposed English texts can be found.

With newly-added concerns found by analyzing the collocations, some implications were put forward for potential future seafarers and maritime industry in order for us to try to answer 'a big question' such as why human beings are willing to use **no sexist language** such as gender-inclusive words in a linguistic perspective view.





Procedures of the present study

To find answers to the first and second questions

To find an answer to the third question

1st Step:

The frequency of seamen and seafarers in the Corpus of Historical American English(COHA) and the Corpus of Maritime Conventions(COMC) is analyzed both diachronically. and synchronically.

2nd Step:

A contrastive comparison of gender-related words like a pair of seamen and seafarers between the two corpora (COHA and COMC).

3rd Step:

Collocations of seamen and seafarers are extracted for finding out some differences and similarities of such collocates to provide some linguistic reasons for no sexist language.

- A Corpus-Based Study of Gender-Inclusive Words in the Maritime Conventions of "Seamen" or "Seafarers" -



2. Previous Studies

A lot of studies on gender and language in general linguistics

The interest in the impact of social factors on usage of gender-related language initiated about in the late 1970s. The use of androcentric generics has been a focus of considerable **feminist energy** as well as a subject of **sociolinguistic inquiry** (Bodine, 1975; Dubois and Crouch, 1979; MacKay, 1980).

- Some findings suggest that gender-marked language has implications for assumptions about the gender of a target person (McConnell & Gavanski, 1994, Schneider & Hacker, 1973; Wilson & Ng 1988).
- McConnell & Fazio (1996) gave support to a hypothesis that man-suffix occupation titles (e.g., chairman/chairwoman) would lead perceivers to interpret a social target's personality as more masculine than no-suffix occupation titles (e.g., chair) and that person-suffix occupation titles (e.g., chairperson) would lead perceivers to interpret a social target's personality as less masculine than no-suffix occupation titles.
- Braun, Sczesny & Stahlberg (2005) studied influence of different types of German generics on the
 cognitive inclusion of women. And they found that he inclusion of women is higher with 'nonsexist' alternatives than with masculine generics.



Two main streams on gender and language in corpus linguistics

The way men and women use language

Reference to men and women

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -1



Previous studies

(1) The way men and women use language

- Rayson et al. (1997) used 4.2 million words of transcribed speech from the spoken demographic section of the British National Corpus (BNC), using chi-squared tests to identify which words were distinctive of male and female speech.
- Kilgarriff (2001) used the Mann Whitney test (which compares rankings rather than actual frequencies) on the same data.
- Schmid (2003) took a corpus-based approach to the spoken section of the BNC, looking for differences in particular domains which other authors have claimed to be indicative of gender differences
- However, many corpus studies actually find similarities too, so such research should not be used in order to reify a 'gender differences' model of language use. (Baker, 2010)

(1) Reference to men and women

- Kjellmer (1986) found that terms for males were more frequent than female terms in British and American written corpora from the early 1960s.
- Biber et al. (1999) examined the Longman English Corpus, reporting that 620 nouns end in the suffix –man, whereas only 38 end in –woman.
- Romaine (2001) examined the entire 100 million-word British National Corpus, finding that Mr occurs more often than Mrs, Miss and Ms combined and that terms that potentially erase females such as chairman and spokesman tend to be prevailing.



More recently, Pearce (2008) looked at **collocates** (words which tend to co-occur next to or near another word with statistically significant regularity) of the lemmas MAN and WOMAN.1 He used the corpus analysis tool Sketch Engine (Kilgarriff and Tugwell, 2002) in order to distinguish which **verbs collocated with** MAN and WOMAN when they occurred as subject or object.

Baker (2010) investigated **frequency** and **context** of usage of gender marked language, four equal sized and equivalently sampled corpora of British English in a range of written genres (press, fiction, general prose, learned writing), from 1931, 1961, 1991 and 2006.

The present study seems to lie in **the second stream on gender**. We will follow corpus analysis from a diachronic perspective to investigate **gender-related professional references** in maritime sectors.

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -13



3. Data and Methodology

General English

For the purpose of the present study to make contrastive comparisons between general English corpus and an ESP corpus, we selected Corpus of Historical American English (COHA) with the data of 475 million words of text in more than 100,000 individual texts collected from 1820 to 2019 as a well-balanced reference corpus.

The COHA allows researchers to examine a wide range of changes in English with much more accuracy and detail than with any other available corpus.

(https://varieng.helsinki.fi/CoRD/corpora/COHA/)

Maritime English

In terms of the ESP corpus, maritime conventions were sampled from the IMO and ILO websites. Since we planned to center on the terms, *seamen* and *seafarers*, the conventions related to the two terms either in title or in the content were extracted. The conventions time span covers 1920s, 1930s, 1940s, 1950s, 1960s, 1970s, 1980s, 1990s, 2000s, and 2010s.



Table 1. A list of maritime conventions relating to seamen

	T
TEXT ID	TITLES
M1	Placing of Seamen Convention, 1920
M2	Medical Examination of Young Persons (Sea) Convention, 1921
M3	Labour Inspection (Seamen) Recommendation, 1926
M4	Repatriation of Seamen Convention, 1926
M5	Seamen's Articles of Agreement Convention, 1926
M6	Holidays with Pay (Sea) Convention, 1936
M7	Officers' Competency Certificates Convention, 1936
MO	SHIPOWNERS' LIABILITY (SICK AND INJURED SEAMEN)
IVIO	CONVENTION, 1936
M9	Hours of Work and Manning (Sea) Convention, 1936
M10	Seamen's Welfare in Ports Recommendation, 1936
M11	Accommodation of Crews Convention, 1946
M12	Food and Catering (Ships' Crews) Convention,1946
M13	Paid Vacations (Seafarers) Convention, 1946
M14	Wages, Hours of Work and Manning (Sea) Convention, 1946
N 4 1 E	Convention (No.74) concerning the Certification of Able
IVI15	Seamen,1946
	M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -



Data

Table 2. A list of maritime conventions relating to seafarers

	T.							
1950s	M16	Seafarers' Identity Documents Convention, 1958						
1960s	M17	Convention on Facilitation of International Maritime Traffic, 1965						
1970s	M18	Merchant Shipping (Minimum Standards) Convention, 1976						
	M19	International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978						
1980s	M20	Health Protection and Medical Care (Seafarers) Convention, 1987						
	M21	Seafarers' Welfare Recommendation, 1987						
	M22	Repatriation of Seafarers Convention (Revised), 1987						
1990s	M23	International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978, as amended in 1995						
	M24	Labour Inspection (Seafarers) Convention, 1996						
	M25	Recruitment and Placement of Seafarers Convention, 1996						
	M26	Seafarers' Hours of Work and the Manning of Ships Convention, 1996						
2000s	M27	Seafarers' Identity Documents Convention (Revised), 2003, as amended						
	M28	Maritime Labor Convention, 2006						
	M29	International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, 1978, as amended in 2010						



Two corpus analysis programs for general English and maritime English data

For general English data analysis, we used the web program of COHA available at https://www.english-corpora.org/coha from 26 September, 2022 to 10 October, 2022 for frequency of the data and extract collocates of *seamen* and *seafarers*

For maritime English data analysis, Teleport Ultra was used to help collect the data and Microsoft word was used to clean the data. For data retrieving, **Wordsmith Tools 8.0** was used to make a wordlist for frequency of the data and to extract collocates of *seamen* and *seafarers* as used in COHA.



Figure 1. A Screen Shot of a COHA web program

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -17



Methodology

Collocation in Corpus Linguistics

In corpus linguistics, there is a general agreement that individual words and their co-occurrences contribute to shaping the meaning of words (Lewis, 1997; Nation, 2001; Schmitt, 2000; Sinclair, 1991).

Different types of collocations have been classified. This first type in Sinclair (1998) and Stubbs (2009)'s term is called "semantic preference" which is the relation between a word and semantically related words in a lexical field. The second type is called "colligation" which is the relation between a word and grammatical categories which cooccur frequently with it.

For the purpose of the present study, it is vital to focus on collocation of *seafarers* co-occurring with semantically related words over a left and right window span 5 with a statistic measure like **Mutual Information 3 (MI 3)**, which has been usually used for general English for the COHA and a different statistic measure like **Log-likelihood (LL)**, which has been usually used for maritime English as an ESP for the COMC ((Daille, 1994; Dunning, 1993; Manning & Schutze, 2000).



4. Results and Discussion

4.1 Frequency Comparison for 1st and 2nd Questions

Table 3. Basic information of the dataset of COHA

Data	1920s	1930 s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Total
Tokens	25,700,422	27,707,879	27,399,750	28,661,274	29,122,676	28,829,225	29,851,580	33,149,318	34,821,812	35,452,806	207,194, 125

Table 4. Frequency of seamen and seafarers in COHA (per million)

	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s	
seamen	4.47	6.71	7.85	4.19	2.33	2.98	1.11	0.81	1.23	0.76	100001
seafarers	0.27	0.36	0.36	1.81	0.31	0.24	0.20	0.06	0.98	0.17	1 4 1

— Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" –

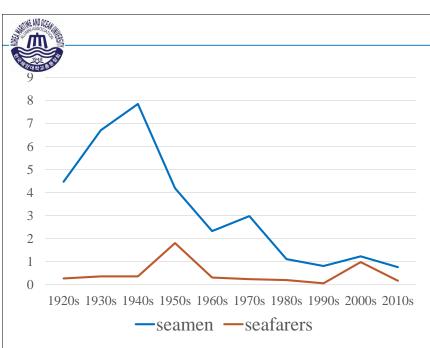


Figure 2. Frequency of seamen and seafarers in COHA (per million)

As for *seamen*, this table shows that the frequency of *seamen* in COHA was increasing from 1920s to 1940s but **fell to a extreme low point in 1950s**. Since then, the frequency of *seamen* has decreased continuously and relatively steadily **to a extremely low point until 2010s**.

Results

In terms of *seafarers*, the frequency is relatively low throughout all the decades. But it can be noticed that the frequency **increased to the highest in 1950s and 2000s.**

Throughout all the decades, seafarers are less used than seamen.



Table 5. Basic information of the dataset of COMC

Data	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Total	_
Texts	5	5	5	1	1	4	3	2	2	1	29	
Tokens (running words)	6,696	9,997	14,905	1,414	9,113	48,879	6,836	24,568	61,777	118,497	302,682	

Table 6. Frequency of seamen and seafarers in COMC (per million)

	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
seamen	6869.77	3601.08	134.18	0.00	0.00	20.46	146.28	325.63	97.12	42.20
seafarers	0.00	0.00	2549.48	4950.50	0.00	1309.36	12141.60	6227.61	9647.60	2076.00

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -

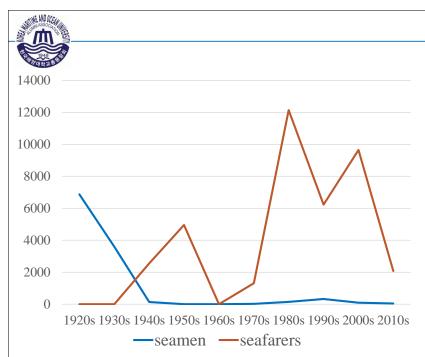


Figure 3. Frequency of *seamen* and *seafarers* in COMC (per million)

Results

As for *seamen*, this table reveals that the frequency of *seamen* in COMC was decreasing from the starting observed decade 1920s all the way to 2010s. However, it can be clearly seen that the frequency tended to be unexpectedly rising in some decades after it has dropped to zero in 1950s.

In terms of *seafarers*, the frequency was zero or nearly zero from 1920s to 1930s. However, the frequency tended to be **rising**, **especially after 1960s**. The frequency **reached the highest in 1980s and 2000s**.



Discussion

Table 7. Concordances of *seamen* in 1970s,1980s, and 1990s

O d d T D		
	Time	Concordance lines
	1970s	Requirements are not those for certification of able seamen nor, except for ship of limited
To clarify the unexpected		size, are
rising of <i>seamen</i> , we searched	1980s	Article 13 This Convention revises the Repatriation of Seamen Convention,1926. Article 14 The formal
for the concordances of		and living conditions since the Labour Inspection (Seamen)Recommendation, 1926, was
seamen. A sub-corpus was set		adopted.
up by including data of 1970s,		with regard to the revision of the Labour Inspection (Seamen)Recommendation, 1926, which is the first
1980s, 1990s and 2010s,		10 This Convention supersedes the <u>Labour</u> Inspection (Seamen)Recommendation,1926.
during which seamen was		Article 11 The on 8 October 1996, and Noting the provisions of the Seamen's Articles of Agreement
used again despite the		Convention, 1926,
frequency of it had dropped to		proposals with regard to the revision of the Placing of Seamen Convention, 1920, which is the third item on
zero in 1950s.		
2010 III 1930s.		Article 7 This Convention revises the Placing of Seamen Convention,1920. Article 8 The formal
The concordances are shown		an act of immediate denunciation of the Placing of Seamen Convention, 1920. Article 10
	1990s	1.A Member
in Table 7.		requirements are not those for certification of Able Seamen as contained in the ILO
		Certification of Able
		Seamen as contained in the ILO Certification of Able Seamen Convention, 1946, or any
		subsequent

—A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -23



Discussion

Table 7. (Cont'd) Concordances of seamen in 1970s,1980s,1990s, and2000s

2010s requirements are not those for certification of Able Seamen as contained in the ILO Certification of Able

Seamen as contained in the ILO Certification of Able Seamen Convention, 1946, or any subsequent

the standards of competence which it required of Able **Seamen** for certificates issued before 1 January 2012

International Labour Organization Certification of Able Seamen Convention, 1946

(No.74) may continue to

International Labour Organization Certification of Able Seamen Convention, 1946 (No.74) may continue to

Table 7 may provide some answers to the unexpectedly use of *seamen*. It reveals that *seamen* adopted in maritime conventions after 1950s seems to be **conventionalized**, **following the existing conventions of 1920s and 1940s.**



Table 8. Comparison of seamen and seafarers in COHA and COMC (per million)

		1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
	СОНА	4.47	6.71	7.85	4.19	2.33	2.98	1.11	0.81	1.23	0.76
seamen	COMC	6869.77	3601.08	134.18	0.00	0.00	20.46	146.28	325.63	97.12	42.20
seafarers	СОНА	0.27	0.36	0.36	1.81	0.31	0.24	0.20	0.06	0.98	0.17
	COMC	0.00	0.00	2549.48	4950.50	0.00	1309.36	12141.60	6227.61	9647.60	2076.00

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" —



Results and Discussion

In terms of *seamen*, Table 8 shows that for both corpora the frequency stared to decrease in **1950s**. But compared with COHA, the frequency of *seamen* **dropped to zero in 1950s and 1960s in COMC**, which never happened in COHA.

As for *seafarers*, the frequency shows a marked increase **in both corpora in 1950s** and **in 2000s**. But the difference can be also noticed, in COHA the frequency of *seafarers* has maintained rather **low** in most decades, whereas in COMC the frequency showed a clear rising trend after 1960s, with the **highest** frequency reached 12141.60.

Based on the comparison result, it can be indicated that:

- the use of **gender-inclusive** words like *seafarers* might **start in1940s in maritime English** and **popular in 1950s** in both corpora. However there seems to be a reducing interest in using *seafarers in general English*. Since the second wave of feminism began in early 1960s, there might be some relations.
- As is mentioned in the background, MLC (2006) has required to avoid using *seamen* but *seafarers*, hence, **the sharp increase in 2000s** seems to show a reflection of such effect.



4.2 Collocation Comparison for 3rd Question

In order to answer the third question, we will compare collocates of *seamen* and *seafarers* in COHA, using MI 3, with those in COMC, using LL, within the right and left five window spans.

Table 9. Top 20 collocates of seamen and seafarers in COHA and COMC

	СОНА	COMC
seamen	IMPRESSMENT(3), NMU(2), A.F.L.(7), LONGSHORE(2), ABLE-BODIED(9), NORESEMEN(2), UNLICENSED(4), SHIPOWNERS(2), NAVIGATORS(2), TORPEDOED(2), MERCHANT(33), SEAMEN(6), MANNED(10), SIXTY-TWO(2), AVERSE(2), C.I.O.(2), NORWEGIAN(8), ADVENTURERS(8) SCURVY(2), DECKS(6)	SEAMEN'S (10), SHIPOWNERS(19), CONDITIONS(19), ABLE(11), WORK (16), FINDING(6), EMPLOYMENT(9), WHICH(22), UNDER(13), CONVENTION (22), AGREEMENT(7), PROTECTION(7), CERTIFICATION(9), FOR(31), INSPECTION(5), SHOULD(14), LABOUR(6), MORE(5), THEIR(6), GENERAL(5)

- A Corpus-Based Study of Gender-Inclusive Words in the Maritime Conventions of "Seamen" or "Seafarers" — 27

19.48 19.48 19.48

Results

Table 9. (Cont'd) Top 20 collocates of seamen and seafarers in COHA and COMC

	СОНА	COMC
seafarers	ADVENTIROUS(1), TELLEPSEN(1), BARBOURS(1), VALERETS(1), VIS-AVIS(1), ANURU(1), INTERNATIONAL(21), S.I.U(8), HALL(6), PORT(5), MTSA(1), AIRCREWS(8), MACASSANS(1), KRAGISH(1), LICENSES(1), SPACEFARERS(1), AIRMAILED(1),	SHIPOWNERS(116), WATCHKEEPING(122), CERTIFICATION(145), EMPLOYMENT(61), ORGANIZATIONS(52), TRAINING(173), FOR(326), LIVING(36), ORGANISATION(33), CONCERNED(67), WORKING(54), SERVING(34), WHO(51), AND(659), RIGHTS (22), WORK (48), CONSULTATION(23), YOUNG(15),
	LONGSHORE(2), UNION(21) INTERNATIONAL(21),	REPRESENTATIVES(18), AGREEMENT(32)

A rough classification of semantic context was made based on references to a website of semantic domains. (https://semdom.org/full-search?keys=international)



- Comparison between COMC and COHA in terms of *seamen*: In COHA it seems that *seamen* occurred with some **sexist words** like *manned*(10) or with somewhat **masculine** word like *body-abled*(9) and *impressment*(3), and it has lexical preference with some negative adjectives, like *scurvy*(2) and *averse*(2). However, in COMC *seamen* didn't appear in such similar contexts.
- Comparison between COMC and COHA in terms of seafarers: Seafarers was involved in broader semantic domains in COMC than in COHA. But in both corpora, seafarers was used with gender-inclusive professional references, such as spacefarers(1) in COHA and representatives(18) in COMC.
- Comparing *seamen* with *seafarers* based on the two corpora together: *Seamen* seems to be used in very specific semantic context. For example it was significantly used together with some **racial references** such as *Norsemen*(2) and *Norwegian*(8).
- In contrast, *seafarers* is not confined in such **specific semantic context** and it seems to be habitually used in a rather **broad semantic context**, involving collocates from various semantic domains.

— A Corpus-Based Study of Gender-Inclusive Words in the Maritime Convention of "Seamen" or "Seafarers" -



5. Conclusion

Question 1

Seafarers seems to be less used in general English throughout all the decades but in 1950s and 2000s seafarers were used much more than other decades. As for seamen, it was popularly used from 1920s to 1940s, but from 1950s there was a reduction tendency in use of it.

Question 2

The word *seamen* showed a reduction trend in both corpora and **it has disappeared** in the maritime conventions except for a crew ranking title such as *able seamen*. *Seamen*, seems to have been replaced as a gender-inclusive term *seafarers* in maritime conventions as well.

In contrast, *seafarers* showed an increasing trend typically in maritime conventions but not in COHA. But *seafarers* showed an outstandingly high frequency in both corpora in the same periods of 1950s and 2000s



Question 3

Seamen, as a sexist term, seems to be habitually used with gender-marked words or descriptions with statistical significance. In contrast seafarers, as a gender-inclusive occupational reference, tends to be habitually used with nonsexist references in other industries. It seems that sexist terms should be avoided in maritime English because it might strengthen the long years of masculinism in marine occupational culture.

Implications for language and culture in maritime:

The word *seafarers* is becoming increasingly popular in the maritime language, which might reflect some tendency of a promising future for women. Language change might rise people's awareness to empower women in the maritime industry, for example, with gender-inclusive words broadly used, a more inclusive culture might be developed so as to gradually break the male-dominated occupational culture in the maritime community.

- A Corpus-Based Study of Gender-Inclusive Words in the Maritime Conventions of "Seamen" or "Seafarers" - $_3$



Thank you very much for listening to our presentation

비판적 담화분석에 기반한 우리나라 고등학교 실용영어 교과서 분석

최예슬 (대덕고등학교) 민수정 (공주대학교)

● 실용영어과목의 특성, 성격, 목표

•실용영어 교과서와 구성

•실용영어 교과서 분석

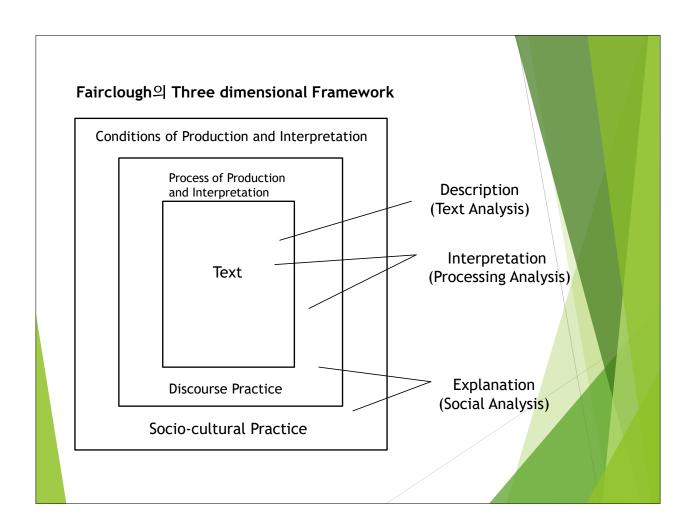
• 대안제시

• 결론 및 제언



실용영어과목의 목표

- 가. 실생활에서 필요로 하는 실용적 의사소통능력을 기른다.
- 나. 실생활 중심의 다양한 주제에 관한 말이나 대화를 듣고 이해한다.
- 다. 실생활 중심의 다양한 주제에 관한 글을 읽고 이해한다.
- 라. 말이나 글을 사용하여 목적, 상황, 형식에 맞게 의사소통을 한다.
- 마. 실생활의 다양한 자료와 정보를 활용하여 진로를 탐색하는 능력을 배양한다.

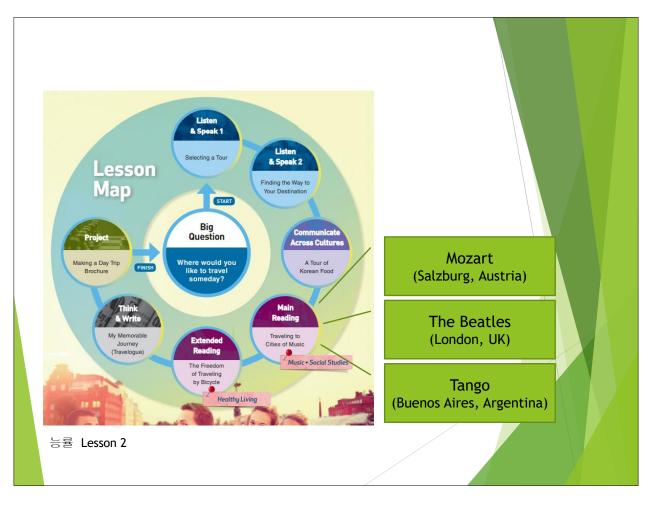




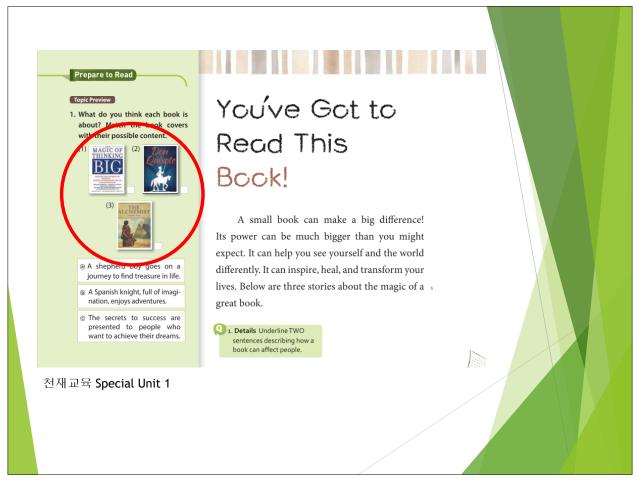
1. Content

		YBM	능률	천재교육
•	학교생활 및 진로교육	Lesson1	Lesson1(E) Lesson 3 Lesson4	Lesson1
	문화교육	Lesson 3	Lesson 2 Lesson1(E)	Lesson 2 Lesson 3 Lesson 5
	세계시민역량	Lesson 2	Lesson5(E)	Lesson 4
	일상생활	Lesson 4 Lesson 5	Lesson2(E)	
	문학, 인물, 역사	Special Lesson	Lesson1 Lesson3(E) Lesson 5 Lesson6 (E)	Lesson 6

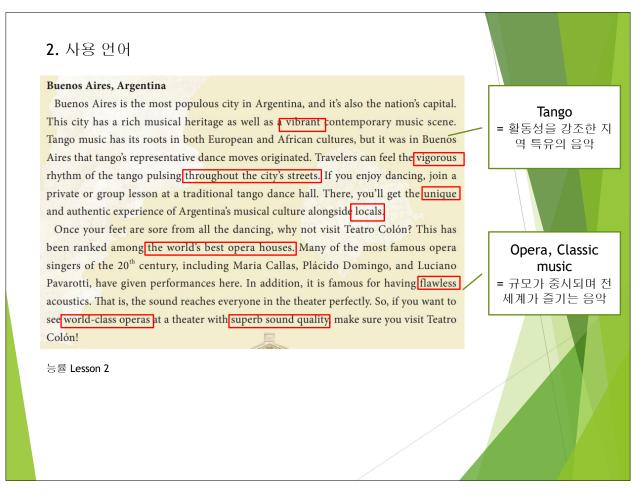
E* : Extended Reading











wanted independence from the West Pakistani government. In March 1971, East Pakistan declared itself the independent nation of Bangladesh. West Pakistan sought to hold onto Bangladesh by force, and war followed. Concerned about the disastrous condition of his country, Yunus returned home to head the Economics Department at Chittagong University. The war and a drought, with its accompanying famine, had devastated the newly formed country. Villagers in Bangladesh were starving without food and fresh water. Parents, with their barefoot children in tow, trudged along dirt roads toward nearby fertile hills in a desperate search for something to eat.

㈜ YBM Special Lesson

지역의 현상태를 나타내는 단어들

The Story of Arinez Gilyard

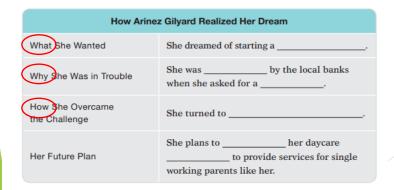
Arinez Gilyard was a single mother with three children. After she was fired from the company where she had worked for 20 years, she decided to start her own business; a daycare center, as it had long been her dream. However, when she asked for a loan, she was rejected by the local banks. She then turned to the Grameen Bank's loan program based on the concept of microcredit. It loans money to lower-income women who want to start their own businesses. She enrolled in training programs and started her first center using the loan money. Gradually,

her business expanded. Soon, she paid back her initial loan to the Grameen Bank.

Now Ms. Gilyard has big plans to expand her daycare facility to operate 24 hours a day, providing services for single working parents who have no family or friends to watch their kids







3. 본문 내 질문 (Comprehension Questions)

Another example of upcycling is in buildings. Here is a picture of the Tubohotel, an imaginative building not far from Mexico City in Mexico. It is made out of sections of discarded concrete tubes! Each of the tube sections measures 3.5 meters long and 2.4 meters wide, large enough to accommodate a queen-sized bed. Some rooms are



stacked in pyramids of three tubes. Though the idea of sleeping inside a concrete tube may not appeal to everyone, the architects have made sleeping inside a pipe not only comfortable but also memorable.

One of the main reasons why people are attracted to upcycled construction is its reduced environmental impact. Like other upcycled items, upcycled buildings use waste products, such as concrete tubes, and therefore reduce the consumption of new raw materials. The benefit of upcycling is not just that it repurposes discarded materials. When new materials are used in construction, they usually emit CO₂ into the environment. On the other hand, upcycled buildings use secondhand materials that no longer emit CO₂. Thus, upcycled buildings become more eco-friendly than new constructions.

Can upcycling be carried out only by those who have professional training or special gifts? Not necessarily. Consider the following interview about two high school students who made and sold their upcycled products for charitable purposes. They say that passion and creativity are all it akes to upcycle.

Q7 What makes upcycled buildings more eco-friendly than new constructions?

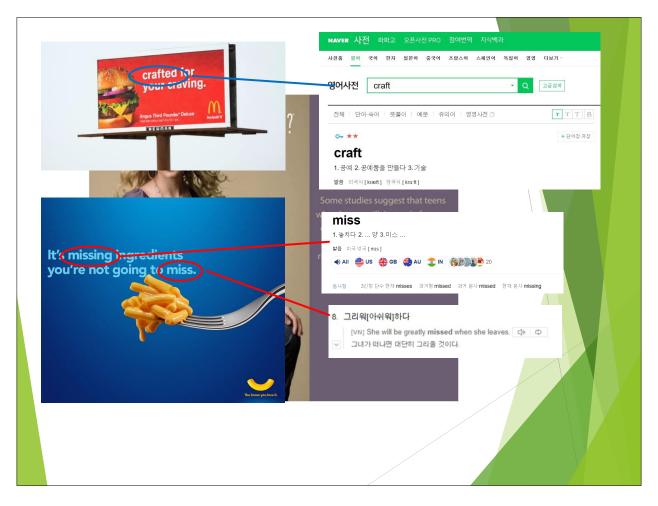
What makes upcycled buildings more eco-friendly than new constructions?

5. 문제점

- 1. 글의 장르가 설명문에 집중되어 있으며 의사소통기반의 새로운 미디어 (SNS, Youtube 등)를 비교적 적게 다루고 있다.
- 2. 내용영역의 소재가 아메리카와 유럽 대륙에 치우쳐있음.
- 3. 2000년대 중반의 자료가 매우 많으며, 새로운 형태의 미디어 (SNS, Youtube등을 활용한 텍스트나 활동이 적은 편.
- 4. 대부분의 질문과 활동이 주로 내용(contents)의 확인에 치우쳐 있으며, 비판적 읽기를 위한 활동이 부족함.

We need a long-term plan to give our team a new home in Oceanside. Their present stadium is too far from the city. Everyone complains it is too cold in the winter and too hot in the summer. Our team needs a world-class stadium to play in. We support the mayor's plan to build a new stadium downtown. "This project will bring in lots of visitors, and that means dollars for restaurants and other local businesses," said Mayor Pat Richardson. Rather than throwing money into repairing the present stadium, we should build the Bisons a new home stadium that we can be proud of.

- is a waste of tax money that should be used for more urgent matters. They the mayor's plan to build a new into repairing the present + Me (pur) in A Yes into repairing the present - Nounce with the mayor's plan to build a new into repairing the present - Nounce with the Nounce wi 을"两环酸Chebr效)을 좌절적으로 조현범. 5. What is the hidden n ets of visitors, and that means that means of the mayor and supports building a new could be the first of the mayor and supports building a new of the side of the side of the side of the mayor and side of the mayor and side of the s 람들에요동의하고 있으며 많은 이익을 가져다준다는 것을 통해 독자가 글쓴이의 논조를 따 ☐Mayor Ent Richardson. •A world-class, be proud of 와 같은 표현을 통하여 독자가 지역팀에 대한 배정과 자긍심을 다겠다고 이를 새 스타디움 건설의 지지로 이어지게 하는 의도를 보임. • 기존 스타디움 보수에는 throwing money라는 부정적 표현을 사용하여, 독자가 보수 공
- 사에 부정적 입장을 가지게 유도함.





7. 결론 및 제언

(1) 실용영어 교과서의 장점:

- 다양한 내용 영역을 다루고 있음.
- 설명문을 중심으로 하고 있지만, 다양한 글의 종류를 다루고 자항.
- 학생 수준에 맞는 과업과 프로젝트를 제공하고 있으며 이는 학생의 흥미를 끌기에 충분할 것으로 보여짐.

(2) 실용영어 교과서의 단점:

- 자료의 시의성이 부족함.
- 정보획득에 초점을 맞춘 독해활동이 대부분이며 비판적 읽기를 장려하지 않음.
- (3) 교사의 역할: 교과서의 단점을 보완할 수 있는 방법을 적극적으 연구하고 도입해야 함.

북한 영어교과서에 나타난 사회주의 교육 테제

김정렬(한국교원대학교)

북한식 사회주의 교육은 전반적으로 1977년 9월 5일에 조선노동당 중앙위원회 제5기 제 14차 전원회의에서 김일성에 의해서 발표된 사회주의 교육에 관한 테제에 기반하고 있다 (차승주, 2013). 북한은 이 날을 기념하기 위하여 9월 5일을 교육절로 지정하고 남한의 스승의 날과 유사한 여러 가지 교육적 행사를 개최하고 있다. 이와 같은 교육테제의 비중을 생각하면 비교적 최근에 발표된 2013년 신교육 강령도 그 이론적 기반은 사회주의 교육에 관한 테제의 틀 속에서 일부 내용을 시대에 맞도록 조율한 정도이다. 교육테제를 발표한 9월 5일을 '교육절'로 정하여 매년 기념행사를 개최할 정도로 중요한 북한의 사회주의 인간교육을 위한 강령이다. 2013년 신교육 강령을 발표할 때도 사회주의 교육에 관한 테제는 북한의 이념적 토대인 주체사상을 구현하는 교육대강이며 김일성의 교육 경험과 업적을 과학적 교육이론으로 집대성한 것으로 사회주의 건설의 전체 역사적 기간을 통틀어서 교육에 적용해야 할 이론적 무기이고 지침이라고 역설하고 2013년 신교육 강령도 사회주의교육 테제의 바탕위에 서 있다(교육신문사, 2012).

북한정권은 1949년 정권수립부터 교육에 최우선을 두고 1967년에 초등학교 4년제 의무교육을 실시하고 1972년에 유치원 1년, 초등학교 4년, 중학교 6년의 전체 11년제 의무교육을 실시하였고 2013신교육 강령을 통해서 현재의 초등학교 4년을 5년으로 늘리면서 전체 12년 제 의무교육을 완성하였다(차승주, 2013). 의무교육의 기간만을 보면 북한이 남한의 9년제 의무교육 기간보다 3년이 더 길다고 볼 수 있다. 그러나 북한의 경우는 여름 모심기철과 가을 추수기에 노동시간이 있고 학교청소 시간, 자기비판과 상호비판과 같은 총화시간, 사상교육 등을 제하면 한 학년의 수업시간은 남한에 비해서 현저하게 짧다.

남북한 교육은 모두 지덕체를 고루갖춘 인간을 길러내는 것이 교육의 목표이다. 다만 북한은 남한의 민주주의적 가치인 학생 개개인의 다양성과 자유보다는 사회주의 인간성을 가진 "전체는 하나를 위하여, 하나는 전체를 위하여"라는 유일사상 체제에 순응하는 인간으로 키워내는 것이다. 북한의 교육은 크게 유일체제를 뒷받침할 수 있는 사상교육, 과학기술교육과 체육교육을 근간으로 하고 있다. 사회주의 교육 테제에서 외국어교육은 과학기술교육을 위한 도구교과이다.

북한에서는 1958년부터 중학교에서 외국어 교육이 시작되었으며, 최근에는 변화하는 시대에 적응하기 위하여 영어, 중국어 등의 실용적인 외국어 교육을 한층 강조하고 있다. 이러한 변화는 영어가 국제어로 통용되는 세계적인 추세를 거스를 수 없다는 김정일의 교육원칙에서 나온 것으로, 김정일은 외화벌이를 많이 하기 위해서라도 영어를 알아야 한다며 영어 교육의 중요성을 강조하였다(한만길, 2001).

2008년 9월부터 초등학교에서도 새로이 영어 과목의 수업을 실시하게 된 것은 최근 북한에서 영어 교육을 얼마나 중요히 여기는지에 대한 반증이라고 할 수 있다(고성호외, 2015). 김정은시대 북한 교육과정 개편에 관한 조정아(2014)의 연구에 따르면, 북한은 중등학교의

영어교육에 있어 EU(Europe Union) 지역의 교수 인력을 초빙하여 강좌를 개설하는 등의 다각도의 노력을 하고 있다. 북한의 이러한 교육 방침은 시대적 요구에 맞추어 영어에 대한 중요성을 인식하고 국가적인 차원에서 영어 교육을 강화하기 위한 조치로 해석된다. 이 와같이 과학기술교육의 도구로서 실용적인 목적을 가지고 시작된 영어교육에서 사회주의 교육테제는 그대로 적용되고 있다. 김일성은 사회주의 교육테제에서 외국의 과학기술도 우리의 실정에 맞도록 적용해야 한다면서 자주 인용하는 말로서 다른 나라에서 대포를 사용하면 그 기술을 가져와서 우리나라와 같이 산악지형이 많은 곳에서는 곡사포로 변형해서 적용할 수 있어야 한다고 말하면서 교육에서 사대주의와 교조주의를 배격하고 있다. 이 말은 교과서 집필시에 서양의 영어 원문을 그대로 가져와서 실는 경우를 극히 제한하는 형태로 실제로 영어교과서에서 나타나고 있다. 이는 남한의 영어교과서가 대부분의 본문을 원문 그대로 실으면서 일부 수정보완하는 것과 달리 북한의 교과서 본문은 대부분 자작된 글이다.

영어교과서는 북한 정권의 교육과정과 밀접한 연계성을 가지며, 북한의 교과서가 "국가가 규정하는 공식적인 교육내용으로서 절대적인 권위"를 내포하고 있다고 볼 때에 사회주의 교육테제가의 바탕위에 영어교육이 서 있다고 보는 것이 타당하다. 이는 북한의 유일사상인 주체사상과 반제혁명세력의 중심인 백두혈통을 중심으로 한 3대 세습 체제를 확립하고 유지하는데 근간을 두고 영어교과서도 만들어지고 학습되어야 한다는 것을 의미한다. 실제로 2001년 북한의 고등중학교 영어 교과서 2학년의 머리말에는 '학생들은 다른 나라말도 조선혁명을 위하여 배운다'는 문구가 나와 있다(임중관외, 2001).

2012년 9월, 북한은 기존의 11년이었던 의무교육제를 12년으로 변경하는 법령을 발표하고 교육과정과 교과서를 개편하고 있다. 북한의 교육과정과 교과서 정책 변화 방향을 3대세습의 지도자 체계를 강화하기 위하여 초·중등교육 전체에서 지도자 관련 과목 및 수업시수가 강화되었다는 것이다(김정원, 김지수, 한승대, 2015). 영어 교육과 관련한 주요 변화를 살펴보면, 정치적 목적 외에 과학기술 교육에 복무하는 실용적 목적이 가미되었다는 정황이 여러 가지 면에서 포착된다. 영어 교과의 수업 시수가 확대되었고, 기존의 '외국어'라는 포괄적인 교과목명 대신에 '영어'라는 특화된 교과를 지정하였다.

제한된 자료의 양과 접근성에도 불구하고, 1990년대 이후 북한이탈주민의 증가와 북한의 영어 교과서 관련 자료의 확보에 따라 북한의 영어 교육 및 교과서에 관한 연구는 꾸준히 진행되어 왔다. 북한의 영어 교과서를 분석한 선행 연구들은 시대적 배경과 연구자의 연구목적에 따라 그 연구 대상이 각각 다르게 나타났으며, 이는 크게 북한의 영어 교과서만을 집중적으로 연구한 것과 남북한 영어 교과서를 비교한 연구로 구분 지을 수 있다.

북한 영어교과서에 관한 그동안의 양적 연구에서 과학기술 내용어휘, 자작된 본문, 참여적 영어교육이라는 특징들을 밝히고 그 구체적인 내용을 밝혔다.[북한영어교과서 양적연구들] 본 연구의 목적은 계량적인 방법으로 현상을 기술하는 것을 넘어서 그동안 밝혀진 북한 영어교과서의 특징이 어디에서 시작되는지 그 원인이 되는 북한의 유훈통치를 통해서알아보고자 하였다. 마르크스 레닌의 사상에서 북한은 사회주의에서 공산주의로 넘어가는 과도기에서 일시적으로 허용되는 독재정권을 사회주의 혁명완성을 한반도의 완전한 공산화

로 정하고 그동안 백두혈통의 세습정권을 무한정 연장하고 있다. 백두혈통 정권은 이미 고인이 된 김일성을 위대한 수령, 김정일을 친애하는 지도자로 우상화하고 그들이 남긴 소위 말씀을 북한의 모든 교과서, 연구물, 기고문에 인용하면서 그 인용문에 근거해서 내용을 전개하는 방식으로 유훈통치를 실시하고 있다. 본 연구에서는 김일성의 사회주의 교육테제에서 주장된 내용을 구조적으로 분석하고 그동안의 영어교과서에서 발견된 양적 현상에 대한원인을 찾아보려고 하였다.

본 연구의 구체적인 목적은 그동안의 북한 영어교과서 연구에서 나타난 특징들을 요약하고 그러한 특징들의 원인을 북한의 사회주의 교육테제에서 찾아보고자 하는 것이다. 이를 위해서 우선 북한영어교과서 연구를 통해서 그동안 밝혀진 특징들을 홍정실, 김정렬(2019)와 김정렬(2020)을 중심으로 정리하면 다음과 같다.

첫째, 김정일 집권기부터 영어교과서 표지에 주체연도를 표시함으로써 김일성 우상화가 나타났다. 둘째, 김정일 집권기부터 서문에 머리말이 등장하였는데 머리말을 통해 북한의 영어교과서가 김정일을 찬양하는 지문과 문법, 독해력을 중시하는 내용으로 이루어짐을 알 수 있다. 그러나 김정은 집권기에는 머리말의 내용이 변화하여 북한의 영어교과서에 언어 의 4기능이 종합적으로 포함되고 과학기술 등과 통합된 다양한 지문이 나타남을 알 수 있 다. 셋째, 김일성과 김정일 집권기 영어교과서의 구성체계는 문법, 번역식 교수법에 기초하 여 구성되었다. 두 교과서의 차이점은 Dialogue부분이 김일성 집권기에는 독립적인 문장을 반복해서 말하는 연습이었지만 김정일 집권기에는 두 사람간의 대화 형식으로 바뀌었다. 그리고 김정은 집권기에는 언어의 4기능이 모두 명시적으로 교과서에 나타나고 그에 따른 다양한 활동이 이루어졌다. 넷째, 교과서의 내용은 김일성, 김정일 집권기의 영어교과서 모 두 우상화 및 애국심 고취, 미 제국주의 비판, 과학기술과 관련된 내용 등으로 이루어졌으 나 김정은 집권기에는 세습체제 유지 내용에서 벗어나 일상생활과 과학, 컴퓨터 등 더 폭 넓은 주제의 내용으로 이루어졌다. 다섯째, 북한 영어교과서의 경우 남한 영어교과서에 비 해서 전문학술 어휘가 좀 더 많이 나타났다. 자본주의 시장경제를 따르는 남한과 사회주의 계획경제를 따르는 북한 영어교과서에는 이와 관련된 어휘의 사용과 의미적 차이가 발견되 었다. 아울러, 종교와 문법 용어의 사용에 있어서 차이가 나타났다.

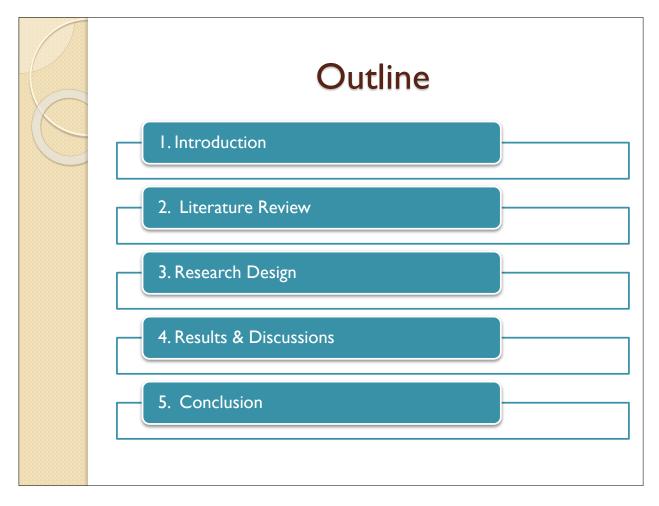
참고문헌

- 고성호, 권영경, 김정수, 이미경, 정은찬, 차문석, 최은석. (2015) 북한 이해. 서울: 통일교육 원. 발간등록번호 11-1250054-00004-10.
- 김정렬. (2020). 남북한 영어교과서 어휘의 차이. 한국콘텐츠학회 논문집, 20(1), 107-116. DOI: 10.5392/JKCA.2020.20.01.107
- 김정원, 김지수, 한승대(2015). 북한 초·중등 교육과정 및 교과서 정책 변화 방향, 한국교육, 42(4), 29-55.
- 박약우, 박기화, 김진철, 고경석, 정국진. (2001). 북한 영어 교과서 분석. 서울: 한국문화사.

- 임중관, 박태선, 김창순, 안혜순. (2001). 영어: 고등중학교 2. 평양: 교육도서출판사.
- 조정아. (2014), 북한의 새로운 세대와 교육: 김정은시대 북한 교육정책 방향과 중등교육과 정 개편. 세계북한학 학술대회 자료집, 제 1권, pp.72-97.
- 차승주. (2013). 김정은 시대 북한 교육의 변화에 관한 소고: 1970년대와의 비교를 중심으로. 북한연구학회보, 17(1), 179-203.
- 한만길. (2001). 북한에서는 어떻게 교육할까. 서울: 우리교육.
- 홍정실, 김정렬. (2019). 북한 영어교과서의 통치권력별 통시적 분석. 학습자 중심 교과교육, 19(6), 599-619. DOI: 10.22251/jlcci.2019.19.6.599.

Thematic Choice in Argumentative Essays by Vietnamese EFL Learners

Presented by Kevin S. Parent Quynh Thuy Bui



I. Introduction

EFL learners in general, Vietnamese EFL learners in particular have difficulties in composing a coherent essay which meets native English speakers' expectation

To make their essays more coherent and cohesive, they use cohesive devices such as coordinating conjunctions, based on Halliday and Hasan's approach (1967)

Halliday and Hasan's theory has been criticized by other researchers since this theory just focused on solving surface cohesion of text

I. Introduction

Green&Christopher & Mei, 2000; Jalilifar, 2010 • Studies on sematic organization on theme and thematic structure have since emerged

Halliday & Matthiessen, 2014

- Thematic structure is constructed by theme and rheme in a clause.
- Theme is "the element that servers as the point of departure of the message; it is that which locates and orients the clause within its context"
- By choosing the theme as the starting point of message, writers guide their audience in processing and interpreting message.

Witt & Faigley, 1981

 To compose effective essay, EFL learners should learn how to place information in the theme and rheme to create a coherent message

I. Introduction

Many researchers analyzed the role of theme in textual organization and problems in using theme in EFL learners' writing (Alonso&McCabe, 1998; Chang& Lee, 2019; Hawes & Thomas, 2012; Wei, 2015).

Following this line, the study focuses on the thematic choices in argumentative essays written by Vietnamese EFL learners and those written by American professional writers to determine the similarities and differences in theme usage between the two groups.

A thorough analysis of the Vietnamese learners' essays is conducted to explore potential factors and some pedagogical implications will be suggested to help Vietnamese EFL learners address the problems of using theme and improve coherence in their writing.

Research Questions

I. What are the similarities and differences between the usage of themes by Vietnamese EFL learners and by American professional writers?

2. Based on the differences found in the study, on which aspects of theme do EFL learners need further instruction?

2. Literature review

2.1 Theoretical background

2.2 Previous studies on Themes in EFL/ESL writing

2. I Theoretical background

Halliday's framework (2014)

Theme is the point of departure of the message, it is that which locates and orients the clause within its context

Types of theme

Topical theme: The clause presents experiential elements: the process itself, the participants in the process, and circumstances related to process: time, manner

- Unmarked and marked themes

Interpersonal themes function to explicitly reflect speaker/writer intentionality

- Vocatives, modal/comment adjuncts and finite verbal operators

Textual theme is any combination of continuative, conjunction, or conjunctive elements, which links ideas

2. I Theoretical background

Thompson (2008)

To identify Theme, a concept of grammatical metaphor is adopted, in which interpersonal and textual Theme may be "experientialised" and recognized in a separate clause or norminalization

- I. It is true that **demand for cocoa** had been rising
- 2. I am becoming suspicious of the whole situation. *The reason is*, **she** wants me to pay for a visa.

This view will be used in my study

2.2 Previous studies on Themes in ESL/EFL writing

 Numerous researchers compare similarities and differences in thematic choices between English and another language (Alonso & Mc Cabe, 1998; Chang & Lee, 2019; Green et al., 2000; Hawes, 2015; Hawes & Thomas, 2012; Herriman, 2011; Wei, 2013a, b, 2015)

2.2 Previous studies on Themes in ESL/EFL writing

Previous studies

- Give similar conclusions: EFL/ESL learners code inappropriate information in the theme, lack awareness of how themes function in an unfolding text
- => Lack coherence in their writing

3. Study Design 3.1 Participants Data A national university second-year 52 VE from writing students, who test completed 4 writing classes 20 AE from - 52 female students Editorials of The at the age of 19-20 **New York Times** (2021)- Intermediate level (lelts score: 5.5-6.0)

General features of datasets

Figures	VE	AE
No. of essays	52	20
No. of words	15367	19735
No. of T-units	880	880

3.2 Procedures of analysis

The unit of analysis T-unit

The T-unit is a clause complex, containing one main independent clause and all the hypotactic clauses.

- Be used in the analysis of written and spoken discourse, particularly in Theme and TP patterns (Ebrahimi & Ebrahimi, 2012; Fries, 1981; Herriman, 2011; Mc Cabe, 1998; North, 2005; Thompson, 2008; Wei, 2016)

Theme identification

Based on Halliday's framework (2014)

Incorporate Eggin's view on complicated cases: thematic equative or non-equative themes

The "experientialized" concept by Thompson (2008) is employed.

Interview

To have further understanding about the reasons why the learners prefer to use some specific themes in their writing, an online interview is conducted.

- Why did you use two interpersonal themes in a T-unit?
- Why did you use more than one textual theme in a T-unit?
- What kind of cohesive devices you guided students to write a coherent essay?

4. Results and Discussions

Table 4. Distribution of Theme Types in VE and AE

Thoma types	Groups		
Theme types	VE	AE	
No. of Topical	880	880	
No. of Marked	170	168	
No. of Unmarked	710	712	
No. of Interpersonal	132	25	
No. of Textual	405	166	

4. I Theme markedness

Theme types	Groups		
	VE	AE	,
No. of Marked	170	168	
No. of Unmarked	710	712	

Both groups have similar preferences in using marked and unmarked themes

Unmarked Theme: overuse of "existential *there*", increase new Themes (Similar with Mc Cabe's (1998) findings).

Use numerous unmarked Themes continuously, in monotonous ways.

4.1 Theme markedness

Marked Theme: helps highlight circumstances of time or place (Butt et al, 2000) or signals structural shifts as the text unfolds

In many cases, using marked Themes disrupt the flow of text, create more New Theme These EFL learners do not recognize the function of Themes in organizing text, and might intuitively select something as Themes

4.2 Interpersonal theme

EFL learners use more interpersonal themes, over five times higher than American writers (Green et al., 200; Wei, 2016)

Difference in using types of interpersonal themes between two groups

Table 5. Frequency of Different Types of Interpersonal Themes

Interpersonal Types	VE	AE
Modal adjunct/expressions	132	18
Finite operator	0	7
Vocative	0	0

Table 6 Frequency of Types of Modal Adjuncts/Expressions in VE and AE

Type	Examples	VE	VE%	AE	AE%
Opinion	I think, I tend to think, personally	38	28.8	1	5.6
Persuasion	I believe, I firmly believe, believe me	40	30.3	2	11.1
Obviousness	It is obvious, it is clear, it is apparent, clearly, obviously	20	15.2	3	16.7
Other	I hope, fortunately, perhaps, indeed, etc.	34	25.7	12	66.6
Total		132	100	18	100

Hyland's (2005) Stance and Engagement framework of interpersonal discourse

Stance refers to the ways writers present their voices, opinions or judgments,

- It contains hedges, boosters, attitude makers, and self-mention

Hedges: I think, likely, possible, etc

Boosters are words such as obviously, clearly, which enables writers to present their ideas with certainty

4.2 Interpersonal Theme

The frequent use of hedges and boosters in VE might indicate that Vietnamese EFL learners try to show that they have mastered a specific discourse pattern

American professional writers do seldom use hedges or boosters in their writings, this might indicate that hedges and boosters used in VE are not frequently used outside of academia itself.

4.2. Interpersonal Theme

2nd problem: Vietnamese EFL learners use more than one interpersonal Theme in unnecessary cases



In my opinion, I think we should separate disobedience students with positive one like an example for them to change, they use interpersonal theme redundantly. (VE 26)

Make their essays more colloquial

4.2 Interpersonal Themes

EFL learners are not sufficiently familiar with argumentative essays in English.

Interview with teachers and EFL learners, they explain that sentential translation is a writing strategy they sometimes employ.

In Vietnamese, the use of two or more interpersonal themes with the same meaning in one sentence conveys modesty in expression.

4.3 Textual Theme

Overuse of textual Themes

-Causing grammar incorrectly, or lack of grammatical mastery may cause inflation of textual Themes

Because it will take someone an amount of time to travel to the libraries and identify the studying resources **so** now many people do not want to use library (VE 43)

4.3 Textual Theme

Using inappropriate textual Themes

 Therefore, they are unwilling to boost learning skills and improve themselves. (13) **Eventually**, they have no chance to acquire knowledge and righteous behaviors from wellmannered students (VE 7) (Causal relationship)

Using conjunctions to link two sentences which are irrelevant

 Putting and teaching separately in one class could cause disobedient students isolated from others, consequently it is an injustice for them due to the equality of opportunity in education. <u>Moreover</u>, obedient and smart children in the class will be role models for disruptive school students. (VE 17)

4.3 Textual Theme

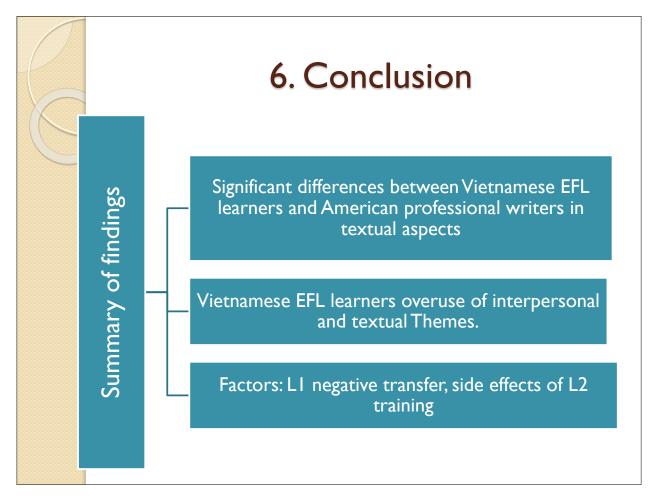
- Use a similar form to organize the essays
- Firstly, the first reason, etc
- Secondly, the second reason, etc
- Conclusion: In conclusion, to sum up Why?

They are taught this type of template in writing course => side effects of L2 training

5. Pedagogical Implications

EFL instructors should pioneer the revision of materials and teaching methods

The concept of themes and its functions should be instructed in EFL writing class



Limitation

Data: American professional writers' essays was used as a baseline, limit comparison of writing quality and organizational structure in the whole discourse. The result might be more interesting if comparing VE to essays by American/ British university students



A Critical Discourse Analysis on the Presidential Discourse: Focusing on Volodymyr Zelensyy's Linguistic and Discoursal Strategies for Persuasion

Ahnyung Choi(Yonsei Univ.)

Abstract: The purpose of this study is to analyze the presidential discourse from the perspective of critical discourse analysis (CDA). The address delivered by Volodymyr Zelenskyy, the President of Ukraine, during the Russia-Ukraine war especially have drawn the attention of the world and been considered persuasive. This study explores what linguistic and discoursal strategies he used to persuade others to achieve his goals, and how those could be regarded as persuasive strategies using the framework of Aristotle's theory of rhetoric in terms of three elements 'logos,' 'ethos,' and 'pathos.' The data consist of three speeches made by Zelenskyy to the European Parliament, the U.S. Congress, and the Japanese Parliament. The findings show that President Zelenskyy used various linguistic strategies (e.g., positive self-presentation and negative other-presentation, GWT, rhetorical questions, evoking history) in his approach and achieved partial success in receiving other countries' financial, political, and humanitarian support.

Keywords: CDA, persuasion, presidential speech, war speeches, Volodymyr Zelenskyy

A Critical Discourse Analysis on the Presidential Discourse: Focusing on Volodymyr Zelensyy's Linguistic and Discoursal Strategies for Persuasion Ahnyung Choi(최아녕)

Introduction

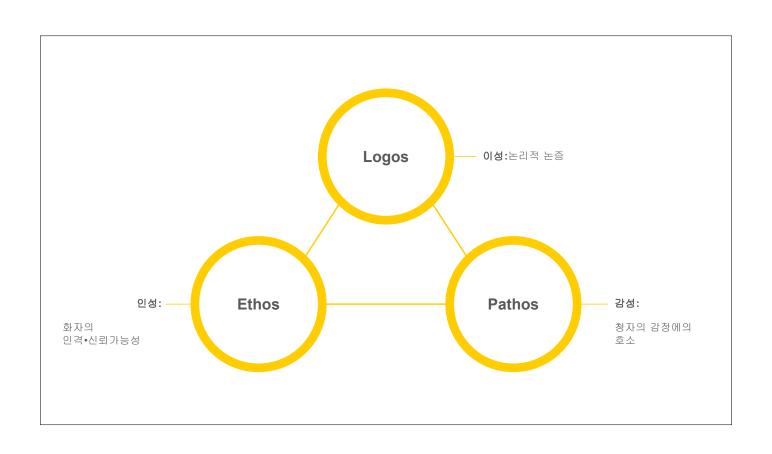
- → 2022년 2월 24일, '러시아군의 침공'으로 인한 '러시아-우크라이나 전쟁' 발발
- → 정치적 배경
 - . 우크라이나
 - . 1992년 이후, 소련의 영향에서 벗어난 **자주권**을 획득하기 위해 서방 세계와 보다 연결된 **개방적인 정책** 추구
 - . NATO & EU에 가입 신청에 대한 지속적인 수락 요구
 - . 러시아
 - . 우크라이나의 결정에 **강력히 반대**
 - 푸틴 (2021) → "역사적 문화적으로 **한 민족(one people a single whole)**"

Introduction

- → '정치 지도자(political leader)'의 역할 및 중요성
 - 특정 상황에서 '정치 지도자' 역할은 더욱더 중요해짐
 - **전시 상황**일 경우.
 - . 정치 지도자의 결정 → 전반적인 **전쟁 상황과 결과**를 **바꿀 수** 있음
 - . 정치 지도자의 연설 \rightarrow **다른 나라**를 **설득**하여 **지원을 받는데 중요한 역할**을 함
- → 볼로디미르 젤렌스키 (Volodymyr Zelenskyy)
 - 우크라이나 대통령 (2019.05~)
 - 젤렌스키의 연설은 많은 사람들에게 **긍정적**으로 평가되어지며, '설득적'으로 여겨짐
 - . 연설 후, 많은 나라들의 지지와 도움이 증가

Background

- → '설득'이란,
 - ◆ **타인의 생각** 및 **행위를 변화**시키거나 **영향**을 끼치는 것 (Halmari & Virtanen, 2005)
 - 수사학: 상대의 생각 및 행위를 효과적으로 바꿀 수 있는 설득적인 언어 선택에 초점
- → '설득'의 중요성
 - ◆ 정치가 및 정치 기관
 - '설득적(persuasive)' 또는 '조작적(manipulative)' 언어 사용으로 유지됨 (Chilton & Schäffner, 2011)
 - 언어적 담화 전략을 사용하여 **청자의 믿음과 생각을 설득**해야 함
- → '설득'에 관한 연구
 - ◆ 정치적 맥락에서 '설득'에 관한 연구를 진행
 - ◆ **아리스토텔레스**의 **'수사학의 세 가지 요소'**를 사용한 분석



젤렌스키의 3가지 연설문에서 어떤 <mark>언어적•담화적 전략</mark>이 사용되었으며, 어떤 점에서 이를 '설득적'으로 볼 수 있는가?

Data and analysis

→ 유럽 연합, 미국 의회, 일본 의회에서의 3가지 연설 분석

Name	Date	Place	Source
Volodymyr Zelensky	March 1,	The	https://www.nationalworld.com/news/wo
speech: full transcript	2022	European	rld/volodymyr-zelensky-speech-today-
of Ukraine president's		Parliament	fulltranscript-of-ukraine-presidents-
statement to European			statement-to-european-parliament-
Parliament			<u>3592244</u>
Address by President	March	The U.S.	https://www.president.gov.ua/en/news/pr
of Ukraine Volodymyr	16, 2022	Congress	omova-prezidenta-ukrayini-volodimira-
Zelenskyy to the US			zelenskogo-pered-kong-73609
Congress			
Speech by President of	March	The	https://www.president.gov.ua/en/news/pr
Ukraine Volodymyr	23, 2022	Japanese	omova-prezidenta-ukrayini-volodimira-
Zelenskyy in the		Parliament	zelenskogo-v-parlamen-73769
Parliament of Japan			

- → 젤렌스키의 연설문 속 설득을 위해 사용된 주요 언어 및 담화 전략들을 여러 가지 측면에서 질적으로 분석
- → 아리스토텔레스의 '수사이론'을 사용해 어떻게 주요 전략들이 '설득'으로 해석될 수 있는가를 재분석

Linguistic and discoursal strategies

- 1) positive self-presentation and negative other-presentation
 - o reference to Ukraine
 - o reference to other countries (e.g., Europe, the U.S., Japan)
 - o reference to Russia
- 2) lexical choices (GWT)
- 3) rhetorical questions
- 4) evoking history

Positive self-presentation and negative other-presentation

- → Dijk (2011): 사회 인지적(socio-cognitive) 이론
 - · 'Us' vs. 'Them'
 - 자신이나 자신이 속한 집단을 긍정적이게 표현하며 타인(them)을 부정적으로 표현
- → 젤렌스키의 연설문
 - 다른 나라의 지원 요청을 위한 설득적인 전략으로 '긍정적 자기제시• 부정적 타인 제시' 전략을 사용
 - 자신의 나라(우크라이나)를 긍정적인 단어 및 이미지로 묘사
 - 도움을 받아야하는 나라를 자신의 편으로 끌어들이며 '우리 집단(us-group)을 형성
 - 자신의 적(러시아)을 부정적인 단어 및 이미지로 묘사

Reference to Ukraine

- → 우크라이나를 긍정적인 단어 및 이미지로 묘사하며, 자신의 적(러시아)을 부정적인 단어 및 이미지로 묘사
 - (1) We are giving away our best people, the strongest ones, the most value-based ones. (The European Parliament)
 - (2) Ukrainians are **incredible**. And very often we love to say that we **win** over everyone, and I'm very happy that you are not only talking about it, but you can see that. And we indeed, we will **overcome** everyone. (The European Parliament)
 - (3) Brave and freedom-loving people, who for eight years have been resisting the Russian aggression. (The U.S. Congress)
- → 우크라이나는 지킬 가치가 있을 뿐만 아니라 적에게 맞서 싸울 수 있는 힘을 가지고 있는 강력한 나라임을 주장
- → 러시아는 그 가치를 무자비하게 파괴하고 무너뜨리는 악으로 표현
 - ◆ 우크라이나를 돕는 것이 가치있는 것임을 어필

Reference to other countries (Europe & the U.S. & Japan)

- → 유럽, 미국, 일본 등 도움을 받아야하는 나라를 자신의 편으로 끌어들이며 '우리 집단(us-group)을 형성
 - (4) I'm very happy that what I've seen here now and heard here now, I'm very glad to sense this mood, **the uniting, unifying** mood. I'm happy that we have unified today, all of you, all the countries of the European Union. (The European Parliament)
 - (8) Ladies and gentlemen, friends. Ukraine is grateful to the United States for its overwhelming support. For everything that your government and your people have done for us, for weapons and ammunition, for training, for finances, for leadership in the free world, which helps us to pressure the aggressor economically. I am grateful to President Biden for his personal involvement for his sincere commitment to the defense of Ukraine and democracy all over the world. (The U.S. Congress)
 - (10) We have similar values with you despite the huge distance between our countries. A distance that doesn't really exist. Because we have equally warm hearts. Thanks to joint efforts, thanks to even greater pressure on Russia, we will come to peace. And we will be able to rebuild our country. Reform international institutions. (The Japanese Parliament)

Reference to Russia

- → 우크라이나를 긍정적인 단어 및 이미지로 묘사하며, 자신의 적(러시아)을 부정적인 단어 및 이미지로 묘사
 - (11) But I did not know that this is the **price** they will have to pay. And this is a **tragedy** for us. for me, for every Ukrainian **tragedy** for our state. I mean, this this **high price**. **Very high price**. **Thousands of people who were killed**. Two revolutions, one war and five days of **full scale invasion of the Russian Federation**. (The European Parliament)
 - (13) I am grateful to you for the resolution which recognizes all those who commit crimes against Ukraine, against the Ukrainian people, as war criminals. However now, it is true in the darkest time for our country, for the whole of Europe, I call on you to do more new packages of sanctions are needed constantly, every week until the Russian military machine stops. (The U.S. Congress)
 - (14) And it would be fair if it ended in a day, in 24 hours, that evil would be punished immediately. (The U.S. Congress)
 - (15) Russia has used more than a thousand missiles against peaceful cities in Ukraine. Countless bombs. Russian troops destroyed dozens of our cities. Some were burned to the ground. In many towns and villages that have come under Russian occupation, our people cannot even bury their murdered relatives, friends and neighbors with dignity. They have to bury them right in the yards of broken houses, near roads, anywhere where it is possible. (The Japanese Parliament)

Lexical choices (GWT)

- → 지역 전쟁을 글로벌한 전쟁으로 표현하면서 다른 나라에게 책임을 공유하게 함
 - (18) We are fighting for our rights. For our freedoms. For life. For our life. And now, we're fighting for survival. And this is the highest of our motivation. But we are fighting also to be equal members of Europe I believe that today we are showing everybody that's exactly what we are. (The European Parliament)
 - (24) Today, the Ukrainian people are not only defending Ukraine. We are fighting for the values of Europe and the world, sacrificing our lives in the name of the future. That's why today the American people are helping not just Ukraine, but Europe and the world to keep the planet alive. (The U.S. Congress)
 - (25) You are the leader of your great nation. I wish you to be the leader of the world. **Being the leader of the world means to be the leader of peace**. (The U.S. Congress)
 - (28) Today, it is Ukraine, the partner states and our anti-war coalition that can guarantee that world security will not be completely destroyed. That in the world there will be a foothold for the freedom of nations. For people and for the preservation of diversity in societies. For security of borders. To make sure that we, our children, our grandchildren still have peace. (The Japanese Parliament)
- → 전쟁은 우크라이나만을 겨냥한 것이 아닌, 인간의 기본권, 자유, 민주주의에 대한 침공임을 주장

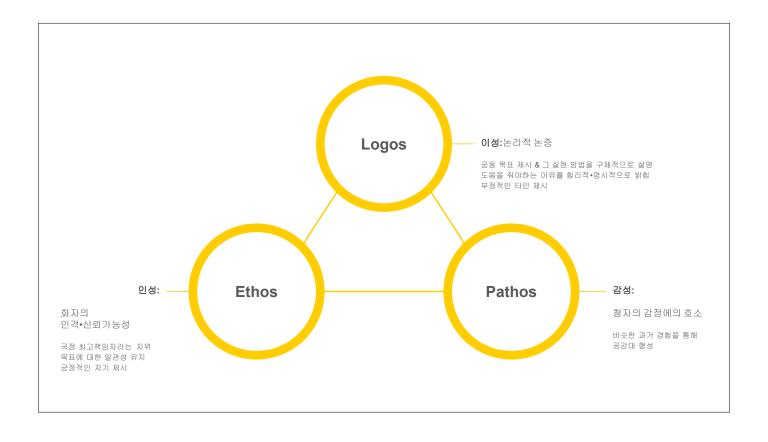
Rhetorical questions

- → 수사적 의문(설의법)을 사용하여 자신의 말을 강조 및 정해진 대답을 유도
 - (29) Can you imagine this morning two cruise missiles hit this freedom square. Dozens of dead ones. This is the price of freedom. (The European Parliament)
 - (30) **Is this a lot to ask for?** To create a no-fly zone over Ukraine to save people. **Is this too much to ask?** A humanitarian no-fly zone, something that Russia would not be able to terrorize our free cities. (The U.S. Congress)
 - (31) You see that international institutions have not worked. Even the UN and the Security Council. **What can they do?** They need reform. They need an injection of honesty. To become effective. To really decide and really influence, not just discuss. Due to Russia's war against Ukraine, the world is destabilized. The world is on the verge of many new crises. And **who is not sure what tomorrow will be like?** (The Japanese Parliament)

Evoking History

- → 역사 속 비슷한 경험들을 끌어내 일종의 공감대 형성을 만들어 냄
 - (32) And we have not even thought about it for a second, just like many other cities and communities in our beautiful country, which found themselves in the worst war since World War II. (The U.S. Congress)
 - (34) Remember Pearl Harbor. Terrible morning of December 7, 1941, when your sky was black from the planes attacking you. Just remember it. Remember September the 11th, a terrible day in 2001 when evil tried to turn your cities, independent territories in battlefields. Our country experience the same every day. Right now, in this moment, every night, for three weeks now, various Ukrainian cities of Odesa, Kharkiv, Lviv, Sumy, Mariupol, Zaporzhzhia and Dnipro. (The U.S. Congress)
 - (36) Each of you knows what Chornobyl is. Nuclear power plant in Ukraine, where a powerful explosion occurred in 1986.

 Radiation release. The consequences of which have been recorded in different parts of the planet. (The Japanese Parliament)



Con	clusi	on
\rightarrow	젤렌:	스키가 3가지 연설문 에서 사용된 언어적∙담화적 전략
	1.	positive self-presentation and negative other-presentation
	2.	lexical choices (GWT)
	3.	rhetorical questions
	4.	evoking history
\rightarrow	아리:	스토텔레스의 '수사학의 세 가지 요소'
	•	젤렌스키가 사용한 언어적•담화적 전략은 설득에서 중요한 역할을 하는 것으로 분석되는 세 가지 수사 요소로 재해석
		될 수 있으며, 이는 왜 젤렌스키의 연설이 설득적으로 여겨지는 지와 긍정적인 영향을 미쳤는지 를 설명할 수 있음

Thank you

Investigation of Engineering-Major Students' Perception of English Learning in a Chinese EMI Context

Yoo Young Ahn, PhD

Lecturer

Chonnam National University

Definition of EMI

Figure 1: Continuum of EMI in practice (adapted from: Thompson & McKinley, 2018)



English as a medium (EMI); Content and language integrated language (CLIL); Content-based learning (CBI); English for specific purposes (ESP) From Rose et al. (2020, p. 4)

Growth of EMI in Asia

- Top-down policies to a) internationalize their institutions, b) attract more international students, & c) develop capabilities to read state-of-art information (esp. in STEM fields) (Macaro et al., 2018)
- In China, multiple government initiatives accelerated the spread of EMI programs (Rose et al., 2020)
- Problems with the pedagogical practices (Macaro & Han, 2019)
- Limited research at the implementation level

Concerns regarding language

- 1. Effects of EMI on students' language development were controversial (does not happen naturally)
- 2. Both instructors' and students' concerns
- 3. A lack of training for faculty
- 4. Compromise disciplinary knowledge due to language issues; or cannot follow content courses

Research questions

- 1. How do first-year Chinese engineering-major students' perceptions of English learning change after a semester in an EMI college?
- 2. How do first-year Chinese engineering-major students view the institutional language support (two English composition courses) mandated by an EMI college?

Methods

- 1. Participants and setting
 - 1st year Chinese students in EMI institute (Western China)
 - Three engineering majors (IE, ME, & ESE)
 - Two mandatory English composition courses, assisted by the Writing Center
- 2. Collected data
 - 86 Reflection papers from Fall 2021 (open coding)
 - 176 students' 345 metaphors from Spring 2022 (Elicited Metaphor Analysis, Jin et al., 2004)
 - "English learning before college was like ______ because _____"
 - "English learning [in this college] is like ______ because ______"

Findings

RQ1. How do first-year Chinese engineering-major students' perceptions of English learning change after a semester in an EMI college?

Three new categories of metaphors for English learning in the EMI college, in addition to English learning as *Work, Journey, Leisure activities*, and *Necessity*

- 1. Learning (learning together, learning faster)
- **2. Water activities** (Boating on a rushing river toward a destination; A ship at sea guided by a lighthouse)
- **3.** Challenges (A challenge; a risk; staying on an island alone)

Findings (cont'd)

- 1. Acknowledgement of English-using environment as facilitating
 - [the institute] created an English [learning and speaking] environment, I should learn English every day just like I should drink water every day, because everyday, you need to speak English and learning English, therefore your English level will improve.
- 2. Feeling guided and supported in the community (with faculty and the Writing Center)
 - Swimming in the swimming pool, sailing in the sea by ship, walking with [a] wise man, boating in the right direction
- 3. EMI as a safe space (with classmates/roommates and senior students)

Findings (cont'd)

RQ2. How do first-year Chinese engineering-major students view the institutional language support mandated by an EMI college?

- 1. Sense of achievement and confidence
 - To be honest, it is really struggling to find a proper topic and start. I can still remember how painful I was when I had to write 1000 words for my first narrative essay. However, I felt proud the moment I finished it. I really appreciate these experiences. These writing experiences not only help me improve my writing skills but also help me strengthen my mind.
- 2. Increased agency in English learning
 - Being a boss of my own company, here I can choose the topic I am interested in and search for information through the Internet and Writing Center. I can also discuss my ideas with classmates and professor to revise my essay.

Findings (cont'd)

- 3. Feeling behind and stressed
 - Being immersed in hot water at 60 degrees for a long time because it is not comfortable, however it will not kill me and I will get used to it gradually
- 4. Limited awareness of effective strategies
 - Returning to the same, meaningless methods (vocabulary and grammar; reading more books)

Conclusion

- 1. Develop a sense of belongingness in the community (students from all years, and faculty in both English and science/engineering)
- 2. Provide explicit English support and resources for general proficiency and content areas
- 3. Composition courses are unlikely to offer language helps

Thank you!

For suggestions and questions,

please contact yooyahn@jnu.ac.kr

General Information

Presentation Language: English

• Presenter: 손 지연 / Jeeyeon Sohn

• Affiliation: 연세대학교, 대한민국 / Yonsei University, Republic of Korea

• Presentation Title: The Development of Generative Grammar: Focus on

Minimization

• Email address: <u>irin369@naver.com</u>

The Development of Generative Grammar :Focus on Minimization

Jeeyeon Sohn (Yonsei University)

I. Introduction

- The subject of syntax is a sentence.
- Discussions on sentences have been made since Plato and Aristotle.
 - In Plato's Cratylus and Sophist
 - In Aristotle's De interpretatione
- Such discussions continued until Chomsky 1957, advancing the Generative Grammar.
- Whether we support it or not, the Generative Grammar is the dominant theory now.
- Therefore, this paper aims to provide an overview of the development of Generative Grammar and understand the Minimalist Program advanced by Chomsky (1995), with specific focus on the concept of minimization.

II. Philosophical Backgrounds

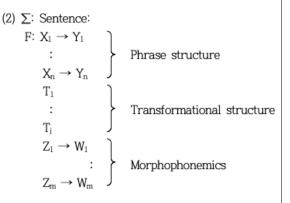
- We, even Chomsky, are the descendants of Plato and Aristotle, the first linguists.
- Several concepts of Generative Grammar were influenced by several philosophers.
 - Plato and Aristotle: Lexical categories and Functional categories
 - Descartes: Reason, Innate ideas
 - Kant: Innate category system
 - Darwin: "Instinctive tendency to speak"
- To sum, the Generative Grammar started with innate ideas that had been suggested by many philosophers and even scientists.

A. Standard Theory

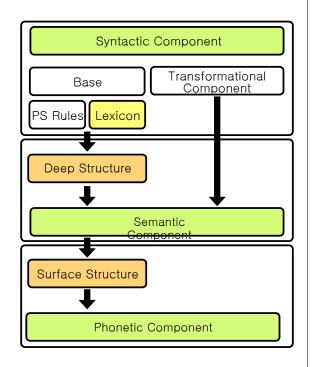
- Empiricism vs. Rationalism
 - Empiricists collected linguistic data first and then analyzed them (Induction).
 - Rationalists framed a hypothesis and then collected linguistic data (Deduction).
- Chomsky 1957
 - Presenting phrase structure rules (PS rules) and transformational rules.
 - Initial state Rewriting rules (PS rules) Final state

(1) a. Sentence
$$\rightarrow$$
 NP + VP d. T \rightarrow the b. NP \rightarrow T + N e. N \rightarrow man, ball, etc. c. VP \rightarrow Verb + NP f. Verb \rightarrow hit, took, etc.

- Still, phrase structure rules cannot effectively describe all the sentences we make.
 - Explaining subject-verb agreement, making interrogative or negative sentences
- To solve this problem, Chomsky presents "Grammatical Transformation."
- To sum,
 - Finite state grammar used before changed to be phrase structure grammar.
 - Those phrase structure rules needed to explain agreement, interrogative, negative sentences are minimized to few transformational rules.

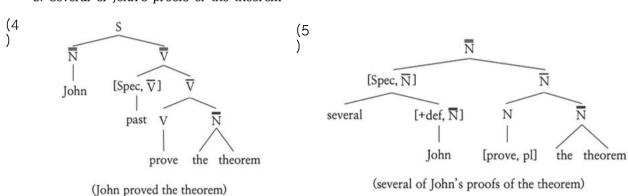


- Chomsky 1965
 - Universal Grammar (UG)
 - Syntactic, Semantic, Phonetic components
 - Deep structure & Surface structure
- In addition.
 - In Chomsky 1957, lexical items were inserted by PS rules but in 1965, they are managed by Lexicon with word insertion rules.
 - In Chomsky 1957, a sentence had Terminal morpheme strings, Kernel sentence strings, and Derived sentence strings, whereas in 1965, a sentence had only



B. Extended Standard Theory

- Chomsky 1970
 - The prototype of X-bar Theory
 - (3 a. John proved the theorem.
 - b. Several of John's proofs of the theorem



- Minimization of PS rules into a few generalized rules (6) a. $X' \rightarrow X$...
 - Category-specific rules to Category-neutral rules b. $X'' \rightarrow [Spec, X'] X'$
- Intermediate levels

- c. $S \rightarrow N'' V''$
- Using PS rules, we cannot distinguish between specifers, complements, and adjuncts
- By postulating intermediate levels,
 - Complements and heads form intermediate levels, and intermediate levels and specifiers form phrase levels, whereas adjuncts and intermediate levels combine to become the same intermediate levels.
- Limitation
 - Category-neutral rules have overgeneralization problem.
 - → The theory become more sophisticated by postulating many constraints.

C. Government and Binding Theory

- Chomsky 1981
 - Grammar: Rules and Principles
 - X-bar theory, Government theory, Theta theory, Binding theory ...
 - Universal Grammar → Principles, Particular Grammar → Parameter
 - Transformational rules are minimized into a single α movement.
 - Before α movement: V movement, I movement, NP movement, WH movement ...
 - These rules can be generalized into X movement and XP movement.
 - In addition, these generalized rules can be minimized into α movement as well.
 - Overgeneralization problem → Constraints decide where and when to move.

- Chomsky 1986
 - S(entence) as a semantic category violates Endocentricity Constraint.

(7) a.
$$S \rightarrow C$$
 S' (8) a. John comes
b. $S \rightarrow NP$ I VP b. For John to come

- In addition to subjects and verbs, tense information is also needed to form a sentence.
- X-bar theory got extended with S and S' being generalized as XP (S to IP, S' to CP)
- To sum,
 - Explaining Language system became simpler with principles and parameters.
 - Several transformational rules are minimized into a single α movement.
 - X-bar theory truly generalized its categories with I and C.

D. Minimalist Program (1990s)

- Chomsky 1995
 - Economics and minimality / Biolinguistic perspective and Conceptual necessity
 - First, no more deep structures and surface structures
 - GB: D-structures, S-structures, Logical Form (LF), and Phonetic Form (PF)
 - MP: Conceptual-intentional system and Articulatory-perceptaul system.
 - Second, top-down to bottom-up generation of sentences
 - GB: Top-down process based on X-bar theory
 - MP: Bottom-up process based on Merge

- Third, features are the basic unit of any operation
 - Categorial features, φ-features, case features, strong F
 - Interpretable Feature (iF) and uninterpretable Feature (uF)
 - Full interpretation principle and Feature checking
 - The work of syntax is thus checking and deleting uFs, making them interpretable.
- To sum,
 - Sentence generation is no more explained by X-bar theory but by feature checking system, generating bottom-up structures with Merge and Movement.

D. Minimalist Program (2000s)

- Chomsky 2001
 - In 1990s, the subject of syntax had been whole sentences, but in 2000s, it became phases, smaller units.
 - Syntactic operations are based on cyclicality, abandoning LF postulated in 1990s.
 - Phase Impenetrability Condition (PIC): The domain of H is not accessible to operations at ZP; only H and its edges are accessible to such operations.
 - Instead of Feature checking and Attract F Theory, movements are explained by Agreement Theory, where uFs are unvalued and thus look for goals to be valued, having Probe-goal association.
- To sum,
 - Less cognitive costs with phase, cyclicality, and PIC.
 - Simpler language system with Agreement theory instead of Feature checking and Attract F.

IV. Conclusion

- This paper introduces the philosophical and theoretical background of Generative
 Grammar and presents the core concepts of syntactic theories existing under the
 Generative Grammar framework. It also discusses which aspects of each theory have
 been changed to achieve minimality.
- Since minimization revisions had already been made to the Standard Theory, the
 Extended Standard Theory, and the Government and Binding Theory, this paper
 concludes that the Minimalist Program is not completely different from other theories
 but rather an extension of them, which particularly emphasizes the economics and
 minimality of theory.

V. Works cited

강명윤. 2004. 생성언어학의 역사, 현황과 전망. *숭실어문* 20:5-32.

고희정. 2018. 생성통사이론의 변천과 발전 전망. *한국어학* 81:1-33.

박현석. 2005. 구범주를 통한 영어 문장 분석. *영어영문학21* 18,1:201-220.

아리스토텔레스. 2005. 범주들·명제에 관하여. 김진성 옮김. 서울: 이제이북스.

조중걸. 2019. 플라톤에서 비트겐슈타인까지: 서양철학사 인식론적 해명. 서울: 지혜정원.

플라톤. 2021. *크라튈로스*. 김인곤, 이기백 옮김. 서울: 아카넷.

피터 애덤슨. 2017. *빈틈없는 철학사2: 소크라테스와 플라톤*. 김은정, 신우승 옮김. 서울: 전기가오리.

Chomsky, Noam. 1957. Syntactic Structure. Berlin: Mouton.

Chomsky, Noam. 1965. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In *Reading in English Transformational Grammar*, ed. by R. Jacobs and P. Rosenbaum, 184-221. Waltham: Ginn.

Chomsky, Noam. 1981. Lectures on Government and Binding. Dordrecht: Foris.

Chomsky, Noam. 1986. Barriers. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. The Minimalist Program. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*, ed. by Michael Kenstowicz, 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. Three Factors in Language Design. Linguistic Inquiry 36,1:1-22.

Darwin, Charles. 1875. The Descent of Man. New York: D. Appleton.

Jackendoff, Ray. 1977. x Syntax: A Study of Phrase Structure. MIT Press, Cambridge, Mass.

Radford, Andrew. 1988. Transformational Grammar. Cambridge University Press.

Roeper, Thomas. 2005. Chomsky's Remarks and the Transformationalist Hypothesis. In *Handbook of word—formation*, ed. by Pavol Stekauer and Rochelle Lieber, 125–146. Dordrecht: Springer.

Shaffer, David E. 1983. Word Order and Generative Grammar. English21 2:213-234.

A New Type of Resultative Construction in English

Juwon Lee

Jeonju University, Republic of Korea juwonlee@jj.ac.kr https://sites.google.com/view/juwonlee19

> 한국영어학회, 한국코퍼스언어학회 가을 공동학술대회 부산대학교 2022.10.22 (토)

> > 1

1. Introduction (1/3)

• Topic:

English sentences like John built the house big

• Question:

Are they a resultative or depictive construction?

A proposal:

They would constitute a third category distinct from resultative or depictive constructions.

1. Introduction (2/3)

- The secondary predicates can serve as a resultative predicate as in (1a) or depictive predicate as in (1b)
- (1) a. Bill hammered the metal <u>flat</u>.
 - b. Bill ate **the fish** raw.
- The sentences like (2) have not been studied in the literature to my best knowledge.
- **Boas** (2003) provides a comprehensive classification of English resultative constructions extracted from **the BNC** (British National Corpus), but sentences like (2) are not reported in the study.
- (2) a. Tom drilled a hole deep.
 - b. Tom built a house big.

3

1. Introduction (3/3)

- *Build-big* constructions:
- (2) a. Tom drilled a hole deep.
 - b. Tom built a house big.
- The sentences in (2) appear to have **the same structure** as that of the resultative and depictive constructions: roughly, $[NP + V + NP + \underline{AP}]$.
- These commonalities between (1) and (2) suggest that they all belong to the same category (namely, **secondary predicate construction**).
- A question is whether the sentences in (2) are resultative or depictive.
- They have some **properties** of **resultative** or **depictive constructions**.

2. Corpus Data (1/4)

- Build-big constructions in the Web:
- (3) a. And I *carved* a hole <u>big enough</u>. (https://www.bruwriter.com/the-little-green-penguin.html)
 - b. ... and they *carved* a ship <u>big enough for fifty penguins</u>. (https://www.bruwriter.com/the-little-green-penguin.html)
 - c. It's a consequence of having *invented* **a technology** <u>powerful enough to alter an entire species</u>.

(https://www.media.mit.edu/articles/gene-drives-could-stop-the-world-s-oldest-problems-kevin-esvelt-wants-to-make-sure-they-don-t-start-any/)

d. Only when you have *digged* a hole <u>deep enough for Shihong</u> that would make his escape impossible, you should contact me.

(https://www.wuxiaworld.co/LEGEND-OF-THE-GOLDEN-

TIGRESS/1939972.html)

e. I *brewed* the coffee strong, the way I like it in order to get the full flavor of the coffee and to see if it had a bitter flavor to it.

(https://www.walmart.com/reviews/product/23555013?reviews_limit=10&reviewId=569 44877)

2. Corpus Data (2/4)

- Build-big constructions in the COCA:
- (4) a. ? Garfield: I don't think they've ever *baked* a cake <u>big enough</u>. I love to get cake-faced on my birthdays.
 - b. When setting concrete pier forms in the ground, *dig* the holes <u>large enough</u> to allow room for side-to-side adjustment.
 - c. They haven't invented the hat big enough to pull that rabbit out of
 - d. I design illusions and *construct* the apparatus <u>necessary for perfoming them.</u>
 - e. Brew the coffee strong please.
 - f. ...and she figures they must be *baking* **the pie** <u>fresh</u>, because it's taking forever.
 - g. ...an Englisch man who was building the buggies cheaper than anyone else.

2. Corpus Data (3/4)

- Build-big constructions in the iWeb:
- (5) a. *Dig* a hole <u>larger</u> and <u>deeper</u> than the <u>container</u>, add a handful of bonemeal and some rotted manure and mix into the soil.
 - b. Dig a hole large enough to hold the plants root system.
 - c. Use a trowel to dig a hole big enough to accommodate the roots of the plant.
 - d. Mourinho is likely to make multiple defensive signings this summer as he looks to *build* a squad capable of challenging for the Premier League title.

7

2. Corpus Data (4/4)

- Three patterns of *build-big* constructions:
- (6) a. V NP AP
 - b. V NP AP enough (to V)
 - c. V NP A-*er* (*than* ...)
- V: drill, carve, dig, build, invent, brew, bake, construct (creation verbs)
- **NP**: a hole, a house, a ship, a technology, the coffee, a cake, the holes, the hat, the apparatus, the pie, the buggies, a squad (**product**)
- **A**: deep, big, powerful, strong, large, necessary, fresh, cheaper, larger and deeper, large, capable (**states**)

3. Property #1: Change of State (1/2)

- Resultative constructions: change of state
- (7) a. He hammered the metal flat. \Rightarrow The metal became flat.
 - b. He wiped the table clean. \Rightarrow The table became clean.
 - c. He followed Lassie <u>free of his captors</u>. ⇒ He became free of his captors.
- Depictive: no change of state
- (8) a. Bill ate the fish raw. \Rightarrow The fish became raw.
 - b. **Bill** ate the fish naked. *⇒* Bill became naked.
- BB: no change of state
- (9) a. Tom drilled **a hole** deep. ⇒A hole became deep.

9

3. Property #1: Change of State (2/2)

- Resultative constructions: change of state
- (10) a. The ice froze solid. \Rightarrow The ice became solid.
 - b. The water froze solid.

 ⇒ The water became solid.
- Are BB constructions like those in (10)?
- But a fundamental difference:

Resultative: the same entity: water (not solid) = ice (solid)

"Bring the water" = "Bring the ice"

"The president was poor at that time."

BB construction: **not the same entity**: ground \neq hole (deep)

3. Property #2: [V N A] vs. [V A N] (1/1)

- Resultative constructions: $V N A \neq V A N$
- (11) a. He hammered the metal flat. \neq He hammered the flat metal.
 - b. He wiped the table <u>clean</u>. ≠ He wiped the <u>clean</u> table.
- Depictive: V N A = V A N
- (12) a. Bill ate the fish \underline{raw} = Bill ate the \underline{raw} fish.
 - b. **Bill** ate the fish <u>naked</u>. = **Bill**, who was <u>naked</u>, ate the fish.
- **BB**: V N A = V A N
- (13) a. Tom drilled **a hole** deep. = Tom drilled a deep **hole**.
 - b. Tom built a house big. = Tom built a big house.

11

3. Property #3:AP only at the end of event (1/2)

- Resultative constructions: AP = only at the end of event
- (14) a. He hammered the metal flat.
 - \Rightarrow The metal was flat only at the end of event.
 - b. He wiped the table clean.
 - \Rightarrow The table was clean only at the end of event.
- Depictive: $AP \neq only$ at the end of event
- (15) a. Bill ate **the fish** raw.
 - \Rightarrow The fish was raw only at the end of event.
 - b. **Bill** ate the fish naked
 - ⇒ Bill was naked only at the end of event.

3. Property #3:AP only at the end of event (2/2)

- BB: AP = only at the end of event
- (16) a. Tom drilled a hole deep.
 - \Rightarrow The hole was deep only at the end of event.
 - b. Tom built a house big.
 - \Rightarrow The house was big only at the end of event.

13

3. Property #4: [A N] Creation (1/1)

- Resultative constructions: [A N] Creation
- (17) a. He hammered the metal flat. \Rightarrow The flat metal became to exist.
 - b. He wiped the table clean. \Rightarrow The clean table became to exist.
- Depictive: No [A N] Creation
- (18) a. Bill ate the fish raw. \Rightarrow The raw fish became to exist.
 - b. Bill ate the fish <u>naked</u>. \Rightarrow Naked Bill became to exist.
- BB: [A N] Creation
- (19) a. Tom drilled a hole deep. \Rightarrow A deep hole became to exist.
 - b. Tom built a house big. \Rightarrow A big house became to exist.

3. Property 5: Object existence (1/2)

- Resultative constructions: Object existence
- (20) a. He hammered the metal <u>flat</u>. \Rightarrow The metal existed during the hammering event.
 - b. He wiped the table clean. \Rightarrow The table existed during the wiping event.
- Depictive: Object existence
- (21) a. Bill ate the fish raw. \Rightarrow The fish existed during the eating event.
 - b. Bill ate the fish naked. \Rightarrow Bill existed during the eating event.
- BB: No object existence
- (22) a. Tom drilled a hole deep. \Rightarrow The hole existed during the drilling event.
 - b. Tom built a house big. \Rightarrow The house existed during the building event.

15

3. Property 5: Object existence (2/2)

- Resultative constructions: Object existence
- (23) a. The ice froze solid. \Rightarrow The ice existed during the freezing event.
 - b. The water froze solid. ≠ The water existed during the freezing event.
- But in (20) **the same entity** is referred to as 'the ice' or 'the water.' That is, the entity existed during the freezing event.

4. Pseudo-resultative construction (1/1)

- Pseudo-resultative constructions (Levinson 2010):
- (24) a. Mary **braided** her hair <u>tight</u>. \Rightarrow tight **braid**
 - b. Mary **piled** her cushions <u>high</u>. \Rightarrow high **pile**
 - c. Mary sliced the bread thin. \Rightarrow thin slice
- Pseudo-resultative predicates do not modify any 'word' in the syntax. They modify the **root** of the verb (Levinson 2010).
- But the secondary predicates of BB constructions modify **the objects**.
- The roots of the creation verbs like *drill* or *build* cannot be modified by the secondary predicates.

17

5. Conclusion: Properties (1/3)

	Resultative	ВВ	Depictive
1. Change of state	1	0	0
2. $[V N A] = [V A N]$	0	1	1
3. AP only at the end of event	1	1	0
4. [A N] Creation	1	1	0
5. Object existence	1	0	1

5. Conclusion: Aspectual differences (2/3)

• Resultative constructions: *Bill hammered the metal <u>flat</u>*.

hammering hammering the metal the metal

• Depictive constructions: Bill ate the fish <u>raw</u>.

eating eating the fish raw raw

• BB constructions: *Tom built a house big*.



19

5. Conclusion: A third category (3/3)

- BB constructions share **some properties** with resultative or depictive constructions.
- BB constructions would constitute a **third category** distinct from resultative or depictive constructions.

Resolving some puzzles in Locative Inversions of English

Suyoung Bae and Myung-Kwan Park (Dongguk University)

In this paper, we first review Bruening's (2010) and Choi & Park's (2021) analysis of locative inversions (LIs) and pseudo-locative inversions (PIs) and show that neither of the analyses can account for the difference between LIs and PIs with a particular focus on the compatibility with polarity emphasis. Then, we argue that the difference between LIs and PIs results from the interplay between the syntax and pragmatics. In particular, we argue that the initial information structure of LIs and PIs derived via fronting a locative phrase is identical; neither structure is information-wise binarily partitioned (e.g. topic/comment or focus/background (Leonetti & Ethere scandell-Vidal 2009 for Spanish). However, we argue that the two constructions end up as diverging both in syntax and information structure from each other because different syntactic mechanisms are employed to remedy the non-split information structure. Given that every sentence needs to have a focus, there are two remedies; 1) an insertion of the expletive 'there' which triggers re-establishing the subject-predicate relation and subsequently, a formation of the new information structure either with a polarity focus 2) or a narrow focus on the markedly post-posed subject. To account for the LI and the PI. we thus suggest subject-predication relation is successfully established in TP (predication-external) level where applying the predicate to the grammatical subject in Spec,TP position gives the sentence at hand the value: true or false. If such a requirement is satisfied, the focus is assigned on the polarity of the sentence, thereby giving rise to the emphatic interpretation on the polarity. This captures the emphatic nature of the PI construction. On the other hand, if the requirement is not satisfied, the narrow focus is imposed on the markedly post-posed subject. The current analysis provides the further support in favor of the interaction between syntax and pragmatics.





한국영어학회(KASELL)-한국코퍼스언어학회(KACL)
2022년 공동 가을 학술대회

Trends and Challenges in Linguistics in Times of New Normal and Big Data

