

2021년 여름학술대회

K₂S 한국언어과학회

4차 산업 시대 언어학의 진화

Evolution of Linguistics in the Fourth Industrial Era

2021년 8월 19(목) **일시**
온라인(Zoom) **장소**

주관

부산대학교 언어정보학과

주최

K₂S 한국언어과학회

[한국언어과학회 2021 여름학술대회] Zoom 접속 주소

회의 주소

<https://pusan.zoom.us/j/85714090668?pwd=eVh1MzVJd1dGR04wSXhVQmVLaIVKdz09>

회의 ID: 857 1409 0668

암호: 123456

한국언어과학회 2021 여름학술대회

“4차 산업 시대 언어학의 진화”

일시: 2021년 8월 19일(목)
 주관: 부산대학교 언어정보학과

장소: 온라인(Zoom)
 주최: 한국언어과학회



12:50 ~ 13:00		개회식 회장 권연진(부산대)			
시간	분과	1분과 (음운/형태/통사)	2분과 (의미/화용)	3분과 (코퍼스)	4분과 (응용/언어교육)
		좌장(사회): 김동국(영산대)	좌장(사회): 박기성(부산대)	좌장(사회): 장세은(한국해양대)	좌장(사회): 김양희(안동대)
13:00 ~ 13:30	제 목	<i>Who left?</i> Construction and EPP-Puzzle	은유적 표현에 의한 부각과 은폐에 관한 연구 -코로나19 사태 관련 신문 사설을 중심으로-	A Corpus-based Study of Specialized Vocabulary Lists for Maritime English ㉠	Research on Covid-19 English Education: Students' Perspectives ㉠
	발 표	권기양(영산대)	김철규(경성대)	슈 린(대련해사대) & 장세은(한국해양대)	허선민 & 이강영 (충북대)
	토 론	장경철(부산대)	박기성(부산대)	이제영(전주대)	조윤경(부경대)
13:30 ~ 14:00	제 목	Interaction between Syntax and External System :Evidence from Fragmentary Questions in Korean	Korean Hotels' Responses to Negative Online Reviews: A Genre Analysis	Functional Analysis of Lexical bundles in rhetorical move of research article abstracts in Applied Linguistics ㉠	L2 Learners' Perceptions of Online English Learning in the Post-Corona Era: Focusing on Affective Factors and Language Proficiency
	발 표	전혜원(부경대)	안병길(경성대)	리양 쉬아오레이 & 장세은(한국해양대)	신유선(순천대학교)
	토 론	권기양(영산대)	김소영(동명대)	홍지윤(한국외대)	오준일(부경대)
14:00 ~ 14:30	제 목	동사이동과 짝병합에 대한 소고	Functional Linguistic Perspective on Twitter Hashtag: What the Patient31 Tells Us ㉠	A comparison of maritime-related characteristics in the 19th century English journal log-book corpus of East Asian regions through keyword analyses	글쓰기 과업에서 언어형태의 복잡성이 주목과 학습에 미치는 영향
	발 표	김대익(영산대)	최예슬(Daedeok Highschool) & 민수정(Kongju National University)	장세은, 이성화, 박호민 (한국해양대)	조윤경(부경대)
	토 론	이성용(고신대)	남정미(신라대)	김철규(경성대)	이현정(부산대)
		좌장(사회): 황규홍(동아대)	좌장(사회): 김수태(신라대)	좌장(사회): 이성용(고신대)	좌장(사회): 우길주(부산교대)
14:30 ~ 15:00	제 목	한국 자폐스펙트럼장애 유아의 발화 운율 분석 연구	How to Fall in Love: 은유와 세계의 인지적 해석	A Hybrid Approach to Keywords Analysis of Aviation Accident Investigation Reports Corpus ㉠	한국 자폐스펙트럼장애 유아의 직접 화행 능력의 선별적 장애
	발 표	이수미(Univ. of Arizona, Linguistics), 정세영(Columbia Univ, Communication Sciences and Disorders), 이영인(Univ. of Hawaii at Manoa, Linguistics), 이경숙(한신대 재활의학과), 정석진(세원영유아아동상담 센터), 조숙환(서강대학교 영어영문학과)	김은일(부경대)	치 양(대련해사대), 장세은, 박호민 (한국해양대)	이영인(Univ. of Hawaii at Manoa, Linguistics), 이수미(Univ. of Arizona, Linguistics), 정세영(Columbia Univ. Communication Sciences and Disorders), 이경숙(한신대 재활의학과), 정석진(세원영유아아동상담 센터), 조숙환(서강대학교 영어영문학과)
	토 론	최영이(부산외대)	오준일(부경대)	안병길(경성대)	박희준(부산가톨릭대)

15:00 ~ 15:30	제 목	일본어 설정음 구개음화와 파찰음화에 대한 조음적 이해: 3D 혀 모델 시뮬레이션과 신경망 학습의 활용	직시적 도치 구문의 의미적 특성 (Semantic Properties of Dexis Inversion Construction)	A Corpus-based Analysis of Syntactic Features on Headlines of Maritime News ㉠	수어코퍼스 기반한 농인 초등 언어교과목 교재 개발을 위한 증강현실 적용방법에 대한 방향 모색
	발 표	장하연(부산외대)	남소영(부산대)	왕 메이한(한국해양대)	권순복(부산대)
	토 론	강은지(부산대)	정현경(신라대)	리양 쉬아오레이 (한국해양대)	박희준(부산가톨릭대)
15:30 ~ 16:00	제 목	한국어 겹자음 실현과 운율 강화	한국수어에 나타난 개념의 신체화		
	발 표	김정윤 & 강은지 (부산대)	최영주(조선대)		
	토 론	장하연(부산외대)	장세은(한국해양대)		
16:00 ~ 17:00	초청 특강	박진호(서울대): 인공지능 언어모델에 통사구조와 한국어 특징에 대한 지식을 주입하는 방법 사회: 강은지(부산대), 토론: 장세은(한국해양대)			
17:00 ~ 17:30	연구윤리교육 : 편집위원장(이원빈, 대전대)				
17:30 ~ 18:00	학술상 시상식				
	정기총회 진행: 총무이사(홍신철, 부산외대)				

㉠ = Presentation in English

Contents

메인세션 (개회식, 초청특강, 연구윤리교육 및 총회)

- 인공지능 언어모델에 통사구조와 한국어 특징에 대한 지식을 주입하는 방법 3
/ 박진호

1분과 (음운/형태/통사)

- *Who left?* Construction and EPP-Puzzle 39
/ 권기양
- Interaction between Syntax and External System :Evidence from Fragmentary Questions in Korean 45
/ 전해원
- 동사이동과 짝병합에 대한 소고 55
/ 김대익
- 한국 자폐스펙트럼장애 유아의 발화 운율 분석 연구 63
/ 이수미, 정세영, 이영인, 이경숙, 정석진, 조숙환
- 일본어 설정음 구개음화와 파찰음화에 대한 조음적 이해: 3D 혀 모델 시뮬레이션과 신경망 학습의 활용 .. 85
/ 장하연
- 한국어 겹자음 실현과 운율 강화 99
/ 김정운 & 강은지

2분과 (의미/화용)

- 은유적 표현에 의한 부각과 은폐에 관한 연구 -코로나19 사태 관련 신문 사설을 중심으로 111
/ 김철규
- Korean Hotels' Responses to Negative Online Reviews: A Genre Analysis 131
/ 안병길
- Functional Linguistic Perspective on Twitter Hashtag: What the Patient31 Tells Us 143
/ 최예슬 & 민수정
- How to Fall in Love: 은유와 세계의 인지적 해석 153
/ 김은일
- 직시적 도치 구문의 의미적 특성 (Semantic Properties of Dexis Inversion Construction) 157
/ 남소영
- 한국수어에 나타난 개념의 신체화 159
/ 최영주

3분과 (코퍼스)

- A Corpus-based Study of Specialized Vocabulary Lists for Maritime English 181
/ 슈린 & 장세은
- Functional Analysis of Lexical bundles in rhetorical move of research article abstracts in Applied Linguistics 205
/ 리양 쉬아오레이 & 장세은
- A comparison of maritime-related characteristics in the 19th century English journal log-book corpus of East Asian regions through keyword analyses 219
/ 장세은, 이성화, 박호민
- A Hybrid Approach to Keywords Analysis of Aviation Accident Investigation Reports Corpus ... 241
/ 치양, 장세은, 박호민
- A Corpus-based Analysis of Syntactic Features on Headlines of Maritime News 261
/ 왕 메이한

4분과 (응용/언어교육)

- Research on Covid-19 English Education: Students' Perspectives 293
/ 허선민 & 이강영
- L2 Learners' Perceptions of Online English Learning in the Post-Corona Era: Focusing on Affective Factors and Language Proficiency 301
/ 신유선
- 글쓰기 과업에서 언어형태의 복잡성이 주목과 학습에 미치는 영향 313
/ 조윤경
- 한국 자폐스펙트럼장애 유아의 직접 화행 능력의 선별적 장애 321
/ 이영인, 이수미, 정세영, 이경숙, 정석진, 조숙환
- 수어코퍼스 기반한 농인 초등 언어교과목 교재 개발을 위한 증강현실 적용방법에 대한 방향 모색 337
/ 권순복

메인세션

개회식, 초청특강, 연구윤리교육 및 총회

인공지능 언어모델에 통사구조와 한국어 특징에 대한 지식을 주입하는 방법

How to inject syntactic structure and knowledge of Korean
into language models in artificial intelligence

한국언어과학회

2021. 8. 19.

박진호 (서울대 국문과)

개요

- 자연언어의 구, 문장은 단어들이 통사규칙에 따라 순차적으로 결합하여 형성된다.
- 구의 의미, 문장의 의미는 단어들의 의미 및 이들 사이의 통사관계에 의해 결정된다.
 - 합성성 원리(principle of compositionality)
- 현재 널리 통용되고 있는 인공지능 언어모델은 이러한 통사구조, 합성성 원리를 별로 존중하지 않고 있다.
- 인공지능 언어모델이 통사구조, 합성성 원리를 존중하는 방향으로 발전하려면 어떻게 해야 하는지 알아본다.
- 인공지능 언어모델이 자연언어처리(NLP)에 사용되는 용도의 주된 2가지 구별인 자연언어이해(NLU)와 자연언어생성(NLG)으로 나누어 살펴본다.

모델, 모델링, 언어모델

- 탐구하고자 하는 대상을 컴퓨터로 모사(simulate)하려면, 그 대상의 중요한 속성을 지니되 컴퓨터 내에서 연산(compute)/조작(manipulate)/처리(process) 가능한 형태의 것을 만들어야 함. 이를 model이라 함
- 대상에 부합되고 연산에 적합한 모델의 구조를 디자인하는 일을 모델링(modeling)이라고 함.
- 탐구하고자 하는 대상이 언어일 경우, 그런 모델을 언어모델(language model)이라고 함.
- 언어모델은 다양한 용도로 사용되고 있음.
 - 음성인식: 음파를 음소, 단어, 문장 등의 언어요소로 변환할 때, 음파의 물리적 속성과 으로는 어떤 언어요소로 변환해야 할지 애매한 경우가 많은데, 이때 언어모델을 바탕으로 예측하면 좋음.
 - 자연언어생성(natural language generation): 기계번역, 문서 요약, 이미지 캡션 달기, 자동 기사 작성 등에서 자연스러운 문장 생성을 위해 필요함.

모델의 두 desiderata 사이의 trade-off

- 모델은 데이터/현상의 중요한 점을 포착하되 조작/연산/추론을 용이하게 할 수 있는 성질을 지니는 것이 좋음.
- 데이터에 너무 정밀하게 맞추려고 모델을 너무 복잡하게 하면 후자의 성질이 희생되고 (도메인 전문가가 빠지기 쉬운 함정)
- 모델을 가지고 할 일을 용이하게 하려고 모델을 너무 단순화하면 전자의 성질이 희생됨. (수학자, 통계학자, 컴퓨터과학자가 빠지기 쉬운 함정)
- 두 가지 성질 사이에서 적절한 중용의 지점을 찾아야 함.
- 통계학자 George Box의 유명한 격언: All models are wrong, but some are useful.
- 모델이 데이터와 완전히 부합될 수는 없음.
- 그러나 어느 정도/상당히 부합되면서 유용한 모델을 모색해야 함.

이해를 위한 모델과 생성을 위한 모델

- 하나의 언어모델이 이해(NLU)와 생성(NLG) 양쪽에 다 사용될 수 있는 경우도 있기는 하나
- 이해에 특화된 모델, 생성에 특화된 모델로 나뉘는 것이 보통임.
- 이해를 위한 언어모델은, 단어, 구, 문장, 문서의 의미를 수학적, 전산적 처리에 적당한 형태(대개 벡터)로 변환하는 것(encoding)이 핵심.
 - 언어요소들 사이의 의미관계를 벡터들이 잘 반영해야 함.
- 생성을 위한 언어모델은, 나타내고자 하는 의미(대개 벡터) 및 prompt(대개 앞부분의 단어(들))가 주어지면, 문장의 나머지 부분을 만들어내는 것(decoding)이 핵심임.
 - prompt로 주어진 단어를 바탕으로 하여 그 다음 단어를 잘 예측해야 함.
 - 이를 조건부확률로 모델링함.

이해를 위한 언어모델의 층위: 단어, 문장

- 언어모델이 사용되는 용도(downstream task)를 위해서는 대개 문장을 적절히 인코딩(벡터화)하는 것이 중요함.
 - 의미가 비슷한 문장들은 비슷한 벡터로 표상되어야 하고(유사성 기준)
 - 의미가 꽤 다른 문장들은 꽤 다른 벡터로 표상되어야 하고(차이점 기준)
 - 두 문장 A와 B의 의미관계와 두 문장 c와 d의 의미관계가 평행하다면, 벡터 A와 B의 관계와 벡터 c와 d의 관계도 평행해야 함. (유추 기준)
- 우선 단어를 벡터화하고, 이 벡터들을 적절히 결합하여 구 벡터, 문장 벡터를 만들어 가는 것이 언어학자들의 상식에 부합될 터이나
- 현재 널리 통용되는 언어모델은 그렇게 하지 않고, 하나의 신경망 모델로 단어 벡터와 문장 벡터를 한꺼번에 구함(예: BERT).
 - 단어에 비해 단어 결합(구, 문장)은 더 sparse한데
 - 이런 data sparseness problem을 극복하고 품질 좋은 벡터를 얻으려면
 - 엄청나게 많은 양의 데이터와 매우 복잡한 구조의 신경망 모델이 필요함.
 - 데이터와 언어모델의 거대화에 따라 엄청난 전기와 전산자원이 필요함.

통사구조와 합성성을 고려한 인코딩 전략

- 구, 문장에 비해 단어를 벡터화하는 것은 훨씬 적은 데이터와 전산자원만 가지고도 가능함. 예: Word2Vec
- 이 값싼 방법으로 단어 벡터를 얻은 다음에
- 이 단어 벡터를 통사적 결합 순서에 따라 적절히 결합하여 점점 더 큰 구의 벡터를 얻고 결국 문장 벡터를 얻는 것이
- 훨씬 경제적이고, 언어학의 이론적 상식에도 부합됨.
- 물론 자연언어의 구, 문장이 완전히 합성적인 것은 아니므로(예: 속어), 이런 비합성성에도 대비해야 하지만
- 자연언어 문장의 기본 원리인 합성성을 모델링하는 것이 우선임.
 - 합성성에 대한 모델링이 어느 정도 이루어진 뒤에, 비합성성을 위한 조치가 추가될 수 있음.

형식의미론과 벡터의미론의 상황 대비

- 형식의미론
 - 합성성 원리를 잘 포착하나(술어-논항 관계 등)
 - 어휘의미를 나타낼 淸속한 방법이 없음.
 - John loves Marry: Love(john, marry)
 - John loves a girl: $\exists x$ (Girl(x) & Love(john, x))
- 현재의 벡터의미론
 - 어휘의미를 벡터로 나타내는 데는 꽤 성공했으나
 - 단어 벡터들을 결합하여 구 벡터를 만드는 방법은 아직 확립되지 못했음.
- 앞으로 나아가야 할 방향
 - 합성적 벡터 의미론(compositional vector semantics)

단어의 벡터화 전략

- 통상적인 벡터화 방법(예: Word2Vec)은 벡터화하고자 하는 타겟 단어의 앞뒤 문맥에 출현하는 단어들을 바탕으로 함.
 - 이때 앞뒤 문맥은 그냥 linear order를 바탕으로 하여 몇 단어 이내에 있다는 식으로 정의됨.
- 타겟 단어의 주위 문맥을 단순히 linear order가 아니라 통사구조(특히 의존구조)에 따라 정의하는 것이 더 바람직.
- 예: **Australian scientists discovered star with telescope.**
 - 타겟 단어: discovered
 - 통상적인 Word2Vec 알고리즘에서 window 크기가 2인 경우, 타겟 단어의 벡터화에 참여하는 문맥은 [Australian, scientists, star, with]가 됨.
 - 'Australian'은 타겟 단어와 별 상관 없이 있는데도 가까이 있다는 이유로 이 문맥에 포함되었음.
 - 'telescope'는 타겟 단어와 관련이 있는데도, 거리가 좀 멀어 제외되었음.

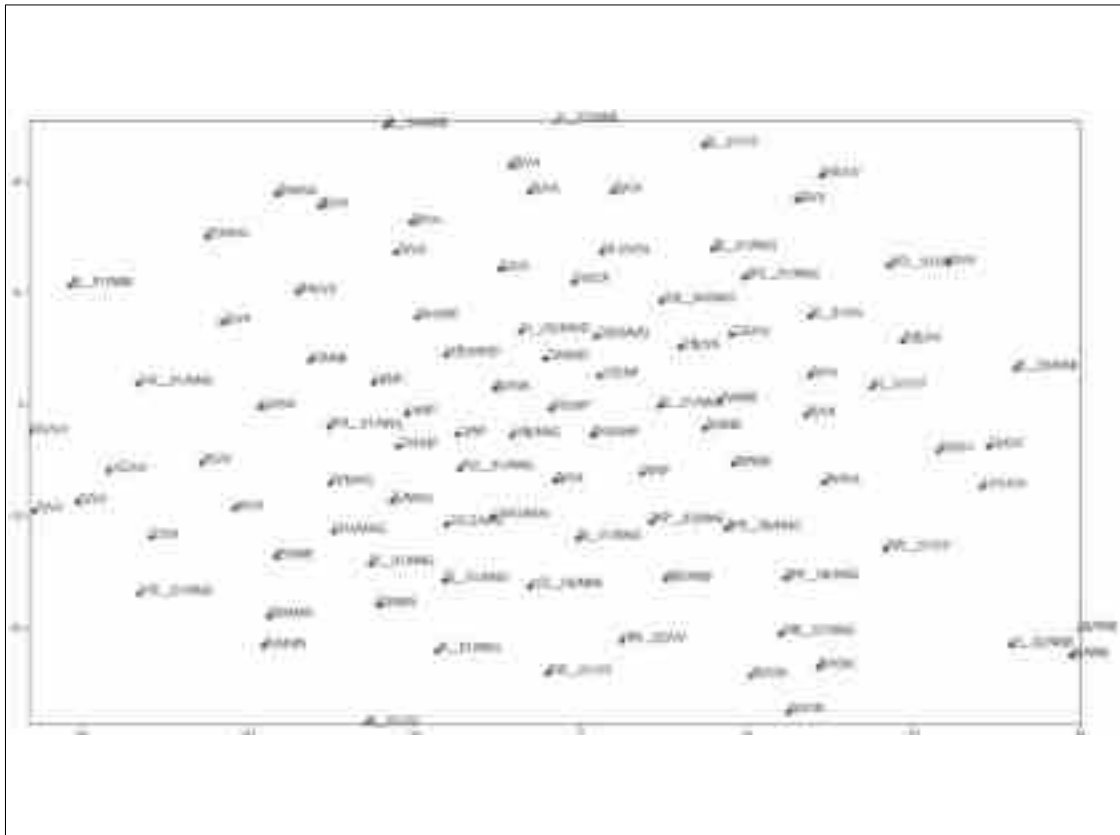
단어 벡터를 구할 때 통사구조의 중요성

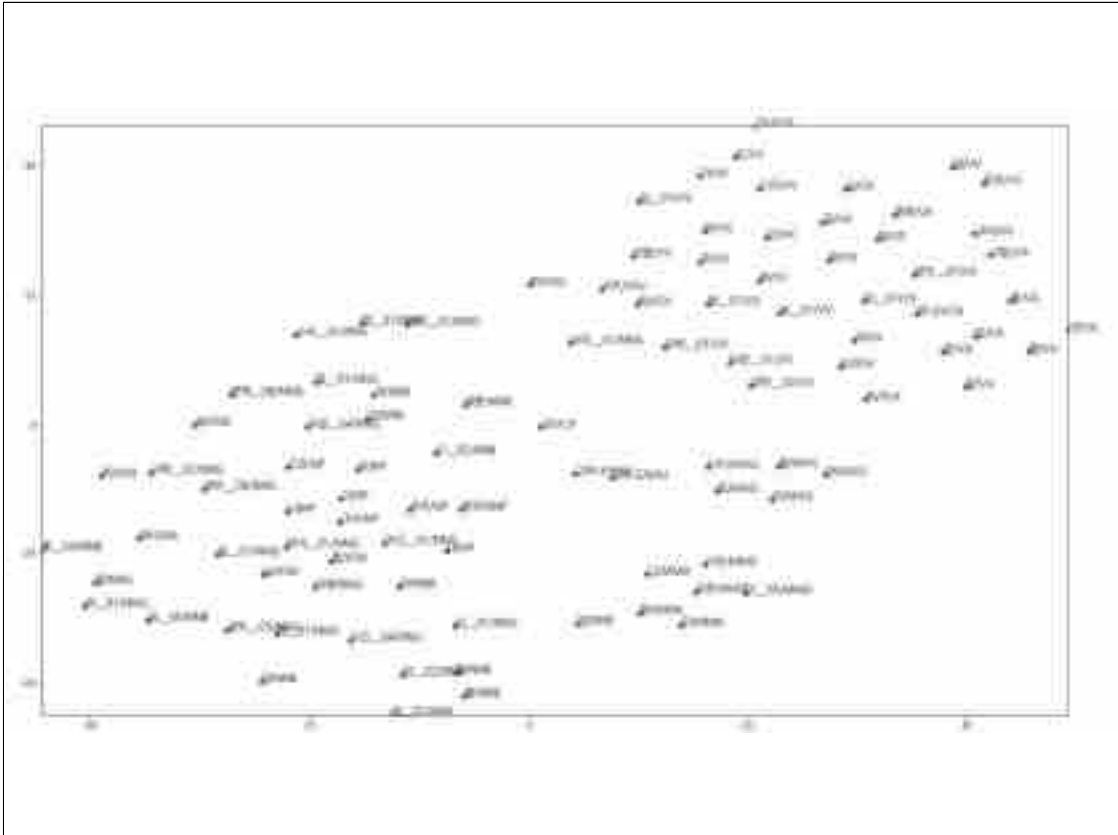
- 눈+을 잘 감+아라.
- 실+을 잘 감+아라.
- 머리+를 잘 감+아라.
- 타겟 단어가 '감'이고 window 크기가 2일 경우
- 타겟 단어의 벡터화에 참여하는 문맥에 '-을/를', '잘'은 포함되지만
- 정작 동형어 '감'의 구별에 핵심적인 역할을 할 목적어 '눈', '실', '머리'는 참여하지 못하게 됨.
- 타겟 단어와 선조적으로 좀 멀리 있더라도 통사적으로 긴밀하면 문맥에 포함시켜야 하고
- 타겟 단어와 선조적으로 가깝더라도 직접적인 통사 관계를 맺고 있지 않으면 문맥에서 제외시킬 필요가 있음.

Levy and Goldberg (2014)

- 바로 이러한 문제의식을 반영하여 Word2Vec 알고리즘의 수정을 제안함.
- 통상적인 방법에 따르면 타겟 단어와 품사가 다른 단어들도 유사 단어로 판정되는 데 비해
 - 주제적 관련성(topical relatedness)
- 통사적 의존관계를 반영한 벡터화 방법을 쓰면 타겟 단어와 문법적/통사적/기능적으로 유사한 단어들만이 엄선되는 경향을 보임.
 - 단어들 사이의 (의미적분 아니라) 문법적 유사성을 포착하는 데 더 적합.

Target Word	BoW	HoW1	HoW2
jetson	lightning	expensive	expensive
	expensive	expensive	expensive
	corrosion	expensive	expensive
	corrosion	expensive	expensive
jetsons	dark-kelco	expensive	expensive
	hollow	expensive	expensive
	full-sized	expensive	expensive
	valley	expensive	expensive
jetting	weather-resistant	non-deterioration	padding
	non-deterioration	non-deterioration	padding
	compatibility	non-deterioration	padding
	deterioration	non-deterioration	padding
jetted	jet	jet	jet
	jet	jet	jet
	jet	jet	jet
	jet	jet	jet
jetted-verb	expensive	expensive	expensive
	expensive	expensive	expensive
	expensive	expensive	expensive
	expensive	expensive	expensive
jetting-verb	jetting	jetting	jetting
	jetting	jetting	jetting
	jetting	jetting	jetting
	jetting	jetting	jetting





두 단어 벡터를 결합하는 방법: 더하기

- 두 단어 벡터를 u, v 라고 할 때 이 두 단어를 결합한 구의 벡터를 p 라 하면, $P = u + v$ 의 식으로 p 를 구하자는 것. (가장 간단한 방법)

- 벡터 u 와 v 의 차원 n 일 때: $P_1 = u_1 + v_1 \dots P_i = u_i + v_i \dots P_n = u_n + v_n$

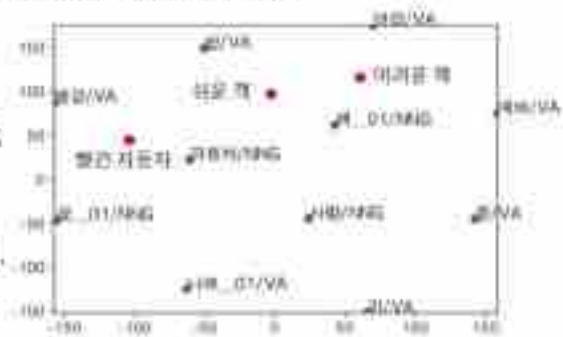
- 벡터 공간에서 벡터 p 의 위치는 벡터 u 와 v 의 중간이 됨.

• 장점

- 연산이 매우 단순함.
- 벡터의 차원이 그대로 유지됨.
 - 결합시 차원이 늘어나면 연산이 어려움.

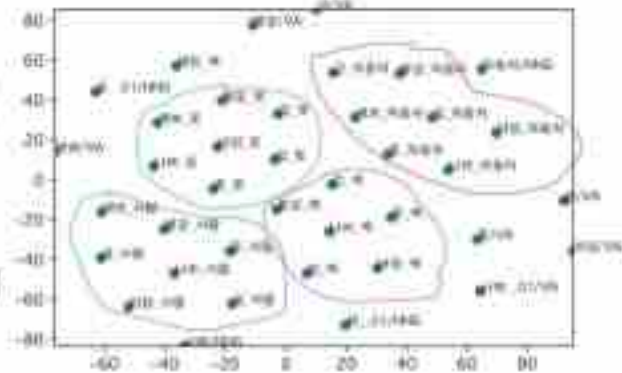
• 단점

- 어순의 차이를 반영하지 못함.
- 핵-비핵(지배자-의존소) 관계도 무시.



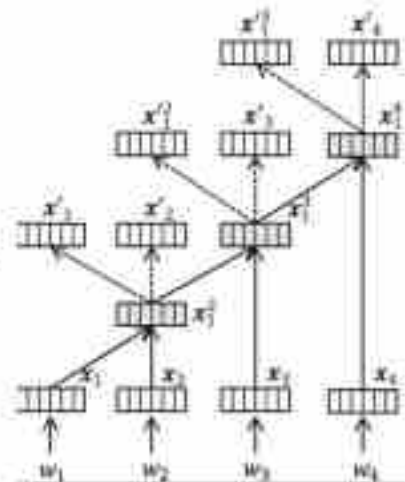
두 단어 벡터를 결합하는 방법: 가중합

- 핵 벡터와 수식어 벡터에 가중치를 다르게 부여하여 더하는 것이 더 합당함: $P = (1 - \beta) * u + \beta * v$ (β 는 실수/스칼라)
- 핵에 0.6, 수식어에 0.4의 가중치를 부여하여 더하면 다음과 같아 됨.
- 핵 명사에 형용사 수식어가
- 첨가되면
- 형용사 쪽으로 좀 이동하지만
- 핵 명사 근처에 머무르게 됨.
- 비정상적인 수식관계는
- 특별한 처리가 필요함.
 - 가짜/잠난갑 총, 종이 학/비행기
 - 둘 사자, 모험 비행기



Recursive auto-encoder: Socher et al (2011)

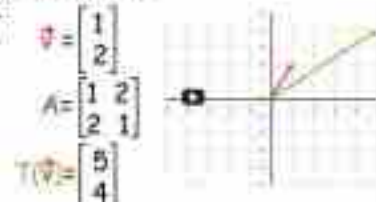
- 두 단어 w_1, w_2 의 벡터 x_1, x_2 (각각 d -차원)를 이어붙여 $2d$ -차원의 벡터를 만들: $[x_1; x_2]$
- 여기에 가중치 행렬 $W^{(1)}$ 을 곱하고 절편 $b^{(1)}$ 을 더한 뒤 비선형함수(\tanh)를 적용하여 w_1, w_2 의 벡터 x_1^* 를 만들. $x_1^* = f(W^{(1)} * [x_1; x_2] + b^{(1)})$
 $W^{(1)} \in \mathbb{R}^{d \times 2d}, b^{(1)} \in \mathbb{R}^d, f(\cdot) = \tanh(\cdot)$
- x_1^* 로부터 x_1^*, x_2^* 을 복원함.
 - 복원 오차가 가급적 적게끔 $W^{(1)}$ 과 $b^{(1)}$ 을 학습함.
- 이런 과정을 w_3, w_4 등이 하나씩 결합할 때마다 반복함.
- 결국 문장 전체의 d -차원 벡터 표상이 얻어짐.
- 이 벡터는 모든 단어에 관한 정보를 포함함.



행·수식어 차이 적극 반영: 곱하기

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$$

- 행렬*벡터=벡터: 행렬은 벡터를 변환하는 연산으로 이해할 수 있음.
 - 특정 축(차원) 방향으로 높이거나 줄임, 회전시킴, 축의 반대편으로 반사시킴.
- '형용사+명사' 결합에서 명사는 벡터이고, 형용사는 행렬로서
- 명사 벡터에 형용사 행렬을 적용하면(곱하면), 구의 벡터가 얻어짐.
- 형용사 행렬은 명사 벡터를 일정한 방식으로 변환하는 작용을 함.
- Baroni and Zamparelli (2010). Nouns are Vectors, Adjectives are Matrices: Representing Adjective-Noun Constructions in Semantic Space.
- 장점: 형용사 의미가 명사에 의존하는 현상 포착
 - good lawyer, good neighbor
 - 명사 벡터의 각 차원에 상이하게 작용할 수 있음.
- 문제점: 형용사 행렬을 어떻게 구할 것인가?



텐서 곱과 convolution을 통한 차원 축소

- n-차원 벡터 u와 v에 텐서 곱(tensor product, outer product)을 적용하면 $n \times n$ 행렬이 만들어짐.
 - practical = [0, 6, 2, 10, 4], difficulty = [1, 8, 4, 4, 0]
 - practical \otimes difficulty =
- | | | | | | |
|---|----|----|----|----|---|
| • | 0 | 0 | 0 | 0 | 0 |
| • | 6 | 48 | 24 | 24 | 0 |
| • | 2 | 16 | 8 | 8 | 0 |
| • | 10 | 80 | 40 | 40 | 0 |
| • | 4 | 32 | 16 | 16 | 0 |
- 0이 $n \times n$ 행렬에 circular convolution을 적용하여
 - n-차원 벡터를 구함,
$$p_j = \sum_i u_i \cdot v_{(i-j)}$$
 - practical \odot difficulty = [116, 50, 66, 62, 80]

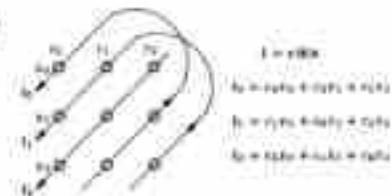


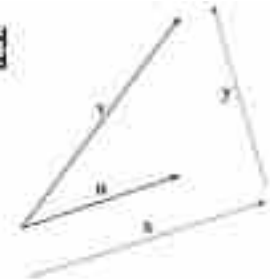
Figure 1: Circular convolution represented as a compressed outer product.

스칼라 가중치를 구하는 선형대수적 방법

- 벡터 v 를 서로 직교하는 두 벡터 x 와 y 로 분해.
- x 는 u 와 평행.

$$x = \frac{u \cdot v}{u \cdot u} u$$

$$y = v - x = v - \frac{u \cdot v}{u \cdot u} u$$



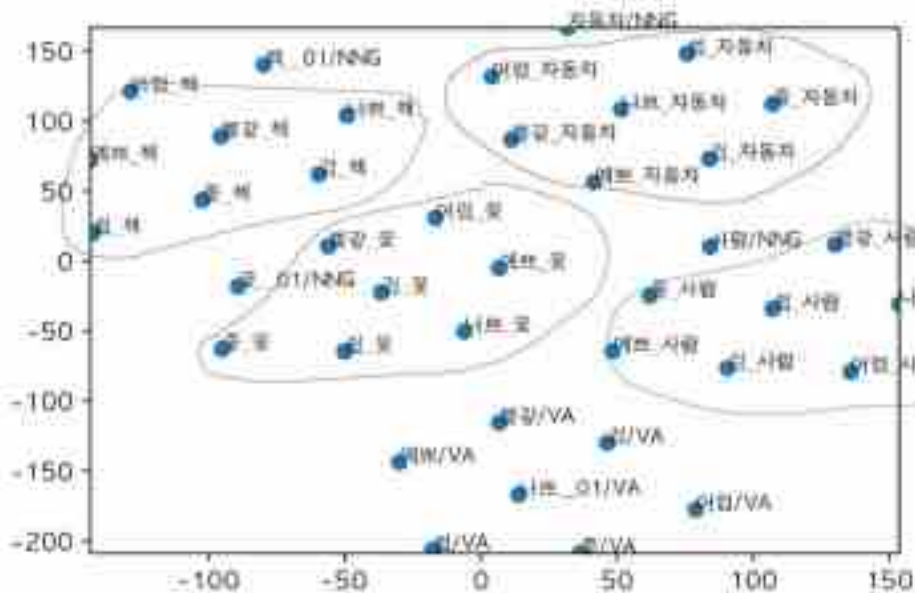
y 는 그대로 두고 x 를 λ 만큼 늘려 v' 을 구함.

$$v' = \lambda x + y = \lambda \frac{u \cdot v}{u \cdot u} u + v - \frac{u \cdot v}{u \cdot u} u = (\lambda - 1) \frac{u \cdot v}{u \cdot u} u + v$$

- 어떤 벡터공간 내의 벡터에 스칼라를 곱하여 벡터의 크기를 변화시켜도, 벡터와 벡터 사이의 각도(코사인 유사도)에는 영향이 없으므로, 원하는 스칼라를 곱하여 식을 단순화시킬 수 있음.

$$p = (u - u)v + (\lambda - 1)(u \cdot v)u$$

- 위의 두 가지 방법으로 p 를 구하여 plotting해 보니 동일한 결과가 나왔음.



행렬 곱셈 모델

- Rudolf and Giesbrecht (2010): Compositional Matrix-Space model (CMS)
 - 단어 의미를 벡터가 아니라 행렬로 나타냄.
 - 두 벡터의 선형결합에 비해, 행렬 곱셈이 두 요소 사이의 상호작용을 더 잘 포착할 수 있음. cf. 다중회귀분석: $y = w_1 * x_1 + w_2 * x_2 + w_3 * x_1 * x_2$
 - 문제점: 행렬 곱셈은 associative하므로 $(AB)C=A(BC)$, 구조적 중의성이 있는 경우에 ('착한 철수의 동생') 두 의미를 다르게 나타낼 수 없음.
- Socher et al (2012): Matrix-Vector RNN (MV-RNN)
 - 각 단어 벡터에 결합하는 단어 행렬을 곱함. (이 행렬은 신경망 학습으로 얻음)
 - 문제점: 결합하는 두 요소 중 어느 쪽이 연산자인지는 상대적임. 밥 먹 vs. 빨리 먹.
 - 연산자 성격이 있는 요소는 적절한 행렬을 할당받고, 그렇지 않은 요소는 단위행렬을 할당받는 것을 목표로 함.
 - 모든 단어에 대해 $d \times d$ 행렬을 신경망 학습을
 - 풀해 얻어야 한다는 것이 큰 부담.
 - 동사결합이 non-commutative, non-associative
 - 한 성격을 포착.

$$(6) \quad \vec{k} = f(B\vec{a}, A\vec{b}) = \sigma\left(W \begin{bmatrix} B\vec{a} \\ A\vec{b} \end{bmatrix}\right)$$

$$(7) \quad H = f_M(A, B) = W_M \begin{bmatrix} A \\ B \end{bmatrix}$$

where $\vec{a}, \vec{b}, \vec{k} \in \mathbb{R}^{d-1}$, $A, B, H \in \mathbb{R}^{d \times d}$, and $W, W_M \in \mathbb{R}^{d \times 2d}$.

행렬 곱셈 모델

$$(8) \quad \vec{k} = \tanh\left(W \begin{bmatrix} \vec{h}_1 \\ \vec{h}_2 \end{bmatrix}\right) + \vec{b} + \vec{h}_1^T \mathbf{V} \vec{h}_2$$

where $\vec{h}_1, \vec{h}_2, \vec{b} \in \mathbb{R}^{d-1}$, $W \in \mathbb{R}^{d \times 2d}$, and $\mathbf{V} \in \mathbb{R}^{(d-1) \times (d-1)}$. RNTN improves on MV-RNN in that

- Socher et al (2013): Recursive Neural Tensor Network (RNTN)
 - 행렬 곱셈의 장점을 유지하면서 MV-RNN의 부담을 경감시킴.
 - 결합되는 두 단어 벡터(h_1, h_2)와 rank-3 텐서 \mathbf{V} 를 곱함. (이 텐서를 구하는 게 부담)
- Tai et al (2015): Tree-LSTM
 - Tree-RNN의 RNN을 LSTM으로 대체하여 vanishing gradient problem 해결.
 - tree 구조 모델들 가운데 최고의 성능 달성.
- Chung et al (2019): Lifted Matrix-Space model (LMS)
 - d -차원 단어 벡터를 $vd \times vd$ 행렬로 변환(lift).
 - 좌측 행렬에 가중치 행렬 W 곱하고 절편 행렬 b 더하고 비선형변환하여 H_{inner} 얻음.
 - H_{inner} 를 가중치 행렬로 삼아 우측 행렬에 동일한 연산을 하여 H_{outer} 얻음.
 - 그 뒤에 행렬들을 다시 d -차원 벡터로 변환하여
 - LSTM 레이어에 잡아넣음.

$$(13) \quad H = \tanh(W_{inner}\vec{c} + B_{inner})$$

where $\vec{c} \in \mathbb{R}^{d \times d}$, $B_{inner} \in \mathbb{R}^{d \times d}$, and $W_{inner} \in \mathbb{R}^{d \times (d \times d)}$. The resulting H matrix

$$(14) \quad H_{inner} = \tanh(W_{inner}H_1 + B_{inner1})$$

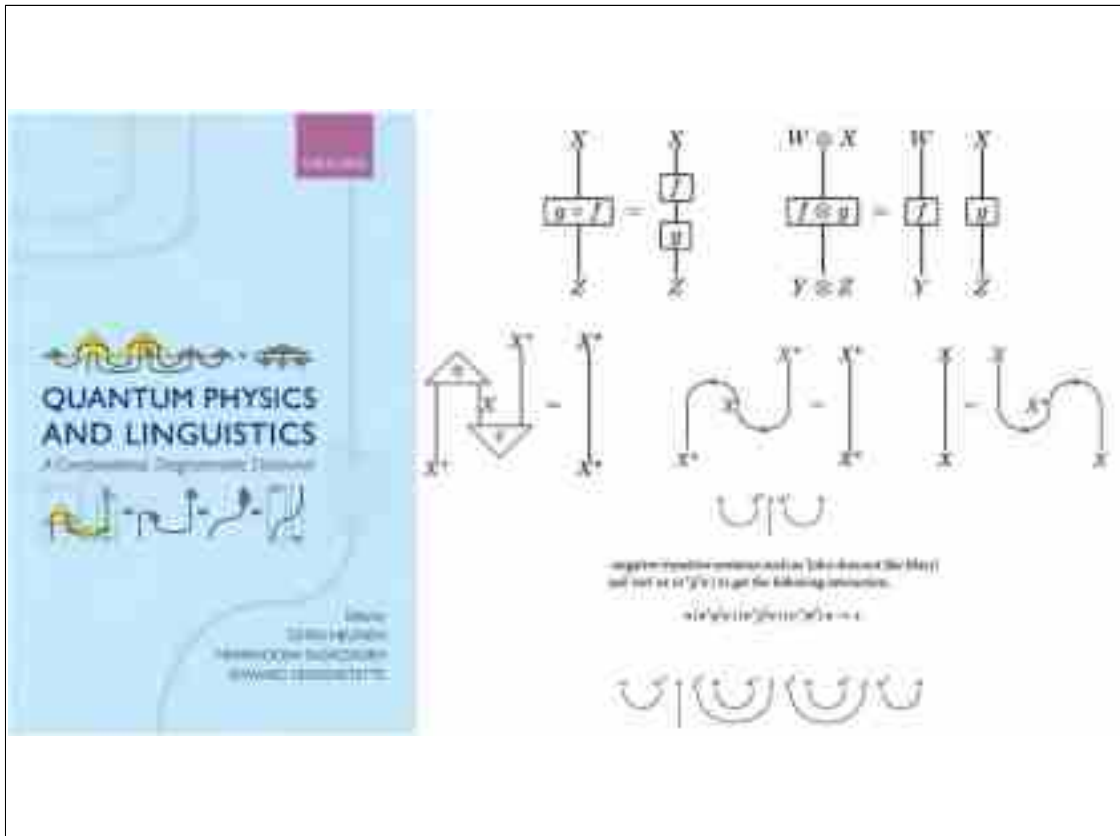
$$(15) \quad H_{outer} = \tanh(H_{inner}H_2 + B_{inner2})$$

행렬 곱셈 모델 소결

- 벡터의 선형결합보다는
- 행렬 곱셈이 통사결합의 다양한 성격을 더 풍부하게 모델링할 수 있는 잠재력이 있음.
- 벡터 대신 행렬, 텐서를 쓰면 parameter 수가 exponential하게 증가하므로, 이 부담을 경감시킬 수 있는 방법 필요.
- 모델을 너무 복잡하게 하기보다는 통사결합이 일어날 때 의미결합이 일어나는 양상의 핵심적 속성을 포착하는 데 집중할 필요 있음.

앞으로의 과제

- 문법요소의 처리 방식을 모색해야 함.
 - 순수한 문법관계 표지는 벡터 연산에 포함시키지 않고, 두 어휘요소 사이의 관계 유형을 나타내는 역할을 하는 것으로 보는 게 좋을 듯.
- '수식어+핵' 결합과 '논항+술어' 결합은 성격이 다르므로, 벡터 결합 방식도 달라야 함.
- 두 언어요소가 결합할 때 이 둘 사이의 통사관계/의미관계는 사실 세분화하면 상당히 많은 부류로 구분할 수 있음.
 - 관형어+계연, 부사어+용언, 주어+용언, 목적어+용언, 부사어+절, 종속절+주절, 대응/병렬 관계
- 한국어에서 이런 관계를 적절하게 분류할 필요 있음.
 - 참고: Igor Mel'čuk의 *Surface Syntax of English*에서 영어의 통사관계 분류
- 이에 상응하는 벡터 결합 방식을 고안할 필요 있음.
 - 선형대수학에서 다루는 벡터, 행렬, 텐서의 연산들을 잘 알아둘 필요 있음.
 - 통계물리학, 양자역학에서 양자계의 상태를 벡터로 나타내고 복합계(complex system)의 상태를 벡터 조합 연산으로 나타냄. 이를 참고할 필요 있음.



텍스트 구조 및 생성 과정 모델링

- 하나의 텍스트는 문단(paragraph)들로 이루어져 있고
 - 텍스트는 문단의 연쇄(sequence of paragraphs)
- 하나의 문단은 문장(sentence)들로 이루어져 있고
 - 문단은 문장들의 연쇄(sequence of sentences)
- 하나의 문장은 단어(word)들로 이루어져 있다고 할 수 있음.
 - 문장은 단어들의 연쇄(sequence of words)
- 단어, 문장, 문단, 텍스트의 의미를 어떤 형식으로든 표상(represent)할 수 있다고 가정할 수 있음.
 - 현재의 기계학습(딥러닝 포함)에서는 수치 벡터로 표상.
- 앞의 언어단위를 바탕으로 하여 그 뒤의 언어단위가 생성된다고 가정.
 - linear, sequential model of language
 - 조건부 확률로 모델링: $p(\text{뒤 언어단위} | \text{앞 언어단위})$

언어단위의 위계적 구조

- 한 문장 안에 있는 단어들끼리의 관계는 밀접하나, 문장 경계를 넘어서는 단어들 사이의 관계는 상대적으로 소원함.
- 한 문단 안에 있는 문장들 사이의 관계는 밀접하나, 문단 경계를 넘어서는 문장들 사이의 관계는 상대적으로 소원함.
- 한 문장 안에서는 앞 단어를 바탕으로 그 뒤 단어를 예측하는 식을 모델링하는 것이 적절하나
- 앞 문장의 (끝 부분의) 단어를 바탕으로 다음 문장의 (첫 부분의) 단어를 예측하는 것은 부적절.
- 문장이 끝날 때마다 그 문장 전체의 의미 표상을 도출하고
- 그 다음 문장의 단어를 예측할 때는 이 앞 문장의 의미 표상을 바탕으로 하는 것이 적절.
- 문단이 끝날 때도 그 문단 전체의 의미 표상을 도출하고, 이를 바탕으로 그 다음 문단의 문장을 예측하는 것이 적절.

생성을 위한 언어모델의 성격

- 언어모델은 보통 단어 연쇄에 대한 확률 분포(probability distribution over sequences of words)로 정의됨.
- 문장을 단어들의 연쇄로 간주하고, 출현 확률을 부여함.
- 어떤 문장 s 가 "BOS $w_1 w_2 w_3 w_4 w_5$ EOS"와 같이 구성되어 있다면
- $P(s) = P(\text{BOS}, w_1, w_2, w_3, w_4, w_5, \text{EOS})$ 와 같은 결합확률(joint probability)로 이해할 수 있고, 이는 다시
- $P(\text{BOS}) \times P(w_1|\text{BOS}) \times P(w_2|\text{BOS}, w_1) \times P(w_3|\text{BOS}, w_1, w_2) \times P(w_4|\text{BOS}, w_1, w_2, w_3) \times P(w_5|\text{BOS}, w_1, w_2, w_3, w_4) \times P(\text{EOS}|\text{BOS}, w_1, w_2, w_3, w_4, w_5)$
- 와 같은 조건부확률(conditional probability)들의 곱으로 이해할 수 있음.
- 즉 한 문장 내에서 앞에 이미 출현한 단어들을 전제로 해서 그 다음 단어가 출현할 확률을 계산함.
- 언어가 linear하게 산출됨을 고려하면 어느 정도 일리 있음.

Markov assumption

- 앞에 출현한 단어가 뒤에 나오는 단어의 출현 확률에 영향을 미치는 것은 하지만 그 거리가 멀어질수록 영향력이 감소함.
- 그래서 앞의 몇 개의 단어만이 영향을 미친다고 단순화하여 가정할 수 있음(simplifying assumption).
- 영향을 미치는 앞 단어가 1개라고 가정하거나(bigram 언어모델), 2개라고 가정(trigram 언어모델)하는 것이 보통임.
- 이런 가정은 말뭉치로 실제 확률을 계산할 때 거의 필수적임.
- 특정 단어들의 연쇄는 그 길이가 길어질수록 말뭉치 출현 빈도가 낮아지기 때문에 신뢰할 만하게 확률을 추정하기가 어려움.
- bigram이나 trigram 정도는 출현 빈도가 어느 정도 되기 때문에 신뢰할 만한 확률 추정이 가능함.
- LSTM 같은 모형은 훨씬 더 긴 단어 연쇄를 기억할 수 있으나, 빈도가 매우 낮은 데이터로 추정하면 과적합(overfitting)의 위험이 있음.

linear language model의 문제점

- 인접한 바로 앞의 단어가 그 다음 단어에 별로 영향을 미치지 않는 경우가 꽤 있음.
- The king of England's hat
 - 토큰 England는 토큰 's와 선조적으로 인접해 있지만(linearly adjacent), 사실은 별로 관련이 없음. 's의 출현에 가장 영향력 있는 토큰은 king임.
- The girl loved by many boys is my sister.
 - boys는 그 뒤의 is와 별 관련이 없을 뿐더러, 오히려 is의 출현 확률에 노이즈로 작용한다고 볼 수 있음. 토큰 is의 출현에 가장 영향력 있는 토큰은 girl임.
- 가난한 거지가 비싼 옷을 사기는 어렵다.
 - 토큰 '거지'는 토큰 '비싸'에 매우 가까이 있지만 별 관련이 없을 뿐더러, 오히려 '비싸'의 출현 확률에 노이즈로 작용한다고 볼 수 있음.
 - 토큰 '비싸'의 출현에 가장 영향력 있는 토큰은 '옷'임.
- 선조적으로 좀 멀리 떨어져 있더라도 통사적으로 가까운 요소가 중요함.

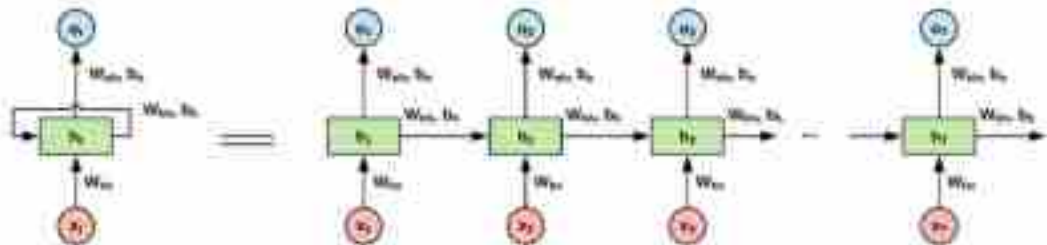
언어모델의 정교화1: 통사구조 고려

- 언어는 의사소통을 위한 도구일 뿐이고
- 의사소통은 발신자의 머릿속 메시지/의미를 수신자에게 전달하는 것.
- 언어모델은 문장이 지닌 메시지/의미의 구조를 잘 반영해야 함.
- 선조적 순서(linear order)는 의미구조를 어느 정도 반영하지만 어느 정도 왜곡하기도 함.
- 통사구조는 의미구조를 상당히 충실히 반영함.
 - 단, 완벽히 반영하는 것은 아님. 속어는 통사적으로 몇 개의 토큰이지만 의미상 하나의 덩어리임. 'a lot of books'의 통사구조는 'a lot [of books]'이지만 'a lot of books'처럼 묶어주는 게 의미상 더 좋음. 속어의 처리는 별개의 문제.
- 의미는 위계적 구조(hierarchical structure)를 지니며 통사구조도 그러하나, 이것이 선형화(linearization)되면서 정보가 소실됨.
- 위계구조 정보가 소실되고 의미구조로부터 좀 떨어진 선조적 언어모델을 (조작이 쉬우므로) 계속 사용할 것인가. (조작이 좀 더 복잡하더라도) 언어현상의 진실을 더 잘 반영하는 언어모델을 적극 채택할 것인가?

언어모델의 정교화2: Attention 메커니즘

- 앞에 출현한 여러 단어들 중 다음 단어 예측에 영향력이 큰 단어를 찾아낼 때, attention 메커니즘을 쓰는 것이 현재 대세임.
- 예컨대 다음 단어 w_5 가 무엇인지를 예측할 때, 그 앞의 단어들 w_1, w_2, w_3, w_4 가 갖는 weight를 신경망 학습을 통해 알아내는 것임.
- 이 역시 모델을 좀 더 복잡하게 함으로써, 언어모델의 예측을 더 정교하게 하려는 시도임.
- 언어학자는 통사구조/통사관계를 고려하여 다음 단어의 예측에 가장 영향력이 있는 것이 무엇인지 알아낼 수 있음.
- 따라서 영향력 있는 단어를 찾기 위해 굳이 신경망 학습이라는 우회적 경로를 밟을 필요가 없음.
- attention 메커니즘과 통사정보라는 도메인 지식을 이용하는 것 가운데 후자가 자원을 적게 쓰므로 경제적, 환경친화적임.

신경망 기반 언어모델의 변천1: Vanilla RNN



- 입력과 출력이 단위들의 sequence라고 가정. $h_t = \sigma(W_h h_{t-1} + W_x x_t + b)$
- sequence 내 각 단위의 위치 또는 time step을 t 라고 할 때
- $t-1$ 에서의 hidden state(h_{t-1})와 t 에서의 입력 단위(x_t)로부터 t 에서의 hidden state(h_t)를 도출하고, 이로부터 t 에서의 출력(y_t)을 뱉어냄.
- h_t 는 연쇄의 처음부터 t 까지의 정보를 총괄하여 지니게 됨.
- 연쇄의 끝까지 가면 h_t 는 연쇄(문장) 전체의 정보(의미)를 지니게 됨.

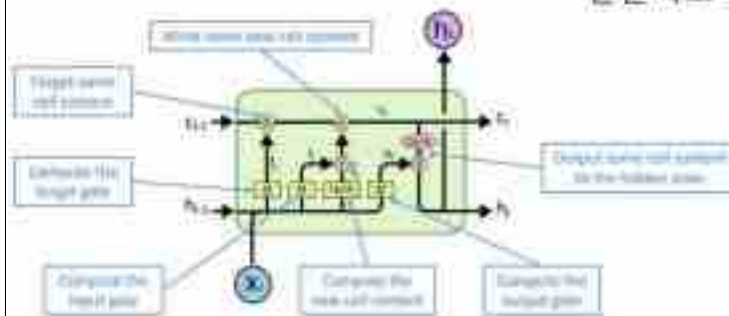
Vanilla RNN의 약점: Vanishing gradient problem

- 각 노드를 연결하는 가중치(weight) 행렬 w 의 요소들은 대개 0에 가까운 매우 작은 실수임.
- 이런 행렬을 반복해서 곱하다 보면 가중치가 점점 더 0에 가까워짐.
- 즉, 연쇄 내에서 거리가 멀어질수록, 한참 앞에 있는 요소에 관한 정보는 잘 보존/전달되지 못하고 소실되는 경향이 있음.
- h_t 가 연쇄 처음부터 거기까지의 모든 정보를 총괄하리라고 기대했지만
- 사실은 가까운 과거만 기억하고 먼 과거는 잊어버리는 문제점이 나타나는 것.

신경망 기반 언어모델의 변천2: LSTM (Long short-term memory)



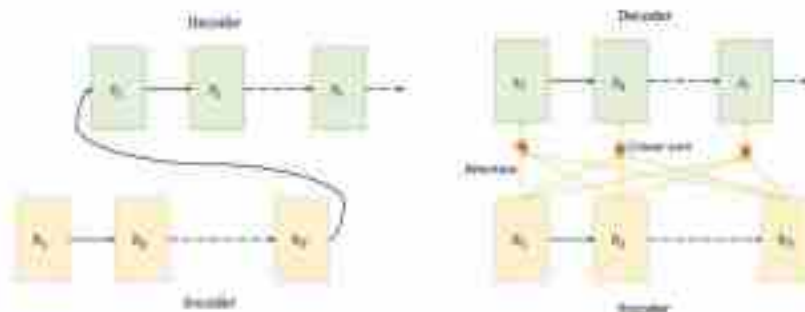
- h_{t-1} 과 x_t 로부터 h_t 를 도출할 때, 그 사이에 3개의 gate(forget, input, output)를 추가
- 앞노드의 정보가 뒤로 멀리까지 전달되도록 함.



- 양방향 LSTM을 사용하면, 앞노드의 정보도 뒤로 멀리까지 전달되고 뒤노드의 정보도 앞으로 멀리까지 전달됨.
- h_t 의 정보를 산출할 때 앞뒤 요소 모두 참고.

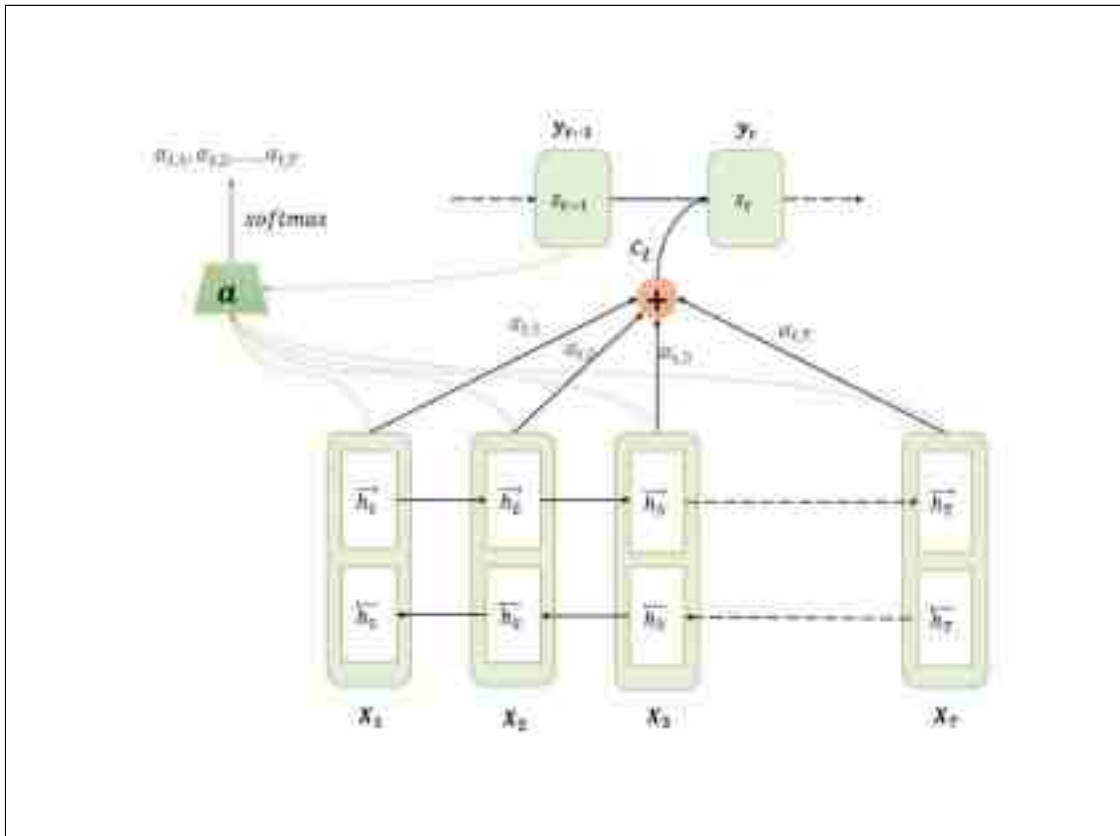
신경망 기반 언어모델의 변천3: Attention

- 기존 모델에서는 encoder에서 정보를 sequentially 축적하여 h_T 에 집약.
 - decoder에서는 h_t 만 참고하여 output 생성.
- output의 각 요소를 산출할 때, input/encoder의 각 요소를 참고하되
- 어느 요소를 중시할 것인지(attention score)를 신경망 학습으로 결정.



(a) Vanilla Encoder-Decoder Architecture

(b) Attention Mechanism



Attention의 두 종류

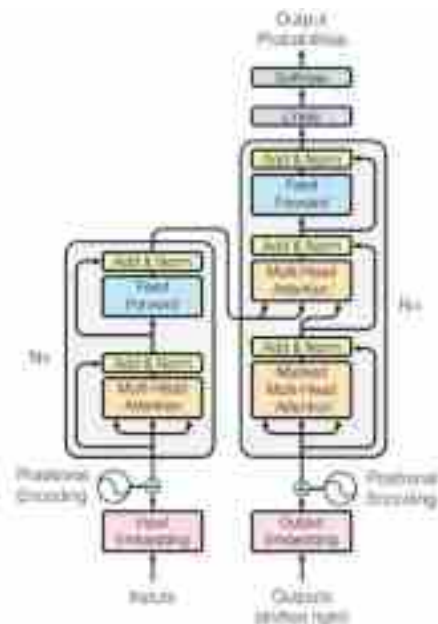
- **cross-attention**
 - decoder에서 output을 산출할 때, output sequence의 각 단위가 input/encoder sequence의 단위들 중 어느 것에 집중하는가를 나타냄.
 - 기계번역 같은 과제에서 사용하기에 적합함.
- **self-attention**
 - 하나의 sequence(문장) 내에서 각 단위가 그 sequence(문장) 내의 어느 단위에 집중하는가를 나타냄.
 - coreference/anaphora resolution("John shot Bill, and he fell down."에서 he의 선행사가 무엇인가?) 같은 과제에서 사용하기에 적합.
- **multi-head attention**
 - 하나의 sequence 내에서 두 요소가 맺을 수 있는 관계는 여러 종류가 있을 수 있음.
 - 예: 금지시 관계, 수식 관계, 주술 관계, 병렬/대조/덧구 관계 등
 - attention을 담당하는 head를 여러 개 마련하면, 이러한 다양한 종류의 관계를 포착할 수 있음.
 - 각 head의 동작은 병렬적으로 처리할 수 있음.

RNN(LSTM 포함)의 약점: 병렬화가 어려움

- 현대 기계학습(딥러닝 포함)은 대량의 훈련 데이터로 신경망을 학습시킴으로써 이루어짐.
- 훈련/학습에 엄청나게 많은 전산 자원, 전기, 시간이 소요됨.
- GPU를 이용하여 연산을 병렬적으로 처리함으로써 훈련에 소요되는 시간을 많이 단축할 수 있음.
 - 전산 자원과 전기의 소모는 여전히 엄청나게 큼.
- 그런데 RNN은 앞 요소의 정보를 뒤 요소로 step-by-step식으로 전달해야 하기 때문에 병렬화하여 처리하기가 어려움.
- CNN 등의 기타 신경망에 비해 훈련에 훨씬 더 많이 시간이 소요됨.

신경망 기반 언어모델의 변천3: Transformer

- RNN의 sequential한 정보 흐름을 아예 없애 버리고 attention 메커니즘만 사용.
 - Attention is all you need.
- encoder, decoder 각각 6개의 layer로 구성
- encoder의 1개 layer는 attention layer와 Feed Forward layer로 구성
 - attention으로 입력 연쇄 내 요소들 사이의 관계를 포착
- decoder의 1개 layer는 self-attention layer, cross-attention layer, FF layer로 구성
 - self-attention layer: 현재 위치 앞만 봄.
 - cross-attention layer: encoder의 마지막 layer가 내뿜은 각 요소 정보를 (차별적으로) 참고.
- <https://jalammr.github.io/illustrated-transformer/>



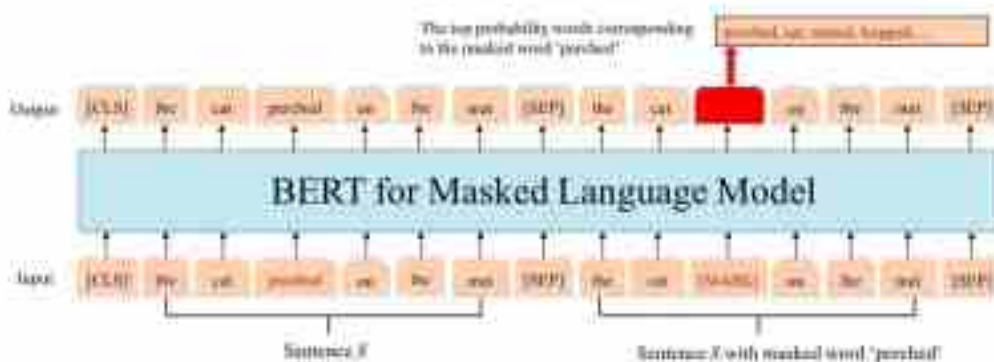
Transformer의 특징과 장점



- 기존 모델은 RNN을 주축으로 하고 거기에 attention을 부가적으로 첨가한 구조였던 데 비해
- Transformer는 조역이었던 attention을 주역으로 바꾸었으며, RNN은 아예 퇴출시킴.
- RNN을 완전히 없앴으로써 병렬화의 장애물 제거.
- self-attention layer에서 multi-head attention(8개)을 사용함으로써 요소들 간의 다양한 관계를 포착하고
- 이 8개 head의 작동은 병렬적으로 이루어짐.
- decoder의 self-attention layer에서는 input/encoder 정보와 output의 앞 요소들을 참고하여 그 다음 output 요소를 예측해야 하므로, 뒤 요소는 masking함.

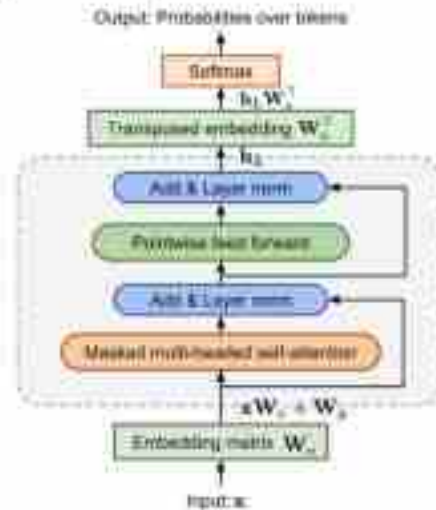
Transformer 기반 모델1: BERT

- Transformer에서 encoder 부분만 따옴. (NLU에 적합, 생성에 부적합.)
 - Bidirectional Encoder Representations from Transformers
- input sequence에서 15%의 요소를 random하게 정하여 masking한 뒤
- 이것을 예측하도록 신경망을 학습시킴.



Transformer 기반 모델2: GPT, GPT-2, GPT-3

- GPT: Generative Pre-Training
- Transformer에서 decoder 부분만 따옴.
 - 따라서 당연히 encoder-decoder cross-attention은 없음.
 - masked multi-head self-attention layer가 가장 중요한 역할을 함.
- 생성에 적합.
- NLU 과제에서는 BERT류 모델에 미치지 못함.
- GPT-2, GPT-3으로 갈수록 모델의 크기가 커져서 parameter의 수가 늘어나고 훈련 데이터의 양이 늘어났을 뿐, 본질적으로는 구조와 성격이 동일함.
- <https://jalanmar.github.io/illustrated-gpt2/>



한국어 GPT 모델

- GPT-2를 바탕으로 하여 SK-텔레콤에서 KoGPT-2 모델 공개
 - 현재는 공개/배포 중지
 - 유사 모델: <https://github.com/SKT-AI/XoBART>
- 이를 바탕으로 약간 수정한 모델을 여러 개인, 기관이 공개
 - <https://github.com/kajee/KoGPT>
 - <https://static.ksjit.com/KoGPT-v0.2.1/>
 - 소설로 fine-tuning하여 한국어 소설을 생성하는 모델도 있음.
 - <https://github.com/shbictai/narrativeKoGPT2>
 - 노래 가사로 fine-tuning한 모델
 - <https://github.com/gyunggyung/XoGPT2-FineTuning>
- 2021년 5월 네이버에서 초거대 한국어 모델(Hyper-Clova) 발표.

한국어 언어모델 만들 때 고려할 사항

- 영어는 띄어쓰기 단위 하나하나를 그대로 토큰으로 간주하면 별 문제가 없지만
- 한국어는 조사, 어미가 많아서 띄어쓰기 단위(어절)를 더 쪼개어야 함.
- 조사, 어미를 모두 별개의 토큰으로 분리하면, 서로 의미상 관계가 깊은 토큰들 사이의 거리가 꽤 멀어짐.
 - 아버지가 예전에 **일**하신 적이 있다고 하는 **직장**. 토큰 '일하'와 '직장'이 의미상 관계가 깊으나 9개 토큰(시, 니, 적, 이, 있, 다고, 하, 는)이 사이에 있음.
 - 맛있었다던 음식 : '맛있'과 '음식' 사이에 3개 토큰(었, 다, 던)이 개재함.
- 특별한 배려 없이 bigram이나 trigram 언어모델로 한국어를 모델링하려 하면, 긴밀한 관계에 있는 두 토큰이 조사, 어미로 인해 꽤 멀리 떨어져서 이들의 관계를 모델링하는 데 실패하게 될 수 있음.
- 각 어절의 어휘요소(실사)들 사이의 관계를 포착할 수 있도록 신경 쓸 필요가 있음.

한국어에 맞는 tokenizer

- 영어는 띄어쓰기 단위를 token으로 간주하면 거의 다 해결됨.
- 널리 쓰이고 있는 Transformer 라이브러리 속에 tokenizer가 포함되어 있음: BPE(byte-pair encoding), Word piece, Sentence piece
- 이 tokenizer는 language-agnostic함.
 - 어느 언어든 (그 언어에 대한 지식이 없어도) 사용 가능함.
 - token 내 글자 전이확률과 token간 글자 전이확률의 차이를 이용.
- 그러나 이 범용의 tokenizer는 한국어에 대한 지식이 없으므로, 한국어 문장을 tokenize할 때 (특히 자소 분리, 생략, 축약, 불규칙활용과 관련하여) 바람직하지 않은 결과를 낳음.
 - 간: 길+ㄴ vs. 긴 학교다: 학교+이+다 vs. 학교+다
 - 갔다: 가+았+다 vs. 갔+다 들은: 들+은 vs. 들+은
- 한국어 형태소분석기를 tokenizer로 사용하는 게 바람직.
- 단, pre-trained model을 가져다 쓸 때에는, 그 모델에서 사용한 tokenizer를 그대로 쓸 수밖에 없음.

의존소-지배소 관계의 모델링

- 통사적으로 직접 관계를 맺고 있는 요소들 사이의 관계를 그래프로 모델링하면, 의미구조를 상당히 충실히 포착할 수 있음.

•  The girl loved by many boys is my sister.

- 가난한 거지가 비싼 옷을 사기는 어렵다.
- 의존소(dependent)-지배소(governor) 관계를 조건부화율로 모델링할 때 어느 쪽을 given으로 볼 것인가?
 - 지배소를 given으로 보는 것이 일리가 있음. (반대 순서도 가능)
 - 한국어는 지배소가 뒤에 오므로, 대체로 문장의 뒤쪽으로부터 앞쪽으로 나아가는 식으로 모델링하게 됨.

다형태소 어절과 의존소-지배소 관계의 동시 고려

- 지배소 어절 '먹었다'가 있고, 그 앞에 의존소 어절 '밥을'이 있다고 치자.
- '밥'을 예측할 때 가장 중요한 단서/천제는 '먹'임.
 - 동사 '먹' 뒤에 붙는 어미는 목적어 예측 시 도움이 되는 경우도 있을 수 있으나, 영향력이 그리 높지 않음.
- 즉 P(밥|먹)을 계산하면 됨.
- 동사 지배소 '먹'과 그 의존소 '밥'이 주어졌을 때, '밥' 뒤에 붙는 조사를 예측할 때는 '밥'과 '먹' 둘 다 필요함. 즉 P(을|먹,밥)를 계산하면 됨.
- '맛있는 밥'의 경우 예측 순서: 지배소 명사 '밥' → 의존소 '맛있' → 어미 '는'. 즉 P(밥), P(맛있|밥), P(는|밥,맛있)의 순서로 계산하면 됨.
- 일반적으로, 지배소 어휘형태소 → 의존소 어휘형태소 → 의존소 문법형태소의 순서로 계산하면 됨.

하나의 어휘형태소 뒤에 복수의 문법형태소가 붙어 있을 때

- ‘**작** **았** **던** 구두’, ‘**밥** **만** **을** 먹었다’처럼 하나의 어휘형태소 뒤에 문법형태소가 둘 이상 붙어 있을 수 있음.
- 이런 경우에는 앞에서부터 순서대로 그 다음 요소의 출현 확률을 조건부확률로 모델링하면 됨.
- $P(\text{작}|\text{구두})$, $P(\text{았}|\text{구두,작})$, $P(\text{던}|\text{구두,작,았})$, $P(\text{EOW}|\text{구두,작,았,던})$
- $P(\text{밥}|\text{먹})$, $P(\text{만}|\text{먹,밥})$, $P(\text{을}|\text{먹,밥,만})$, $P(\text{EOW}|\text{먹,밥,만,을})$
- 통사구조는 지배소→의존소 순서라서 뒤→앞이 되지만
- 어절 내의 형태구조는 앞→뒤의 순서로 모델링하는 게 더 좋음.

복수의 의존소가 하나의 지배소에 의존할 때

- ‘**영희**가 **철수**를 **때**렸다’, ‘**철수**의 **작**은 **구두**’처럼 하나의 지배소에 둘 이상의 의존소가 달려 있을 수 있음.
- 지배소에 선조적으로 더 가까운 요소가 동시적으로도 더 가까우므로, 모델링 순서도 그에 따르면 됨.
- $P(\text{철수}|\text{때리})$, $P(\text{를}|\text{때리,철수})$, $P(\text{EOW}|\text{때리,철수,를})$
- $P(\text{영희}|\text{때리,철수,를,EOW})$, $P(\text{가}|\text{때리,철수,를,EOW,영희})$, $P(\text{EOW}|\text{때리,철수,를,EOW,영희,가})$
- 조건부확률의 조건 부분이 너무 장황해지므로, 여기서도 Markov assumption를 고려해 볼 수 있음.
- $P(\text{작}|\text{구두})$, $P(\text{은}|\text{구두,작})$, $P(\text{EOW}|\text{구두,작,은})$, $P(\text{철수}|\text{구두})$, $P(\text{의}|\text{구두,철수})$, $P(\text{EOW}|\text{구두,철수,의})$
- 처럼 의존소의 출현 확률을 고려할 때 그 지배소만 고려하고, 그 지배소에 딸린 다른 의존소는 고려하지 않을 수 있음.
 - 동사에 목적어가 결합한 후에 명사 의존소가 결합한다면 주어일 터. 향후 고려 필요.

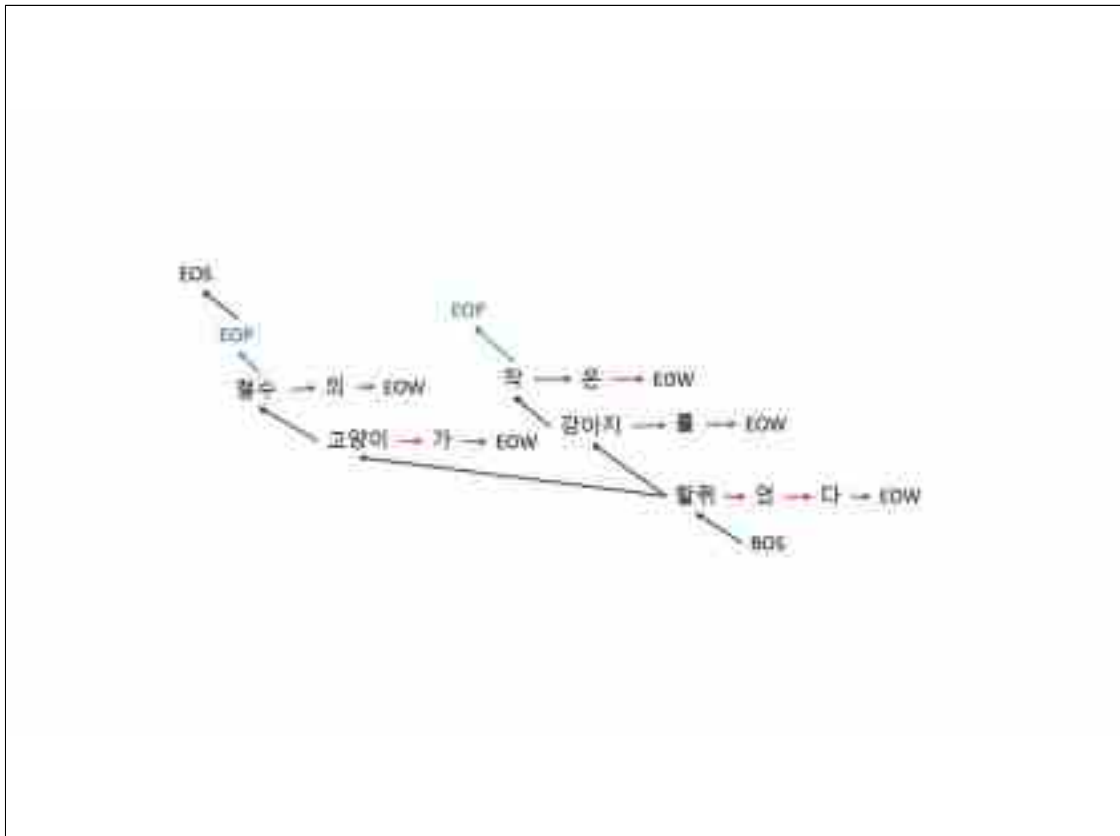
철수의 고양이가 작은 강아지를 핏귀였다.

- 조건화(conditionalization)의 순서
- BOS → 핏귀 → 있 → 다 → EOW
- | → 강아지 → 를 → EOW
- ↳ ↳ 작 → 은 → EOW
- ↳ ↳ 고양이 → 가 → EOW
- ↳ 철수 → 의 → EOW
- ↳ EOS

- 조건부확률 계산 순서
- P(핏귀|BOS), P(있|BOS, 핏귀), P(다|BOS, 핏귀, 있), P(EOW|BOS, 핏귀, 있, 다)
- P(강아지| 핏귀), P(를| 핏귀, 강아지), P(EOW| 핏귀, 강아지, 를)
- P(작| 강아지), P(은| 강아지, 작), P(EOW| 강아지, 작, 은)
- P(고양이| 핏귀), P(가| 핏귀, 고양이), P(EOW| 핏귀, 고양이, 가)
- P(철수| 고양이), P(의| 고양이, 철수), P(EOW| 고양이, 철수, 의)
- P(EOS| 철수)

어절 종결과 句 종결

- 어떤 어절 내에서 형태소가 하나하나 붙어서 어절이 확장되다가 그 다음 요소로 EOW가 선택되면 그 어절의 확장이 종결됨.
 - 어절 종결(word closure), EOW(end of word) 토큰
- 어절이 종결된 다음, 그 어절의 첫머리의 어휘형태소가 자신에게 의존하는 의존소를 가질 수도 있고 안 가질 수도 있음.
- 전자의 경우, 조건화에 따른 지배소-의존소 모델링이 계속됨.
- 후자의 경우, 해당 句의 더 이상의 확장이 일어나지 않고 종결됨.
- 이를 句 종결(phrase closure)이라고 부를 수 있으며, 조건화 모델링에 포함되어야 함. EOP(end of phrase) 토큰도 추가해야 함.



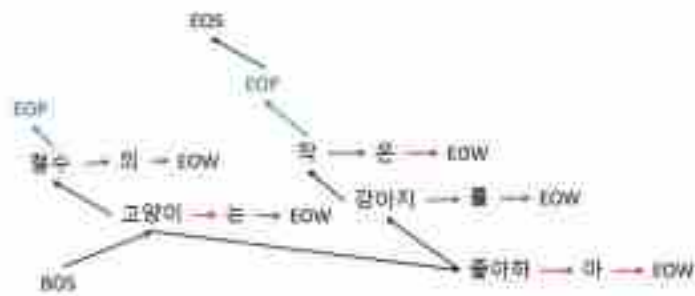
제언문과 정언문

- 정보구조(information structure)의 관점에서 문장은 다음의 두 유형으로 분류됨.
- 제언문(thetic sentence): 화제(topic)와 평언(comment)으로 구분되지 않는 문장.
 - 한국어에서 주어에 조사 '이/가'가 붙음. 문장 전체가 신정보임.
 - A: 무슨 일이야? B: **철수가 갑자기 쓰러졌다.**
- 정언문(categorical sentence): 화제와 평언으로 구분되는 문장.
 - 한국어에서 주어(또는 문두 성분)에 조사 '은/는'이 붙음.
 - A: 철수는 어떤 사람이야? B: **철수는 정직하고 성실한 사람이야.**
- 제언문은 지금까지 설명한 대로 모델링하면 됨.
- 정언문의 경우, 화제가 먼저 주어지고, 평언이 그에 대한 추가 정보로서 제시되므로, 조건부확률의 given을 화제로 보는 것이 실체에 더 부합함.

화제-평언 관계의 모델링

- BOS 다음에 화제句의 head 어절의 확장이 일어남.
 - 철수의 고양이는 작은 강아지를 좋아해.
 - $P(\text{고양이}|\text{BOS})$, $P(\text{는}|\text{BOS, 고양이})$, $P(\text{EOW}|\text{BOS, 철수, 의})$
- 이 head 어절에 의존소가 달려 있으면, 그句 확장은 앞의 설명대로 일어남.
 - $P(\text{철수}|\text{고양이})$, $P(\text{의}|\text{고양이, 철수})$, $P(\text{EOW}|\text{고양이, 철수, 의})$, $P(\text{EOP}|\text{고양이, 철수, 의, EOW})$
- 화제句가 종결되면, 화제句를 바탕으로 평언句의 head인 문말의 서술어에 대한 조건화가 일어남.
 - $P(\text{좋아하}|\text{고양이, 는})$
 - $P(\text{아}|\text{좋아하})$, $P(\text{EOW}|\text{좋아하, 아})$
- 문말 서술어에 딸린 의존소들의 모델링은 앞서의 설명과 같음.
- 단, BOS의 위치가 다름.

철수의 고양이는 작은 강아지를 좋아해.



맺는 말

- 인공지능 언어모델을 만드는 산업계, 학계의 연구자, 개발자들이 개별 언어의 특성과 언어학의 도메인 지식을 경시하는 풍조가 있음.
 - 그렇게 하고도 꽤 좋은 성과를 거두고 있기 때문에 자만에 빠진 측면도 있음.
 - 그러나 그렇게 해서도 성능 향상의 한계에 부딪히게 될 것.
- 이해를 위한 언어모델에서, 단어 벡터를 값싸게 얻은 뒤 이들 벡터를 결합하여 구, 문장의 벡터를 얻는 전략이 바람직함.
 - 더 경제적, 환경친화적이고, 언어학의 상식에도 부합.
- 생성을 위한 언어모델에서, 선조적 순서뿐 아니라 통사구조와 정보구조를 반영한 모델링 전략이 필요함.
- tokenizing algorithm, 조사와 어미의 처리 등에서 한국어의 특성에 맞는 처리 방식을 고안할 필요 있음.
- 자연언어처리에서 언어학자, 통사론 연구자의 역할을 적극적으로 모색할 필요 있음.

참고문헌

- 박진호(2020). 문장 벡터를 이용한 동형어 구분. <한국(조선)어교육연구> 16. pp. 7-48.
- Akyürek, Ekin, Afra Feyza Akyürek and Jacob Andreas (2021). Learning to recombine and resample data for compositional generalization. In: *ICLR 2021*. <https://arxiv.org/abs/2010.03706>
- Andreas, Jacob and Zoubin Ghahramani (2013). A generative model for vector space semantics. In: *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*. pp. 91-99.
- Andreas, Jacob and Dan Klein (2014). How much do word embeddings encode about syntax? In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. pp. 822-827.
- Andreas, Jacob et al (2016). Learning to compose neural networks for question answering. <https://arxiv.org/abs/1601.01705>
- Andreas, Jacob (2019). Measuring compositionality in representation learning. In: *ICLR 2019*. <https://arxiv.org/abs/1902.07181>

- Baroni, Marco and Roberto Zamparelli (2010). Nouns are Vectors, Adjectives are Matrices: Representing Adjective-Noun Constructions in Semantic Space. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. pp. 1183–1193.
- Blacoe, William and Mirella Lapata (2012). A comparison of vector-based representations for semantic composition. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. pp. 546–556.
- Chung, Woonjin, Sheng-Fu Wang and Samuel R. Bowman (2019). The lifted matrix-space model for semantic composition. <https://arxiv.org/abs/1711.03602>
- Coecke, Bob et al (2018). Uniqueness of composition in quantum theory and linguistics. In: *14th International Conference on Quantum Physics and Logic*. pp. 249–257.
- Dyer, Chris et al (2016). Recurrent neural network grammars. In: *Proceedings of NAACL 2016*. <https://arxiv.org/abs/1602.07776>

- Fyshe, Alona et al (2013). Documents and dependencies: An exploration of vector space models for semantic composition. In: *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*. pp. 84–93.
- Frandsen, Abraham and Rong Ge (2019). Understanding composition of word embeddings via tensor decomposition. <https://arxiv.org/abs/1902.00613>
- Grefenstette, Edward et al (2010). Concrete sentence spaces for compositional distributional models of meaning. In: *International Conference on Computational Semantics 2011*. <https://arxiv.org/abs/1101.0309>
- Grefenstette, Edward et al (2014). New directions in vector space models of meaning. In: *ACL 2014 Tutorial*.
- Heunen, Chris, Mehrnoosh Sadrzadeh and Edward Grefenstette (ed) (2013). *Quantum physics and linguistics: A compositional, diagrammatic discourse*. Oxford University Press.

- Kanerva, Pentti (2009). Hyperdimensional Computing: An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors. In: *Cognitive Computing*. pp. 139-159.
- Krishnamurthy, Jayant and Tom M. Mitchell (2013). Vector space semantic parsing: A framework for compositional vector space models. In: *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*. pp. 1-10.
- Lake, Brenden and Marco Baroni (2018). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In: *Proceedings of the 35th International Conference on Machine Learning*.
- Liu, Zhiyuan, Yankai Lin and Maosong Sun (2020). *Representation Learning for Natural Language Processing*. Springer.
- McCoy, R. Thomas et al (2019). RNNs implicitly implement tensor-product representations. In: *ICLR 2019*. <https://arxiv.org/abs/1812.08718>

- Mitchell, Jeff and Mirella Lapata (2010). Composition in distributional models of semantics. In: *Cognitive Science* 34. pp. 1388-1429.
- Mitchell, Jeffrey (2011). *Composition in distributional models of semantics*. University of Edinburgh Ph.D. dissertation.
- Plate, Tony (1991). Holographic reduced representations: Convolution algebra for compositional distributed representations. In: *Proceedings of the 12th international joint conference on Artificial intelligence (IJCAI'91, August 1991)*, Volume 1. pp. 30-35.
- Radford, Alec, Karthik Narasimhan, Tim Salimans, Ilya Sutskever (2018). Improving Language Understanding by Generative Pre-Training. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- Recchia, Gabriel et al (2010). Encoding Sequential Information in Vector Space Models of Semantics: Comparing Holographic Reduced Representation and Random Permutation. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 32.
- Rudolph, Sebastian and Eugenie Giesbrecht (2010). Compositional matrix-space models of language. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. pp. 907-916.
- Socher, Richard et al (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. pp. 151-161.

- Socher, Richard et al (2012). Semantic compositionality through recursive matrix-vector spaces. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. pp. 1201–1211.
- Socher, Richard et al (2013). Recursive Deep Models for Semantic Compositionality over a Sentiment treebank. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. pp. 1631–1642.
- Tai, Kai Sheng, Richard Socher and Christopher D. Manning (2015). Improved semantic representations from tree-structured Long Short-Term Memory networks. In: *ACL 2015*. <https://arxiv.org/abs/1503.00075>
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin (2017). Attention is All You Need. <https://arxiv.org/abs/1706.03762>
- Widdows, Dominic (2018). Semantic vector products: Some initial investigations. In: *Second Conference on Quantum Interaction*.
- Wittek, Peter and Sándor Darányi (2011). Spectral composition of semantic spaces. In: *5th International Symposium on Quantum Interaction*. pp. 60-70.
- Wu, Stephen (2011). Structured composition of semantic vectors. In: *Proceedings of the Ninth International Conference on Computational Semantics*.
- Zong, Chengqing, Rui Xia and Jiajun Zhang (2021). *Text data Mining*. Springer.

음운/형태/통사

Who left? Construction and EPP-Puzzle

Evan Chung (Yongnam University)

1. Introduction

(1) Who left?

(2) *Analysis 1* : Pesetsky & Torrego (2001)

$\{_{CP} \text{ who } \{_{C'} \{_{TP} \text{ t } \} \} \text{ left } \} \}$

(3) *Analysis 2* : Cheng & McCloskey (1985), Chomsky (1986), Aghayani (2000), Carstairs et al. (2016), Gallego (2017)

$\{_{CP} C \{_{TP} \text{ who } \{_{TP} \text{ t } \} \text{ left } \} \}$

(4) *Analysis 3* : Rizzi (1986), Rizzi & Cinque (1998), Campos (1997), McCloskey (2001), Holmberg & Hovav (2002), Fitzpatrick (2006), Boković (2016), Ilićević (2016)

$\{_{CP} \text{ who } \{_{C'} \{_{TP} \text{ t } \} \text{ left } \} \} \} \rightarrow$ **EPP-Puzzle**

2. Empirical Evidence for Analysis 1

I. Pesetsky (1987), Ginzburg & Sag (2000), Pesetsky & Torrego (2001) : *only D-linked wh-structures*

- (5) a. [What the hell] did you buy?
 b. *Who bought [what the hell]?
 c. [Who the hell] bought a car?

II. Aghayani (2000) : *Ellipsis phenomena after wh-movement*

(6) John bought something, but I don't know what. [John-bought-t] : TP-ellipsis after wh-movement

- (7) A: Someone bought a car.
 B: Who [t-bought-a-car]?

III. Boković (2016) : *Coordinate Structure Constraints*

(8) *Who left and John went to the store?

- (9) a. [_{CP} [who left] & [_{CP} John went to the store]] → wrong prediction
 b. [_{CP} who [_{CP} t left] & [_{CP} John went to the store]] → right prediction

(10) a. Who left and went to the store?
 b. [_S who [_S I_S t_S left] & [_S t_S went to the store]]

IV. McCloskey (1990), Foxpatrick (2006), Henry (2012) : **Placing Quantifier Phrases**

(11) West (Inner English (McCloskey 2000:77))
 a. Who was throwing stones **all** around Butcher's Gate?
 b. *They were throwing stones **all** around Butcher's Gate.

V. see Dikken & Griffiths (2010) : **Do-Deletion in British English**

(12) British English (see Dikken & Griffiths 2010)
 a. John might seem to enjoy that, and Peter might (do) seem to enjoy that too.
 b. I know who John will like and who Peter will (not) like t.

(13) British English (see Dikken & Griffiths 2010:22)
 A: Sue wouldn't like Peter, just right?
 B: Well, who would (+do) t, like him.

3. Analysis 2 & Anti-Locality Constraints

(14) Anti-Locality Constraint (Chomsky 2014, 2015, 2000) → **equation**
 Movement of a phrase from the Specifier of XP must cross a maximal projection other than XP.

(15) a. who left?
 b. [_S who [_S I_S t_S left] III]

4. Who left? & EPP-Puzzle

I. Labeling Algorithm & EPP-requirement 1 : **The Speed of Moving Element**

(16) Labeling by Minimal Search
 a. Merge (H, XP)
 b.

(17) Labeling with Maximal Projection
 a. Merge (XP, YP)
 b.

000 Chomsky (2013, 2015)

- labeling through movement
- labeling through feature-sharing

II Labeling Algorithm & EPP-requirement II: **The Case of *Who?***

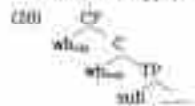


- *[EPP] operator
- *Who do you think that t_i will leave?

(18) EPP-Puzzle

- ↳ *who*_{CP} | *you*_{TP} ≠ *you*_{TP} | *who*_{CP} III

III. Contrastive Approach of EPP: Bošković (2001) → (I) **reduction of rich-ness conditions** (II) **reduction**



010 Construction with Auxiliary

- Which man's leaving first?
- *Which man's the founder of? (Kosme 1983)

020 *wh*-object > topic

- Mary wonders *which book*, for Jim, Peter should buy.
- *Mary wonders *which student*, for Jim, should buy that book. (Bošković 2001:12)

030 a. I wonder *what* Mary bought.

- I wonder *who* left.
- I think *Mary* left.

040 Hierarchy: *what*(*wh*-object) > *who*(*wh*-subject) > *Mary*(*wh*-topic)

IV. Null Expletive (Q) & EPP-Puzzle

050 *Pro* > null *Q*

- **Q*elle studente, crede-tu *che* *già* l, ve parte?
'Which student' do you believe that is going to leave?'
 - Q*elle studente, crede-tu *che* l, ve parte?
'Which student' do you believe *Q* is going to leave?' (Rizzi and Rizzi 2000: 136)

- (20) Quelle étudiante crois-tu **qu'elle** va partir?
 'Which student do you believe (th) is going to leave?'
 (21) **Quelle** étudiante crois-tu **qu'elle** va partir?
 'Who do you think that s. will leave?' (Oftst and Stosny 2007: 127)
- (22) **Explicite Agent Focus (AF)**
 a. Achise ***x-p-u-it** / **x-p-u-it** it why?
 who COM-3sg-A.3sg-acc / COM-3sg-acc-AP the tortile
 'Who ate the tortile?'
 b. Achise **x-p-u-it** / ***x-p-u-it** it a hunt?
 what COM-3sg-A.3sg-acc / COM-3sg-acc-AP hunt
 'What did hunt eat?' (Grewer 2000: 12)

- (23) a. [_{CP} Achise(wh) [C] [_{TP} **AP**] [T] [_{VP} Achise(wh) [_{VP} ritse] why(the tortile)]]]]
 b. [_{CP} Achise(wh) [C] [_{TP} a hunt] [T] [_{VP} x hunt [_{VP} ritse] Achise(wh)]]]]

V. Determinacy Principle & Null Expletive

- (24) Determinacy Principle (Chomsky 2010a: 270)
 If the structural conditions for a rule hold for some workspace, then the structural change must be unique.

EU who left?

- (25) a. [_{CP} who IT [_{TP} who [_{VP} BELIEVED]]] -
 b. *_{CP} who IC [_{CP} who IT [_{TP} who [_{VP} BELIEVED]]] - : Violation of Determinacy Principle
 (26) [_{CP} who IC [_{TP} **EXP**] [T] [_{VP} who [_{VP} BELIEVED]]] - : Observation of Determinacy Principle

A Conclusions & Further Studies

- The null expletive EXP can solve the EPP-problem in *wh*-subject construction.
- a. *Who there left? → Maybe, sentence (a) is not the unaccusative construction.
 b. Who EXP left? → It is curious that the null expletive EXP can exist in the accusative construction.

References

- Aghajani, Brian. 2009. *Wh-objects in English and the narrowest movement hypothesis*. *Linguistic Inquiry* 40, 769–773.
DOI: <https://doi.org/10.1017/S0000000009000477>
- Baltov, Zdeněk. 2011. On the status of labeling: Reducing wrap-free effects, the left–right condition, the adjunct condition and linking in Czech labeling. *The Linguistic Review Special issue on Labeling* 28, 17–61. DOI: <https://doi.org/10.1017/S0022268110000303>
- Baltov, Zdeněk. 2016. On the coordinate structure constraint, across-the-board movement, phases and labeling. Ms., University of Connecticut.
- Baltov, Z. 2021. *Merge, Move, and Continuity of Syntax: The Role of Labeling, Successor-Cyclicity, and EPP*. ms., University of Connecticut.
- Brath, Lidija & Martina Čukrić. 2005. Two Italian (dis)likes and the null subject parameter. In Corrado (ed.) & Rizzi (ed.), *The null subject parameter*, 111–142. Dordrecht: Kluwer.
DOI: <https://doi.org/10.1007/s10994-005-2049-5-4>
- Chomsky, Noam. 1997. On subject extraction and the anti-agreement effect in Romanian. *Linguistic Inquiry* 28, 10–111.
- Corbett, Vicki, Robert Harizan & T. David Swick. 2005. Head movement in *Problems of Projection: The Linguistic Review Special Issue on Labeling* 22(1), 47–66.
DOI: <https://doi.org/10.1017/S0022268105000304>
- Chomsky, Noam. 1980. *Lectures Cambridge, MA: MIT Press.*
- Chomsky, Noam. 2011. *Problems of Projection*. *Lingua* 120, 23–46.
- Chomsky, Noam. 2016a. *Some Functional Translational Issues: The Reading Program*. *Canadian Journal of Linguistics Special Issue* 261–296.
- Chomsky, Noam. 2016b. *Functional issues in Linguistics*. MIT Lectures, April 28th and 29th.
- Chomsky, Noam. 2020. *ICLA Lectures*, April 28th, 29th, May 1st, 2nd.
- Chomsky, Noam, Angel Gallego, and Isaac Ok. 2010. *Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges*. *Global Journal of Linguistics* 22–31.
- Chung, Sandra & James McCloskey. 1985. On the interpretation of certain island facts in DPSC. *Linguistic Inquiry* 16(4), 756–771.
doi:10.1017/S0022268105000304
- de Cat, Marcel & Annette Grewer. 2002. *From 'wh' to polarity: 'Agree(-)ingly' and 'd'-based' wh-structures in polarity focus*. *Linguistic Inquiry* 33(1), 31–52.
DOI: <https://doi.org/10.1017/S0000000002011982170>
- Edwards, Michael Yoshitaka. 2016. *Anti-locality and optimality in English Agent Focus*. *Natural Language & Linguistic Theory* 34(3), 429–479.
- Edwards, Michael Yoshitaka. 2018. *Extraction and lowering in Type B languages*. *Language* 94(3), 612–637.
- Edwards, Michael Yoshitaka. 2020. *Anti-locality and subject extraction*. *Glenn* 24(4), 1–38.
- de Cat, Marcel & James Gair. 2018. *Clitics and anti-local agreement*. Ms., *Respiration: Analysis of the Substratum and L2/1a*, Theoria.
- Fitzpatrick, Jeffry. 2005. *The structure and semantics of English question/wh-words*. Cambridge, MA: MIT dissertation.
- Gallagher, Aaron. 2007. *The EPP & Labeling Theory*. *PhD thesis from Boston, Syntax* 25(4), 394–229.
DOI: <https://doi.org/10.1111/j.1525-1518.2007.01111.x>
- Geach, Peter. 1962. *Reference and generality in natural language*. Cambridge, MA: MIT dissertation.
- Geach, Peter & Ian A. Lee. 2008. *Interpretive investigations: The MIT/CELI Publications*.
- Geach, Peter. 2011. *These edges, question that and the nature of (locum-) variables*. *Logic* 4(1), 23–28.
- Hendrick, Andrew & Thering Hirschbuhl. 2011. *Agreement and movement in Icelandic raising constructions*. *Lingua* 115, 967–1003. doi:10.1016/j.lingua.2010.09.004
- Isakov, Elen. 1981. *The syntax of auxiliary reduction in English*. *Language* 56(3), 103–112.
- Isakov, E. 2018. *Move-effect and Labeling Algorithm*. *Studies in Russian and American Language and Literature* 22, 127–148.
DOI: <https://doi.org/10.31273/stsl.2018.12.127>
- Lee, Ian A. & Xinyi Chen. 2018. *On the Asymmetry in the PP intervention Effects in English Constructions*. *New Britain Journal of English Language and Literature* 42(4), 285–291.
DOI: <https://doi.org/10.21515/nbe.42.4.285-291>
- McCloskey, James. 2000. *Quantifier float and Wh-movement in an Irish Dialect*. *Linguistic Inquiry* 31(3), 57–84.
DOI: <https://doi.org/10.1017/S0000000000000408>
- Minnikin, Tracy. 2003. *The derivation of subject subject questions and the nature of the EPP*. *Glenn: a Journal of general Linguistics* 31(1), 1–15. DOI: <https://doi.org/10.2234/gll.2003>
- Postal, Paul. 1967. *Wh-in-situ: Movement and operator binding*. In Yip, Toshiaki & Alan Lee (eds.), *The representation of Wh-structures*. MIT Press, Cambridge, MA: MIT Press.

- Penning, David & Esther Torregrossa. 2001. V41-C movement: Cause and consequence. In Michael Ramchand (ed.), *East Asia & the SL languages*, 35-49. Cambridge: MIT Press.
- Tsunenaka, K. T. 2001. Subject extraction, the (de)activation of ergativity and syntactic movement. In *Subject systems in discourse and the theory of universal grammar*, ed. A. Haidou & J. -Y. Jung, 103-142. New York: Oxford University Press.
- Wang, Guoq. 2002. *Issues in Chinese syntax*. Dordrecht: Kluwer.
DOI: <https://doi.org/10.1016/978-1-4020-0111-1>
- Wang, Guo & Yu Gao. 2001. Strategies of subject extraction. In Hans Martin Grewer & UE Grewer (eds.), *Structure + meaning + language?*, 115-141. Berlin: Mouton de Gruyter.

Interaction between Syntax and External System :Evidence from Fragmentary Questions in Korean

Haewon Jeon

(Pukyong National University)

1. Introduction

The aim of this paper is to reveal Fragmentary Questions (FQs) in Korean are characterized by prosodic realization, and to show that FQs are best analyzed in terms of movement of remnants to the edge of root sentences followed by clausal ellipsis. In support of the movement and ellipsis analysis for FQs, we examine connectivity effects such as Case and word order restriction that suggest the existence of an underlying full-fledged clausal structure. Furthermore, we show that a contrastive topic element must precede a *wh*-phrase in multiple FQ. Specifically, we argue that the landing site for a contrastive topic (Top-FocP) is higher than the landing site for a *wh*-phrase (FocP) in the CP periphery. FQs in Korean are sensitive to prosodic constituent structures, which yield ambiguity. Ahn and Cho (2012) present the ambiguity retained in FQs. They observe that *nwukwu* is ambiguous between indefinite interpretation ‘someone’ and *wh*-interpretation ‘who’.

- (1) A. Chelswu-ka nwukwu-lul manna-ss ni? (K)
 Chelswu-nom who-acc meet-past q
 ‘Who did Chelswu meet?’
 ‘Did Chelswu meet anyone?’
- B. Kulssey, kulem Yenghi-nun? (FQ)
 well then Yenghi-top
 ‘Well, then, who did Yenghi meet?’
 ‘Well, then did Yenghi meet anyone?’
- C: Kulssey, kulem Yenghi-nun manna-ss ni?
 well then Yenghi-top meet-past q
 ‘Well, then did Yenghi meet anyone?’
 * ‘Well, then, who did Yenghi meet?’

(Ahn and Cho, 2012)

Ahn and Cho (2012) observes that *nwukwu* is ambiguous between indefinite interpretation ‘someone’ and wh-interpretation ‘who’. For instance, (1A) is interpreted as either a Yes–No interrogative with indefinite object or a wh-interrogative question. The ambiguity is retained in FQ, as shown in (1B). However, the null object construction in (1C) is unambiguous, where the null part is only interpreted as an indefinite noun, not as a wh-phrase. The availability of a wh-phrase interpretation in (1B), which is absent in (1C), shows that FQ in Korean cannot be derived from argument ellipsis of a wh-phrase. In this paper, I present there exists different scope behavior, and if we treat prosodic properties as one that occurs to satisfy an interface need relating to contrastivity and its scope of FQs in Korean, I believe this will lead us to a better understanding of how language works.

2. Properties of Fragment Question (FQ) in Korean

The derivation of FQ in Korean can be assimilated to that of FA in English, involving focus movement and TP ellipsis. Research on fragmentary sentences, specifically Fragment Answers (FAs) for wh-questions, has yielded a number of important observations and consequences for the study of ellipsis in general. Specifically, much attention is paid to how to derive a fragmentary expression that nonetheless yields a full-fledged clausal interpretation. For instance, the FA uttered by the speaker B in (2) consists only of the focused phrase Alex that corresponds to the answer for the wh-phrase who in Speaker A’s utterance, yet the FA has the same interpretation as the full-fledged sentence Beth is bringing Alex.

- (2) A: Who is Beth bringing?
 B: Alex. (FA)

(Merchant 2004: 661)

There have been at least two approaches for FAs: the ellipsis approach and the direct interpretation approach. The ellipsis approach assumes that an apparent fragmentary expression is underlyingly a full-fledged clause and the ellipsis applies to the clause, except for the focused remnant. Specifically, Merchant (2001, 2004, 2006) suggests the movement and ellipsis approach, where FAs are derived by the movement of a remnant to the CP domain, followed by clausal deletion. (see also Morgan 1973, 1979; Park 2005; Ahn and Cho 2006; Nichigauchi 2006, 2010, 2011).

(3) A: Chelswu-nun Yenghuy-eykey mwues-ul cwess ni? (Korean: K)
 Chelswu-top Yenghuy-dat what-acc gave q
 ‘What did Chelswu give to Yenghuy?’

B: Wine ipnida.

wine cop

‘Wine.’

A: So-Yun-eykey-nun? (FQ)

So-Yun-dat-top

‘To So-Yun? (Intended: What did Chelswu give to So-Yun?)’

(Maeda and Jeon 2019: 174)

In (3)¹⁾, Speaker A’s second utterance only consists of So-Yun-eykey with the topic marker nun. Yet, they yield a wh-interrogative interpretation without the presence of an overt wh-phrase. Maeda (2019) presents that such topic-marked remnants are constructed as a contrastive topic of the interrogative clause, and such a fragmentary expression with a contrastive topic remnant is derived by deleting a full-fledged interrogative clause. She argues that the derivation of FQ involves movement of a remnant and clausal deletion. For explicitness, we suggest the structure of the left periphery for FQs in Japanese and Korean, as is schematized in (4).

(4) [_{Top-FocP} **remnant**_i Top-Foc [_{FocP} (wh) Foc [_{TP} ~~t_i (wh)~~]]]

(Maeda and Jeon 2019:174)

A configuration for FQ is illustrated in (4). Since Top-FocP projects above FocP in the left periphery, a remnant moves to a higher position than a (covert) wh-phrase. Maeda and Jeon (2019) following Merchant (2001, 2004) and Aelbrecht (2010), among others, assume that a head with [E]-feature licenses the ellipsis of its complement. The assumption that FQs are derived by clausal deletion is supported by the fact that FQ reflects the interpretation of a linguistic antecedent (cf. Li 2016); For instance, FQ (the second utterance of A) in (3) is interpreted as a wh-interrogative sentence, mirroring the wh-interrogative interpretation of its antecedent (the first utterance of A).

(5) a. Chelswu-nun So-Yun-eykey-nun mwues-ul cwess ni?

Chelswu-top So-Yun-dat-top what-acc gave q

b. PF: **So-Yun-eykey-nun** ~~mwues-ul Chelswu-nun cwess ni?~~

c. LF: So-Yun-eykey-nun mwues-ul Chelswu-nun cwess ni?

1) The Korean data given in (3) are actually based on the Japanese data examined by Maeda (2019).

However, the phenomenon of missing phrasal constituents (known as “ellipsis”), *mwues-ul Chelswu-nun cwess ni?* in (5b) needs to be speculated : If [E]-feature is employed to license the deletion of elided constituents, can it be deleted at PF or at narrow syntax (NS)? If so, why and how? To give a plausible solution, we should focus on how to treat elided part of FQ that is “unpronounced” at Phonetic Form (PF), but there is at Logical Form (LF) like (5b,c). Put differently, what are the licensing conditions on putative missing material? In the minimalist program, the output of the syntactic derivation is shipped to the semantic and phonological interfaces, namely LF and PF, referred to as the Conceptual-Intentional system (CI) and the Sensori-motor system (SM), by the process of Spell-Out (or 'Transfer'). This paper also aims to present the possibility that Transfer play a crucial role in ellipsis phenomena, specially in FQ in Korean.

Fragment Question (FQ) in Korean can be analyzed as FA in English (Merchant 2004) by assuming that FQ involves fragment movement and TP ellipsis.

(6) A: Chelswu-nun Yenghuy-eykey mwues-ul cwess ni?
 Chelswu-top Yenghuy-dat what-acc gave q
 ‘What did Chelswu give to Yenghuy?’

B: Wine ipnida.

wine cop

‘Wine.’

A: So-Yun-eykey-nun? (FQ)

So-Yun-dat-top

‘To So-Yun? (Intended: What did Chelswu give to So-Yun?)’

(7) A: Ken-wa Mari-ni nani-o ageta no? (Japanese)
 Ken-top Mari-dat what-acc gave q
 ‘What did Ken give to Mari?’

B: Hanataba desu.

bouquet cop

‘A bouquet.’

A: Yumi-ni-wa? (FQ)

Yumi-dat-top

‘To Yumi? (Intended: What did Ken give to Yumi?)’

(Maeda and Jeon 2019:174)

In (6) and (7), Speaker A’s second utterance only consists of *Yumi-ni/So-Yun-eykey* with the topic marker *wa/nun*. Yet, they yield a wh-interrogative interpretation without the presence of an overt wh-phrase. These examples show that such topic-marked remnants are construed as a contrastive topic.

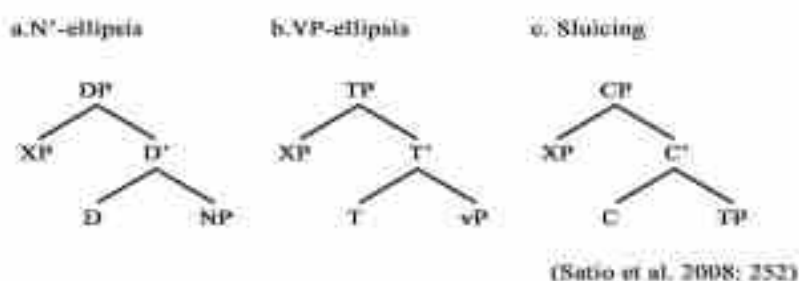
Given Lobeck (1990,1995)'s proposal that there exist common properties among N'-deletion, VP-deletion, and sluicing, they all involve the deletion of the functional head-complement. She further claims that these major cases of ellipsis bear the same structural requirement. In other words, these all involve functional heads (D,T,C), and the deletion of the complement is allowed only when the Spec position is filled. Merchant (2001)'s proposal deploying an ellipsis feature [E] that occurs on the licensing head and triggers deletion at PF of its complement implies that the licenser and the elided part are always in a head-completion relation. The examples are presented below in (8).

- (8) a. John can play five instruments, and Mary can play six [e].
- b. John can play the guitar and Mary can [e] , too.
- c. John can play something, but I don't know what [e].

(Merchant 2013:3)

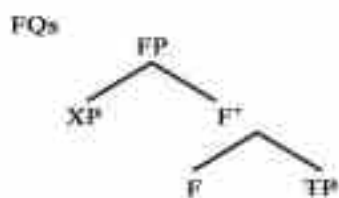
Each case of ellipsis in (8) are illustrated in (9).

(9)



FQs in Korean are depicted as below.

(10)



As sketched in (8-10), with this background in place, including FQs, various types of ellipses can be unified in one single deletion site. In accordance with Takita et al. (2016), den Dikken (2008), and Ko (2011), I propose that there must be a functional head H mediating the remnant (Constrastive Topic) and

elided material in FQs in Korean that can salvage remnants from labeling failure at NS. For expository convenience, let us revisit the generalization of Lobeck (1991, 1995), and present the structure of FQs in Korean in (12).

- (11) a. [vP V° [E] {VP...}] (VP-deletion)
 b. [TP T° [E] {vP—[VP...]}] (vP-deletion)
 c. [CP C° [E] [TP {vP—[VP...]}]] (TP-deletion)
- Lobeck (1991, 1995)

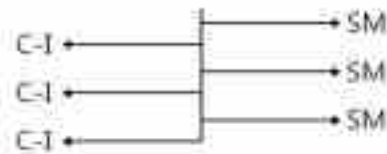
- (12) [FP F° [E] —[TP—[vP—[VP...]]]] (TP-deletion) : FQs in Korean

3. How to Transfer

In the recent Minimalism, the operations that hand SOs generated by syntax to SEM and PHON are called *Interpret*²⁾ and *Spell-Out*³⁾, respectively (cf. Chomsky 1995, Lasnik et al. 2005). In the recent Minimalist framework, the way the structure is mapped to PF and LF has become more refined thanks to the introduction of multiple Spell-Out (Uriagereka 1999, Chomsky 2000, 2001, among others).

Phase Theory (Chomsky 2000, 2001, 2004, 2007, 2008) posits that Spell-Out is cyclic, occurring more than once at structural points called 'phases'. More specifically, upon the completion of a phase, the complement of the head of the phase is removed from the syntactic derivation, and transferred to the interfaces. As widely adopted view of the syntax-interfaces is that narrow syntactic processes (Merge) precede Spell-Out and Transfer⁴⁾. In other words, once a particular domain (phases) is spelled out, it is no longer accessible to syntax.

- (13)



(adapted from Citko 2014)

- 2) Citko (2014) presents a term *Interpret-Out* that describes transfer to C-I interface.
 3) There are two kinds of Spell Out: Single Spell Out and Multiple Spell Out (Uriagereka, 1999). The former has only one point at which syntax could 'communicate' with phonology and semantics called Single Spell Out and the latter occurs more than once per derivation.



Citko (2014)

- 4) In current minimalist program, the notions Spell-Out and Transfer are no longer differentiated in a strict way.

Citko (2014) claims that Non-Simultaneous Multiple Transfer is well designed to handle a situation in which a given element is phonologically visible in one position but semantically visible in another position. Marušič and Zaucer (2006) and Marušič (2005,2007) deployed Non-Simultaneous spell-out as a derivational mechanism, and well explained two distinct yet very similar phenomena, Total reconstruction and quantifier raising. Non-simultaneous Spell-Out to the two interfaces has already been proposed in Megerdooian (2003), Felser (2004), and Wurmbbrand & Bobaljik (2003) is also hinted at in Sauerland and Elbourne (2002) and offered as a possibility.

Let us examine a conventional way of Transfer in Minimalist Program.

(14) $\{_{PH} XP, \{H, YP\}\}$ (Chomsky 2013, Chomsky et al 2019)

(15) a. What does John like?

b. $\{_{CP} \text{What}, \{C, \{_{TP} \text{John}, \{T, \{\langle \text{what} \rangle, \{\langle \text{John} \rangle, v\text{-like}\}\}\}\}\}\}$

If we do Transfer in a conventional way like (14), Complement of Phase head all eliminated, and there only remains $\{_{CP} \text{What}, \{C\}\}$.

Shim (2021) presents an alternative way as depicted below.

(16) WS = $\{_{PH} XP, \{H, YP\}\}$ (Shim 2021)

In previous section, I propose that there must be a functional head H mediating the remnant (Contrastive Topic) and elided material in FQs in Korean

(17) [remnant_i [FP F° [E] [TP t_i (wh)]]]

4. Puzzles : Sensitivity to Prosodic Constituency

In the Minimalist literature, many researchers have independently made numerous proposals that emphasize the role of the interaction between syntax and phonology and showed that such approaches indeed help us account for certain phenomena that have remained problematic for purely syntactic approaches. (Bobaljik 1994, 1995, Bošković 2001, Bošković and Lasnik 2003, Ćavar 1999, Embick and Noyer 1999, to appear, Fox and Pesetsky 2005, Franks 1998a, Franks and Bošković 2001, Halle and Marantz 1993, Harley and Noyer 1999, Ishihara 2003, 2004, Kandybowicz 2006, Kayne 1994, Merchant 1999, 2001, Nunes 1995, 2004, Pesetsky 1997, Selkirk 1995, 1997, Stjepanović 1999, among many others).

Ishihara (2002), for example, attempts to incorporate the phenomenon of the

wh-scope and prosody correspondence in Japanese into the more general patterns of focus phonology, arguing that the wh-intonation contour in Tokyo Japanese exhibits "Focus Intonation." Jung(2010) argues that for the phonological phrasing of a wh-phrase and its associated complementizer, the syntax-phonology interface is sensitive to syntactic features such as [+Q], [+QFoc] and [+Wh], which are derivational residues of the syntactic computation.

An(2007) argues based on a set of novel data from Korean and Japanese that RNR is sensitive to the prosodic constituent structure - in particular, the I-phrase structure - of a sentence.

- (18) Mary-nun wusumyense cilmwunha-nun haksayng-ul ttayri-ess-ta.
M-top with.a.smile ask.a.question-rel student-acc hit-past-dec
 (Kang and Speer 2005)

- (19) a. High reading
 'With a smile on her face, Mary hit the student who was asking a question.'
 b. Low reading
 'Mary hit the student who was asking a question with a smile on his face.'

What is interesting is that disambiguation of this sentence is done by prosody, so that the two readings in (18) correlate with the position of a pause surrounding the adjunct phrase. More concretely, if the adjunct phrase is parsed into the same I-phrase with the subject, the sentence only receives the high reading, as shown in (19a). In contrast, if the adjunct phrase is parsed into the same I-phrase with the object, the sentence only receives the low reading.

Ishihara(2002 et seq.), among others, argues that the domain of the intonation contour of wh-questions in Tokyo Japanese, which always exhibits Focus Intonation (FI), coincides with the scope of wh-questions, and that such FI-scope correspondence is a direct consequence of Multiple Spell-Out (MSO) (see also Kratzer and Selkirk 2007).

Jung(2010) argues the prosodic properties of wh-questions in Busan Korean are particularly interesting in that they reflect syntactic information quite faithfully but with differentiating devices of intonation contour. Based on the experimental investigation of the correlation of wh-scope and prosody in Busan Korean. Let us visit the data (1) here as (20).

- (20) A. Chelswu-ka nwukwu-lul manna-ss ni?
 Chelswu-nom who-acc meet-past q
 'Who did Chelswu meet?'
 'Did Chelswu meet anyone?'

- B. Kulssey, kulem Yenghi-nun? (FQ)
 well then Yenghi-top
 ‘Well, then, who did Yenghi meet?’
 ‘Well, then did Yenghi meet anyone?’

(Ahn and Cho, 2012)

In section 1, I presented there exists scope ambiguity in(19 B) explored by Ahn and Cho(2012). They show two different interpretation of FQ in Korean.

- (21) B. Kulssey, kulem Yenghi-nun? (FQ)
 well then Yenghi-top
 ‘Well, then, who did Yenghi meet?’
 ‘Well, then did Yenghi meet anyone?’
 ‘Din’t you meet Yenghi ?’

(21B) is interpreted as either yes-no interrogative or wh-interrogative. Interestingly, however, it can be interpreted as ‘Din’t you meet Yenghi ?’with lowering intonation. I present there exists one more meaning that is captured from FQ, *Yenghi-nun*. How the nature of the correlation of wh-scope and prosody can be captured in the current minimalist theory? The intonation of the sentence in question simply serves to signal which structure is being used. I believe this will lead us to a better understanding of how language works.

Selected References

- An, Duk-Ho. 2007. *Syntax at the PF interface: Prosodic mapping, linear order, and deletion*. Ph.D. dissertation University of Connecticut,
- Ahn, Hee-Don and Sungeun Cho. 2012. Fragments vs. null arguments in Korean. In *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar*, eds. Stephan Müller, Chungnam National University Daejeon, 369–387. Stanford, CA: CSLI Publications
- Aelbrecht, Lobke. 2010. *The syntactic licensing of ellipsis*. Amsterdam: John Benjamins Publishing.
- Chomsky, Noam. 2013. Problems of Projection, *Lingua* 130,33–49,
- Chomsky, N. 2015. Problems of Projection: Extensions. In E. Di Domenico, C.Hamann & S. Matteini (eds.). *Structures, strategies and beyond: Studies in honour of Adriana Belletti*. Amsterdam: Benjamins, 1–16.
- Masako, Maeda. 2019. Fragmentary Questions and the Left Periphery in Japanese, *proceedings of 2nd FLC & NGC Joint Conference Pukyong National University*.
- Maeda, Masako and Haewon Jeon. 2019. Fragmentary Questions in Japanese and Korean, *proceedings of GLOW-in-Asia XII & SICOGG 21, Dongguk University*.
- Haewon Jeon. 2021. Fragmentary Questions in Korean and Non simultaneous Transfer,

proceedings of Korean Language and Science, Busan National University.

- Ishihara, Shinichiro. 2003. *Intonation and interface conditions*. Ph.D. dissertation, MIT.
- Ishihara, Shinichiro. 2004. Prosody by phase: Evidence from focus intonation-wh-scope correspondence in Japanese. In *Working Papers of the SFB 632, Interdisciplinary Studies on Information Structure 1*, ed. by Shinichiro Ishihara, Michaela Schmitz and Ane Schwartz, 7-19. Potsdam: University of Potsdam.
- Jung, Yeun-Jin 2010, Syntax-Phonology Interface of Wh-questions, *Studies in Generative Grammar 20*, 1:549-576
- Shim, J.-Y. 2021 Transfer and Dynamic Access, *proceedings of Studies in Modern Grammar, Kyungpook National University.*

동사이동과 짝병합에 대한 소고

김대익

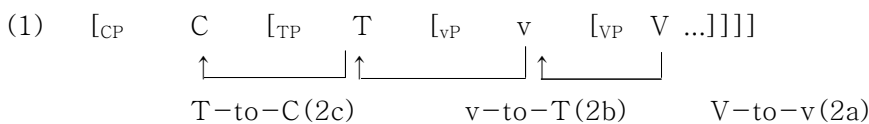
(영산대학교)

1. 서론

최소주의에서 동사이동(verb movement)은 언제나 논란이 되어왔다. 동사이동은 통사부에서 제공하는 여러 가지 제약들을 위반하여 Chomsky (2001)까지 통사부이동이 아니라 PF 이동이라고도 생각되었다. 그러나 최근에 들어와 동사이동 중에 V-to-v이동은 통사부이동으로 인정하지만 별다른 설명 없이, 병합(merge)의 측면에서 인접이동(adjunction)으로, 순서가 없는 일반병합이 아니라, 순서가 있는 짝병합(Pair Merge)이 되는 독특한 이동을 한다고 한다(Chomsky 2015, 2017). 하지만 이런 짝병합의 동사이동은 최근 최소주의(Chomsky 2017, 2019)의 7가지 격식(desiderata)의 측면에서 많은 문제들을 야기한다. 본 연구는 Chomsky (2017)의 7가지 격식측면에서 어떻게 해야 짝병합되는 V-to-v*이동이 통사부에서 적합한 이동이 되는지 설명하는데 그 목적이 있다.

2. 동사이동

2.1 동사이동의 예문들



- ▷ V-to-v* 이동 (2a)
- ▷ v*-to-T 이동 (2b)
- ▷ T-to-C 이동 (V2) (2c)

▷ 동사의 위치는 언어마다 다를 수 있지만 그 해석은 위치에 관계없이 동일하다. 그래서 동사이동은 구문의 해석에는 관여하지 않는다고 생각된다.

2.2 동사이동의 특성

- 동사와 접사(affix)와 의 결합으로 함께 이동을 한다.

□ 핵이동 제약(Head Movement Constraint: 이후 HMC)

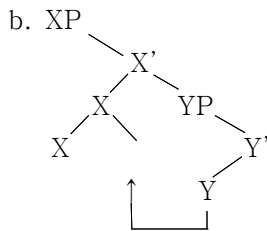
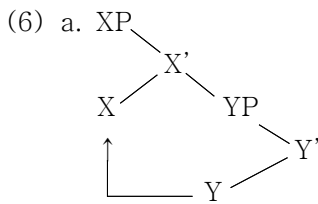
(4) Head Movement Constraint (Travis 1984: 131)

An X^0 may only move into the Y^0 which properly governs it.

2.3 동사이동의 병합 방법

(5) a. 대체 (substitution) 이동

b. 인접 (adjunction) 이동



□ 인접이동방법과 통사부에서 많은 제약들과 충돌.

▷ 확장조건 (extension condition)¹⁾을 위배 (anti-cyclicity).

▷ 성분통어 (c-command)를 위배함²⁾.

▷ 연쇄일률성조건 (chain uniformity condition)을 위배.³⁾

▷ A-over-A 규칙을 위배. A-over-A 원칙이란 A가 같은 성향의 A범주가 간혀있는 경우 ([A..[A..]]), 후자 A는 전자 A의 바깥으로 이동할 수 없다는 규칙이다. 즉 동사기 인접 이동하여 (6b)처럼 [X^0 [X^0 Y^0 ..]]로 된 경우 X^0+Y^0 같은 동사 핵인 X^0 를 벗어나서 이동할 수 없다.

3. 동사이동과 병합

3.1 통사부에서의 동사이동

□ 동사이동은 통사부는 이동

1) The Extension Condition (Chomsky 1993: 23)

GT and Move α extend K to K', which includes K as a proper part.

2) Proper Binding Condition (Fiengo 1977: 45)

In surface structure S, if [e] is not properly bound, then S is not grammatical

3) Chain Uniformity Condition (Chomsky 1995: 253)

A chain is uniform with regard to phrase structure status.

- (8) a. R(V)-to-v*이동
 b. [<R, v*> [t_R DP]]
 c. .. $\begin{array}{c} R \\ / \quad \backslash \\ R \quad v^* \end{array}$

3.2. 동사이동과 짝병합

▷ Chomsky(2013, 2015)의 일반병합.

- (10) Merge (a, b) = {a, b}
 (11) Pair-merge (a, b) = <a, b>

▷ Epstein, Kitahara, and Seely(2016)의 bridge-verb construction

- (13) What do you think that Bob read?
 a. [Bob read what]
 b. C [Bob read what]
 c. [_awhat C [Bob read what]
 d. [_β R [_awhat C [Bob read what]
 e. [v*_{uφ} [_β R [_awhat C [Bob read what]
 f. [<R, v*> [_β t_R [_awhat C [Bob read what]

- (14) External Pair-Merge of R to v* (Epstein, Kitahara, and Seely 2016)
 R and v* can be each taken directly from the lexicon and externally pair-merged together.

- (15) What do you think that Bob liked
 a. <R, v*>
 e. [<R, v*> [_awhat C [Bob like what]]
 c.. [Bob like what]
 b. C [Bob like what]
 c. [_awhat C [Bob like what]]
 d. [_β R [_awhat C [Bob like what]]]
 e. [v*_{uφ} [_β R [_awhat C [Bob like what]]]

3.3. Chomsky의 7 격식

- (17) Seven Desiderata
 a) Descriptive adequacy (DA)
 b) SMT(Strong Minimalist Thesis)

- c) Restricting Computational Resources (RCR)
- d) Determinacy
- e) Stability
- f) Recursion
- g) Strict Binary

(18) What do you think that Bob liked?

- a. $\langle R, v^* \rangle$
- b. $[_a \text{what } C \text{ [Bob like what]}]$
- c. $\langle R, v^* \rangle \text{ } [_a \text{what } C \text{ [Bob likes what]}]$

(19) $\{\{a,b\}, \{c, d\}\}$

(20) John likes Mary

- a. $[R, DP]$
- b. $[DP_\phi [R [t_{DP}]]]$
- c. $[v^*_{u\phi} [DP_\phi [R [t_{DP}]]]]$
- d. $[\langle R, v^* \rangle [DP_\phi [t_R [t_{DP}]]]]$

4. 짝병합과 자질공유

4.1 짝병합과 일반병합의 비교

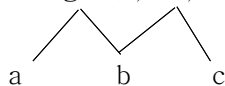
- (21) a. Set-Merge $(a, b) = \{a, b\}$
 b. Pair-Merge $(a, b) = \langle a, b \rangle$

(22) Pair-merge

“But it is an empirical fact that there is also an asymmetric operation of adjunction, which takes two objects β and a and forms the ordered pair $\langle a, \beta \rangle$ a adjoined to β (Chomsky 2004: 117–118)

(24)

- c. Pair-merge $(c, \{a, b\}) = \{a, b\}, \{b, c\} \rightarrow$



(25) Primary plane vs Separate plane (Chomsky 2004: 117–118)

“Given the basic properties of adjunction, we might intuitively think of a as attached b on a separate plane, with b retaining all the properties on the primary plane, “ the simple structure.”

$$(26) \dots \begin{array}{c} R \\ / \quad \backslash \\ R \quad v^* \end{array}$$

$$(27) \langle a, b \rangle \text{ -SIMPL-} \rightarrow \{a, b\}$$

4.2. 짝병합과 자질공유(feature sharing)

$$(28) \langle R, v^* \rangle \rightarrow \text{자질공유} \rightarrow [R-v^*]$$

▷ 이런 $\langle R, v^* \rangle$ 의 짝병합을 자질공유로 단일한 $[R-v^*]$ 로 변환된다고 하면 여러 가지 언어이론에서 좋은 장점들이 생긴다.

- ▶ 첫째는 병합상의 다지배구조를 회피할 수 있다. 짝병합이 되면 $\{a, b\}$, $\{b, c\}$ 와 같은 다지배구조가 되지만 자질공유를 통해 $\{c, \{b, a\}\}$ 와 같은 일반적인 병합이 된다.
- ▶ 둘째로 접사의 상태를 통사부에서 알 수 있다. 이와 관련하여 Chomsky(2004)의 기본 비행기와 기본비 행기와 분리된 비핵기(separate plane)의 개념을 제거할 수 있다. 즉 또 다른 동사와 접사의 이동위치를 알 수 있다. 짝병합으로 v^* 가 접사가 되어 통사부에서 보이지 않으면 어형부나 LF에서 접사가 어디에 위치하는지 알 수 없다. 그래서 통사부에서 접사가 보여야 한다. 또한 $[R-v^*]$ 에서 접사 v^* 가 통사부에서 보이므로 국면성을 계속 가지고 있으므로 Epstein, Kitahara and Seely(2016)가 제시한 다리 동사구문의 분석에서도 잘 설명할 수 있다.
- ▶ 셋째로 Chomsky(2004)의 문자화 과정에서 적용되는 짝병합을 일반 병합으로 바꾸는 단순화과정 (simplification operation)을 제거할 수 있다.
- ▶ 넷째로 Chomsky(2017)의 7가지 격식을 준수한다. $\langle R, v^* \rangle$ 가 자질공유로 단일한 어휘 $[R-v^*]$ 가 되면 $\{a, \{b, \{c, d\}\}\}$ 와 같은 병합이 되어 기술적 타당성이나 RCR 및 determinacy와 stability alc strict binary를 다 준수하는 건강한 병합이 된다.
- ▶ 다섯째 $[R-v^*]$ 가 단일한 SO가 되어 병합되므로 확장조건을 준수하며 a-over-a제약도 $[R-v^*]$ 가 하나의 단일한 요소가 되어 관련이 없게되므로 통사부제약들을 위배하지 않는다.

참고문헌

- 김대익. 2010. 국면과 문자화와 동사이동. 「현대문법연구」 60. 1-20.
- Baltin, M. 2002. Movement to the Higher V is Remnant Movement. *Linguistic Inquiry* 33, 653-669.
- Barnes, M. 1986. Faroese Syntax-achievements, Goals, and Problems. ms. University College London.
- Bobaljik, J. D. 2001. Rich agreement: On the Morphological Consequences of Syntactic Variation. ms. McGill University.
- Boeckx, C. 2011. approaching Parameters from Below. In a. Maria Di Sciullo & C. Boeckx (eds). *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, 205-21. Oxford: Oxford University Press.
- Boeckx, C and S. Stjepanović. 2001. Head-ing towards PF. *Linguistic Inquiry* 32, 345-355.

- Borer, H. 2014. The Category of Roots. In a. alexiadou, H. Borer, & F. Schllfer (eds). *The Syntax of Roots and the Roots of Syntax*. 112–148. Oxford: Oxford University Press.
- Brody, M. 2000. Mirror Principle. Syntactic Representation in Perfect Syntax. *Linguistic Inquiry* 31, 29–56
- Chomsky, N. 1993. Minimalist Program for Linguistic Theory. ms. MIT.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge. Cambridge, Ma: MIT Press.
- Chomsky, N. 2000. Minimalist Inquiries: The Framework, in: Roger. Martin et al, (eds). *Step by step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89–157. Cambridge, Ma: MIT Press.
- Chomsky, N. 2001. Derivation by Phrase, in Michael Kenstowicz, (ed). *Ken Hale: a Life in Language*, 1–53. Cambridge, Ma: MIT Press.
- Chomsky, N. 2004. Beyond Explanatory adequacy In Structures and Beyond: The Cartography of Syntactic Structures. 3. (Ed) adriana Belletti, 104–131. New York: Oxford University Press.
- Chomsky, N. 2005. On Phase. ms. MIT.
- Chomsky, N. 2013. Problems of Projection. *Lingua* 130, 33–49.
- Chomsky, N. 2015. Problems of Projection: Extensions. In E. Domentico, C. Hamann & S. Matteini (eds). *Structures, Strategies and Beyond: Studies in Honor of adrian Belletti*, 3–16. amsterdam: John Benjamins.
- Chomsky, N. 2017. A Talk Given at the University of Reading.
- Chomsky, N., Gallego, a., and Ott, D. (2017) Generative Grammar and the Faculty of Language Insights, Questions, and Challenges. ms. <Lingbuzz/003507>
- Epstein, S., Kitahara, H., and Daniel, S. (2016) Phase–Cancellation by pair–Merge of Heads. *The Linguistic Review* 33. 87–102.
- Goto, N. 2016. Labeling and Parametric Variation in Syntax. ms. Toyo University.
- Groat, E. and O’ Neil, J. 1996. Spell–out at the Interface: achieving a Unified Syntactic Computational System in the Minimalist Framework. In W. abraham, S, D. Epstein, H. Thrainsson & C. J–W Zwart (eds). *Minimal Ideas: Syntactic Studies in the Minimalist Framework*, 113–139. amsterdam: John Benjamins.
- Holmberg, a. and C. Platzak. 1989. The Role of agr and Finiteness in Germanic VO Languages. *Working Papers in Scandinavian Syntax* 43, 51–76.
- Holmberg, a. and C. Platzak. 1995. *The Role of Inflection in Scandinavian Syntax*. Oxford: Oxford University Press.
- Holmberg, a. and I. Roberts. 2014. Parameters and the Three Factors of Language Design. *Linguistic Variation in the Minimalist Framework*. 61–81.
- Hosono, M. 2018. Verb Movement in Narrow Syntax. ms. Keiyo University.
- Kosmeijer, W. 1986. The Status of the Finite Inflection in Icelandic and Swedish. *Working Papers in Scandinavian Syntax*, 26.
- Lasnik, H. 1995. a Note on Pseudo–Gapping. In R. Pensalfini & Hiroyuki Ura (eds). *Papers on Minimalist Syntax*. 143–163. MIT Working Papers in Linguistics 27. Department of Linguistics and Linguistics, MIT.
- Lecher, W. 2005. Interpretive Effects of Head Movement. ms. University of Tübingen.
- Lightfoot, D. and N. Hornstein. 1994. (eds). *Verb Movement*. New York: Cambridge University Press.
- Lockwood, W. B. 1964. *an Introduction to Modern Faroese*. Tórshavn: Føroya

Skólalabkagrunnur.

- Matushansky, O. 2006. Head Movement in Linguistic Theory. *Linguistic Inquiry* 37, 69–109.
- Nomura, M. 2019. Labelling and Pair–Merge of Heads. Ms. Nanzan University.
- Pollock, J. Y. 1989. Verb Movement, UG and the Structure of IP. *Linguistic Inquiry* 20, 365–424.
- Roberts, I. 1993. *Verbs and Diachronic Syntax*, Universite de Genève.
- Roberts, I. 2010. Head Movement and the Minimalist Program. ms. University of Cambridge.
- Rohrbacher, B. 1999. *Morphology–Driven Syntax: a theory of V to I Raising and Pro–Drop*. *Linguistik aktuell* 15. amsterdam: John Benjamins.
- Omune, J. 2019. a MERGE–based approach to Head adjunction. *Journal of Inquiry and Research* 109. 167–185.
- Oseki, Y. 2015. Eliminating Pair–Merge. Proceedings of the 32nd West Coast Conference on Formal Linguistics. somerville, Ma: Cascadilla Proceeding Project.
- Sigurðsson, H. 2011. Uniformity and Diversity: a Minimalist Prospective. *Linguistic Variation* 11, 189–222.
- Vikner, S. 1994. Finite Verb Movement in Scandinavian Embedded Clauses. In D. Lightfoot & N. Hornstein. (eds). *Verb Movement*, 117–148. Cambridge University Press.

한국 자폐스펙트럼장애 유아의 발화 운율 분석 연구

이수희 (University of Arizona, Linguistics), 정세영 (Columbia University,
Communication Sciences
and Disorders), 이영진 (University of Hawaii at Manoa, Linguistics), 이경숙 (한신
대 재활의학과),
정석진 (서울영유아돌보당센터), 조숙환 (서강대학교 영어영문학과)

2021년 12월 한국언어학회 제 47회 학술
대회

개요

- I. 연구 배경
- II. 자폐스펙트럼장애 (ASD) 유아의 발화 운율 특성
- III. 방법론
- IV. 결과 및 논의

I. 연구 배경

연구 목표

영유아 자폐스펙트럼장애의 운율 (prosody) 분석에서 나타난
조기 선별 지표 개발을 위한 예비적인 탐색

연구 배경

- 자폐성장장애의 핵심적인 특징은 "사회적 의사소통의 결함" (APA, 2013)
- 현재까지 영유아기 ASD 선별 위한 의사소통 관련 위험지표는 가리키기, 호명반응, 제스처 등 인데 특정 연령 (1세 이상) 되어야 나타나거나 AI 기술로 응용될 수 있는 정상화된 데이터로 측정하기 어려움
- 의사소통 지표 중 운율 (prosody) 은 생애 초기부터 보고되므로, ASD 선별에서 주요 지표로 될 수 있음

운율 (prosody) 의 정의

What is said (lexical content)

How it's said (prosody)

Prosodic Features of Speech

말의 운율 (prosody) 요소

- Pitch
- Stress
- Pauses
- Loudness
- Tempo

운율 발달

- 사람의 목소리와 운율적 요소는 아주 일찍부터 인지처리가 되기 시작한다

(Grosman et al., 2010)

- 연구에 의하면 출생 직후의 갓나아가(출생 후 2달)도 말의 운율과 소리 요소를 구분할 수 있다

(Perani et al., 2011)

- 대부분의 연구에 의하면 영아가 만 2세에 가까워지면 영아가 내는 운율이 경험과 글썽해지며 이는 폭발적인 단어 수 증가와 동시에 발생한다

(Vihman, Macken, Miller, Simmons, & Miller, 1985; cited in DePaolis et al., 2008).

II. 자폐스펙트럼장애 (ASD) 유아의 발화 운율 특성

ASD의 운율 (prosody) 결함

- “비정상적” 운율: 발성 및 운동 능력의 차이
- 구조적인 언어 결함
- 발화 내용 일관성에 문제

(Boucher, 2012; EIGHTI, Bennetto, & Nadlani, 2007; Kielgaard & Tager-Flusberg, 2001; McCabe, Hillier & Shapiro, 2013; Nadiq & Shaw, 2012).

ASD의 운율 (prosody) 결함

- 자폐 아동은 **과장되거나 단조로운 (monotonous) 운율로** 말을 하며, **말화의 속도나 억양의 경우** 다양한 양상을 보임
(Baron-Cohen and Staunton, 1994).
- 말의 운율은 ASD를 위한 치료 및 훈련 과정에서 흔히 배제되고 하나, 이와 같은 말화의 **운율적 특성이 ASD 인구의 사회적 수용의 가장 큰 장벽**이 될 수 있는 요소
- **운율적 결함**은 다른 언어적 영역과 달리 **선생애에 걸쳐** 이어지는 경우가 많음
(Shriberg et al. 2001)

ASD의 운율 (prosody) 중 Pitch (높낮이)

- **운율을 조정하는 기관의 결함:**
청각적 자극에 대한 인지 및 처리, 음운표상 인식 능력, 운동 기능 등
(Bonneh et al., 2010; Diehl & Paul, 2013)
- 각기 다른 높낮이의 발화를 재생산하지 못함
ASD 아동은 **높낮이를 활용하여 말하는 데에 어려움을 겪고** 과도하거나 잘못된 감세를 사용하기도 함
(Hargrove, 1997)

ASD의 화행 이해 결함

- ASD 아동은 정상발달 아동에 비해 의사전달을 위한 적절한 문장 형성에 어려움을 겪음. 화행에 대한 이해가 상대적으로 낮고 발음이 늦음
(Gran, Ürömi, Goldberg & Mimouni-Bloch, 2021)
- 상대방의 의도와 목적을 읽는 능력은 사회적으로 살아가는 데에 필수적
(Tomasello, Carpenter, Call, Behne & Moll, 2004)
- ASD 아동은 사회적으로 소통하는 것과 관련 표상을 인지하는 데에 결함이 있으며, 언어적인 소통을 통해 의도를 전달하는 것에 어려움을 겪음
(Wetherby & Prizant, 1990)

ASD 운용 및 문말어미 중요성

- 교육 및 임상 언어 평가의 중요성
 - ✓ 교육적 교정기 시작, 학업준비 시작, 직업적 적응을 위한 언어 시작
 - 특히 중요한 원동기 (pragmatic) 언어 중 화행기능을 포함하는 대표적 언어
- ✓ 40%: 개인적 언어로 표현하고 이해하는 능력과 부족하고, 60%정도의 문법적 언어를 사용. 의사소통의 문제 발생함
(Tager-Flusberg, 2000)
- 교정된 언어 형태보다 교정되지 않음 지남
 - 형태, 문법적, 문구적 정보 공유 가능
 - 문법적은 문법적 형태가 유지되는 중요한 언어 요소가 문법-의미적 요소
 - 문법-의미적으로 잘못 사용된 어휘의 사용, 문법적-의미적으로 사용된 어휘, 전문적인 상황 어휘

선행 연구 고찰

3. 언어 능력 통계를 통한 집단간 비교 중요
 - 언어 능력은 코를 표현에 영향을 미치므로, 언어 능력의 통계적 차이 집단과 발화 집단간 비교가 중요하다.
4. 국내 연구의 문제점
 - ① 국내 ASD 아동의 문법 특성에 대한 연구는 만 4-7세 아동의 읽기 과제를 통한 문법 연구가 일부 존재되나, 또 어휘 능력의 유아 대상 연구는 없음.
 - ② 언어능력을 측정하지 않았거나 언어능력을 통제하지 않은 일반 아동과 비교함
 - ③ 문법 특성 분석에 유용한 언어 수집방법만 자발에 연구는 없음
 - ④ ASD 아동의 진단과정의 확실성이 문제시됨.

III. 방법론

연구대상자 등록 및 언어분석 과정



분석할 언어 선정 과정



언어 분석 과정

1. 영상을 보며 아동의 행동과 언어를 전사한다. (1인당 1시간 30분~2시간 소요)
2. 전사 파일과 영상을 보며 아동의 발화 중 분석할 언어만 선별한다
3. (2)의 단계에서 추려진 발화인 wave file을 추출해 Praat에서 pitch 값(범위 75-500 Hz)을 확인한다
4. 발화 중 명서문의 경우 천유(계안)형(예: 초콜릿 먹자) 및 명명문(예: 이거 뭐), 화유형(예: “같이 해요” 제외하고 진술/기술 문장만 포함한다
5. 동일한 발화가 바로 연이어 나오는 경우 두 번째 발화는 제외한다.
6. Praat 프로그램에 업로드한다

언어 분석 단위

- 선행 연구들은 명서문에 명명, 계안, 화유 등 다양한 화형들을 함께 포함하여 분석.
- 하지만 본 연구에서는 명서문에 ‘진술/기술 statement’ 화형만을 포함하여 명서문의 화형기능을 단일화시켰다는 장점
- 한 가지 화형기능을 집중적으로 분석함으로써 여타 화형의 효과나 차이점을 배제하여 결과값을 얻는 장점

언어 분석 단위

- 명사문 364개, 의문문 148개 분할 어미의 pitch를 비교
- 의문문의 경우 "yes no 의문문"과 "wh 의문문" 구분하여 분석

yes-no 의문문

A : 누구 만나요?

B : 네. (친구를 만나려고요)

A : 뭐 먹어요?

B : 네. (점심을 먹으려고요)

wh 의문문

A : 누구 만나요?

B : 수미를 만나요.

A : 뭐 먹어요?

B : 과자 먹어요.

언어 분석 예시

▶ 분할 어미 분석: 항상 유어 사례

12139 (비밀리 - 었디 일기) 어음부는 "일기"

12139 (R) ▶ 00 ▶ 12139 (S*) ▶ 00 ▶ 12139 (S**)

- A와 a- 사이에 (s) 한 문장만 있으므로 a-는 분석하지 않음

- a-와 a* 사이에 (s) 한 문장만 있으므로 a*는 분석하지 않음

- a와 a* 사이에 두 문장 이후 오픈 경우 분석함

- "아름답게" 3회 나오는 것 2회 더 이상의 annotation이 없으면 분석을 않, 3회부 이후는 문장을 분석하지 않는 (다음 문)

12126-12129 "아름답게" "위리" "위리" "일기" "일기"

12140 "아름답게" "일기"는 분석

12151 "일기" "아름답게" <c> 문맥에서 register function의 용어문과 다른 경우는 제외 (s) 소문자

12153 (s-1) 문맥에서 제외

12142, 12156 "이거 뭐야" - 첫 번째만 올리임, 두 번째는 제외.

12152 "다시 해 다시 해" - 같은 의미 두 번 반복할 경우, 첫 번째만 분석

12123 "다시 해" - 문맥문리므로 제외

12154 - 첫 번째 "아름답게" 내외자*만 분석

연구 대상자

자폐 스펙트럼 장애를 가진 12~14세 아동 60명

성별	연령	자폐 스펙트럼 장애, 중등		자폐 스펙트럼 장애, 경증	
		IQ 범위 (M±SD)	ADOS-2 범위 (M±SD)	IQ 범위 (M±SD)	ADOS-2 범위 (M±SD)
남자 (n=48)	12~13	75.5(12.5)	12.5(3.5)	85.5(11.5)	10.5(2.5)
	14	85.5(11.5)	11.5(2.5)	95.5(10.5)	9.5(2.5)
여성 (n=12)	12~13	85.5(11.5)	11.5(2.5)	95.5(10.5)	9.5(2.5)
	14	95.5(10.5)	10.5(2.5)	105.5(9.5)	8.5(2.5)
자폐 스펙트럼 장애 (n=60)	12~13	80.5(11.5)	12.0(3.0)	90.5(11.0)	10.0(2.5)
	14	90.5(10.5)	11.0(2.5)	100.5(10.0)	9.0(2.5)
	12~13	80.5(11.5)	12.0(3.0)	90.5(11.0)	10.0(2.5)
	14	90.5(10.5)	11.0(2.5)	100.5(10.0)	9.0(2.5)
	12~13	80.5(11.5)	12.0(3.0)	90.5(11.0)	10.0(2.5)
	14	90.5(10.5)	11.0(2.5)	100.5(10.0)	9.0(2.5)
	12~13	80.5(11.5)	12.0(3.0)	90.5(11.0)	10.0(2.5)
	14	90.5(10.5)	11.0(2.5)	100.5(10.0)	9.0(2.5)
통제군 (n=12)	12~13	95.5(10.5)	8.5(2.5)	105.5(9.5)	7.5(2.5)
	14	105.5(9.5)	7.5(2.5)	115.5(8.5)	6.5(2.5)
자폐 스펙트럼 장애 (n=60)	12~13	80.5(11.5)	12.0(3.0)	90.5(11.0)	10.0(2.5)
	14	90.5(10.5)	11.0(2.5)	100.5(10.0)	9.0(2.5)

연구 대상자

- 자폐 집단과 정상 집단은 표현언어 수준에서 사전 통질성이 검증됨.
- 자폐 집단과 정상 집단 간에 표현 언어 능력에서 사전 통질성을 검증하기 위해 Mann-Whitney U-Test를 시행한 결과, 말인 초기 학습 가능도가 중 표현 언어 획득에서 통계적 으로 유의한 차이를 보이지 않았다 (각각 $z = -1.66$, $p = .04$, $z = -1.93$, $p = .05$).

자폐 스펙트럼 장애를 가진 12~14세 아동 60명

그룹	자폐 스펙트럼 장애		정상 집단		Mann-Whitney U	p
	말인 초기 학습 가능도	표현 언어	말인 초기 학습 가능도	표현 언어		
자폐 스펙트럼 장애 (n=60)	4.4	21.0	4.0	21.0	22.0	.60
정상 집단 (n=12)	4.0	21.0	4.0	21.0	4.0	.60

분석 문장 수

(단위: 언어 부진이나 지체된 아동)

번호	소아명	성별	진단명	유아 언어 지체 정도	유형 언어 (문장)	유형 문장 (문장)	문장 운용 (%)			총 문장 수
							유사문	개문	다문	
1	유지	남	ASD	21-24	10	10	0	0	0	0
2	유지	남	ASD		11	10	10	0	0	10
3	유지	남	ASD		10	10	10	0	0	10
4	유지	남	ASD	25-28	10	10	10	0	0	10
5	유지	남	ASD		10	10	10	0	0	10
6	유지	남	ASD		10	10	10	0	0	10
7	유지	남	ASD	31-34	10	10	10	0	0	10
8	유지	남	ASD		10	10	10	0	0	10
9	유지	남	ASD		10	10	10	0	0	10
10	유지	남	ASD	35-38	10	10	10	0	0	10
11	유지	남	ASD		10	10	10	0	0	10
12	유지	남	ASD		10	10	10	0	0	10
13	유지	남	ASD	39-42	10	10	10	0	0	10
14	유지	남	ASD		10	10	10	0	0	10
15	유지	남	ASD		10	10	10	0	0	10
16	유지	남	ASD	43-46	10	10	10	0	0	10
17	유지	남	ASD		10	10	10	0	0	10
18	유지	남	ASD		10	10	10	0	0	10
19	유지	남	ASD	47-50	10	10	10	0	0	10
20	유지	남	ASD		10	10	10	0	0	10
21	유지	남	ASD		10	10	10	0	0	10
22	유지	남	ASD	51-54	10	10	10	0	0	10
23	유지	남	ASD		10	10	10	0	0	10
24	유지	남	ASD		10	10	10	0	0	10
25	유지	남	ASD	55-58	10	10	10	0	0	10
26	유지	남	ASD		10	10	10	0	0	10
27	유지	남	ASD		10	10	10	0	0	10
28	유지	남	ASD	59-62	10	10	10	0	0	10
29	유지	남	ASD		10	10	10	0	0	10
30	유지	남	ASD		10	10	10	0	0	10
31	유지	남	ASD	63-66	10	10	10	0	0	10
32	유지	남	ASD		10	10	10	0	0	10
33	유지	남	ASD		10	10	10	0	0	10
34	유지	남	ASD	67-70	10	10	10	0	0	10
35	유지	남	ASD		10	10	10	0	0	10
36	유지	남	ASD		10	10	10	0	0	10
37	유지	남	ASD	71-74	10	10	10	0	0	10
38	유지	남	ASD		10	10	10	0	0	10
39	유지	남	ASD		10	10	10	0	0	10
40	유지	남	ASD	75-78	10	10	10	0	0	10
41	유지	남	ASD		10	10	10	0	0	10
42	유지	남	ASD		10	10	10	0	0	10
43	유지	남	ASD	79-82	10	10	10	0	0	10
44	유지	남	ASD		10	10	10	0	0	10
45	유지	남	ASD		10	10	10	0	0	10
46	유지	남	ASD	83-86	10	10	10	0	0	10
47	유지	남	ASD		10	10	10	0	0	10
48	유지	남	ASD		10	10	10	0	0	10
49	유지	남	ASD	87-90	10	10	10	0	0	10
50	유지	남	ASD		10	10	10	0	0	10
51	유지	남	ASD		10	10	10	0	0	10
52	유지	남	ASD	91-94	10	10	10	0	0	10
53	유지	남	ASD		10	10	10	0	0	10
54	유지	남	ASD		10	10	10	0	0	10
55	유지	남	ASD	95-98	10	10	10	0	0	10
56	유지	남	ASD		10	10	10	0	0	10
57	유지	남	ASD		10	10	10	0	0	10
58	유지	남	ASD	99-102	10	10	10	0	0	10
59	유지	남	ASD		10	10	10	0	0	10
60	유지	남	ASD		10	10	10	0	0	10

IV. 결과 및 논의

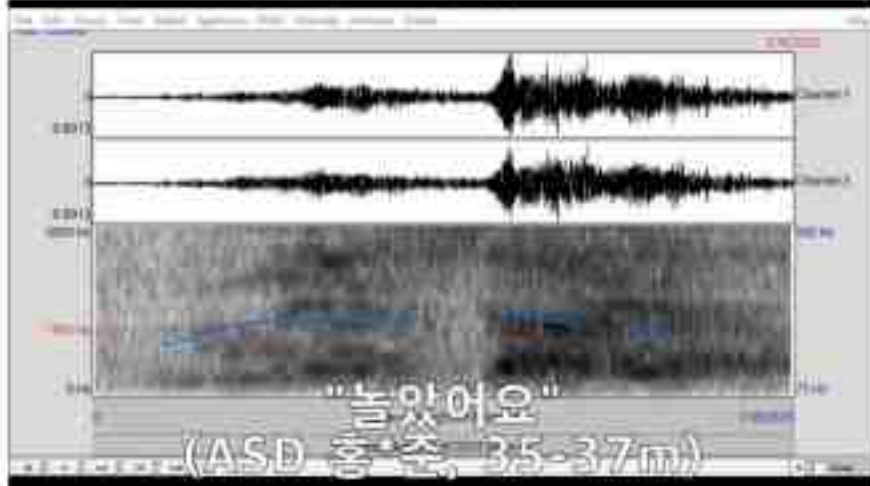
분석 결과: 평서문

- 장애 집단과 정상 집단의 평서문 용절별 용절 차이 검증
- ✓ 장애 유아는 3음절, 4음절, 5음절 평서문에서 음절이 위로 갈수록 pitch가 낮아지거나 오히려 더 높아지는 혼합적인 양상을 보임.
- ✓ 정상 유아는 평서문에서 pitch가 낮아지는 특성이 두드러짐
- (한국어 특성상 평서문은 하향 어조)

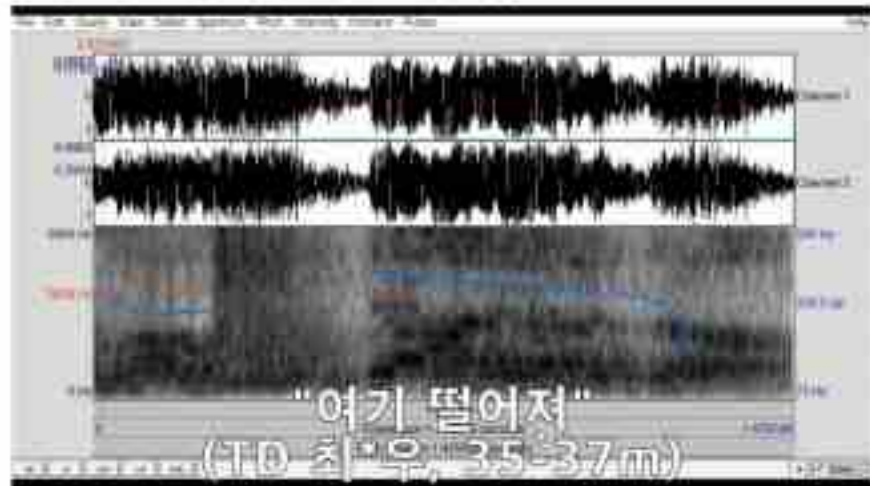
장애 집단과 정상 집단의 평서문 용절별 용절 차이 검증

집단	음절	음절	음절	음절	음절	음절	음절	음절	음절
장애	1	2	3	4	5	6	7	8	9
	10	11	12	13	14	15	16	17	18
	19	20	21	22	23	24	25	26	27
	28	29	30	31	32	33	34	35	36
	37	38	39	40	41	42	43	44	45
	46	47	48	49	50	51	52	53	54
	55	56	57	58	59	60	61	62	63
	64	65	66	67	68	69	70	71	72
	73	74	75	76	77	78	79	80	81
	82	83	84	85	86	87	88	89	90
	91	92	93	94	95	96	97	98	99
	100	101	102	103	104	105	106	107	108
	109	110	111	112	113	114	115	116	117
	118	119	120	121	122	123	124	125	126
	127	128	129	130	131	132	133	134	135
136	137	138	139	140	141	142	143	144	
145	146	147	148	149	150	151	152	153	
154	155	156	157	158	159	160	161	162	
163	164	165	166	167	168	169	170	171	
172	173	174	175	176	177	178	179	180	
181	182	183	184	185	186	187	188	189	
190	191	192	193	194	195	196	197	198	
199	200	201	202	203	204	205	206	207	
208	209	210	211	212	213	214	215	216	
217	218	219	220	221	222	223	224	225	
226	227	228	229	230	231	232	233	234	
235	236	237	238	239	240	241	242	243	
244	245	246	247	248	249	250	251	252	
253	254	255	256	257	258	259	260	261	
262	263	264	265	266	267	268	269	270	
271	272	273	274	275	276	277	278	279	
280	281	282	283	284	285	286	287	288	
289	290	291	292	293	294	295	296	297	
298	299	300	301	302	303	304	305	306	
307	308	309	310	311	312	313	314	315	
316	317	318	319	320	321	322	323	324	
325	326	327	328	329	330	331	332	333	
334	335	336	337	338	339	340	341	342	
343	344	345	346	347	348	349	350	351	
352	353	354	355	356	357	358	359	360	
361	362	363	364	365	366	367	368	369	
370	371	372	373	374	375	376	377	378	
379	380	381	382	383	384	385	386	387	
388	389	390	391	392	393	394	395	396	
397	398	399	400	401	402	403	404	405	
406	407	408	409	410	411	412	413	414	
415	416	417	418	419	420	421	422	423	
424	425	426	427	428	429	430	431	432	
433	434	435	436	437	438	439	440	441	
442	443	444	445	446	447	448	449	450	
451	452	453	454	455	456	457	458	459	
460	461	462	463	464	465	466	467	468	
469	470	471	472	473	474	475	476	477	
478	479	480	481	482	483	484	485	486	
487	488	489	490	491	492	493	494	495	
496	497	498	499	500	501	502	503	504	
505	506	507	508	509	510	511	512	513	
514	515	516	517	518	519	520	521	522	
523	524	525	526	527	528	529	530	531	
532	533	534	535	536	537	538	539	540	
541	542	543	544	545	546	547	548	549	
550	551	552	553	554	555	556	557	558	
559	560	561	562	563	564	565	566	567	
568	569	570	571	572	573	574	575	576	
577	578	579	580	581	582	583	584	585	
586	587	588	589	590	591	592	593	594	
595	596	597	598	599	600	601	602	603	
604	605	606	607	608	609	610	611	612	
613	614	615	616	617	618	619	620	621	
622	623	624	625	626	627	628	629	630	
631	632	633	634	635	636	637	638	639	
640	641	642	643	644	645	646	647	648	
649	650	651	652	653	654	655	656	657	
658	659	660	661	662	663	664	665	666	
667	668	669	670	671	672	673	674	675	
676	677	678	679	680	681	682	683	684	
685	686	687	688	689	690	691	692	693	
694	695	696	697	698	699	700	701	702	
703	704	705	706	707	708	709	710	711	
712	713	714	715	716	717	718	719	720	
721	722	723	724	725	726	727	728	729	
730	731	732	733	734	735	736	737	738	
739	740	741	742	743	744	745	746	747	
748	749	750	751	752	753	754	755	756	
757	758	759	760	761	762	763	764	765	
766	767	768	769	770	771	772	773	774	
775	776	777	778	779	780	781	782	783	
784	785	786	787	788	789	790	791	792	
793	794	795	796	797	798	799	800	801	
802	803	804	805	806	807	808	809	810	
811	812	813	814	815	816	817	818	819	
820	821	822	823	824	825	826	827	828	
829	830	831	832	833	834	835	836	837	
838	839	840	841	842	843	844	845	846	
847	848	849	850	851	852	853	854	855	
856	857	858	859	860	861	862	863	864	
865	866	867	868	869	870	871	872	873	
874	875	876	877	878	879	880	881	882	
883	884	885	886	887	888	889	890	891	
892	893	894	895	896	897	898	899	900	
901	902	903	904	905	906	907	908	909	
910	911	912	913	914	915	916	917	918	
919	920	921	922	923	924	925	926	927	
928	929	930	931	932	933	934	935	936	
937	938	939	940	941	942	943	944	945	
946	947	948	949	950	951	952	953	954	
955	956	957	958	959	960	961	962	963	
964	965	966	967	968	969	970	971	972	
973	974	975	976	977	978	979	980	981	
982	983	984	985	986	987	988	989	990	
991	992	993	994	995	996	997	998	999	
1000	1001	1002	1003	1004	1005	1006	1007	1008	
1009	1010	1011	1012	1013	1014	1015	1016	1017	
1018	1019	1020	1021	1022	1023	1024	1025	1026	
1027	1028	1029	1030	1031	1032	1033	1034	1035	
1036	1037	1038	1039	1040	1041	1042	1043	1044	
1045	1046	1047	1048	1049	1050	1051	1052	1053	
1054	1055	1056	1057	1058	1059	1060	1061	1062	
1063	1064	1065	1066	1067	1068	1069	1070	1071	
1072	1073	1074	1075	1076	1077	1078	1079	1080	
1081	1082	1083	1084	1085	1086	1087	1088	1089	
1090	1091	1092	1093	1094	1095	1096	1097	1098	
1099	1100	1101	1102	1103	1104	1105	1106	1107	
1108	1109	1110	1111	1112	1113	1114	1115	1116	
1117	1118	1119	1120	1121	1122	1123	1124	1125	
1126	1127	1128	1129	1130	1131	1132	1133	1134	
1135	1136	1137	1138	1139	1140	1141	1142	1143	
1144	1145	1146	1147	1148	1149	1150	1151	1152	
1153	1154	1155	1156	1157	1158	1159	1160	1161	
1162	1163	1164	1165	1166	1167	1168	1169	1170	
1171	1172	1173	1174	1175	1176	1177	1178	1179	
1180	1181	1182	1183	1184	1185	1186	1187	1188	
1189	1190	1191	1192	1193	1194	1195	1196	1197	
1198	1199	1200	1201	1202	1203	1204	1205	1206	
1207	1208	1209	1210	1211	1212	1			

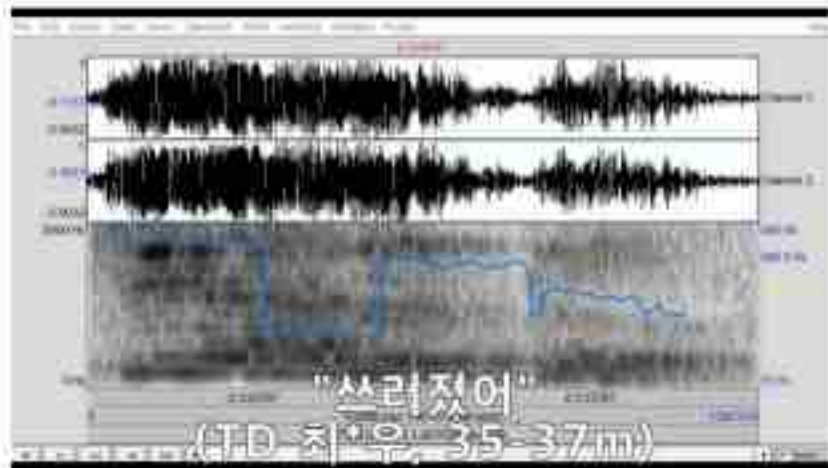
연구 결과 (음성 예시)



연구 결과 (음성 예시)



연구 결과 (음성 예시)



결과 및 논의

ADD유아(23-39개월)는 이른 시기부터 일반아동과 운용의 활용에 차이가 나는 경향이 있음.

- **광서훈**의 경우 일반아동이 문말어미로 갈수록 하향어조인 것에 비해 ADD 아동은 혼합적인 어조 양상을 보임.
하향어조와 상향어조가 혼합적으로 사용됨.
- **의문포**의 경우 일반아동이 문말어미로 갈수록 상향어조 (특히 yes-no의문문에서는 두드러짐), ADD아동은 혼합적인 어조 양상을 보임.
하향어조와 상향어조가 혼합적으로 사용됨.

결과 및 논의

- ▷ 일반 유아는 적중부터 의도를 갖고 의문문, 말서론으로 발화하기 시작하는 것에 비해 장애 유아는 의도성의 구분이 분명하게 드러나지 않음.
- ▷ 문술이 다른 의미를 변환시키거나 상상시키려고 할 때 사용되는 주요 의사소통 기능인데 ASD는 상황에 따른 적절한 화용론적 의사소통 기능의 부족으로 나타나는 것으로 해석 (Conger-Kirklin, 1998)
- ▷ ASD는 일반 아동보다 개인별 차이가 더 큼. (Ballas et al., 1999)

추후 분석 제언

- ASD 초기 증세이상성 치료 찾기를 위해 표집 수 확대를 통해 한국어 문장형태에 따른 문발어미의 특징 지표 재확인 필요
- 장애 유아집단과 일반 집단의 문장유형별 `gitok zagep` (문초문초문초) 비교
- 장애 유아집단과 일반 집단의 문장유형별 감세 비교
- 장애 유아집단과 일반 집단의 문장유형별 음성의 크기 비교
- 지체스펙트럼장애 선별을 위한 의사소통 관련 지표를 개발하기 위한 참고 지표로 사용

한국 자폐스펙트럼장애 유아의 발화 운율 분석 연구

이수희 (University of Arizona, Linguistics), 정세영 (Columbia University,
Communication Sciences

and Disorders), 이영진 (University of Hawaii at Manoa, Linguistics), 이경숙 (한신
대, 재활리학과),

정석진 (서울영유아발달센터), 조숙환 (서강대학교 영어영문학과)

2021년/21 한국언어과학회 하계 학술
symposium



일본어 /t/의 실현

(ex) /kat/ 'win' (Itō & Mester 2015)
Imperative: kat-e
Present: kats-u
Infinitive: katʃ-i

[t] before /e, o, a/
[ts] before /u/
[tʃ] before /i/

일본어 /t/의 실현: 기존 연구

- Crazy rule (Bach & Harms 1972)
 - [-sonorant, +coronal] → [+delayed release, +strident, **anterior**] / _
[V, +high, **aback**]
 - 부자연스러운 규칙들은 역사적인 음 변화들과 뒤따르는 일반화를 통해 만들어질 수 있다.
- Aerodynamic motivation (Kim, H. 2001; Kim, Y. 2019)
 - 고모음 앞에 위치하는 파열음의 개방 단계에서 발생하는 짧은 난기류가 청각적인 소음을 만들어 [+strident] 속성을 추가하게 된다.
- Articulatory motivation (Kim, Y. 2019)
 - 앞선 파열음의 개방 직후에 고모음을 조음하기 위해 혀와 입천장 사이에 좁은 틈을 형성하게 되므로 마찰 소음이 생긴다.

일본어 /t/의 실현: 조음적 이해의 문제

모음의 동시조음적 효과에 의한 /t/의 음성적 실현 변이

기대 효과

- 모음이 위에서 발음될수록 혀가 더 뒤로 움직이며 따라서 /t/도 더 뒤쪽에서 실현될 것이다.

실제 패턴

- 가장 위에서 발음되는 저모음 /w/ 앞에서 변화없이 /t/로 실현된다.
- 중모음 /e, o/ 앞에서도 변하지 않는다.
- 후설 고모음 /u/ 앞에서는 파열음의 조음 위치는 변하지 않는다.
- 전설 고모음 /i/ 앞에서는 구개음화 조음 위치가 뒤로 이동한다.

일본어 /t/의 실현: 쟁점

왜 고모음 앞에서만 /t/의 실현이 달라지는가?

왜 전설고모음 앞에서만 조음 위치가 뒤로 이동하는가?

- 혀는 통채로 연결된 근육 덩어리로 보이지만, 혀의 각 부분물 움직이는 근육들이 구분된다 (Gick et al. 2013).
- 동시조음되는 모음의 혀 운동 위치와 /t/가 실현되는 위치의 불일치를 이해하기 위해 '혀 움직임의 원리'를 이해할 필요가 있다.

본 연구의 구조



조음 정보: 3D 혀 모델을 이용한 근육 활성화 시뮬레이션



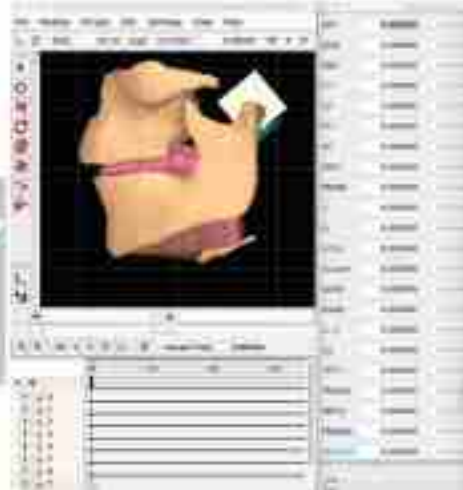
조음 정보와 음운 표상의 연결: 신경망 모델 학습



음운 문법 모델: 조화 문법 (Harmonic Grammar)

1. 조음 정보: 3D 혀 모델 근육 시뮬레이션

- 3D 모델링 플랫폼 Artisynth (Lloyd et al. 2012; artisynth.org)
- Static jaw-hyoid-tongue model
 - 턱-상악 모델(Hannam et al. 2008)과 3D 혀 모델(Vogt et al. 2006) 결합
 - 11개의 근육 활성화 따로 조절 가능 → 혀의 변형, 움직임 제어
 - 타임라인 패널: 혀 근육 활성화 지속 시간, 타이밍 조절



1. 조음 정보: /t, ts, tʃ/ 시뮬레이션

- MRI images (sail.usc.edu/span/rtmri_ipa/mg_2015.html)

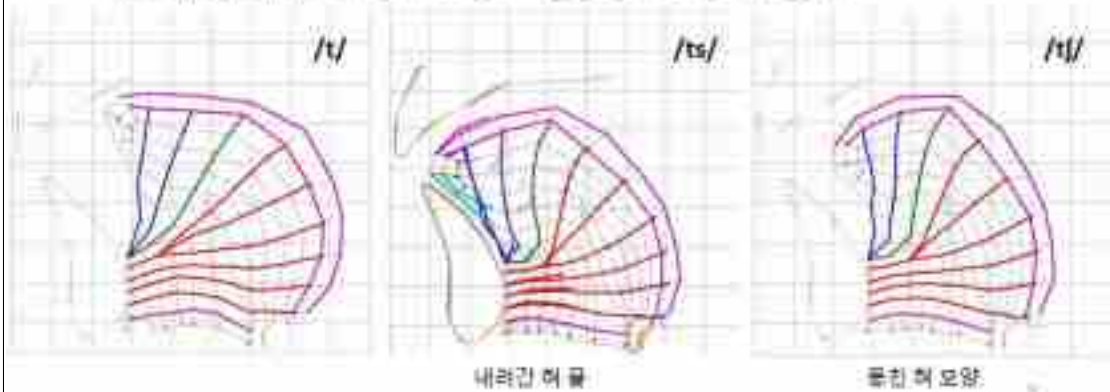


내려간 혀 끝

더 많이 내려간 혀 끝 → 뭉친 혀 모양

1. 조음 정보: /t, ts, tʃ/ 시뮬레이션 결과

- MRI image에서 드러난 혀 모양 특성을 살리는 근육 활성화



1. 조음 정보: /t/+V 시뮬레이션

- 동시 조음되는 모음이 /t/의 조음에 미치는 효과 확인
- /t/와 모음의 조음은 동시에 시작되고, 모음의 조음 지속 시간이 /t/보다 2배 길게 설정되었다. (Fowler 1980; Löfqvist & Gracco 1999; Goldstein et al. 2007; Nam 2007).



1. 조음 정보: /t/+V 시뮬레이션 결과

- 모음 환경에 따라 구분되는 혀 끝 내려간 정도 (i > u > a)



1. 조음 정보: 시뮬레이션 결과 해석

혀 끝 내려간 정도

- $i > u > a$
- 모음환경에 따른 $i > u > a$

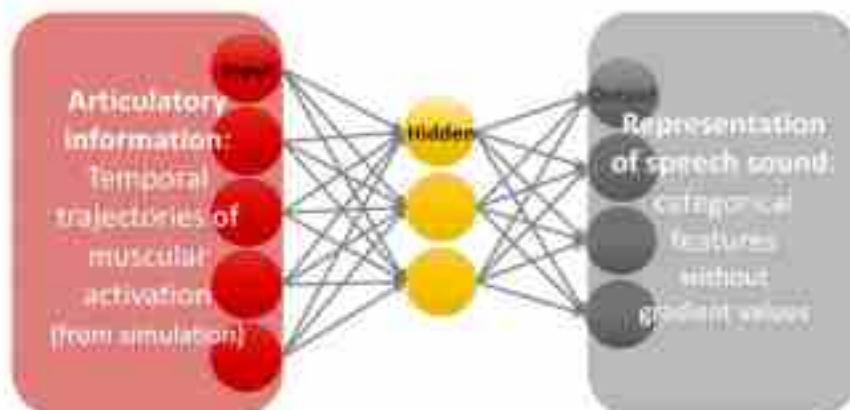
혀 끝 하강의 효과

- 자음 조음에 사용되는 혀 부위가 혀 끝에서 혓날-앞혀로 변화
- 입천장과의 접촉 면적이 넓어짐과 동시에 거리는 더 멀어지므로 조음 방법이 파열음에서 파찰음으로 변화
- 혀 끝이 더 많이 내려갈수록 혀의 가장 높은 위치 (조음 위치)가 좀 더 뒷쪽으로 이동

1. 조음 정보: 음운자질 표상

- [distributed] in Chomsky and Halle (1968)
 - 혀끝소리 (apical) = [-dist]
 - 혀날소리 (laminal) = [+dist]
- 본 연구의 [distributed]: 혀 끝의 방향
 - 혀 끝을 올려 조음되는 말소리들 = [-dist]
 - 자음 /r/
 - 혀 끝을 내려 조음되는 말소리들 = [+dist]
 - 자음 /ts, tʃ/
 - 모음 /i, u/ (& /e, o/)
 - 모음 /a/의 경우 [dist] 자질 비명세

2. 조음 정보와 음운 표상 연결: 신경망 모델



2. 조음 정보-음운 표상: 신경망 모델 학습

Training input

- 시간의 흐름에 따라 변화하는 근육 활성화값을 표현한 11x18 matrix
- individual segments: t, ts, tj, a, e, l, o, u
- CV sequences: ta, ts, tl, to, tt
- 이 때 모든 환경에 따라 t 조음을 완성하기 위해 근육값들이 포함되었다.

Training output

- input의 시간축 크기를 유지한 5x18 matrix
- Corresponding featural representation: strident, distributed, high, low, back
- +는 1, -는 -1 값을 갖도록 설정하고 underspecification은 0으로 설정하였다.
- CV sequences의 t는 근육값이 달라도 동일한 음성 자질로 표상되었다. [-strid, -dist]



2. 조음 정보-음운 표상: 신경망 모델 테스트

Test input

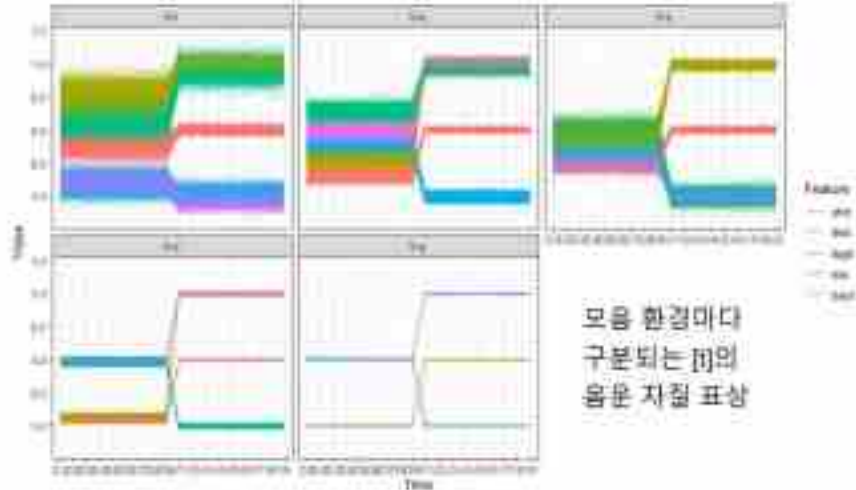
- Training input과 같은 형태의 matrix
- CV overlap: ts, ts, ts, ts, ts
- 이 때 t의 근육 활성화는 동일하게 유지되었다.
- 모든 환경의 phonetic context가 t의 음운 자질에 다르게 표상되는지 확인하고자 함

Test output: 예측

- 모든 환경마다 t의 음운 자질 표상에 미치는 영향이 다를 것이고, A의 실험과 유사한 패턴을 보일 것이다.
- [t] to +c [t]의 음운 자질인 [+strid, +dist, +high]에 가까운 표상값을 보일 것이다.
- [t] in +c [ts]의 음운 자질인 [+strid, -dist]에 가까운 표상값을 보일 것이다.
- [t] in ts, ts, ts: [t]의 음운 자질인 [-strid, -dist]에 가까운 표상값을 보일 것이다.



2. 조음 정보-음운 표상: Test output (1000 seeds)



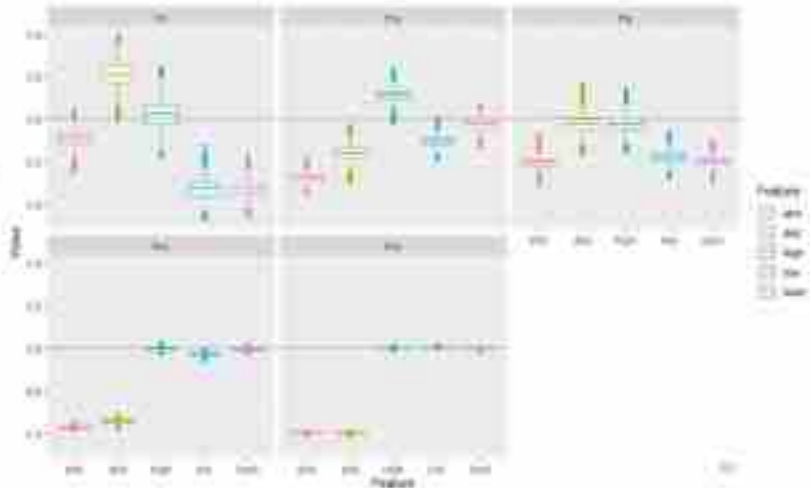
2. 조음 정보-음운 표상: Test output (1000 seeds)

• [~str], ~str

- 모음의 경우 [str] 자질이 억제되어 일치 없음
- 정도값의 차이는 없으나 모음 환경에 상관없이 모든 [~str] 유지

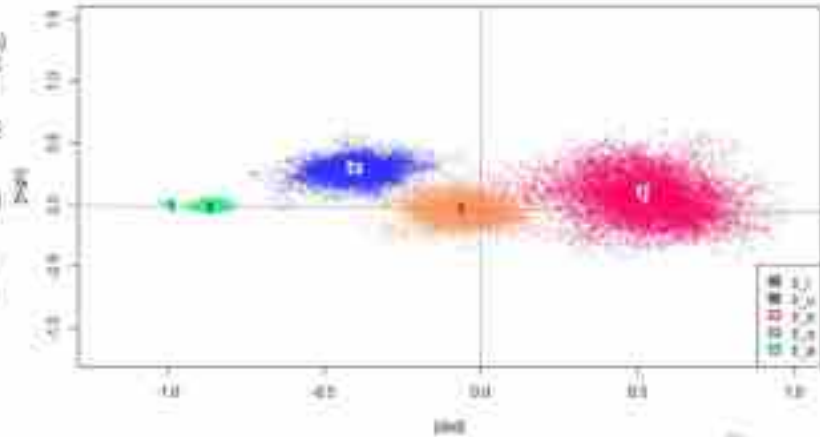
- /r/를 제외한 모음들은 [+dist] 자질 영세
- /r/ 환경에서만 [+dist]

- 고모음 /i, u/ [+high]
- /i, u/ 환경에서 [+high]



2. 조음 정보-음운 표상: Test output (1000 seeds)

- 제1형성 (vs. [t], [d]) (high)
- 이 글의 내리간 상태에서 일
전장과의 높은 원을 만들어
피장음을 초장하기 위해서
이 글들이 뒤로 올라가는 것
으로 이해할 수 있다.
- 피장음 치장 연쇄에 (+high)
후기
- t (-strut) -dist
- ts (+strut) -dist +high
- tʰ (+strut) +dist +high



3. 음운 문법 모델: 조화 문법

- Harmonic Grammar (HG: Legendre et al. 1990a, 1990b; Smolensky & Legendre 2006).
 - Optimality Theory (OT: Prince & Smolensky 1993/2004)과
동일하게 언어적 적합성을 설명하기 위해 위한 가능한 문법 제약들 사용
 - 제약 간의 순위를 매겨야 하는 OT와 달리 HG는 제약 조건에 수치적으로
가중치를 부여
 - HG에서 후보의 최적성은 각 후보에 대한 제약 조건 위반의 가중치
합으로 계산되는 Harmony 점수로 정의
 - Sum(각 제약의 위반값 x 해당 제약의 가중치)

3. 음운 문법 모델: 문법 제약

음운차별 연관에 대한 제약

- IDENT-IO[dist]: 가저형의 [dist] 자질 표상이 표면형에서 유지되어야 한다.
- IDENT-IO[cont]: 가저형의 [cont] 자질 표상이 표면형에서 유지되어야 한다.
- DEP-IO[high]: 표면형의 [high] 자질이 가저형에 포함되어 있어야 한다.

발소리 연관에 대한 제약

- *Tl: 전설고모음 // 앞에 설형음 //가 올 수 없다. (→ 구개음화)
- AGREE-CV: 연속하는 C와 V의 음운 자질(극성값±)이 일치해야 한다.

3. 음운 문법 모델: 일본어 문법 모델

- 파찰음화된 output candidate [ts]의 [dist]값을 신경망 모델에서 얻은 비정수 값으로 대체
 - 모음 환경의 동시조음효과 (허플 내림 정도) 반영
- 제약 가중치 (OT-Help 2.0 사용)
 - AGREE-CV 6.667 >> IDENT-IO[dist] 5.167 >> *Tl IDENT-IO[strd], DEP-IO[high] 1

Weights	6.667	5.167	1	1	1	Weighted total
Input: /t/ → /s/	AGREE-CV	IDENT-IO[dist]	*Tl	IDENT-IO[strd]	DEP-IO[high]	
→ a. [ts] [-strd1, -dist1]	0	0	0	0	0	0
b. [tsa] [+strd1, -dist1, +high1]	-2	0	0	-3	-1	18.333
c. [ts] [+strd1, +dist1, +high1]	-2	-3	0	-3	-1	-38.667

3. 음운 문법 모델: 일본어 문법 모델

Weights	6.567	5.167	1	1	1	Weighted total
Input: /t/ → /t/	AGREE-CV	IDENT-ID[dist]	*T1	IDENT-ID[stnd]	DEP-ID[high]	
a. [t] [-stnd1 -dist1]	-2	0	-1	0	0	-14.333
b. [ts] [+stnd1 -dist1 +high1]	-1	-1	0	-1	-1	-14.833
⇒ c. [tʰ] [+stnd1, +dist1, +high1]	0	-1	0	-1	-1	-13.333

Weights	6.567	5.167	1	1	1	Weighted total
Input: /t/ → /t/	AGREE-CV	IDENT-ID[dist]	*T1	IDENT-ID[stnd]	DEP-ID[high]	
a. [tʰ] [-stnd1 -dist1]	-2	0	0	0	0	-13.333
⇒ b. [ts] [+stnd1 -dist1 +high1]	-1.4	0	0	-1	-1	-12.333
c. [tʰ] [+stnd1, +dist1, +high1]	0	-1	0	-1	-1	-13.333

95

3. 음운 문법 모델: 일본어 문법 모델

Weights	6.567	5.167	1	1	1	Weighted total
Input: /t/ → /e/	AGREE-CV	IDENT-ID[dist]	*T1	IDENT-ID[stnd]	DEP-ID[high]	
⇒ a. [e] [-stnd1 -dist1]	-2	0	0	0	0	-13.333
b. [ts] [+stnd1 -dist1 +high1]	-3.1	0	0	-1	-1	-23.667
c. [ʰe] [+stnd1, +dist1, +high1]	-2	-1	0	-1	-1	-26.667

Weights	6.567	5.167	1	1	1	Weighted total
Input: /t/ → /u/	AGREE-CV	IDENT-ID[dist]	*T1	IDENT-ID[stnd]	DEP-ID[high]	
⇒ a. [u] [-stnd1 -dist1]	-2	0	0	0	0	-13.333
b. [ts] [+stnd1 -dist1 +high1]	-3.8	0	0	-1	-1	-29
c. [ʰu] [+stnd1, +dist1, +high1]	-2	-1	0	-1	-1	-26.667

96

한국언어과학회 2021 여름학술대회 발표자료

한국어 겹자음 실현과 운율 강화



김정윤 (jonnykim@pusan.ac.kr)

강은지 (ejkang@pusan.ac.kr)

음절말 겹자음 단순화

▶ 한국어의 음절구조: (C) V (C)

◀ 음보의 /C₁C₂/ → [C₁] 혹은 [C₂]

▶ 모음 앞에서: C₁만 실현되고 C₂는 제음절화

◀ /w/ > [w] > [palma]

◀ /m/ > [m] > [palma]

◀ /n/ > [n] > [palma]

▶ 자음 앞에서: 경우에 따라 다른 표준발음 규정

◀ /w/은 어휘에 따라 예외 허용

• [w] > [palma], [malma], [malma], ... → C₁만 발음

• 예외: [malma] → [malma] (C₂만 발음)

◀ /r/은 항상 C₁만 실현

• [malma], [malma], [malma] → 예외 없이 C₁만 발음

◀ /n/은 후행자음에 따라 예외 허용

• [malma], [malma], [malma], [malma] → C₂만 발음

• 예외: [malma] → [malma], [malma] → [malma] (C₁만 발음)

실제 음성형

➤ 유음 /ㄹ/로 시작하는 접미형·후형자를 앞에서 여러 실현형이 공존 (Kim-Rohwedder 1976, Kim 1987, Jun 1988, Sohn 1988, 김민호 2001, 김민호 외 2004, Cho & Kim 2008, 박지은 2011, 김민호 외 2014)

✓ 주의할 선택: 화자 간, 화자 내 차이가 모두 존재 (Cho & Kim 2008)

구어형	문어형	실제 음성형		
		[C1]	[C2]	[C3]
/lɥ/	원기	[원기]	[원기]	[원기]
/lɥ/	방사	[방기]	[방기]	[방기]
/lɥ/	상차	[상기]	[상기]	[상기]

➤ 하나의 자음으로 실현될 경우

✓ /m/은 C₁ 단독형만 가능

✓ /n/은 C₁ 단독형을 선호하는 반면, /w/은 C₁ 단독형을 선호 (Kim-Rohwedder 1976, Jun 1988)

✓ /m/과 /n/ 모두 C₁ 단독형 [m]을 선호 (Kim 1987, 김민호 2001, 김민호 외 2004, Cho & Kim 2008)

➤ 둘 다 실현 가능 (Kim-Rohwedder 1976, Kim 1987, Jun 1988, Sohn 1988, 김민호 외 2004, Cho & Kim 2008)

✓ 유음 C₁ 안개 C₂에 대한 종극단기(종말음 현이)가 발생할 수 있음 (Jun 1988) (예, [lɥ] → [lɥ]만 가능)

연구의 목적

➤ 변이형 선택의 양상

✓ /m/과 /n/이 단독형으로 실현시 C₁과 C₂의 실현 빈도

• [C₁] 형으로 변화 중립 가능성

✓ 두 자음 실현의 빈도

• [C₁C₂] 형으로 변화 중립 가능성

➤ 두 자음이 모두 실현되는 경우에 영향을 미치는 음운 및 음운 인식의 면밀한 분석

✓ 추가요인: 어휘 빈도, 발음자의 성별

두 자음 실현 정도에 대한 예측

- 음운환경(C₁의 종류)에 따른 두 자음 실현 정도: /tʰ/ > /t/ > /tʰ/
 - ↳ 앞순음 C₁를 가진 /tʰ/ > /t/ > 연구개 C₁를 가진 /t/ (Chang 2003, Cho & Kim 2009)
 - 어금(/r/)의 앞순음(/s, ɰ/)의 독립된 조음동작 vs 어금(/r/)내 어금(/s/)의 독립된 연속동작
 - ↳ 앞순음 끝에서도 /tʰ/ > /t/
 - /t/과 /tʰ/의 음향도의 차이 (Fujita & Mitsuhashi 1987)가 적은 자음군은 음향 내 자음과 모음의 경계가 모호
 - 두 자음 모두 코딩되기 때문에 시공제 오류(Bernard et al. 1995, Chittaman 1995, Ohn 1999, Ohn 2000, Ohn 2002, Wang and Berlin 2005)
- 운율환경에 따른 두 자음 실현 정도
 - ↳ 억양구 내 위치(prosodic position)의 영향: 억양구 초 강화 (Cho & Keating 2001, Ferguson 2001, Cho et al. 2011b)
 - 일반적인 조음 강화원리에 의해 기대됨에 강화 → 두 자음 실현 정도 증가
 - 후음의 기능(contagious function)에 의해 반대로 음절구조 제약을 준수 → 두 자음 실현 정도 감소
 - ↳ 강조(prominence)의 영향: 억락 상 변별이 필요한 음운에 대조초점(contrastive focus) (Cho et al. 2011a)
 - 어간말 단리음과 다른 음운환경에 의한 대조(contrast)에 의해 억락에 억락자의 겹자음을 강화 → 두 자음 실현 정도 증가
 - 실험적 선택의 순의성을 고려하면 억락을 형성 할수록 억락화 가능성도 존재 → 두 자음 실현 정도 감소

실험 참가자

- 서울말의 억양 (Lee 2011)을 자연스럽게 구사하는 부산대학교 재학생 11명
 - ↳ 만 20-27세
 - ↳ 남도 여 6

실험단어

➤ 어간말에 유음술 C,으로 하는 자음군이 포함된 16개의 용어를 실험단어로 선정

- ✓ C₁ = 8개의 연구자들: 말-뽀-밭-뽕-뽕-뽕-뽕-뽕
- ✓ C₂ = 8개의 양순음
 - 양순 비성음: 뽕-뽕-뽕-뽕-뽕
 - 양순 비음: 뽕-뽕-뽕-뽕-뽕

• 어미: -자
• C₁, C₂ 지의 세 자음 연속 환경

실험 문장 세트

➤ 각 실험단어는 네 가지 구어체 문장 안에서 등장

- ✓ 참가자가 빈칸을 채워 말하는 형식

역양구 1	역양구 2	위치	음운환경	선택지 1	선택지 2
___이 일한 시험하고,	이 일한데 뽕 의 일해서 뽕이	역양구 1	•	계곡물	계곡물
계곡물이 담긴 시험하고	다. 용기에 ___ 일해서 뽕이	역양구 2	○	밭지	밭지
___이 일한 시험하고, 다. 용기에	뽕 의 일해서 뽕이	역양구 3	•	계곡물	계곡물
계곡물이 담긴 시험하고, 다. 용기에	___ 일해서 뽕이	역양구 4	○	밭지	밭지

➤ 두 조건을 가지는 두 가지 실험변인의 조합(2×2)

- ✓ 역양구 내 위치: 역양구 ① ② vs. 역양구 ③ ④
 - 음표가 없는 곳에서 반드시 음운 위치를 두어 읽도록 지시
- ✓ 음운경쟁을 일으키는 어휘: 있음 ② ④ vs. 없음 ① ③
 - 두 선택지 중 한개 보 자음스러운 단어 삽입해 읽도록 지시

➤ 역양구 2의 음절수 통제

- ✓ 실험단어 앞 = 4음절 (11 음절까지)
- ✓ 실험단어 뒤 = 8음절 (19 음절까지)

필러 항목(filler items)

- ▶ 겹자음 단어들로부터 주의를 분산
 - 빈칸 채우기 과제와 자연스러운 역할에 집중하도록

선제문 1	제정문 2	선제문 1	선제문 2
① _____의 일한 사람이고,	이 문장에 적절하게 잘라서 보아야.	시종	년월
② 내용이 일단 서중이고,	다. 용문대 _____ 있어서 보아야.	계급까지	계급까지
③ _____의 일한 사람이고 다. 용문대,	계급까지 있어서 보아야.	년월	시종
④ 내용이 일단 서중하고 다. 용문대,	_____ 있어서 보아야.	계급까지	계급까지

- ▶ 실험문장 세트마다 꼭지어진 네 개의 문장
 - 실험단어 받(미)과 받(미) 대신 유리어를 이용한 받(미)에서 받(미)까지를 사용
 - 겹자음이 없는 받(미)과 받(미)도 유리어를 이용한 받(미)과 받(미)으로 교체

두 자음 실현 분석 방법

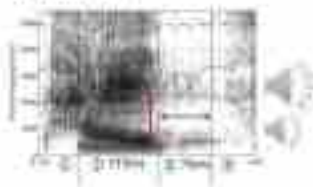
- ▶ 두 자음 함수 두 자음 실현에 대한 음향단서의 총합을 기준으로 네 단계 점수화

- 0: [미]와 [미]의 양방향으로 다른 음절로 음절
- 1: [미]만 음절
- 2: [미]와 함께 C₂가 실락 음절
- 3: [미]와 C₂가 분명히 음절

▶ 주요 음향단서

- 음절 구간 내 제 1, 2, 3) 포인트
- 최대음계(stop gap)의 상대 길이

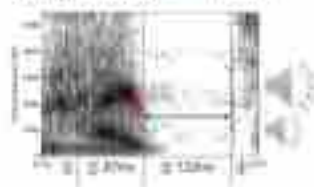
• [미]만 음절



• [mi]의 실현(수리)으로 정보 포함



• [mi]의 실현(수리)으로 수리적 점수화



분석 대상과 인원

- ▶ 필리 항목 제외, 총 1,408개의 녹음자료(16 단어 × 4 실험조건 × 2회 반복 × 11명)
 - 그중 8건 제외: 녹음상태 불량, 발화 실패
 - 최종 분석대상: 총 1,400개의 녹음자료
- ▶ 두 개의 분석조(각 3명)가 700개씩 나눠 두 자음 정수를 각자 측정
 - 제 1저자: 음성용문 전문 대학원생, 언어학 관련 전공 학부생
 - 제 2저자: 음성용문 전문 대학원생, 언어학 관련 전공 학부생

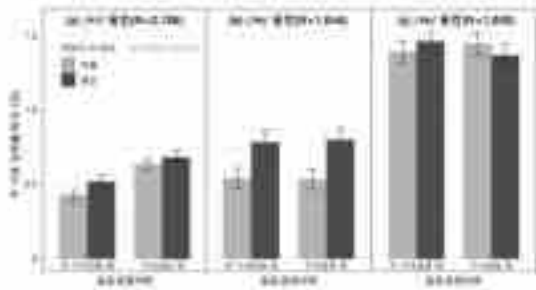
두 자음 실현 점수의 빈도 분석

- ▶ C_1 에 따른 두 자음 실현 점수의 빈도 (단어 단위)

개사별	*합(단어당)으로 실현		두 자음 실현의 조합이 아닌 형태		
	C_1 만 실현	C_2 만 실현	1점	2점	3점
/m/	1125(81%)	134(9.6%)	117(15.1%)	23(3.1%)	134(8.4%)
/n/	770(73.8%)	3(0%)	39(3.7%)	49(4.7%)	198(17.8%)
/ŋ/	50(3%)	528(37.0%)	163(11.5%)	345(24.9%)	21(2.1%)

- ▶ 단자음으로만 실현된 경우
 - /m/: 항상 C_1 (0)으로만 실현
 - /n/과 /ŋ/: 대부분의 경우 C_1 (=)로 실현
- ▶ 두 자음이 함께 실현된 경우
 - /m/ (66.5%), /n/ (32.6%), /ŋ/ (26.2%) 순으로 빈번하게 나타난다
 - /m/과 /n/ 보다 빈도는 적지만 3점으로 돌린 경우(17.8%)가 많음 (> 공명도 높은 [n]의 영향)
 - /m/도 /n/에 비해 2점, 3점이 많음 (> 양순음의 순자조음)

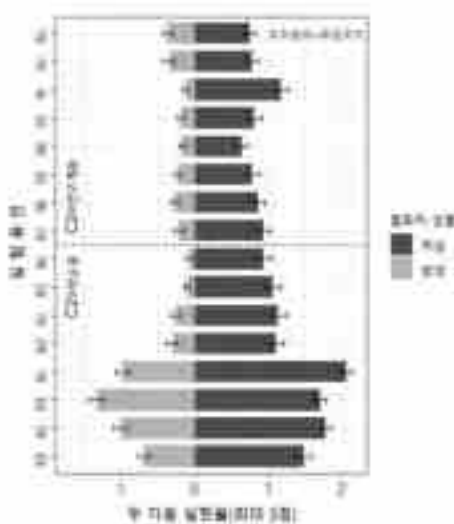
두 자음 실현을 분석



	말안배	어휘	말안배 < 어휘
C1의 조음위치	0.000	0.000	0.000
말안배 내 위치: C1	0.000	0.000	0.000
말안배 내 위치: C2	0.000	0.000	0.000
말안배 내 위치: C3	0.000	0.000	0.000
어휘 내 위치: C1	0.000	0.000	0.000
어휘 내 위치: C2	0.000	0.000	0.000
어휘 내 위치: C3	0.000	0.000	0.000

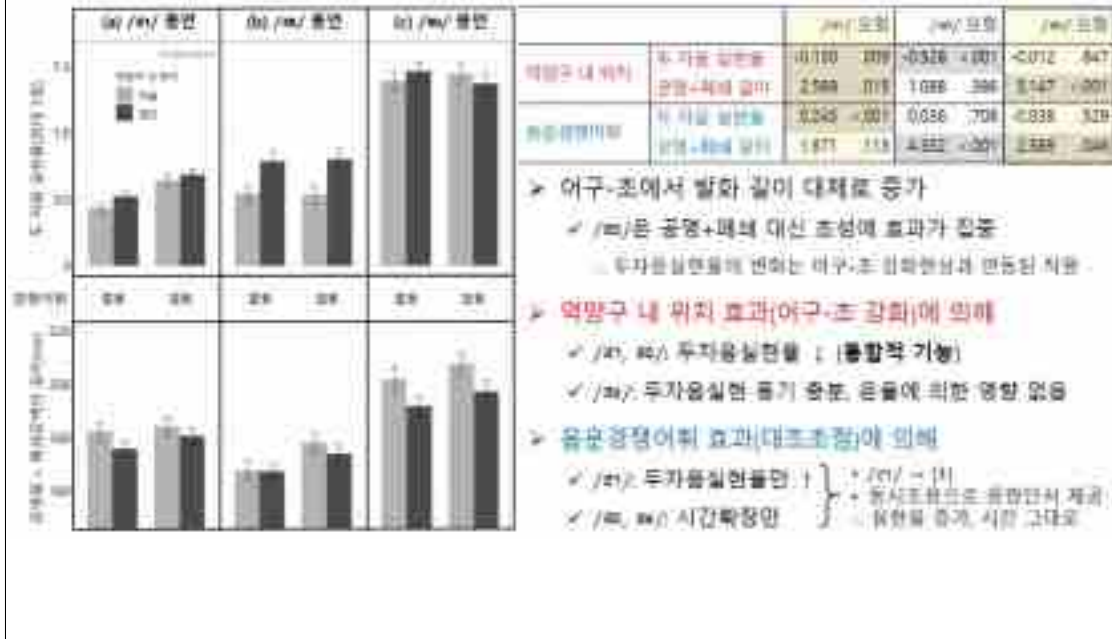
- > 두 자음 실현율 (N=11)
 - ✓ 분석자마다 측정된 검수의 평균값(최대 3)
 - 두 자음 실현시 양도어 정도가 최대 반영
 - ✓ 전체 데이터의 두 자음 실현율 = 0.004
 - ✓ 통합 순서별 표집 정규화된 분석 실시
 - ordinal 대카지(Chowen 2011: dmm) 검수 적용
- > C₁의 조음위치 효과 (N=11)
 - ✓ 말안배 > 어휘 때 두 자음 실현율 증가(p<0.01)
- > 문법의 효과
 - ✓ 억양구 내 위치의 주효과 (N=11)
 - 말안배에서 두 자음 실현율 감소(p<0.01)
 - ✓ 음운경쟁어휘의 주효과 (N=11)
 - 음운경쟁어휘 사용 때 두 자음 실현율 증가(p<0.01)
 - C₁ > 말안배일 때 효과 감소(p<0.01)

기타 요인



- > 발화자 성별의 영향
 - ✓ 여성 발화자(N=6)일 때 두 자음 실현율 증가(p<0.01)
- > 개별 어휘의 영향
 - ✓ 자음군의 종류가 같은 어휘 중에도 차이가 존재
 - ✓ 어휘마다의 사용빈도를 고정효과로 넣어 모형에 반영함
 - 세운빌음치에 현대국어 텍스트인 자음(Oh et al. 2019)에서 추출
 - ✓ 어휘사용빈도가 높을수록 두 자음 실현율이 증가(p<0.01)
- > 임의효과
 - ✓ 참가자 효과(N=11) } 각 임의효과에 설명함
 - ✓ 분석자 효과(N=6)

자음군별 운동효과



결과 요약(1)

- > 실현형의 선택의 변이
 - < 단독형 실현
 - 고개 /w/은 어휘적 폐쇄, /n/은 후행지음에 따라 c, c₂ 존재하면 실현
 - 동양음계 (c, [n]형 증가세) > [n] > [m], 유안양 2002, 유안양 & 오 2004, Oh & Kim 2008, 지복
 - < 두 자음 실현
 - c, c₂가 모두 실현되는 양상(유안양 2004, Oh and Kim 2008) 지속
- > 두 자음 실현의 음운환경
 - < 조음의 용이성: c₂가 연구개음일 때보다 경순음일 때 실현률 증가
 - < 두 자음 간 공명도 차이: c₂가 앞순음 중에서도 /w/일 때가 /n/일 때보다 실현률 높음
- > 두 자음 실현과 음운강화
 - < 통합적 기능 강화: /m/에 비해 두 자음 실현 동기가 약한 /n/과 /w/은 어구-초에서 두 자음 실현률 감소
 - < 음운대조에 의한 강화: /w/과 /m/에 비해 순차적 음향단서가 뚜렷한 /n/은 공명-폐쇄 구간의 길이는 유지, 두 자음 실현률은 증가

결과 요약(2)

- ▶ 발화자 성별의 영향: 여성 발화자가 두 자음을 더 충실하게 실현할 가능성
 - 작은 샘플(N=6+5)
 - 젊은 여성 집단이 이끄는 한국어의 사회언어학적 변이(Bong, 2014; Bong et al., 2014)의 일환으로 주목할 필요
- ▶ 어휘사용빈도의 영향: 사용빈도가 높을 어휘일수록 두 자음을 더 충실하게 실현할 가능성(?)
 - 작은 샘플(N=16):
 - 자음군별 차이: /r/ : 평균 48회, /rʷ/ = 95.25회, /rʷ/ 135.25회 } 개별어휘 효과를 반영하기 위한 통계적 도구로서 적당

결론 및 제언

- ▶ 한국어의 유음으로 시작하는 겹받침 발음의 실현 양상
 - 특정적인 요소에 의해 복수의 실현방식이 확률적으로 경쟁
 - 운율요소, 운율표지, 어휘표지 등의 언어적 요인
 - 성별 등 사회적 요인
- ▶ 말소리 변이에 대한 이론적 시사점
 - 운율 요인과 음운의 변이형 선택 간의 연관성에 대한 논의 부족
 - 변이 중간 말소의 실현에서 음운구분이 가진 역할에 대한 논의 필요

의미/화용

은유적 표현에 의한 부각과 은폐에 관한 연구 - 코로나19 사태 관련 신문 사설을 중심으로 -*

김철규

(경성대학교)

1. 연구의 사회적 배경과 연구 질문

이 논문에서는 2020년 ‘신종 코로나바이러스 감염증’(이후 ‘코로나19’라고 함) 사태를 다룬 신문기사에 나타난 은유적 표현들은 어떤 것들이 있으며 이러한 은유적 표현들이 어떤 개념적 은유를 구체화시킨 것인지 알아보겠다. 그리고 이 개념적 은유들에 의해서 활성화될 수 있는 인식의 틀인 프레임(frame)에 대한 논의를 해보겠다. 최종적으로는 이러한 개념적 은유와 프레임을 통해서 부각(highlighting)되는 것과 은폐(hiding)되는 것은 무엇인지 비판적으로 살펴보고자 한다.

코로나19는 SARS-CoV-2라는 코로나바이러스에 감염되어 발생하는 질병으로 중국 우한에서 2019년 12월에 집단 발병한 폐렴의 원인을 밝히던 중 그 존재가 확인되었고¹⁾ 2021년 현재도 바이러스 감염이 진행 중이다.²⁾ 우리나라에서는 감염 예방을 위한 단계별 사회적 거리두기가 1년 이상 시행됨으로 인하여 정상적인 경제활동에 많은 제약을 주었고, 그 결과 경기 침체 등 사회 전반에 부정적인 영향을 미치고 있다. 이와 같은 상황에서 코로나19 감염 상황 및 이에 대한 대응은 전 세계인들에게 초미의 관심사가 되었고 신문도 국내외 코로나19 상황에 대한 보도에 많은 지면을 할애하였다. 그 결과, 코로나19 상황을 은유적으로 표현한 다양한 예들을 찾을 수 있었다.

- (1) 계속되는 신종 코로나바이러스 감염증(코로나19)과의 전쟁에 의료진이 지쳐가고 있다. 코로나 전투 최일선에서 뛰는 의료진의 번아웃(Burnout·소진)을 우려하는 목소리가 나온다. (중앙일보, 2020년 3월 30일)
- (2) 국내 코로나19 상황은 큰 불길을 잡고 잔불 끄기에 들어간 상태라고 할 수 있다. 남아 있는 불씨는 점화력이 강력하다. 불씨가 큰불로 이어지지 않으려면 집단감염을 차단하고 통제하는 것이 중요하다. (경향신문, 2020년 3월 22일)

* 이 논문은 2021년 5월 『언어과학』 28권 2호(pp.49-78)에 게재되었습니다.

** 경성대학교 영어영문학과

1) 출처: 대한감염학회 http://www.ksid.or.kr/rang_board/list.html?num=3414&code=ncov_faq

2) 세계 보건 기구(WHO)에 따르면 2021년 5월 11일까지 전 세계적으로 162,177,376건의 발병이 확인되었고 그 중 3,364,178명이 사망하였다(출처: 세계보건기구 코로나19 현황판 <https://covid19.who.int/>) 대한민국도 현재 누적 확진자 132,290명, 사망 1,903명에 이르고 있다. (출처: 중앙방역대책본부 2021년 5월 17일자 보도자료)

(3) 전 세계 경제가 코로나19라는 강력한 쓰나미에 떨고 있다. 방과제가 될 600억 달러 규모의 한·미 통화 스와프가 성사된 것은 천만다행이다. (중앙일보, 2020년 3월 23일)

개념적 은유 이론(conceptual metaphor theory)의 관점에서 보았을 때, 우리에게 낯선 코로나19 집단감염 상황이라는 목표영역(target domain)이 ‘전쟁’, ‘불’, ‘쓰나미’와 같이 우리에게 보다 익숙한 근원영역(source domain)으로 사상(mapping)되어 우리 안에 개념화되어 있다는 것을 알 수 있다. 또한 개념적 은유를 위해 어떤 근원영역을 선택하느냐에 따라 목표영역(target domain)의 특정 측면은 부각되는 반면, 다른 측면은 은폐되는 일이 발생한다(Lakoff and Johnson 2003: 10). 정치나 경제 분야 관련 보도에서는 개념적 은유를 통해서 특정 이념이 부각되거나 은폐되는 일이 발생한다(김주식, 2011; 체민경·구본관, 2014; 최영주, 2018; 나익주, 2019). 이와 달리 코로나19 사태와 같은 전 지구적 감염상황 보도에서는 코로나19가 개념적 은유를 통해서 표현되는 과정에서 이념보다는 과학적 사실들과 사회적 현실이 불균등하게 부각되거나 은폐되어질 수 있다. 따라서 이번 연구에서는 아래의 연구질문에 대한 답을 찾을 것이다.

연구질문 1: 신문 보도에서 코로나19 사태는 구체적으로 어떤 은유적 표현을 통해서 보도되는가?

연구질문 2: 이러한 은유적 표현들은 어떤 개념적 은유들을 구체화시킨 것인가?

연구질문 3: 이러한 개념적 은유들에 의해서 활성화되는 프레임은 무엇이며 이로 인해 부각되거나 은폐되는 과학적 사실들과 사회적 현실은 무엇인가?

이 연구질문들에 대한 답을 찾기 위해서 먼저 2절에서 개념적 은유이론, 3절에서 질병 관련 은유표현에 관한 선행 연구를 소개하고 이를 기초로 하여 4절에서 연구방법론을 구성하겠다. 이후 5절에서는 분석 자료에서 유추되는 개념적 은유들과 연관된 프레임들이 어떤 과학적 사실들과 사회적 현실을 부각시키고 은폐시키는지 논의하겠다.

2. 개념적 은유 이론과 프레임

아리스토텔레스의 *Poetics* 이후 은유는 언어의 전달력을 높이기 위해 일상적 언어사용 방법에서 벗어난 새롭고 극적인 언어 표현 방법의 하나라고 여겨져 왔다(권연진, 2017). 하지만 1980년대 이르러 이러한 고전적인 은유관은 Lakoff & Johnson(2003)의 *Metaphors We Live By*에서 제기된 개념적 은유 이론에 의해서 큰 변화를 맞는다. 그들은 일상생활 속 언어에서 빈번히 사용되는 은유적 표현들 안에 담겨 있는 개념적 은유에 주목하고 은유의 영역을 단순한 수사법 차원이 아니라 우리가 생각하고 행동하는데 영향을 미치는 인지적 기제를 이루는 정신 작용으로 확대시킨다. 개념적 은유 이론의 핵심은 우리가 인지하려고 하는 추상적인 목표영역을 보다 구체적인 근원영역에 사상하여 목표영역을 이해하려는 정신적 기제가 은유라는 것이다. 예를 들어 은유적 표현인 ‘They have never won an argument with her.’는 개념적 은유인 [논쟁은 전쟁이다](ARGUMENT IS WAR)가 언어적으로 나타난 것이고, 이를 통해서 ‘전쟁’이라는 근원영역의 여러 양상(aspect)들에 목표영역인 ‘논쟁’의 특정 양상들이 사상되는 과정이 수반된다. 즉 이 개념적 은유는 논쟁에

참여한 사람들이 논쟁 상대자에 대항하여 반론을 제기하고 상대의 논리를 비판하는 일련의 과정이 전쟁에서 적을 막고 공격하는 양상과 대응된다는 우리의 사고 과정을 담고 있다.

Lakoff와 Johnson에 따르면 개념적 은유를 위해서 어떤 근원영역을 선택하느냐에 따라서 목표영역의 특정 양상은 부각시키고 그 이외의 양상은 은폐될 수 있다고 주장한다. 앞서 언급한 [논쟁은 전쟁이다]라는 은유는 논쟁 과정에서 발생하는 경쟁적인 양상은 부각시켜 보여주지만, 논쟁 과정에 결론을 도출하기 위해서 엄연히 존재하는 논쟁 당사자들 사이의 협조적인 관계들은 은폐시키는 결과를 낳는다. 은유를 통한 개념화 과정에서 불균형적인 부각과 은폐로 인해 발생하는 편향성은 우리가 현상을 이해하는 방식인 프레임의 형성에 영향을 미친다. 프레임 개념은 사회학, 인공지능, 의미론 등 이미 여러 분야에서 널리 사용되어지고 있는 개념이다³⁾. Lakoff(2014)는 인간이 현실을 이해하게 하고 때로는 현실이라고 받아들이는 것을 창조하는 정신적인 구조물들(mental structures)을 프레임이라고 하였다. 그는 인간이 어떤 단어를 들으면 인간의 뇌 안에서 그와 관련된 프레임이 활성화되어 우리가 세상을 바라보는 방식을 결정하는 정신적 구조물의 역할을 한다고 하였다. Semino *et al.*(2018)은 프레임을 “세상의 특정한 측면과 관계가 있고, 의사소통하고 행동할 때 기대와 추론을 발생시키며 특정한 어휘문법적 선택과 연결이 되는 약간의 배경지식”이라고 정의한다. 나익주(2019: 5-6)는 “프레임이란 우리의 마음속에 자리 잡은 구조화된 체계로서 세상을 바라보는 창”이기에 “우리의 아이디어와 개념을 구조화하고 우리의 사유 방식을 형성하며 심지어는 우리의 지각 방식과 행동 방식에도 영향을 준다.”라고 하였다.

Lakoff(1991)는 이라크의 쿠웨이트 침공으로 시작된 걸프전에 미국의 참전을 정당화하기 위해 사용된 은유의 분석을 기초로 하여 은유가 현실을 위험한 방식으로 은폐하게 된다면 은유는 치명적일 수 있다고 하였다. 왜냐하면 은유는 현실의 특정한 측면을 부각하고 은폐하는 것을 통해서 우리가 실재(實在)를 제대로 볼 수 없게 만들 위험성이 있기 때문이다. 이후 Lakoff(2014)는 프레임을 통해서 은유가 가질 수 있는 선별적 부각과 은폐의 부정적인 측면을 강조한다. 그는 뇌가 이해하라고 허락한 것만 우리가 이해한다고 지적하면서 우리의 뇌는 쉽게 변하지 않으며 대부분의 시간 동안 뇌가 하고 있는 일에 대하여 의식하지 못한다고 하였다. 그는 우리가 세상을 보는 방식을 결정짓는 정신적인 구조가 프레임이며 (p. xi) 우리는 언어를 통해 프레임의 존재를 알기 때문에 프레임은 우리가 특정한 언어를 사용하도록 이끈다고 하였다. 대부분 우리가 일상생활에서 자주 사용하는 은유적 표현들은 단순한 수사적 표현으로만 작용하는 것이 아니다. 만약 특정 은유적 표현에 우리가 반복적으로 노출된다면 은유적 표현 이면에 있는 개념적 은유는 우리의 사고과정에서 무의식적으로 프레임으로 작용하게 되며⁴⁾ 결과적으로 사람들의 추론, 평가, 감정에까지 특정한 방식으로 영향을 미친다는 것이 여러 실험적 연구에서 밝혀졌다(Potts & Semino, 2019 참조).

개념적 은유에 관한 연구는 다양한 주제로 활발하게 진행되었지만(Kim, 2013; 조영순, 2015; 권연진, 2016; 최영주, 2018; 최재영 · 권연진, 2018) 개념적 은유가 발생시키는 부

3) 이 용어는 사회학자 Erving Goffman(1974)이 그의 책 *Frame Analysis*에서 처음 사용하였는데, 그는 병원과 카지노와 같은 사회적 기관과 테이트, 쇼핑처럼 관습화된 사회적 상황에서 나타나는 인간의 행동방식은 정신 구조인 프레임에 의해서 결정된다고 주장한다. 예를 들어 병원 프레임은 그에 속한 사람들의 역할(예: 의사, 환자, 간호사, 방문자, 경비, 접수 데스크 종사자 등)과 그들의 행동을 결정한다. 또한 Fillmore(1985)의 틀의미론(Frame Semantics)에서도 인간이 특정한 방식으로 사고하고 행동하며 경험한 것을 이해하는 이유도 우리의 머리 안에 인지구조인 틀이 존재하기 때문이라고 했다.

4) Lakoff(2014)는 일단 우리의 뇌 속에 하나의 프레임이 자리 잡으면 그것은 쉽게 변하지 않으며 그 프레임에 일치하지 않는 사실은 기억되지 않는다고 하였다(p. xiv).

각과 은폐에는 크게 중점을 두지 않았다. 일부 논문들에서 부각과 은폐에 관한 언급을 부분적으로 하고 있다. 신선경(2006)은 은유가 설득을 위한 수단으로 사용될 수 있다고 주장하면서 그에 대한 근거로서 은유가 갖은 부각과 은폐의 기능을 언급한다.⁵⁾ 제민경·구본관(2014)은 경제 신문 텍스트에 사용된 은유의 양상을 통해서 은유가 경제 현상을 담화로 구성할 때 대상을 명명하고 개념화하면서 특정한 관점을 어떻게 부각시키고 다른 측면들은 어떻게 은폐시킬 수 있는지 언급하였고 이를 통해 이데올로기에 있어서 중립적이라 여겨졌던 경제 담화에서도 정치 담화에서처럼 이데올로기의 양극성이 개념적 은유를 통해서 어떻게 나타나는지 밝혔다. 개념적 은유가 갖는 부각과 은폐에 관하여 본격적으로 논의한 연구는 김주환(2015)과 김훈기(2017)가 있을 뿐이다. 김주환(2015)은 은유를 통해서 통해 표현하고자 대상의 특정 부분은 보여주고 나머지 부분은 보여주지 않는 것을 “걸러내기”(filtering)라고 규정하고 구약성경 호세아서에서 은유를 통해서 무엇이 걸러지며 강조되는지 고찰하였다. 김훈기(2017)는 과학기술 담화에서 유전자 변형기술로 생산된 농산물의 안정성을 대중들에게 설득하기 위해 다양한 은유법을 통해서 유전자변형 농산물의 긍정적 측면은 어떻게 강조되고 부정적인 측면이 어떻게 은폐될 수 있는지를 보여주었다. 이번 연구에서도 코로나19 집단감염 관련 신문사설에서 은유적 표현으로 인하여 부각되거나 은폐될 수 있는 과학적 사실들과 사회적 현실은 무엇인지 살펴보는 것을 통해서 개념적 은유에 대한 연구의 외연을 확장해 보겠다. 이에 앞서 그동안 이루어졌던 질병에 대한 은유적 표현에 관한 선행연구들을 3절에서 정리해보겠다.

3. 질병과 은유적 표현

개념적 은유이론을 중심으로 질병에 대한 표현을 분석한 연구가 이미 다수 존재한다. 연구의 초점은 질병을 어떤 근원영역을 통해서 은유적으로 표현했는지 그리고 그러한 은유적 표현이 갖는 영향력에 맞추어져 있었다. 특히 암과 1980년대 바이러스성 질병인 후천성 면역결핍증⁶⁾(이후 ‘에이즈’라 함)의 창궐로 인하여 질병에 대한 은유적 표현이 갖는 영향에 관한 연구가 시작되었다. 오랜 기간 동안 영어권에서 질병들은 전투(battle)나 싸움(fight)에 비교되어졌다(Sontag, 1978; Reisfield & Wilson, 2004; Granger 2014). Sontag(1978)은 암투병이 ‘war on cancer’와 같이 군사적 은유(military metaphor)를 통해서 표현되는 것과 ‘Stalinism is the cancer of Marxism.’처럼 부정적인 것이 암에 빗대어 표현되는 것을 비판했다. 왜냐하면 암이 계속적으로 악의 전형으로 표현되는 것을 암환자들이 듣는 것은 치료에 전혀 도움이 되지 않을 뿐만 아니라, 환자들이 숙명론과 가혹한 치료법을 수용하게끔 권유하는 것이기 때문에 매우 악한 일이라고 했다(p.83-85). 또한 Sontag(1989)은 공공보건교육 현장에서 에이즈는 규칙적으로 사회를 침공하는 적으로 기술되고 에이즈로 인한 사망률을 낮추기 위한 노력들은 싸움, 투쟁, 전쟁으로 묘사되기 때문에 에이즈 치료와 치료법 발전을 논하는 의료담화에 군사 어휘가 다수 사용된다고 하였다. 또한 의료담화 이외의 영역에서도 에이즈가 외부의 적으로 표현되는 은유가 공동의 위협에 대응하기 위해서

5) 신선경(2006)은 화자의 발화의도에 따라 은유가 원 개념과 보조 개념이 일반적으로 함의하고 있는 것 중에서 특정한 부분만 부각시키고 나머지 부분은 은폐시키는데, 이를 통해 원 개념의 부분적 속성이 전체적인 속성이 것처럼 일반화되어, 청자도 화자의 시각으로 원 개념을 바라보게 된다는 점에서 은유가 설득의 수단인 된다고 주장한다.

6) AIDS (Acquired Immune Deficiency Syndrome)

대중이 긴급히 동원되어야 하는 정서를 유발시키는데 일조한다고 주장한다. 또한 Sontag는 언어와 오명(stigma) 사이에 상호연관성이 있다고 보았다. 그녀는 에이즈를 표현하기 위해서 사용된 역병 은유(plague metaphor)와 오염원 은유(pollutant metaphor), 죄악(sin metaphor) 은유가 에이즈 환자들에게 오명을 씌운다고(stigmatize) 하였다. 그러므로 Sontag는 질병을 개념화하는데 이러한 은유를 사용하면 안 된다고 주장한다. Ross(1986)도 질병과 관련된 군사적 은유가 환자들이 그들의 권리를 희생시키는 것을 보다 쉽게 여기도록 할 수 있다고 주장한다(p.18).

Wallis와 Nerlich(2005)는 2003년 중증 급성 호흡기 증후군(이후 ‘사스’라고 함.)⁷⁾ 유행 당시 영국의 5대 일간지 보도에서 사스가 은유적 표현들을 통해서 어떻게 개념화되는지 살펴보았다. 세계 다른 지역에서 사스 보도에 군사적 은유(military metaphor)를 가장 많이 사용했던 것과 다르게, 영국의 일간지들은 사스의 성질에 대해서 언급할 때에 ‘killer virus, deadly bug, hit list, it claims victims, it lingers on door handles’ 등과 같은 은유적 표현을 사용하여 SARS IS A KILLER 라는 은유를 통해서 사스를 가장 빈번하게 개념화하며, 사스가 사람들과 국가경제에 미친 충격을 언급할 때에는 ‘it slams shares, it hurt businesses, it hammered corporations, it knocked profits’와 같은 표현처럼 물리적으로 힘을 행사하는 존재로 개념화한다는 사실을 밝혔다. Chiang과 Duann(2007)은 대만과 중국의 신문에 나타난 사스 관련 보도에서 사용된 개념적 은유인 DISEASE IS WAR, DISEASE IS A NATURAL DISASTER/A DISASTER IN GENERAL을 통해서 신문이 가진 서로 다른 정치적 주제(agenda)와 이념이 은유의 사용에 어떻게 스며들어 있는지 분석하였다.

국내에도 질병과 관련된 개념적 은유 연구들이 다수 존재한다. 전해영·유희재(2016)는 2015년 전 세계적으로 유행한 중동호흡기질환(이후 ‘메르스’라고 함.)⁸⁾에 관한 KBS와 JTBC의 보도에서 질병을 나타내기 위해서 사용된 개념적 은유의 종류와 방송국이 가진 이념이 개념적 은유의 사용에 어떻게 반영되는지를 알아보았다. 이 연구는 두 방송국에서 공통적으로 메르스가 전쟁, 물결, 인간, 생물, 사물, 불의 개념을 통해서 은유적으로 표현되었으나 KBS에서는 JTBC보다는 정부에 책임이 덜 가도록하는 전략을 취한다는 것을 밝혔다. Shin(2016)은 메르스 감염 관련 소식을 보도하는 국내 5대 일간지 헤드라인에 사용된 개념적 은유에 관해서 살펴 본 결과 메르스는 ‘전쟁’, ‘자연재해’, ‘유기체’ 및 기타 개념으로 프레임이 형성된다는 사실과 신문의 이념적 성향에 따라 선호하는 프레임이 다르다는 사실을 발견하였다. 송현주는 암(2018), 우울증(2019), 코로나19(2020)와 관련된 은유적 표현들 속에 어떤 개념적 은유가 자리 잡고 있는지 연구하였다. 국내 질병 관련 은유연구에서는 공통적으로 질병이 어떤 은유를 통해서 개념화되는가에만 초점이 맞추어져 있는 반면에 은유로 인하여 부각되거나 은폐되는 것이 무엇인지 살펴보지 않았다. 따라서 본 연구에서는 여기에서 한 단계 더 나아가서 코로나 19를 보도하기 위해 사용된 개념적 은유뿐만 아니라 이러한 개념적 은유를 통해서 형성될 수 있는 프레임으로 인하여 부각되고 은폐되는 과학적 사실들과 사회적 현실이 무엇인지까지 살펴보겠다.

4. 연구자료 및 분석방법

7) SARS (Severe Acute Respiratory Syndrome)

8) MERS (Middle East Respiratory Syndrome)

이번 연구를 위한 분석자료는 뉴스 빅데이터 분석서비스인 빅카인즈(www.kinds.or.kr)를 통해서 수집하였다.⁹⁾ 본 연구에서는 아래의 조건에 따라 자료를 수집하였다(<표 1> 참조).

<표 1> 분석 자료 수집을 위해 설정한 빅카인즈 검색 조건

검색 조건	설정된 조건
(가) 기간	2020년 1월 1일-2020년 12월 31일
(나) 언론사	경향신문, 중앙일보 (가나다 순)
(다) 검색되는 텍스트 유형별 분류	뉴스, 인용문, 사실 중에서 사실에 한정
(라) 분야별 분류	모든 분야 (정치, 경제, 사회, 문화, 국제, 스포츠, IT/과학)
(마) 사건사고별 분류	모든 분야 (범죄, 사고, 재해, 사회)
(바) 검색어 분포 범위	제목 또는 본문
(사) 검색어	코로나 (정확히 일치하는 단어)

‘검색기간’은 코로나 유행이 시작되어 전 세계에 유행한 2020년 1년간으로 하였다. 그리고 언론사의 이념이 은유의 사용방법에 영향을 미칠 수 있으므로 국내 대표적인 좌·우 성향의 일간지인 경향신문과 중앙일보를 고르게 선택하였다.¹⁰⁾ 사건을 객관적으로 보도하는 뉴스 텍스트 보다는 기자와 논설위원의 의견이 보다 직접적으로 피력되는 사실에 은유의 사용이 빈번할 것이라 판단하여 ‘검색되는 텍스트 유형’은 사실로 선택하였다. 또한 ‘검색어’는 ‘코로나’로 설정하여 사실의 제목 또는 내용에 ‘코로나’가 정확히 한 어휘로 사용된 적이 있는 텍스트를 선택하여 코로나와 관련된 텍스트가 선택되도록 하였다. 이러한 과정을 통해서 경향신문에서 총 161개, 중앙일보에서는 총 135개의 사실을 수집하였다.

코로나19와 관련된 은유적 표현의 식별은 기계적으로 할 수 없으므로 수집된 모든 텍스트를 직접 들어가면서 분석하였다. 은유의 분석은 비판적 담화분석론(CDA), 개념적 은유이론(CMT), 코퍼스 언어학 이론을 적용한 Charteris-Black(2004)의 비판적 은유 분석법(Critical Metaphor Analysis)에 따라 수집된 자료를 분석하였다. 비판적 은유 분석법은 크게 제 1단계 ‘은유 식별’, 제 2단계 ‘은유 해석’, 제 3단계 ‘은유 설명’으로 이루어진다.

제 1단계인 은유 식별 단계에서는 질적 연구 단계로서 직접 읽고 은유적 표현이라 생각되는 후보 은유(candidate metaphor)들을 선별하였다. 이때 어떤 특정 표현이 ‘인지적’, ‘언어적’, ‘화용적’ 기준에 부합하면 후보 은유로 판단하였다. 첫째, 인지적 기준에서 보았을 때 근원영역이 갖는 속성과 목표영역이 갖는 속성 사이에 심리적 연관성이 있어서 둘 사이에 사상이 있는지 여부를 판단하였다. 둘째, 언어적인 기준에서 보았을 때 추상적인 것을 다른 맥락의 구체적인 것으로 언급하는 과정을 통해서 구체화시키거나, 무정물을 다른 맥락의 유정물로 언급하여 의인화시키거나, 유정물을 다른 맥락의 무정물로 언급하여 비의인화(depersionification)시키면 언어적 기준을 충족하는 후보 은유로 보았다. 셋째, 어떤 표현이

9) 빅카인즈는 국내 신문, 방송 등 총 54개 언론사에서 생산되는 텍스트와 사진을 제공한다. 텍스트는 기간별, 언론사별, 분야별, 사건사고별, 텍스트 유형별(뉴스, 인용문, 사실)로 검색할 자료의 범주를 설정할 수 있다. 또한 검색어가 분포하는 범위(제목, 본문, 제목+본문)도 설정하여 자료를 수집할 수 있다.

10) 연구 초반 자료수집 시, 좌·우성향의 일간지에서 사용된 은유적 표현을 보다 폭넓게 수집하기 위하여 조선일보와 한겨레신문까지 포함하여 검색하였다. 그러나 검색되는 자료의 수가 너무 많아서 한정된 시간에 네 개 신문들의 자료를 다 읽고 분석하는 것이 불가능하리라 판단되어 중앙일보와 경향신문만으로 자료수집 대상을 한정하였다.

목표영역과 근원영역 사이의 인지적 사상에 그치지 않고 담화를 생산하는 주체의 이념을 담고 있으며 설득적인 기능까지 수행하면 화용적 기준을 충족한다고 보았다. 본 연구를 위해 수집한 자료인 신문 사설은 신문의 텍스트 중에서 신문사의 이념을 가장 극명하게 전달한다. 따라서 사설의 은유적 표현들은 독자들의 의견과 판단에 영향을 미치려는 내재적 의도를 가지고 사용된다고 볼 수 있다. 그러므로 사설에 사용된 은유적 표현들은 세 번째 화용적 기준을 모두 충족한다고 여겼다.

제 2단계인 은유 해석 단계에서는 1단계에서 발견한 은유적 표현들을 기초로 하여 개념적 은유 및 그 개념적 은유가 일관성(coherence)있게 함의(entailment)할 수 있는 “하위 은유(sub-metaphor)”(Kövecses, 2010: 127)들도 식별한다.

마지막 3단계인 은유 설명 단계에서는 개념적 은유에 의해서 활성화될 수 있는 프레임이 코로나19 사태의 어떤 측면이 부각되어지고 은폐되어 독자들의 사고에 어떤 영향을 미칠 수 있는지 설명해 보겠다.

5. 결과 및 논의

수집한 자료에서 찾아낸 은유적 표현과 여기에서 추론되는 개념적 은유 그리고 이들로 인하여 활성화되는 프레임은 <표 2>와 같이 정리된다.

<표 2> 자료 분석 결과

은유적 표현	개념적 은유	프레임	빈도수 (%)
감염병과의 전쟁, 코로나 의병, 코로나 국난(國難), 장기전, 전시(戰時) 재정, 전쟁 난이도, 코로나의 공격, 코로나19 전시체제, 코로나19 전시 상황, 전면전, 방역의 최대 승부처, 코로나19 확산의 북병, 코로나19 최전선, 총력전, 코로나 전사(戰士), 선전(善戰), 영웅, 승자, 총알받이, 코로나19와 악전고투, 코로나 직격탄, 코로나19 확산의 뇌관, 집단감염의 시한폭탄, 코로나 재확산의 기폭제, 확진자 대폭발, 재확산의 촉발점, 감염 폭발	[코로나19 바이러스는 침략자] [코로나19 바이러스 방역활동은 전투] [코로나19 바이러스 감염자를 치료하는 의료인력은 군인] [코로나19 바이러스로 인한 충격은 전장에서 포탄으로 인한 충격]	전쟁	48 (50%)
코로나 국난의 불씨, 코로나 확산의 불씨, 코로나의 새 발화점, 코로나가 번지는 상황, 코로나의 급한 불, 불길	[코로나19 바이러스는 불씨]	화재	14 (14.58%)
집단 감염의 진앙, 신종 코로나의 진원지, 확진자가 급증하는 제2차 파도, 코로나 쓰나미, 충격파, 일파만파	[코로나19 감염사태는 지진] [코로나19 감염사태는 쓰나미]	자연재해	11 (11.56%)
검역망 뚫려, 검역 시스템에 구멍이 숭숭, 촘촘한 방역망, 지역 방역망 확실히 쳐야, 방역망에 큰 구멍, 방역 고삐를 죄어야, 매섭게 살아나다.	[코로나19 바이러스 방역조치는 그물] [코로나19 바이러스 방역조치는 고삐]	짐승	10 (10.42%)
중소기업에 ‘피니시 블로’ 를 날리다, 방역의 허점을 노리다. 경제의 발목을 잡다, 북병	[코로나19 바이러스는 사회발전과 안녕을 저해하는 인물]	사람	7 (7.29%)
터널, 긴 터널	[코로나19 감염사태는 어두운 터널]	터널	3 (3.13%)
총 빈도수			96 (100%)

5.1. 코로나19 관련 개념적 은유들과 프레임

첫째, <표 2>에서 소개한 은유적 표현들을 통해서 [코로나19 바이러스는 침략자]라는 개념적 은유를 도출할 수 있었다. 이는 코로나19 바이러스를 국경선을 넘어오는 침략자로 보는 것을 통해서 눈에 보이지 않는 바이러스에 의한 감염과정을 보다 쉽게 이해할 수 있게 만든다. 은유가 정합적 체계(coherent system)를 이루고 있다는 점을 생각할 때(Lakoff & Johnson, 2003: 87), [코로나19 바이러스 방역활동은 전투]라는 은유적 함의(entailment)를 생각할 수 있다. 이러한 추론은 ‘침략자’라는 근원영역에 대한 배경지식을 통해서 가능하다. 침략자는 우리의 영토, 가족, 자유, 재산을 약탈하기 위해서 쳐들어온다. 이를 지키기 위해서 우리는 침략자에 맞서 싸운다. 따라서 코로나19 바이러스가 침략자라면 코로나19 바이러스는 우리 자신의 생명, 사랑하는 사람, 우리의 일상 등을 앗아갈 수 있다. 또한 [코로나19 바이러스 방역활동은 전투]는 [코로나19 바이러스 감염자를 치료하는 의료인력은 군인]이라는 또 다른 개념적 은유를 함의할 수 있다. 코로나19 바이러스가 침략자라는 근원영역에 사상된다면 이에 감염된 환자들을 치료하는 의료인력은 침략자에 맞서 싸워 우리의 영토, 가족, 자유와 재산을 지키는 군인에 사상된다. 이렇게 활성화되는 ‘전쟁’ 프레임은 코로나19 바이러스에 감염된 환자를 치료하는 의료인력은 침략자를 맞서 싸우는 전사(戰士) 또는 전쟁영웅으로(예(4)), 코로나19환자를 치료에 참여하거나 방역활동에 참여하는 행위는 전장에서 적군과 싸우는 전투(戰鬪)행위로, 환자를 치료하는 병원은 침략자를 맞서 싸우는 최전선(最前線)으로(예(5)), 코로나19가 유행하는 시기를 전쟁을 하는 전시(戰時)로(예(6)), 의료인력은 아니지만 방역활동에 도움을 주는 시민들은 의병(義兵)으로(예(7)), 코로나19 방역을 전 국민이 참전해야하는 전면전(全面戰), 총력전(總力戰)으로 묘사하는 은유적 표현들이 가능하게 된다(예(8)).

- (4) 그런데 최근 보건의료노동자들 사이에서 더 이상 ‘코로나 전사’로 부르지 말아달라는 하소연이 나오고 있다. (경향신문, 2020년 11월 10일)
- (5) 이런 평가의 이면에 최전선에서 분투한 보건의료노동자들의 헌신이 있다는 점을 누구도 부인할 수 없다. (경향신문, 2020년 11월 10일)
- (6) 지금은 경제 비상 상황이다. 문재인 대통령의 진단처럼 코로나 사태 여파로 전시체제나 다름없다. (중앙일보, 2020년 6월 12일)
- (7) 어려운 시기에 희생정신을 발휘한 ‘코로나 의병(義兵)’의 행동은 모두를 숙연하게 한다. (중앙일보, 2020년 2월 28일)
- (8) 정부는 단기에 모든 수단을 집중하는 총력전을 펼쳐야 한다. (경향신문, 2020년 12월 6일)

또한 개념적 은유 [코로나19 바이러스 방역활동은 전투]에 기초하여 전투에 대한 배경지식을 통해서 전장에서는 총탄과 포탄이 사용되는 것을 추론할 수 있다. 이를 바탕으로 [코

로나19 감염사태의 충격은 전장에서 포탄에 의한 충격]이라는 추가적인 은유적 함축이 가능하며 이는 직격탄(예(9)), 시한폭탄(예(10)) 등 총탄과 포탄의 작용과 관련된 다양한 은유적 표현을 가능하게 만든다. 이러한 일련의 개념적 은유들을 통해서 우리는 코로나19 바이러스 집단감염의 상황을 모두 ‘전쟁’이라는 프레임을 통해서 인식하고 있다는 것을 알 수 있다.

(9) 정부가 잣대로 삼은 건강보험은 코로나 직격탄을 맞은 자영업자를 2018년 자료로 판단하고, 소득·재산·부양가족·맞벌이 여부를 복합적으로 따지며 형평성 시비가 일었다. (경향신문, 2020년 4월 6일)

(10) 쿠팡의 집단감염은 언젠가는 터질 수밖에 없는 시한폭탄이었다는 점이 드러나고 있다. (경향신문, 2020년 5월 30일)

둘째, [코로나19 바이러스는 불씨]라는 개념적 은유를 발견할 수 있었다. 작지만 큰 화재를 일으킬 수 있는 불씨의 속성과 작지만 큰 전염력을 갖는 코로나19 바이러스가 갖는 전염력 사이에 사상이 이루어지는 것을 통해 이 개념적 은유가 가능해진다. 즉 작은 불씨가 옮겨 붙어서 대형 화재를 유발하는 과정은 지름이 0.1에서 0.2 마이크로미터로 추정되는 코로나19 바이러스¹¹⁾가 사람들에게 삼시간에 전염되어 의료대란을 일으키는 과정에 대응된다. 또한 화재현장에서 불씨가 갖는 여러 가지 속성들에 대한 우리의 배경지식이 코로나19 바이러스 전파가 일어날 수 있는 대규모 집회는 코로나19 확산의 불씨 또는 발화점으로, 코로나19 전염을 막는 각종 방역행위는 화재의 확산을 막기 위해 불길을 잡는 행위, 불을 끄는 행위 등 다양한 은유적 표현을 가능하게 한다. 이러한 은유적인 표현들과 개념적 은유를 통해 우리가 코로나19 사태를 ‘화재’ 프레임으로 바라보고 있다는 것을 알 수 있다.

(11) 젊은층의 유흥업소·클럽, 코로나 2차 확산 불씨 안돼야 (경향신문, 2020년 4월 8일)

(12) 부천 쿠팡 물류센터 관련 코로나19 확진자가 29일 100명을 넘어서면서 비슷한 사업장이 코로나 확산의 새로운 발화점으로 부각되고 있다. (경향신문, 2020년 5월 30일)

(13) 현 단계에서 코로나19 불길을 잡으려면 마스크를 착용하고, 이동·접촉을 최대한 줄인 ‘집콕 생활’이 최선이다. (경향신문 2020년 8월 25일)

셋째, 코로나19 감염 사태의 영향이 우리 사회에 무차별적, 불가항력적으로 미친다는 점이 지진과 쓰나미가 우리에게 무차별적, 불가항력적으로 영향을 미치는 점과 서로 대응된다는 점에서 [코로나19 감염사태는 지진]과 [코로나19 감염사태는 쓰나미]와 같은 개념적 은유를 생각할 수 있다. 따라서 코로나19 감염자가 발생한 곳은 진앙, 진원지, 그리고 코로나19 감염자 수가 갑작스럽게 증가하는 상황을 파도가 몰려오는 것에 비유하는 은유적 표현들이 가능하다. 이러한 개념적 은유들은 우리들이 코로나19 감염사태를 ‘자연재해’ 프레임을 통해서 인식하고 있다고 볼 수 있다.

11) 출처: 「시민 유일무기 ‘마스크’ ... 지름이 0.1 마이크로미터 바이러스 막을 수 있을까」 중앙일보 2020년 1월 27일자 기사

(14) 특히 이번 집단 감염의 진앙이 온라인 쇼핑업체라는 점에서 충격이 크다. (중앙일보, 2020년 5월 29일)

(15) 나 하나쯤이야 안 지키면 어찌라 하는 방심이 올 최대의 ‘코로나 쓰나미’를 부를 수 있다. (경향신문, 2020년 11월 26일)

넷째, 코로나 19 바이러스 확산을 막기 위한 조치들이 짐승을 포획하거나 잡아서 통제할 때 사용하는 도구들인 그물(예(16))과 고삐(예(17))에 비유되는 은유적 표현들을 발견할 수 있고, 이러한 표현들은 [코로나 19 바이러스 방역조치는 그물]과 [코로나 19 바이러스 방역조치는 고삐]라는 개념적 은유에서 비롯되었다고 보인다. 또한 이러한 개념적 은유들의 존재를 통해서 우리가 현재 유행 중인 코로나19 바이러스를 잡아서 통제해야 하는 ‘짐승’ 프레임으로 바라보고 있다고 할 수 있다.

(16) 현재의 유증상자 위주의 방역에서 벗어나 확진자가 ‘무증상 상태’에서 접촉한 이들도 자가격리나 능동감시 대상에 포함시키는 등 방역망을 촘촘히 짜야 한다. (경향신문, 2020년 2월 3일)

(17) 개개인이 ‘거·마·손’ 기본방역(거리 두기·마스크·손 씻기)을 엄수하고, 대면활동 최소화 빠른 진단으로 방역 고삐를 단단히 죄어야 한다. (경향신문, 2020년 12월 6일)

다섯째, 코로나19 바이러스와 집단감염에 의해 야기된 사회적 충격이 사람으로 구체화되어 [코로나19 바이러스는 사회발전과 안녕을 저해하는 인물]로 개념화 된다. 이를 통해서 우리는 코로나19 바이러스에 의해서 벌어지는 상황을 인간적인 동기를 가진 개체에 의해서 벌어지는 사건으로 이해하게 된다(Lakoff & Johnson, 2003: 33). 코로나19 바이러스에 의한 집단 감염이 사회 전 영역에 부정적인 영향을 미치고 이로 인하여 개인의 생활도 전반적으로 영향을 받고 있는 상황을 마치 인격화된 코로나19 바이러스에 의해서 사람과 사람 사이에 일어나는 다양한 유형의 상호작용으로 구체화시키고 있다(예(18), 예(19)). 이는 ‘인간’이라는 프레임이 작동한 결과라 볼 수 있다.

(18) 코로나 사태는 이미 비틀대는 중소기업에 ‘피니시 블로’를 날린 것이나 다름없다. (중앙일보, 2020년 6월 12일)

(19) 이러다가 신종 코로나가 세계 경제의 발목을 잡을 복병이 되지 않을까 우려된다. (경향신문, 2020년 1월 28일)

여섯째, 코로나19 감염사태를 어두운 터널에 비유하는 은유적 표현들이 발견되었으나 빈도가 적어서(3회) 유의미한 분석이 불가능하여 특별히 더 논의하지는 않겠다.

5.2. 부각과 은폐

5.2절에서는 <표2>에서 보인 여섯 개의 프레임들 중에서 부각과 은폐가 뚜렷하게 나타나 는 ‘전쟁’, ‘자연재해’, ‘침승’ 프레임과 그와 관련된 개념적 은유들만 선별하여 논의하겠다.

5.2.1. ‘전쟁’ 프레임과 개념적 은유 [코로나19 바이러스는 침략자]의 경우

코로나19 바이러스가 국가의 영토를 침범하는 침략자로 개념화되는 것은 인간이 바이러스에 감염되는 과정과 외국군대와 같은 침략자가 영토를 침범하는 과정 사이의 유사성에서 비롯된다. 이를 논의하기 위해서는 바이러스와 바이러스 감염과정에 관한 배경지식이 필요하다. 브리태니커 온라인 백과사전에 따르면 바이러스는 다른 생물체 내부에서만 복제를 할 수 있는 감염성 병원체이다. 세포막 구조가 없고 유전물질(DNA 또는 RNA)과 이를 감싼 단백질로 이루어져 있기 때문에 생물과 무생물의 중간적 형태라고 여겨진다. 따라서 바이러스의 경우 생명력의 유무를 논하지 못하고 감염성 혹은 활동성의 유무를 논할 수 있을 뿐이다. 그리고 바이러스는 세포막 구조를 갖지 못하기 때문에 세포막 구조를 갖는 세균(bacteria)과 구분된다. 또한 바이러스는 스스로 증식하는 능력이 없기 때문에 숙주 세포의 표면에 있는 수용체(受容體)를 통해서 숙주 세포 내부에 침투하여 숙주 세포의 복제시스템을 활용해서 자신의 유전체를 복제하여 증식한다. 바이러스가 침투한 숙주 세포는 정상적으로 기능할 수 없기 때문에 감염된 세포는 죽게 된다. 감염된 세포가 죽게 되면 증식된 바이러스는 죽은 세포 밖으로 나아가서 다른 세포에 침투하는 동일 과정을 반복하는 것을 통해서 숙주는 바이러스에 감염된다.

따라서 [코로나19 바이러스는 침략자]라는 개념적 은유는 바이러스가 숙주인 인간의 세포에 침투하는 과정은 잘 부각시키지만 숙주 세포 안에서 바이러스가 자체로 생명력이 없기 때문에 스스로는 증식될 수 없는 점은 은폐된다. [코로나19 바이러스는 침략자]라는 은유에 나타나는 사상은 <표 3>과 같이 정리된다. 이를 통해 바이러스에 생명력이 없다는 속성이 사상될 침략자의 속성은 없다는 점을 쉽게 알 수 있다.

<표 3> [코로나19 바이러스는 침략자] 은유에서 발생하는 근원영역과 목표영역의 사상

근원영역: 침략자	목표영역: 코로나19 바이러스
국경선 또는 방어선	숙주 세포의 세포막
침략자에 뚫린 국경선 또는 방어선	숙주 세포 표면의 수용체
침략자의 국경선 또는 방어선 돌파	바이러스의 세포 내 침투
?	바이러스가 자체로 생명력이 없는 점

5.2.2. ‘전쟁’ 프레임과 개념적 은유 [코로나19 바이러스 감염자를 치료하는 의료인력은 군인]의 경우

[코로나19 바이러스 감염자를 치료하는 의료인력은 군인] 은유 또한 의료인력이 하고 있는 역할 중 특정 측면을 부각 또는 은폐시킨다. 감염자를 치료하는 의료인력과 전쟁에 징집된 군인은 그 역할이 엄연히 다르다. 의료인력은 환자를 살리는 역할을 하지만 동시에 똑같이 자기 자신도 바이러스에 감염되지 않도록 보호받아야 하며 자신이 감염되는 희생을 강요당할 수는 없다. 하지만 군인의 경우 무력으로 침략해오는 적을 무력으로 제압하는 역할을 해야 하며 이 과정에서 자신의 희생도 요구되어진다. 따라서 의료인력을 군인에 비유한다면 의료인력에게도 전장의 군인에게 요구되어지는 희생을 강요하는 것을 은연중에 당연하게 여

길 수 있다.

또한 [코로나19 바이러스 감염자를 치료하는 의료인력은 군인] 은유는 코로나19 감염자 치료과정이 담고 있는 여러 측면들 중에서 특정 측면을 부각시키고 은폐시킨다. 이를 논의하기 위해서 코로나19 감염자 치료과정을 두 가지 관점에서 바라볼 필요가 있다. 첫째 사회시스템 전체를 아우르는 거시적인 관점에서 바라볼 때 의료인력이 감염자를 발견하고 격리하여 의료시설에 수용되어 일어나는 사회적 치료 과정이 있다. 둘째 미시적인 관점에서 감염자의 신체만 한정해서 코로나19 바이러스 감염자에게 바이러스 치료제¹²⁾ 사용을 통해서 감염자 몸속에서 일어나는 여러 생화학적 작용에 의해서 일어나는 치료과정이 있다. 환자의 몸 밖에서 사회시스템에 의해서 이루어지는 감염자의 발견과 격리 조치 그리고 의료인력이 감염자의 증상을 관찰하고 대증적으로 조치하는 치료행위도 필수적이지만, 코로나19 바이러스와의 실질적이고도 직접적인 전투는 미시적인 수준에서 감염자의 몸 안에서 생화학적으로 일어나고 있다고 볼 수 있다. 따라서 미시적 수준에서 일어나는 생화학적 작용을 통한 증상 완화과정은 육안으로 확인할 수 없기 때문에, 개념적 은유 [코로나19 바이러스 감염자를 치료하는 의료인력은 군인]에 의한 은유적 표현들은 코로나19 바이러스 치료 과정 중에서 의료인들에 의해서 수행되는 사회적인 치료 과정만 부각시키며 코로나19 바이러스 감염자 몸속에서 일어나는 생화학적 기제는 은폐시키게 된다.

5.2.3. ‘전쟁’ 프레임과 개념적 은유 [코로나19 바이러스 방역활동은 전투]의 경우 — 방역 방식의 관점에서 바라본 부각과 은폐 —

개념적 은유 [코로나19 바이러스는 침략자]에 함의된 [코로나19 바이러스 방역활동은 전투] 은유도 방역 과정이 담고 있는 한 가지 중요한 과학적 사실을 은폐시킨다. ‘방역’의 사전적 정의는 ‘전염병이 발생하거나 유행하는 것을 미리 막는 일’(국립국어원 표준대국어사전)이다. 앞서 언급한 바이러스 감염자를 발견, 격리하여 바이러스 치료제로 치료과정이 사후 조치라면, 방역은 병의 발생을 사전에 막는 조치라고 볼 수 있다. 사전 조치인 방역은 감염과정의 관점에서 보았을 때 코로나19 바이러스 자체에 어떤 방식으로 대응하느냐에 따라서 ‘소극적’ 또는 ‘적극적 대응’으로 구분할 수 있다. 감염학적 관점에서 보았을 때 마스크 착용과 사회적 거리두기와 같은 조치는 코로나19 바이러스 접촉을 회피하는 것에 역점을 두는 소극적 대응의 방역활동이라 볼 수 있다. 이와 대조적으로 백신접종은 바이러스를 회피하지 않고 바이러스가 우리 몸에 침투하였을 경우를 대비하여 바이러스를 이길 수 있는

12) 바이러스 치료제의 종류는 치료제의 작용 기전에 따라 아래와 같이 구분된다.

치료제의 종류		작용 기제(mechanism)
항(抗)바이러스제		바이러스가 세포 안으로 침투하는 경로를 차단하거나, 바이러스가 침투한 세포 안에서 유전물질을 만드는 과정을 방해해서 바이러스 증식을 막아 치료효과를 나타냄.
면역 조절제	항염증제	바이러스의 침입으로 신체에서 일어나는 면역작용과 이에 수반되는 염증반응이 과도하게 발생하는 것을 억제하여 장기손상 예방.
	면역증강제	감염자가 면역력이 약화되어 면역세포(T세포)에 의한 면역반응이 잘 안 일어날 경우, 면역증강제가 바이러스 증식의 진행을 더디게 하거나 바이러스에 감염된 세포를 제거하여 감염자의 회복을 돕는 방식으로 감염자의 증상이 악화되는 것을 막는다. 여기서 말하는 면역증강은 우리가 일반적으로 알고 있는 백신접종을 통해서 백신의 항원이 면역세포(B세포)를 자극하여 형성된 항체로 바이러스를 중화시키는 것과는 다른 기제를 말하는 것이다.

출처: 식품의약품안전처 2020년 8월 11일자 보도자료

항체를 우리 몸속에 먼저 형성시켜 놓는 것이다. 따라서 백신접종은 바이러스에 적극적 대응을 하는 방역활동의 일환이라고 볼 수 있다. ‘전쟁’ 프레임에서 보자면 마스크 착용과 거리두기는 아군의 상황이 침략자인 적군을 이길 수 없기 때문에 적군과의 전투를 애써 회피하는 소극적 전략이라고 볼 수 있다. 반대로 백신접종은 적군과 싸울 준비를 미리 해놓고 적군과 전투를 하게 하는 적극적 전략이라고 볼 수 있다. 전시에 외교적인 방법에 의존하는 것을 제외하고 적군과 교전을 피하는 소극적 대응으로는 적군을 결코 격퇴할 수 없으며, 회피전술은 아군이 불리한 상황에서 한시적으로 필요한 전략일 수는 있지만 궁극적으로 전쟁을 종결할 수 있는 최종 전략은 결코 될 수 없다.¹³⁾ 하지만 [코로나19 바이러스 방역활동은 전투] 은유는 마스크 착용과 거리두기와 같은 코로나19 바이러스에 대한 소극적 방역활동까지도 [전투] 개념에 사상시킴으로써 마스크 착용과 사회적 거리두기만으로는 코로나19 사태를 종결시킬 수 없다는 과학적 사실을 은폐시킬 수 있다. 그러므로 [코로나19 바이러스 방역활동은 전투] 은유는 마스크 착용과 거리두기와 같은 소극적 방역활동까지도 코로나19 바이러스를 적극적으로 없애는 전투를 하는 것이라는 잘못된 메시지를 대중들에게 전달할 수 있다. 실제로 방역을 책임지는 정부기관 책임자들, 전국의 지방자치단체들, 방송매체들은 “마스크 착용은 최고의 백신”과 같은 표어를 반복적으로 사용해서 마스크 착용을 독려했다(다음 페이지 <그림1>에서 <그림4> 참조). 하지만 이런 표어는 소극적 방역활동 수단인 마스크 착용과 적극적 방역활동 수단인 백신 접종을 동일시하는 논리적 오류가 있는 잘못된 메시지이다. 왜냐하면 ‘전쟁’ 프레임에서 보자면 마스크 착용은 바이러스와 전투를 전혀 하지 않는 방역수단이고, 반대로 백신 접종은 바이러스와 적극적으로 전투를 하는 것인데 이를 동일시하는 오류를 범하고 있는 것이기 때문이다.

따라서 개념적 은유 [코로나19 바이러스 방역활동은 전투]가 여론의 관심을 마스크 착용과 같은 소극적 방역활동에만 집중하게 하고 적극적 방역활동인 백신 접종에는 관심이 집중되지 못하게 하는 효과를 낳았을 수도 있다는 합리적 추론을 할 수 있다. 실제로 신문에 보도된 백신, 마스크, 거리두기에 대한 언론의 관심도가 이런 추론이 가능할 수 있다는 점을 뒷받침한다.



<그림1> 중앙방역대책본부장 브리핑 영상 (연합뉴스, 2020년 9월 17일)



<그림2> 중앙방역대책본부 부분부장 브리핑 영상 (대한민국 정부 유튜브, 2020년 11월 10일)

13) 백신이 개발되어도 코로나19 바이러스들은 계속 새로운 변이형을 만들어내기 때문에 백신이 바이러스 감염 중에 대한 완전한 해결책이 될 수 없다는 주장도 있다. 하지만 기존의 코로나19 바이러스용 백신들이 변이형 바이러스에도 역시 항체를 생성하는 경우도 있고 또 새로운 백신개발로 다시 변이 바이러스에 맞설 수 있다. 그리고 무엇보다도 변이 바이러스도 마스크 착용과 사회적 거리두기와 같은 소극적 전략만으로는 퇴치할 수 없다. 그러므로 궁극적으로는 백신접종과 같이 체내에 항체를 형성해서 바이러스를 퇴치하는 적극적 전략이 여전히 필요하다.



<그림3> 국무총리 주재 코로나19 중앙재난대책본부회의 보도자료 (2020년 11월 8일)



<그림4> 광명시 대국민 홍보자료 (광명시 뉴스, 2020년 8월 25일)

이를 확인하기 위해 2020년 1월 1일부터 2021년 3월 31일까지 우리나라 6대 일간지 (조선, 동아, 중앙, 한국, 경향, 한겨레)에서 발견된 백신, 마스크, 거리두기 관련 신문기사의 수를 신문자료 데이터베이스인 빅카인즈를 통해 검색해 본 결과는 <그림 5>와 같다.¹⁴⁾ 백신 관련 기사는 2020년 1월에는 178건에 불과했으나, 2020년 12월 세계 최초로 코로나19 예방 백신이 영국에서 주사되고 나서 백신 관련 기사의 수가 급증하였다.¹⁵⁾ 반면 같은 기간 동안 마스크 착용 관련 기사의 출현빈도는 백신 접종 관련 기사 빈도수보다 항상 많았으나, 최초 백신 접종 이후에는 백신 관련 기사보다 발생빈도수가 크게 뒤떨어지게 된다. 이는 영국에서 최초로 코로나19 백신이 접종되기 전까지 여론의 관심이 코로나19를 최종적으로 종결할 수 없는 소극적 방역대책인 마스크에 더 집중되었을 수도 있다는 점을 잘 뒷받침하는 증거라고 할 수 있다.¹⁶⁾

14) 이를 위해서 다음과 같이 빅카인즈 검색 조건을 구체적으로 설정하였다.

검색 조건	선택한 조건
(가) 기간	2020년 1월 1일-2021년 3월 31일
(나) 언론사	경향신문, 동아일보, 조선일보, 중앙일보, 한국일보, 한겨레신문 (가나다 순)
(다) 검색되는 텍스트 유형별	뉴스, 인용문, 사실 중에서 뉴스에 한정
(라) 분야별	모든 분야 (정치, 경제, 사회, 문화, 국제, 스포츠, IT/과학)
(마) 사건사고별	모든 분야 (범죄, 사고, 재해, 사회)
(바) 검색어 분포 범위	제목 또는 본문
(사) 검색어	백신, 마스크, 거리두기 (정확히 일치하는 단어)

15) 출처: 「영국, 이번 주 코로나 백신 대량 접종 시작...세계 최초」 중앙일보 2020년 12월 7일자 기사

16) 우리나라는 마스크 착용과 사회적 거리두기에 국민들이 전적으로 협조한 덕분에 바이러스의 확산을 성공적으로 통제할 수 있었다. 하지만 백신 조기 조달을 위한 구매 시기를 놓쳐서 OECD 회원국 가운데에서 백신 접종 개시 시점이 가장 늦었으며, 2021년 5월 현재 인구 100명당 백신접종 건수 순위는 세계 115위에 그치고 있다(출처: 조선일보 2021년 3월 22일자 사실과 5월 22일자 기사 「접종 115위 한국, 백신 가뭄 넘어 반전 이룰까」).



<그림 5> 2020년1월부터 2021년 3월까지 국내 6대 일간지에 나타난 ‘백신’, ‘마스크’, ‘거리두기’ 관련 기사 빈도수

5.2.4. ‘전쟁’ 프레임과 개념적 은유 [코로나19 바이러스 방역활동은 전투]의 경우 — 방역 활동에 대한 국민적 협조의 관점에서 바라본 부각과 은폐 —

[코로나19 바이러스 방역활동은 전투] 은유는 방역활동에 대하여 국민적 협조의 관점에서 보았을 때 한 가지 중요한 사회적 현실을 은폐시킨다. 지난 1년 3개월 동안의 코로나19 방역활동을 위해서 단계별 사회적 거리두기와 이를 구체적으로 실현하기 위한 집합금지 행정명령을 시행하였다. 그리고 정부와 정치권은 ‘코로나19 바이러스 유행상황이 전시(戰時)에 준하므로 전쟁에 온 국민이 동참해서 국난을 극복해야 하듯이 코로나19 방역활동에도 온 국민이 적극적으로 동참해야 한다.’라는 메시지를 계속 전달한다(<그림6>, <그림7> 참조).



<그림6> 더불어민주당

코로나19국난극복위원회 · 선거대책위원회
연석회의 장면 (MBC, 2020년 3월 20일)



<그림7> 보건복지부 장관, 국무총리,

중앙방역대책본부장 기자 간담회 영상(연합뉴스,
2020년 11월 29일)

이러한 일련의 메시지들은 [코로나19 바이러스 방역활동은 전투] 은유로 인해 활성화된 ‘전쟁’ 프레임을 통해서 국민들로 하여금 방역활동에 동참하는 것이 침략자에 맞서는 국가총력전에 동참하는 것이라 생각하게 만들면서 이로 인한 희생을 마치 군인들이 전쟁에서 자신을 희생하는 것처럼 당연하게 여기도록 만들 우려가 있다. 실제로 우리나라에서는 1에서 3

단계까지 각 단계별로 사회적 거리두기와 이를 위한 집합 금지 행정명령을 시행하였고 다수의 자영업자들은 영업에 제한을 받음에도 불구하고 방역활동 초기에는 사회적 거리두기에 자발적으로 동참하였다. 하지만 코로나19 사태가 1년 이상 장기화됨에 따라 집합 금지 행정명령으로 영업에 제한을 받은 자영업자들의 매출감소가 누적되어 금전적으로 큰 손실을 입었고, 이에 대한 손실보상을 요구하기에 이른다. 하지만 정부는 자영업자들의 피해에 적절한 보상은 하지 못하고 있다.¹⁷⁾ 따라서 [코로나19 바이러스 방역활동은 전투] 은유에 의해 활성화된 ‘전쟁’ 프레임은 자영업자들이 집합금지 명령에 따르는 것을 국가총력전에 동참하는 거룩한 행위로 인식하게 만들어 집합금지 명령으로 인한 매출감소와 같은 금전적 손해가 발생하는 측면을 우리 사회가 잘 보지 못하게 만들 우려가 있다.

5.2.5. ‘자연재해’ 프레임과 개념적 은유 [코로나19 감염사태는 지진], [코로나19 감염사태는 쓰나미]의 경우

작금의 코로나19 사태를 표현하기 위한 개념적 은유 [코로나19 감염사태는 지진]과 [코로나19 감염사태는 쓰나미]는 ‘자연재해’ 프레임을 활성화시킨다. 코로나19 바이러스는 중국 우한에 서식하는 관박쥐(冠박쥐 horseshoe bat) 몸속에서 수십 년에 걸쳐 일어난 변이를 통해 만들어진 것으로 박쥐를 숙주로 삼은 바이러스가 사향고양이, 낙타, 천산갑, 뱀 등 중간 숙주를 거쳐 인간에게 감염되었다(Boni *et al.*, 2020). 이러한 사실에서 코로나19 감염사태는 인간에게는 불가항력적인 자연재해라고 할 수 있고 개념적 은유 [코로나19 감염사태는 지진]과 [코로나19 감염사태는 쓰나미]를 통해 활성화되는 ‘자연재해’ 프레임은 코로나19 감염사태가 갖는 불가항력적 측면을 부각시킨다.

하지만 이러한 코로나19 감염사태가 전적으로 자연재해라고 볼 수 없다는 주장도 있다. 현재의 코로나19 사태는 도시지역의 확대와 대규모 농축산업과 같은 인간의 자연파괴로 인해 야생동물의 서식지가 파괴되어 인간의 거주지와 야생동물의 서식지 사이에서 경계역할을 해온 녹지가 파괴되어 인간과 야생동물이 직접적으로 접촉하게 되는 경우가 더 빈번해짐에 따라 야생동물을 숙주로 하는 바이러스에 인간이 감염되었다는 것이다. 『인수공통(人獸共通) 모든 전염병의 열쇠』를 쓴 콰먼(2020)에 따르면 지구상의 모든 동물종(種)이 새롭게 출현하는 바이러스들의 숙주가 될 수 있지만 그중에서 특히 호모 사피엔스가 바이러스들의 숙주가 될 가능성이 가장 높은데 그 이유는 바로 인간이 동물의 생태계를 가장 많이 침범하고 있기 때문이라고 한다. 최근 20년간 발생한 신종감염병의 약 75%는 인수공통 혹은 매개체 관련 감염병(김우주, 2016)이라는 사실이 이러한 주장을 잘 뒷받침한다. 이뿐만 아니라 세계보건기구(WHO, 2020)도 인간이 촉발한 기후변화와 인수공통전염병의 증가에는 상관관계가 있음을 경고하면서 도시화 및 화석연료 사용 증가로 인한 탄소배출량 증가가 기후변화를 야기한 결과 서식지가 파괴된 야생동물과 인간의 접촉이 증가하여 인수공통전염병이 증가하였다고 한다. 따라서 코로나19 감염사태를 ‘자연재해’ 프레임을 통해 바라보면 코로나19 바이러스의 창궐이 갖고 있는 불가항력적인 측면은 부각되는 반면 인간이 바이러스 확산에 기여한 측면들은 은폐될 수 있다.

17) 이러한 주장에 대한 근거가 되는 관련기사는 다음과 같다.

(1) 「영업제한 소상공인 지원에 ‘인색’한 한국…미국 10분의1」 한겨레신문 2021년 1월 28일자 기사
 (2) 「작년 매출 20조 줄어...소상공인 “보상하라” vs 정부 “재정상 곤란”」 조선일보 2021년 3월 17일자 기사

5.2.6. ‘짐승’ 프레임과 개념적 은유 [코로나19 바이러스 방역조치는 그물], [코로나19 바이러스 방역조치는 고삐]의 경우

코로나19 바이러스의 전파를 막는 방역조치들을 그물과 고삐에 빗대어 표현하는 것을 통해서 코로나19 바이러스는 우리가 포획, 통제할 수 있는 ‘짐승’ 프레임이 작동하고 있다는 것을 알 수 있다. 코로나19 바이러스 전파속도는 방역활동과 감염자 치료 및 백신 접종을 통한 항체생성과정을 통해서 어느 정도까지 통제 가능하다. 코로나19 바이러스를 ‘짐승’이라는 프레임을 통해서 바라본다면 바이러스 전파 속도에 대해서 인간이 어느 정도까지는 통제가 가능하다는 측면을 부각시킬 수 있다.

하지만 ‘짐승’ 프레임은 코로나19 바이러스가 갖고 있는 본질적인 특성을 은폐시킬 수 있다. 바이러스는 독립된 생명체가 아닌 감염성 입자로서 생명력의 유무를 논할 수 없고 단지 활동성의 유무만 논할 수 있는 존재이다. 따라서 어느 다른 생명체와는 달리 독자적으로 대사와 증식을 할 수 없기 때문에 살아있는 숙주세포에 침투하여 증식해야 한다. 바이러스가 숙주세포 안으로 침투하면, 바이러스의 단백질 껍질을 숙주세포의 단백질과 합쳐 버리고, 단백질 껍질 안에 감싸고 있던 자신의 유전자 DNA 혹은 RNA를 방출한다. 그 후에 숙주세포의 효소를 이용하여 자신의 DNA나 RNA를 다량으로 복제하고, 이 유전자의 유전 정보를 이용하여 자신의 단백질 껍질을 합성한다. 이러한 과정을 통해 자신과 같은 새로운 바이러스들이 만들어져 숙주세포 밖으로 방출되고, 숙주세포는 파괴된다. 방출된 바이러스는 새로운 숙주세포 안으로 침투하여 동일한 방법으로 복제를 한다. 바이러스 복제 시, 자신의 DNA나 RNA를 그대로 복제해 다음 세대에 전달해야 하는데 복제하는 과정에서 염기서열에 변화가 일어나서 돌연변이가 발생하여 변이 바이러스가 된다. 바이러스가 DNA와 단백질 또는 RNA와 단백질로 이루어졌느냐에 따라 DNA 바이러스와 RNA 바이러스로 나뉘는데, DNA 바이러스에는 돌연변이를 막기 위한 교정 기능이 있어 돌연변이들이 원래 바이러스 형태로 복귀되기 때문에 변이 바이러스가 발생하지 않는다. 하지만 RNA 바이러스에는 이와 같은 교정 기능이 없으므로 RNA 바이러스에 변이 바이러스가 많이 발생한다. 코로나19 바이러스는 RNA 바이러스의 일종이라서 변이가 많이 발생한다.¹⁸⁾ 따라서 항바이러스제나 백신개발을 하여 코로나19 바이러스 감염을 통제하는 것은 쉬운 일이 아니다. 그러므로 코로나19 바이러스를 우리가 포획, 통제 가능한 ‘짐승’이라는 프레임으로 바라본다면 바이러스 감염에 대하여 통제가 가능한 측면만을 부각시킬 수 있다. 하지만 바이러스는 짐승과 같은 독립된 생명체가 아니라는 점과 RNA 바이러스인 코로나19 바이러스가 갖고 있는 변이형 때문에 통제가 어려울 수 있다는 과학적 사실을 은폐시킬 수 있다.

6. 결론

이번 연구에서는 2020년 한 해 동안 우리나라의 대표적인 보수와 진보 성향의 일간지인 중앙일보와 경향신문의 사설에서 코로나19 바이러스 감염사태를 언급하는 하는데 사용된 은유적 표현들을 수집하여 이러한 표현들을 가능하게 만든 개념적 은유들과 이로 인해 작동한 프레임에 의해서 부각되거나 은폐될 수 있는 과학적 사실들과 사회적 현실이 무엇인지 논의하였다. 가장 빈번하게 사용된 은유적 표현들은 모두 ‘전쟁’ 프레임과 관련이 있는 개념

18) 출처: 「앗! 바이러스의 실수...복제 중 빼앗하면 ‘돌연변이’가 나온다.」 매일경제 2021년 3월 19일자 기사

적 은유들에서 비롯된 것이었다. 이러한 현상은 2절에서 언급했던 국내외 질병 관련 은유표현 관련 선행연구의 결과들과 일치하는 결과였다. 그 다음으로 자주 사용된 은유적 표현들은 ‘화재’, ‘자연재해’, ‘침승’의 프레임과 관련이 있는 개념적 은유들에서 비롯되었다고 유추할 수 있었다.

‘전쟁’의 프레임과 관련이 있는 개념적 은유는 전시에 침략하는 적과 인체로 침투하여 감염시키는 코로나19 바이러스의 유사성, 위험을 무릅쓰고 나라를 지키는 군인과 환자를 돌보는 의료인의 역할의 유사성, 적을 막기 위한 전투와 코로나19 바이러스의 전염을 막는 방역의 유사성을 강조하였다. 하지만 바이러스가 생명력을 가진 존재가 아니라는 점, 전시에 군인에게 요구되는 만큼의 희생이 의료인력들에게 강요되고 있으며, 이들의 희생을 당연시하고 있는 사회적 현실, 마스크 착용과 거리두기 같은 소극적 대응으로 코로나19 사태를 완전 종결시킬 수 없다는 과학적 사실, 거리두기로 인한 자영업자 손실을 전시상황에서 내려지는 국가총동원령의 일환으로 생각하여 당연시 하려는 시각들이 은폐될 수 있다.

마지막으로 방역에 책임이 있는 정치권, 중앙정부기관, 지방자치단체에서 마스크 착용을 독려하기 위해 전달하는 “마스크 착용은 최고의 백신”이라는 대국민 메시지는 마스크의 효과의 백신의 작용에 대한 명백한 과학적 오류를 담고 있는 은유이다. 왜냐하면 코로나19 상황을 ‘전쟁’의 프레임으로 본다면 마스크 착용은 바이러스와의 전투를 회피하는 소극적 전략인 반면, 백신접종은 바이러스와 직접 전투를 준비하는 적극적 전략인데 “마스크 착용은 최고의 백신”이라는 은유는 이 두 가지 상반된 전략을 동일시하는 논리적 오류를 범하고 있기 때문이다. 또한 “코로나 전쟁 반드시 승리하겠습니다.” 또는 “함께 이기리라.”와 같은 대국민 메시지도 코로나19 바이러스 일체의 방역활동에 국민들이 협조하는 것을 마치 전시에 국가총동원령에 국민들이 의무적으로 따라야 하는 것으로 보게 만들어서 방역협조에 따른 국민들의 희생을 당연하게 여기게끔 만들 우려가 있다. 따라서 정부와 지자체는 대국민 메시지를 전달할 때 이러한 점들은 재고(再考)할 필요가 있으며 방송매체들도 그동안 이러한 메시지를 무비판적으로 확대하고 재생산하지는 않았는지 성찰할 필요가 있다고 본다.

참고문헌

- 권연진(Kwon, Y. J.). 2016. 경제담론 상에 나타난 은유표현의 인지언어학적 연구 (Cognitive Linguistic Study of Metaphorical Expressions in Economic Discourses). 『언어과학』 (Journal of Language Sciences) 23.2, 1-19. DOI: 10.14384/kals.2016.23.2.001
- 권연진(Kwon, Y. J.) 2017. 『인지언어학에서 은유의 보편성과 상대성』 (The Universality and the Speciality of Metaphors in Cognitive Linguistics). 서울: 한국문화사.
- 김우주(Kim, W. J.). 2016. 신종인수공통전염병의 출현 전망과 대응 전략 (The Prospect of the Emergence of New Zoonosis and its Counterstrategy). 대한인수공통전염병학회 학술발표 초록집 (Proceedings of the Korea Society for Zoonoses Conference) 2016.1, 13-14.
- 김주식(Kim, J. S.). 2011. 은유의 이데올로기 분석 (An Analysis on Metaphoric Ideology). 『언어과학연구』 (The Journal of Linguistic Science) 56, 29-52.
- 김주환(Kim, J. H.). 2015. 호세아서 1장 2절-2장 15절에 나타난 은유의 부각과 은폐 (Emphasis and Suppress in the Metaphor of Hosea 1:2-2:15) 『신학논단』 (Theological Forum) 79, 195-224. DOI: 10.17301/tf.2015.79..009
- 김훈기(Kim, H. G.). 2017. 유전자변형 농산물의 안전성 설득을 위한 은유 — 크리스퍼/카스9에 대한 대중적 표현에서의 부각과 은폐 — (Metaphor for Persuading the Safety of

- Genetically Modified Agricultural Products: Emergence and Concealment in Popular Expressions of CRISPR/CAS9). 「수사학」 (Korean Journal of Rhetoric) 29, 55-80.
- 나익주(Na, I. J.). 2019. 도덕성 은유와 프레임 전쟁 — 적폐 청산을 중심으로 — (Metaphors for Morality and a Framing War: Centered on the Settlement of Deep-rooted Evils). 「담화와 인지」 (Discourse and Cognition) 26.2, 1-25. DOI: 10.15718/discog.2019.26.2.1
- 송현주(Song, H. J.). 2018. ‘암’ 담화의 비유 사용 연구 (Figurative Expressions in Cancer Discourse). 「Journal of Korean Culture」 43, 7-36.
- 송현주(Song, H. J.). 2019. 의료 담화의 비유 연구 — ‘우울증’ 을 중심으로 — (A Study of Figurative Language in Medical Discourse with a Focus on Depression Metaphors). 「담화와 인지」 (Discourse and Cognition) 26.2, 27-48. DOI: 10.15718/discog.2019.26.2.27
- 송현주(Song, H. J.). 2020. 전염병의 개념화 양상 — 코로나 19를 중심으로 — (Conceptualizing Pandemic Diseases: The Case of Covid-19). 「동서인문」 (Journal of East-West Humanities) 14, 103-129. DOI: 10.22856/jewh.2020.14.103
- 신선경(Shin, S. K.). 2006. 설득의 수단으로서의 은유 — 은유의 생성과 수용에 대한 일고찰 — (Metaphor: A Way of Persuasion). 「한국어 의미학」 (Korean Semantics) 20, 139-159.
- 전혜영, 유희재(Jeon, H. Y. and H. J. Yu). 2016. <메르스>에 대한 은유와 이데올로기적 함축 — KBS와 JTBC 뉴스 보도를 중심으로 — (Metaphors for MERS and Their Ideological Meaning: Focusing on the News Reports from Korean Media KBS and JTBC). 「한국어학」 (Korean Linguistics) 72, 199-225. DOI: 10.20405/kl.2016.08.72.199
- 제민경, 구본관(Je, M. K., and B. K. Koo). 2014. 경제 현상의 언어화를 통해 본 은유의 담화 구성력 (Metaphor's Discourse Constructing Power through the Verbalization of Economic Phenomena). 「한국어 의미학」 (Korean Semantics) 43, 1-31.
- 조영순(Cho, Y. S.). 2015. 은유의 문화 다양성 기술 — 한국어와 영어의 [얼굴] 은유를 중심으로 — (Description of Cultural Diversity of the Metaphor: With Reference to Korean and English [FACE] Metaphor). 「언어과학」 (Journal of Language Science) 22.2, 77-97. DOI: 10.14384/kals.2015.22.2.077
- 최영주(Choi, Y.). 2018. 신문기사 제목에 나타난 은유와 프레임 전쟁 — 삼성 이재용 부회장 구속영장청구 사례를 중심으로 — (How Do Metaphors in Newspaper Headlines Contribute to Frame War? With the Case of Arrest Warrant for Samsung's Lee Jae-yong). 「언어와 언어학」 (Language and Linguistics) 80, 147-166. DOI: 10.20865/20188006
- 최재영, 권연진(Choi, J. Y. and Y. J. Kwon). 2018. 독립선언서에 나타난 도덕적 은유의 개념적 양상에 관한 연구 (A Study on the Conceptual Aspects of Moral Metaphor in the Declaration of Independence). 「언어과학」 (Journal of Language Sciences) 25.3, 231-248. DOI: 10.14384/kals.2018.25.3.231
- 콰먼, 데이비드(Quammen, D.). 2020. 『인수공통 모든 전염병의 열쇠』 (Spillover: Animal Infections and the Next Human Pandemic) (개정판), 강병철 (Kang, B. C.) 역. 서울: 꿈꿀 자유-서울의학서적 (Seoul Medical Books).
- Boni, M. F., P. Lemey, X. Jiang, T. T.-Y. Lam, B. W. Perry, T. A. Castoe, A. Rambaut and D. L. Robertson. 2020. Evolutionary Origins of the SARS-CoV-2 Sarbecovirus Lineage Responsible for the COVID-19 Pandemic. *Nature Microbiology* 5, 1408-1417. <https://doi.org/10.1101/2020.03.30.015008>
- Charteris-Black, J. 2004. *Corpus Approaches to Critical Metaphor Analysis*. New York, NY: Palgrave Macmillan.
- Chiang, W.-Y. and R.-F. Duann. 2007. Conceptual Metaphors for SARS: “War” between Whom?. *Discourse and Society* 18(5), 579-602. <https://doi.org/>

10.1177/0957926507079631

- Fillmore, C. 1985. Frames and the Semantics of Understanding. *Quaderni di Semantica* 6, 222–254.
- Granger, K. 2014. ‘Having Cancer is not a Fight or a Battle’ . *The Guardian* 25 April. <http://www.theguardian.com/society/2014/apr/25/having-cancer-not-fight-or-battle>.
- Goffman, E. 1974. *Frame Analysis: An Essay on the Organization of Experience*. New York, NY: Harper Colophon Books.
- Kim, C. K. 2013. The Metaphorical Meanings of the Korean Adjective chakhan in Korean Newspapers from 1990 to 2011. *Discourse and Cognition* 20(3), 89–109. DOI: 10.15718/discog.2013.20.3.89
- Kövecses, Z. 2010. *Metaphor: A Practical Introduction* 2nd. New York, N. Y.: Oxford University Press.
- Lakoff, G. 1991. Metaphor and War: The Metaphor System Used to Justify War in the Gulf. *Peace Research* 23(2/3), 25–32.
- Lakoff, G. and M. Johnson. 2003. *Metaphors We Live by*. Chicago: University of Chicago Press.
- Lakoff, G. 2014. *The All New Don't Think of an Elephant: Know Your Values and Frame the Debate*. White River Junction, Vermont: Chelsea Green Publishing.
- Potts, A. and E. Semino. 2019. Cancer as a Metaphor. *Metaphor and Symbol* 34(2), 81–95. <https://doi.org/10.1080/10926488.2019.1611723>
- Reisfield, G. and G. Wilson. 2004. Use of Metaphor in the Discourse on Cancer. *Journal of Clinical Oncology* 22(19), 4024–4027.
- Ross, J. W. 1986. Ethics and the Language of AIDS. *Federation Review* 9(3), 15–19.
- Semino, E., Z. Demjén and J. Demmen. 2018. An Integrated Approach to Metaphor and Framing in Cognition, Discourse, and Practice, with an Application to Metaphors for Cancer. *Applied Linguistics* 39(5), 625–645. <https://doi.org/10.1093/applin/amw028>
- Shin, J. W. 2016. Metaphorical Analogies for ‘MERS’ in Korean Newspaper Headlines. *Discourse and Cognition* 23(2), 1–19. DOI: 10.15718/discog. 2016.23.2.1
- Sontag, S. 1978. *Illness as Metaphor*. New York, NY: Farrar, Straus & Giroux.
- Sontag, S. 1989. *AIDS and Its Metaphors*. New York, NY: Farrar, Straus and Giroux.
- Wallis, P and B. Nerlich. 2005. Disease Metaphors in New Epidemics: the UK Media Framing of the 2003 SARS Epidemic. *Social Science and Medicine* 60(11), 2629–2639. <https://doi.org/10.1016/j.socscimed.2004.11.031>
- WHO(World Health Organization). 2020. WHO Climate Change and Infectious Diseases. <https://www.who.int/globalchange/summary/en/index5.html>

Korean Hotels' Responses to Negative Online Reviews: A Genre Analysis

August 19, 2021

Byeong-ki Ahn
byeongki08@gnu.ac.kr
(GNetech)

An Outline of Presentation

- I. Introduction
- II. The Previous Studies and Conceptual Framework
- III. Methodology
- IV. Results and Discussion
- V. Conclusions

Introduction

1.1 The **Purposes** of the paper are to

- 1) To identify surface **linguistic features** such as native structure, lexical choices and English expressions
- 2) To provide Korean students with **authentic framework of hotel responses** to reviews in English business writing
- 3) To understand how Korean Hotels' responses to customers' reviews are composed
- 4) To be beneficial to professionals in **service industries**

II. Literature Review and Conceptual Framework

2.1 Electronic word-of-mouth

- a) **About 75% of tourists** have taken online reviews into consideration
- b) Due to the development of social network services such as the Internet and SNS over the past decade, travelers have become very convenient to make their travel plans
- c) User-generated content on these sites can be viewed as a form of **electronic word of mouth (eWOM)**. (Cox, Burgos, Sellino, & Ibrahim, 2009) The internet has changed "word-of-mouth" into a massive form of communication with predefined groups, friends or thousands of strangers connected in online communities(O'Connor, 2010)

2.2 Hotel responses

- a) 70% of interviewed consumers indicated that they **read online consumer reviews** (Nielsen 2013)
- b) **Consumer-generated online travel reviews** have become a widely used critical information resource. (Chung & Bulutlu 2018)
- c) Approximately 55% of readers consult online comments during their decision-making process. (Zhang, Vom, and Kirtland 2009)
- d) Over 70% of customers checked out other's reviews before booking a hotel. (PloCuoWright 2014)

2.3 ESP-based genre analysis

- **Moves**
 - a) Discourse or rhetorical unit (Swales 2004)
 - b) Realizable communicative unit
 - Each move can also contain steps (Thunwächt 2016)
 - c) A section of a text
 - Each move also represents the bigger picture of the communicative purpose of the genre (Biber et al 2007)
 - d) A text is broken down into **distinctive moves and steps**
 - e) The success and effectiveness of customer review section turn this discourse community into distinct genres (Skalicky 2013)
- **A genre** comprises a class of communicative events, the members of which share some set of communicative purposes

*** Move analysis**

- a) Discourse analysis used in the research and teaching of genres within the area of language for specific purposes
- b) RNRs (responses to negative reviews) by scholars in the field of business
 - language dimension of RNRs interests linguistic analysts
 - establishment of RNRs as a genre and sub-genre
- c) Swales' TDP-based genre analysis
 - business letter of negotiation, replies to customer inquiries, text completion letters, purpose statement in graduate school applications
- d) Composing responses to negative hotel reviews (Thunwacht and Gruninger 2019)
- e) 두브론스키를 통한 네러티브 텍스트의 문법 기능 고찰(유인태 2014)
- f) Hotel responses to positive reviews (Thunwacht 2016)

2.4 Move-step approaches in genre analysis

- a) Move structure of fundraising letters (Hilber et al., 2007)

Move Type 1	Get attention
Move Type 2	Introduce the cause and/or establish credentials of org.
Move Type 3	Solicit response
Move Type 4	Offer incentives
Move Type 5	Relevance (cost)
Move Type 6	Express gratitude
Move Type 7	Conclude with pleasantries

b) Moves and steps in hotels' responses to positive reviews on TripAdvisor (Thunwisch 2016)

Move	Step
Move 1	Salutation
Move 2	Thanking Step 2a: Choosing the hotel Step 2b: Sharing experience
Move 3	Acknowledging message Step 3a: Expressing gratitude Step 3b: Re-stating the compliment Step 3c: Declaring mission accomplished Step 3d: Informing associated staff
Move 4	Ending Step 4a: Giving second thank Step 4b: Expecting a return visit Step 4c: Offering a direct contact
Move 5	Closing Step 5a: Signing-off Step 5b: The author's full name Step 5c: Position

d) Move structure of BNRs (Thunwisch & Ganppes 2019)

Move	Step
1. Salutation	
2. Acknowledging feedback	a. Expressing gratitude b. Valuing feedback c. Expressing regret/concern/apology
3. Brand positioning	a. Stating hotel's commitment b. Confirming hotel's standard
4. Dealing with complaints	a. Explaining causes of the incident b. Reporting action taken c. Admitting mistakes
5. Concluding remarks	a. Expressing gratitude(2) b. Expressing regret/concern/apology c. Asking for a return visit d. Soliciting direct contact e. Promising to improve service
6. Closing	a. Sign off b. Signature c. Job title d. Contact information e. Affiliation

• A Genre Study of *Doctoral Dissertation Acknowledgments*

by Native and Korean Students. (Ahn & Cho 2017)

- some significant differences in the length of the text, rhetorical moves, frequency of occurrence, individuals acknowledged, etc. More important was to find a new move which native speakers do not use in Korean writers' acknowledgments

III. Methodology

3.1 Data collection

- a) 100 responses from 5-star hotels in Korea
- b) Terrible(1 point) and Poor(2 points) section from each hotel
- c) 5 hotel, L hotel, and H hotel in Seoul

3.2 Model of analysis

a) Move structure of RNBs by *Timmins & Gunipeta 2019*

Move	Step
1. Satisfaction	
2. Acknowledging feedback	a. Expressing gratitude b. Valuing feedback c. Expressing regret/concern/apology
3. Brand positioning	a. Stating hotel's commitment b. Confirming hotel's standard
4. Dealing with complaints	a. Explaining causes of the incident b. Reporting action taken c. Admitting mistakes
5. Concluding remarks	a. Expressing gratitude(2) b. Expressing regret/concern/apology c. Asking for a return visit d. Soliciting direct contact e. Promising to improve service
6. Closing	a. Sign off b. Signature c. Job title d. Contact information e. Affiliation

IV. Results and Discussion

4.1 Linguistic features

a) Statistics

Word Count	3,585
Character Count	17,506
Lexical Density	48.6
Unique Words	324
Number of Paragraphs	44
Syllable Count	6,383
Sentence Count	312
Characters per Word	5.0
Syllables per Word	1.8
Average Sentence Length	11.5

b) Readability

Automated Readability Index	7.8
Flesh-Kincaid Grade Level	7.8
Coleman-Liautaud	13.8
SMOG Grade	18.9
Average GradeLevel	10.3
Flesch Reading Ease	58.7

c) Mark Words

Frequency Word List (FWD)	188
Count of Occurrence List (COL)	3,173
Word Occurrence List (WOL)	4,288
Wordnet & Following Language File	228
Wordnet & Following Synonym	277
Wordnet & Following Morph	257
Wordnet & Following, Social Media	88

d) Cloud styles



a) Wordlist (1-40)

1	you	12	for	31	back	31	in
3	the	13	will	32	regards	32	working
4	we	13	think	33	warrant	33	financial
8	your	14	also	34	all	34	and
5	to	15	financial	35	from	35	review
6	that	16	marketing	36	are	36	shared
7	on	17	type	37	the	37	and
8	of	18	our	38	strongly	38	all
9	about	19	recent	39	is	39	with
10	with	20	is	40	concerns	40	the

4.2 Move structure

a) Move frequency

Move 1	Initiation	100%
Move 2	Elaborating feedback	100%
Move 3	Expressing gratitude	100%
Move 4	Making feedback	41.7%
Move 5	Expressing regret/condemnation/apology	100%
Move 6	Final positioning	100%
Move 7	Ending with a compliment	41.7%
Move 8	Offering help / a reward	4.2%
Move 9	Ending with a question	100%
Move 10	Expressing interest of the addressee	41.7%
Move 11	Requesting action/ advice	41.7%
Move 12	Asking a question	11.7%

Move 3	Expressing gratitude	100%
Step 3a	Expressing gratitude(1)	26.7%
Step 3b	Expressing regret/condemnation/apology	26.7%
Step 3c	Asking for a related task	46.7%
Step 3d	Selecting direct content	0%
Step 3e	Proceeding to improve service	41.7%
Move 6	Closing	100%
Step 6a	Say off	100%
Step 6b	Signature	100%
Step 6c	Job title	100%
Step 6d	Contact information	0%
Step 6e	Affiliation	100%

- Obligatory communicative units: M1-M2Sa-M2Sc-M6Sa-M6Sb-M6Sc-M6Se
- Conventional communicative units: M4Sb-M5Sc-M5Se
- Optional communicative units: M3Sb-M4Sa-M4Sc-M5Sb-M5Sd-M6Sd

Ex)

- a. Thank for so much for taking the time to share your experiences regarding your recent stay with us.
- b. [Your valuable comments] will be shared with our staff members.
- c. We regret to hear that your recent experience at OO hotel was not entirely satisfactory.

b) Frequently used expressions

Move 1	Dear ...
Move 2	
Step 2a	- Thank you so much for choosing the OO hotel
Step 2b	- Your feedback is important to us as we wish to ...
Step 2c	- Please accept our sincere apology for any disappointment
Step 2d	- We appreciate your sincere feedback and regret to hear...
Move 3	
Step 3a	- We informed international hotel booking websites of this to minimize any possible inconvenience or discomfort...
Step 3b	- Despite our efforts, it is important that you had to face some unexpected inconvenience
Step 3c	- We will make sure that each and every staff member serves our guests properly
Move 4	
Step 4a	- Please note that saunas are operated in the manner of public bathhouses in Korea, which necessitates addressing before ...
Step 4b	- We will take your comments into consideration to us ...
Step 4c	- Your valuable comments will be shared with our staff members.

b) Frequently used expressions

Scene 1	
Step 1a	- We greatly appreciate your feedback and ...
Step 1b	- Once again, we apologize for any inconvenience caused!
Step 1c	- We hope to have the opportunity to welcome you back to OO hotel...
Step 1d	
Step 1e	- In the future we will do our best to improve our service
Scene 2	
Step 2a	- Warmest regards - Sincerely
Step 2b	- Wataru / I write on behalf of OO hotel
Step 2c	- Vice President
Step 2d	- Marketing Division
Step 2e	

4.3 Type and Order of Move

Type	Move Order	Frequency
1	M1-M2-M3-M4-M5-M6	42
2	M1-M2-M5-M6	19
3	M1-M2-M3-M4-M5-M6	2
4	M1-M2-M5-M4-M6	4
5	M1-M2-M3-M5-M6	1
6	M1-M2-M3-M4-M5-M6	2
	합계	70

Ex)

- a. [Your valuable comments] [will be shared with our staff members] and [moving forward in the future we will do our best to improve our service]
- b. [We genuinely appreciate your feedback] and [will take your comment into consideration].

4.4 Step and N-gram continuum

Move/Step	4-gram	Freq
Solution(M1)	Dear Sir	79
Expressing gratitude(M25a)	thank you so much	79
	for taking the time	77
Valuing feedback(M25b)	appreciate your feedback and	70
	valuable comments will be	33
Expressing regret/insuenr/ apology (M25c)	your valuable comments about	5
	accept our sincere apology	42
	any disappointment you experienced	42
Stating hotel's commitment(M25d)	experienced in our service	42
	in our service during	42
	providing our customers with	5
Confirming hotel's standard(M25e)	to providing our customers	5
	we are committed to	5
	of our exact faith	3
	in case of certain	3

4.4 Step and N-gram continuum

Reporting action taken/ result of investigation(M3a)	be shared with our	38
	shared with our staff	38
	staff members and meeting	35
Admitting mistakes(M3b)	your comments into consideration	42
	will take your comments	42
Expressing gratitude(M3c)	to hear that some	31
	we genuinely appreciate your	42
Expressing regret/insuenr/ apology (M3d)	your feedback and will	40
	apology for any inconvenience	5
	for any inconvenience caused	3
	once again in apology	3

Asking for a return visit(MSSc)	opportunity to welcome you	25
	another opportunity to welcome	46
	hope that you will	46
	look to the 00	38
Ending direct contact(MSSd)	_____	8
Promising to improve services (MSSe)	better service in the	42
	can provide you with	42
	services in the future	42
	do our best to	42
Sign off(MSSe)	warmest regards	19
Signature(MSSh)	a name	79
Job title(MSSc)	marketing division	79
Contact information(MSSd)	_____	8
Affiliation(MSSe)	marketing division	79

V. Conclusion

Functional Linguistic Perspective on Twitter Hashtag: What the Patient31 Tells Us

Yeseul Choi · Sujung Min

(Daedeok Highschool) (Kongju National University)

1. Introduction

Ever since the first patient of the covid-19 virus in Korea had been detected, 20 Jan 2020, Korea was evaluated as “most prepared countries for rapidly responding to and mitigating the spread of an epidemic and pandemic.” On 18 February, when a patient was confirmed as 31st case of the virus infection, it became a symbolic day for the Coronavirus outbreak in Korea. The patient31—known as a “super spreader,” a believer of a Korean cult religious group Shincheonji—triggered the explosive infection in Daegu and Kyung-buk province and appeared as a world-wide Twitter trending list, which means the item (patient31) was frequently mentioned in a given period. People from different countries gathered on the web and were talking about patient31 of South Korea: they talked about how one irresponsible person ruined the quarantine system, and what each country should do to combat the virus.

The purpose of this presentation is to explain how people living in a global pandemic or crisis build and urge solidarity in web-based communication, by examining the linguistic choice of a tweet with a specific hashtag. The data of the analysis, scraped by using library TwitteR of Rstudio and Twitter API, were 101 random English tweets including the hashtag ‘#patient31’ from February 1st 2020 to March 31st 2020. Commercial accounts were manually excluded. Emojis, video clips, and visual memes or figures were not included as subjects of analysis; rhetorical devices such as changing lines were ignored; and some expressions written in other than English, which were partially used in tweets, were not excluded. Typos and grammatical errors were not corrected. Antconc (Anthony, 2009) was used for analyzing concordance, 3-gram and 4-gram, and collocates of the tweets. The tweets and hashtag will be explained and analyzed within Critical Discourse Analysis, Systemic Functional Linguistics, and Appraisal Framework.

The tweet consists of informational parts such as username, time of the posting,

location, the number of retweets, replies, and hearts. The main body of the tweet, which conveys meaning, includes hashtags. Hashtag is a word or a bunch of words written without blank, with pound sign (#). Twitter users use hashtags to note a subject or event. At first, a hashtag was invented to gather the information easily, only for functional convenience. However, by tagging the specific hashtag, the users come to expect other users to visit their tweets and finally to participate in the new kind of interaction. It means people exchange their opinion, evaluate the value, and express sympathy. This allows users to make ‘agora’ of the virtual era. There are some scholars who showed insight for this interactive nature of hashtag. Murthy(2013) explained “By including a hashtag in one’s tweet, it becomes included into a larger “conversation” consisting of all tweets with the hashtag.” Zappavigna(2015) introduced Searchable talk—” the extension of metadata from a topic-marking function, to encompassing additional impersonal function.” Konnelly(2015) claimed that “hashtags have an unambiguous interpersonal function wherein users affiliate with values related to each hashtag, which are themselves linguistic devices that mark the topic of evaluation in different ways.” However, this “interaction, or virtual conversation” is different from traditional communication. It is more of a “stream,” that anyone can participate in anytime from anywhere.

2. Literature Review

Cho, Jung and Park(2013) analyzed twitter use during Japan’s 2011 earthquake and explained that users posted emotional messages directed toward unspecified individuals and relied more on information from government accounts than from news channels. The authors concluded that in order to recover from the disaster users relied on peer-generated resources as well as peer-to-peer communication.

Inako(2015) also investigated Twitter use and community formation after the 2011 Japan earthquake. Inako analyzed the linguistic choice of tweets from different groups(freelancer journalists and the expertise of scientists) with Appraisal framework, and examined how they built affiliation respectively. The two groups showed different value systems, and chose linguistic strategies according to the different value systems.

The kidnap of Chibok girls by Boko Haram in 2014 and #bringbackourgirls hashtag campaign on Twitter and Facebook showed how a local tragic affair became an international campaign promoting social movement. Chilwa and Ifukor(2015) explained users’ stance toward kidnapper and government and examined the activism based on the hashtag campaign.

3. Background Theory

Going backward, there was a movement to recognize the dynamic nature of language and to interpret the language in a broader domain. This movement is relevant to the academic trend of the time that tried to explain all by-products of human activities within the structure of society. Critical Discourse Analysis (CDA) was born from this atmosphere.

There was another movement focusing on functionality of language; later, this movement was developed as Systemic Functional Linguistics (SFL). Halliday (1978, 2014) claimed that language is a system interacting with the context within which the language is used. SFL explained that any act of communication involves larger linguistic options. Therefore, our linguistic realizations are interpreted in the linguistic system. The main difference between the two fields is that while SFL focuses on the interaction between the functions of language and context, CDA on revealing the structure of power in society and its influence (Young & Harrison, 2004). Moreover, it has been pointed out that CDA does not provide practical frameworks for analysis, meanwhile SFL has been elaborated to provide analyzing tools. However, CDA and SFL have common ground in that “both emphasize cultural and historical aspects of meaning (Graham, 2004: p. 63).” Moreover, both perspectives share the view that the meaning–negotiation and interpretation of meaning is established in context.

One of the important points of SFL is that our everyday language is the realization of three metafunctions of language use. Three metafunctions are: ideational, interpersonal, textual function. According to Martin and White, Ideational metafunction is to construe experience. It is related to questions such as “What’s going on? Who’s doing what to whom, where, when, why and how?” Interpersonal function is concerned with negotiating social relations—how people are interacting, feeling they try to share. Textual function is accounted for information flow—the ways in which ideational and interpersonal meanings are distributed. Sequences of discourse, organizing the discursive flow, and maintaining cohesion and continuity is dealt with.

Appraisal Framework or Theory has been evolved within the concept of interpersonal metafunctions of language. According to Martin and White, meanings of positive or negative are conveyed in discourse, and listeners/ readers are involved in the meaning–making process, interpreting the discourse. The system of appraisal consists of three dimensions: attitude (expressing emotion, judgement based on ethics, and evaluation), graduation (gradability of evaluation), and engagement (other people’s voice represented in ‘say, tell, and etc.’), each of which has its subsystems. Graduation is regarding the degree of evaluation (force), and boundary of evaluation (focus). Engagement is explained by the

concept of mono-gloss/ hetero-gloss, which is derived from Bakhtin's notion of heteroglossia. In literature, drawing others' voice is available by using verbs such as 'say' and 'tell.' On Twitter, this is facilitated technologically by bringing other people using '@username' into one's tweet or using retweet (RT) function. Attitude, which is mostly focused in this study, is divided into three regions: affect, judgement, and appreciation. Affect is about people's feelings of positive/negative. Judgement is for assessing behaviors on ethical basis (should/ should not do), and appreciation is related to value evaluation on aesthetic perspective. Judgement and appreciation are "institutionalized" feelings: we all start with the affect system, an innate appraisal system, and as we grow, socialization and institutional education teaches to the participants uncommon, but shared values based on the perception of morality and aesthetics.

4. Analysis

4.1. Functions of Hashtag

The location of hashtags is flexible: it is located in the first or the end of the tweet, and in the middle of the tweet as a linguistic element of the sentence in the tweet. However, the location of hashtags does not always decide the function of each hashtag.

Hashtags, located at the first or end of the tweet, usually do ideational function as a topic marker of the tweet. The tweet below shows that the subject of the tweet is patient31 of South Korea, and her deeds during the quarantine.

- (1) If you are a church leader and are still debating whether or not to stream/cancel services, do have a read. I have many relatives and loved ones that live in Daegu. #Patient31 #SocialDistance

Hashtag, a seemingly topic marker, conveys more complicated intentions of users by being interpreted in a larger context or situation. These are examples of interpersonal function of hashtag:

- (2) Holy moly! Listen to how one covidiot can infect thousands. Don't be #Patient31
- (3) What scares me most about the packed beaches in Florida is that they're all tourists. So they'll head home, likely many of them asymptotic carriers, or pre-symptomatic themselves, & infect entire communities because "they wanted to party." This is #Patient31 x 1,000.

Lexical choices such as covidiot (a combination of covid and idiot), infect, scare, and phrase 'don't be' give patient31 a negative connotation. This will be explained more in the next section.

Textual function is related to how linguistic items are organized in the discourse. Cohesion and the information unit of 'theme and new' are achieved by syntactic construction, which applies to tweets and hashtag as well.

(4) #Patient31 in South Korea infected over 1,000 people. she visited a church. A man in infected 400 people. he visited a mosque. A priest in (U.S flag icon) who infected over 100 people during Sunday communion. It will be difficult to stop religious people from congregating. Many are nuts.

(5) Patient 215 is our #Patient31 #Bahrain #COVID2019

#patient31 in (4) is a topic marker and notify 'theme' of the discourse. but in (5), the same hashtag is new information. Moreover, using a specific hashtag means that the user would participate in a bigger conversational flow. communication. Hashtag can be a kind of cohesive marker for new type of communication..

4.2 Evaluation on hashtag #patient31 (Appraisal Framework and Critical Discourse Analysis)

However, a hashtag may execute more than one function at a time: a combination of metafunctions is given to a linguistic item, and it is called 'coupling (Martin, 2000).' According to Zappavigna (2011), "coupling of evaluation with other kinds of linguistic meanings in texts is a way of tracking the kinds of values construed in the process of affiliation." Zappavigna indicated that the affiliation is ambient because "the users may not have interacted directly and likely do not know each other, and may not interact again," and suggested a term 'ambient affiliation.' #patient31 and its evaluative languages will be explained within Appraisal Framework in this section.

(6) 'doctors at the hospital first suggested she be tested for the coronavirus, Instead, the woman went to a buffet lunch with a friend at a hotel, and to church gatherings.... #patient31 #korea #coronavirus #irresponsible #stupidity

These tweets intend to explain how covid-19 virus was spread by 'patient31,' and the hashtag 'patient31' shows that the main subject of the tweet is patient31. In addition to the topic marking or indexical function (ideational function), it indicates the subject of the appraisal. There are other hashtags stacked at the end position

of tweet, which realize appreciation (#irresponsible, #stupidity).

Extracted tweets below show how the languages surrounding the hashtag perform evaluative function:

- (7) #Patient31 for your **reckless [negative judgement]** actions which put hundreds more people at risk.
- (8) Last Month, South Korea was doing an excellent job of controlling CoronaVirus Then #Patient31 did not adhere to social distancing and a single rogue [negative judgement] patient caused 2 clusters and those clusters were responsible for 80% of South Korea's infections (Over 6200 cases).
- (9) Everyone is screaming [negative affect] #patient31 don't go to church.. practice social distancing yet we will all go to work tomorrow and still be exposed to this same risks.... its really funny [appreciation]

It is evident that evaluative languages for patient31 and her behaviors convey users'opinion of patient31 being irresponsible and self-centered (reckless, rogue). Moreover, users criticize the consequence of the incident by using words such as 'screaming' and 'funny' in (9).

- (10) Looks like some people in the US are going about like Shincheonji Church of Jesus's #Patient31 (in South Korea), a churchgoer who's lack [appreciation] of collaboration is responsible for most of the country's #coronavirus's outbreak.
- (11) The @WHO are biggest [appreciation] bunch of corrupt, spineless, paper-pushers that you can find today, masquerading around as "Health Professionals". This man #BruceAylward is #patient31 who has furthered their Communist Party's agenda since day 1.

In the case of the phrase 'lack of (10),' the phrase itself is classified into 'appreciation,' however, with the word 'collaboration, it plays a role of judgement condemning an individual (patient31)'s self-centeredness. Likewise, 'biggest (11),' the word of appreciation, shows the user's negative stance toward WHO when used with the words "(bunch of) corrupt, spineless paper-pushers." This meaning-making process grants 'patient31,' whose initial meaning is '31st patient of South Korea', the newly settled meaning of "a reckless, self-centered super spreader of virus."

- (12) #Patient31 is like time bomb, risking the nation just by ignoring minor symptoms.
Respect your life and others. #SocialDistancingNow #CoronavirusOutbreak

The users of Twitter express their negative evaluation of patient31 using the metaphor of ‘time bomb (something dangerous)’ in (12). Musolff(2012) stated that metaphor is crucial for creating social realities. Lakoff and Johnson(1980) claimed that metaphor can guide future action for the readers. Therefore, the metaphors for patient31 in below examples warn that it would be dangerous to go somewhere crowded during the epidemic.

- (13) Looks like some people in the US are going about like Shincheonji Church of Jesus’s #Patient31 (in South Korea), a churchgoer who’s lack of collaboration is responsible for most of the country’s #coronavirus’s outbreak.

Genericization refers to the process in which the generic feature of a subject replaces the specific name or social status of the subject. From the early stage of ‘patient31 case,’ her religious background has been mentioned and regarded as an important element to clarify the starting point of the infection. In some tweets, users call her ‘churchgoer’, instead of patient31.

The negative evaluation toward patient31 can be confirmed in the context in which the hashtag ‘#patient31’ is used. From 101 tweets used for this study, a linguistic item, ‘#patient31’ appeared as an element of a phrase ‘don’t be #patient31’ frequently. The result of 3-gram and 4-gram shows this:

Table 1. 3-grams for 101 Tweets of #patient31

Total No. of N-Gram Types: 2710		
Total No. of N-Gram Tokens: 2888		
Rank	Frequency	3-gram
1	16	don t be
2	14	in south korea
3	12	t be patient
4	8	patient in south

Table 2. 4-grams for 101 Tweets of #patient31

Total No. of N-Gram Types: 2792		
Total No. of N-Gram Tokens: 2887		
Rank	Frequency	4-gram
1	12	don t be patient
2	8	patient in south korea

The Table 1 above is the result of 3-gram, a cluster of three words. Because 3-gram does not recognize punctuation marks as a constituent, apostrophe is not represented in the cluster. From the 2,710 3-gram types, the most frequent cluster is 'don t be,' which appears 16 times. The second most frequent cluster is 'in south korea' (14 times). 't be patient' and 'patient in south' are used 12 times, 8 times, respectively. The result of 4-gram (Table 2) shows that the cluster of 'don t be patient' was used 12 times and 'patient in south korea' was used 8 times. The result of 3-gram 't be patient' and 4-gram 'don t be patient' coincides. Consequently, it is obvious that the phrase 'don't be patient31' is used as a fixed form. Figure 3 below is the concordance line of 'don t be;' it shows that 'don t be' phrase is frequently used with the hashtag '#patient31.'

The phrase 'patient in south korea' is recognized in the 4-gram result. This shows that twitter users recognize patient 31 as a negative being by exclaiming 'do not behave like patient31'; moreover, the fact that this case of infection in South Korea has become widely acknowledged across the world is confirmed through this result.

Some users expose their personal backgrounds to tweet and add the intention of warning to '#patient31,' for their peoples to inform that the case of 'patient31' may occur in the near distance:

(14) Our #patient213 patient could be #Patient31 #SocialDistancingNow for people who are the only bread earners of their families. Think of this nation as your own family. It's not religious or racial matter it's human survival in hard times. #CoronaVirusPakistan

(15) @user14 Patient 215 is our #Patient31 #Bahrain #COVID2019

In this case, #CoronaVirusPakistan, #Bahrain plays both ideational function and interpersonal function simultaneously, by specifying the target of warning.

5. Discussion and Conclusion

This study examined evaluative languages of hashtag '#patient31' and tweets with the framework of SFL and Appraisal Framework, and found that the evaluative language use in the tweets gave patient31 a status of 'a reckless, irresponsible super spreader of virus' through the meaning-making process.

Previous works on SNS only centered on content analysis without a linguistically

analytic framework. Admittedly, SNS research based on the content analysis was considered to be worthwhile in that it showed the mechanism of producing and diffusing information. However, language use is not a simple process of information distribution but interpretation within its context. This study, as a qualitative research, linguistically analyzed the hashtag use in Twitter from the functionalist point of view. It showed how SFL perspective can be used to evaluate the words in the cyber space and how meaning has been constructed.

As for limitations, the data was not big enough and subjected to Twitter only; admittedly, the analysis of Twitter is only a first step. When aiming for a fine-grained description of language use in SNS, it needs to be complemented with a detailed analysis of other types of SNS platforms. Also I'm considering how SFL and this tool for analysis can be introduced to the English class. Even for EFL students, it will be worthwhile to learn analysis framework in order to critical literacy for their understanding target language.

6. Reference

- Chiluwa, I., and Ifukor, P. 2015. 'War against our Children': Stance and Evaluation in #BringBackOurGirls Campaign Discourse on Twitter and Facebook. *Discourse & Society*, 26(3), 267–296. <https://doi.org/10.1177/0957926514564735>
- Graham, P. 2004. Predication, Propagation, and Mediation: SFL, CDA, and the Inculcation of Evaluative–Meaning Systems. In Young, L. and Harrison, C. (eds.), *Systemic Functional Linguistics and Critical Discourse Analysis: Studies in Social Change*. London: Continuum. 53–67
- Halliday, M. A. K. 1978. *Language as Social Semiotic*. Edward Arnold.
- Halliday, M. A. K. (revised by Matthiessen) 2014. *Halliday's Introduction to Functional Grammar*. Abingdon, Oxon: Routledge
- Inako, A. 2015. *Affiliating in crisis: a linguistic perspective on community formation on Twitter after the nuclear accident in Japan in 2011* (Doctoral dissertation).
- Konnolly, A. 2015. # Activism: Identity, affiliation, and political discourse–making on Twitter. *The Arbutus Review*, 6(1), 1–16.
- Lakoff, G., and Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press.
- Martin, J. R. 2000. 'Beyond Exchange: APPRAISAL Systems in English' in S. Hunston and G. Thompson (eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford University Press. 142–175
- Murthy, D. 2013. *Twitter: Social Communication in the Twitter Age*. Cambridge: Polity Press
- Musolff, A. 2012. The Study of Metaphor as Part of Critical Discourse Analysis. *Critical Discourse Studies*, 9(3), 301–310. <http://dx.doi.org/10.1080/17405904.2012.688300>
- S. Cho, K. Jung, and H. Park. 2013. Social media use during Japan's 2011 earthquake: How Twitter transforms the locus of crisis communication
- Young, L., and Harrison, C. (eds.). 2004. *Systemic Functional Linguistics and Critical*

Discourse Analysis: Studies in Social Change. London: Continuum. 1-11

Zappavigna, M. 2011. Ambient Affiliation: A Linguistic Perspective on Twitter. *New Media & Society*, 13(5), 788-806. <https://doi.org/10.1177/1461444810385097>

Zappavigna, M. 2015. Searchable Talk: The Linguistic Functions of Hashtags in Tweets about Schapelle Corby. *Global Media Journal*. Volume 9 Issue 1 <https://doi.org/10.1080/10350330.2014.996948>

How to Fall in Love: 은유와 세계의 인지적 해석

김은일

(부경대학교)

I. 들어가기

(1) 그대를 만날 때면
이렇게 포근한데
이룰 수 없는 사랑을 사랑을
어쩌면 좋아요
미소를 띄워봐도
마음은 슬퍼져요
사랑에 빠진 나를 나를
건질 수 없나요

(2) a. 나는 사랑에 빠졌어요.
b. I fell in love. (Papago 번역)

(3) a. I fell in love.
b. 나는 사랑에 빠졌다. (Papago 번역)

(4) a. 사랑에 빠진 나를 건질 수 없나요?
b. Can't you save me from falling in love? (Papago 번역)

☞ 등가(equivalence)?

II. 'fall' vs. '빠지다'

(5) 빠지다: 물이나 구덩이 따위 속으로 떨어져 잡기거나 잡겨 들어가다 (표준국어대사전)

(6) Q: '사랑하다' vs. '사랑에 빠지다' 두 표현의 차이?

A: '사랑하다'는 '사랑을 하다'를 말합니다. '사랑에 빠지다'는 '사랑에서 헤어나지 못하는 것'입니다. (지식iN)

☞ (사랑에) 빠지다: hard to get out or control

- (7) a. to drop down from a higher level to a lower level
 (tps://www.oxfordlearnersdictionaries.com)
- b. If someone or something falls, they move quickly downwards onto or towards the ground, by accident or because of a natural force.
 (https://www.collinsdictionary.com/)
- c. to go down suddenly and hit the ground
 (https://dictionary.cambridge.org/)
- d. to descend freely by the force of gravity
 (https://www.merriam-webster.com/)

☞ to fall: quickly; uncontrollable

III. How to Fall in Love

3.1. 부사 연어



<그림 1. 부사 연어 검색>

부사	연어	MI
DEFINITELY	16	4.5
PROBABLY	15	4.32
ESPECIALLY	9	3.6
AREW	8	3.45
PROBABLY	8	3.44
ESPECIALLY	57	2.88
ESPECIALLY	48	2.32
ACCIDENTALLY	5	2.97
GRACIOUSLY	18	2.54
GRACE	264	1.88
GRACE	10	1.85
GRACE	8	1.82
GRACE	25	1.54

<그림 2. 부사 연어들>

☞ Collocationality: MI > 1.6 (2.0, 2.4)
 Frequency > 3

3.2. 부사 언어의 의미 유형별 분류

종류	예
time	someday
frequency	anew, again, eventually
manner	(속도) instantly, promptly, immediately, quickly; gradually
	(노력) magically, accidently, easily
	(강도) passionately

☞ types of adverbs: manner, place, and time, frequency, degree; conjunctive, sentence

IV. 은유와 세계의 인지적 해석

4.1. 은유와 유사성

☞ Metaphor is based on similarity.



<그림 1> mouse

☞ 은유적 확장(metaphorical extension)

(8) mouse: ① 쥐 ⇒ ② (컴퓨터) 마우스

☞ 부각(highlighting) 과 은폐(hiding)



<그림 3> tongue

<그림 4> cat' s eye

4.2. 인지적 해석(cognitive construal)



<그림 5> 침 vs. hand

(9) @

나라	원어	의미
한국	골뱅이	골뱅이
네덜란드	apestaart	원숭이 꼬리
덴마크	snabel	코끼리 코
핀란드	kissanhnta	고양이 꼬리
독일어	klammeraffe	매달린 원숭이
헝가리	kukac	벌레
노르웨이	grisehale	돼지 꼬리
러시아	sobachka	작은 강아지

(10) a. He looked back to him.

b. He looked back on her childhood.

cf. 뒤(back)돌아 보다(look)

(11) a. He stepped down from the train.

b. He stepped down from the throne.

cf. 하야(下野)하다

(12) 서양 사람들은 사랑에 빠지는 건 ‘벼락에 한 방 맞은’ 것처럼 온다고 표현합니다. [...] 사람이 익어 가면 사랑도 익어 갑니다. 일상의 연인들이 할 수 있고 해야 할 일은 사랑을 숙성시키는 것입니다. 정성이 가득 담긴 술과 빵과 글처럼 말입니다. (김용석. 사랑은, 포도가 와인이 되는 것. <https://www.donga.com/news/Opinion/article>)

☞ 벼락에 맞음: 강렬함, 피하기 어려움, 순식간
숙성, 익어감: 느낌, 되돌리기 어려움

V. ‘Fall in love’ vs. ‘사랑에 빠지다’ : 사건의 국면

☞ 사건: (① 시작,) ② 과정 (/동작), ③ 결과 (/상태)

(13) a. I fell in love.

b. 나는 사랑에 빠졌다. (Papago 번역)

(14) a. I was in love.

b. 난 사랑에 빠졌어. (Papago 번역)

(15) a. 찾으라 그러면 찾을 것이요 (마태복음 7: 7)

b. Seek and you will find

직시적 도치 구문의 의미적 특성 (Semantic Properties of Dexis Inversion Construction)

남소영
(부산대학교)

1. 시제적 특성 (Aspectual properties)

직시적 도치 구문은 장소의 직시적인 *here, there*을 문두에 두고 주어인 명사구가 동사 뒤에 위치한다. 이러한 AP-V-NP 어순으로 인해 도치 구문 이라 불린다. 다음은 직시적 도치구문의 예시이다.

1. a. There's Harry with his red hat on.
- b. Here comes Harry around the corner. (Shinzawa 2015: 52)

위의 직시적 도치 구문들이 중요한 이유는 어순으로 인한 형식상의 특이성과 더불어 일반적인 어순의 직시 구문과는 다른 의미를 가지며 담화 기능에 있어서도 차이를 보이기 때문이다. 이를 기능상의 차이로 설명하자면, 직시적 도치 구문은 화자의 눈앞에서 일어난 사건에 대해 서술할 때만 사용되며, 발화 시점에 진술 대상이 되는 명사를 지시하는 행위가 동반된다는 점이다. 또한 두 구문은 동일한 시제가 사용되더라도 도치 구문과 비도치 구문으로서 나타내는 의미가 다르며, 이런 기능상의 차이와 의미상의 차이로 인해 문법성에 있어서도 차이를 보이게 된다.

직시적 도치 구문은 위에서 서술 하였듯 반드시 화자가 발화 시점 당시 직접적으로 인지된 사건만을 진술한다. 즉 시제가 현재 시제라도 일반적인 구문의 현재 시제와는 다른 제약들이 따르며, 그 의미 또한 다르다. 이는 다음의 직시 구문이 시간의 부사와 함께 사용되는 예시들을 통해 확인 할 수 있다.

2. a. Harry comes here from time to time.
- b. *Here comes Harry from time to time. (Ojea 2018: 269)

(2a)는 일반적 어순의 현재 시제 문장으로 서술 대상의 반복적이고 습관적인 행위를 나타내고 있다. 그러나 (2b)의 구문은 동일한 의미의 시간 부사와 사용되었는데 비문이 되었다. 그 이유는 직시적 도치 구문은 화자가 발화 당시에 일어난 직접 인지한 사건에 대해서 진술할 때만 사용 가능하고 일반적인 현재 시제가 나타내는 습관적 행위를 진술하는 구문으로는 사용되면 그 의미가 성립하지 않아 비문이 되는 것이다. 즉 직시적 도치 구문의 현재 시제는 일반적인 현재 시제가 나타내는 습관적 행위나 진리적 의미를 나타내지 않는다고 봐

야 한다.

그렇다면 직시적 도치 구문에서 진술되는 사건의 의미는 시제와 더불어 어떻게 해석될 수 있는가? 이를 위해 다음의 방향의 부사와 사용된 직시 구문들을 살펴보자.

c. Here comes David McNeil across the street.

d. *David McNeil comes here across the street. (Prodo-Alonso 2016: 66)

방향 부사의 경우 직시적 도치 구문에서는 사용이 가능하나, 반대로 일반 어순의 직시 구문에서는 불가능하다. 이에 반해 직시적 도치 구문은 서술 대상의 경로를 나타내주는 방향 부사와 사용이 가능하다. 이는 직시적 도치 구문이 행동이 진행되고 있는 과정을 나타내는 것이 가능하다는 의미일 것이다. 그러나, 직시적 도치 구문의 진행형은 성립하지 않는다.

3. a. *Here is coming Harry. (Ojea 2018: 269)

b. *There is going bus. (De wit 2018: 13)

위의 진행형의 직시적 도치 구문이 비문인 것으로 보아 직시적 도치 구문은 진행 중에 있는 동작을 설명하기 위해 의도된 문장이 아니라는 것을 알 수 있다. 즉 동작의 진행 과정을 나타내는 것은 가능하나, 발화 시점에 여전히 진행되고 있는 동작을 나타내는 것이 불가능하다면 이것은 결국 직시적 도치 구문에서 서술되고 있는 동작은 일련의 과정을 거쳐 발화 당시에는 종결된 사건을 서술한다고 보아야 할 것이다. 직시적 도치 구문이 발화 당시의 시점과 일치하는 시제를 가지는 것은, 현재 시제일 경우에, 단지 지시하는 행위가 발화 시간과 동시에 일어났다는 이유일 것이다. 이는 직시적 도치 구문에 대해 Ojea(2018) 이외 여러 학자들이 제시한 상황 진술문의 성격을 가진다는 정의와 부합한다. 이로써 직시적 도치 구문은 De wit(2018)에 의해 제시된 것처럼 과거에 일어나 발화 시점인 현재에는 끝난 사건에 관한 서술이며 이는 단순 시제가 사건의 지속성이라는 상(aspectual)의 의미를 지니며, 완료 시제적인 의미를 가진다고 볼 수 있다.

한국수어에 나타난 개념의 신체화

최영주 (조선대학교)

2021년 한국언어과학회 여름 학술대회 (2021.8.)(목)

- 개념적 은유 이론의 핵심 가정:
 - 개념의 신체화 (Embodiment of Conceptualization)
 - 인간의 신체가 환경과 작용하는 방식에 의해 사고가 결정된다는 가정
 - 개념적 은유의 정의와 직접적으로 연관
- 개념적 은유란 복잡하고 추상적인 개념을 이해하기 위하여 구체적이고 이해하기 쉬운 개념을 이용한다
- 구체적이고 이해하기 쉬운 개념?
 - 우리의 신체와 그 신체를 둘러싼 환경, 그리고 타인과의 상호작용

개념의 신체화에 대한 기존 연구들

- '손', '발', '머리', '눈', '코', '입', '가슴' '배'와 같은 신체부위를 가리키는 어휘가 어떻게 은유적으로 추상적인 개념을 가리키게 되는가에 대한 연구 (우도열 1989, 배도연 2002, 임지용 2007, 2016, Kraska-Selentik 2019)
- 신체어 뿐만 아니라 우리의 신체를 둘러싸고 있는 공간에 대한 연구
- 공간개념이 시간개념으로 확장되는 사례 연구(Radden 2011, Núñez and Tsche 2006, Chen 2007, Shinohara and Pardeshi 2011, Chen 2014, Huang 2016)
- 사람이 물리적으로 경험하는 사물의 크기 확장에 대한 연구 (김해연 2012, 2014, 정상원 2020, 최명주 2020)

수어:

신체 부위나 신체 활동을 통하여 개념화

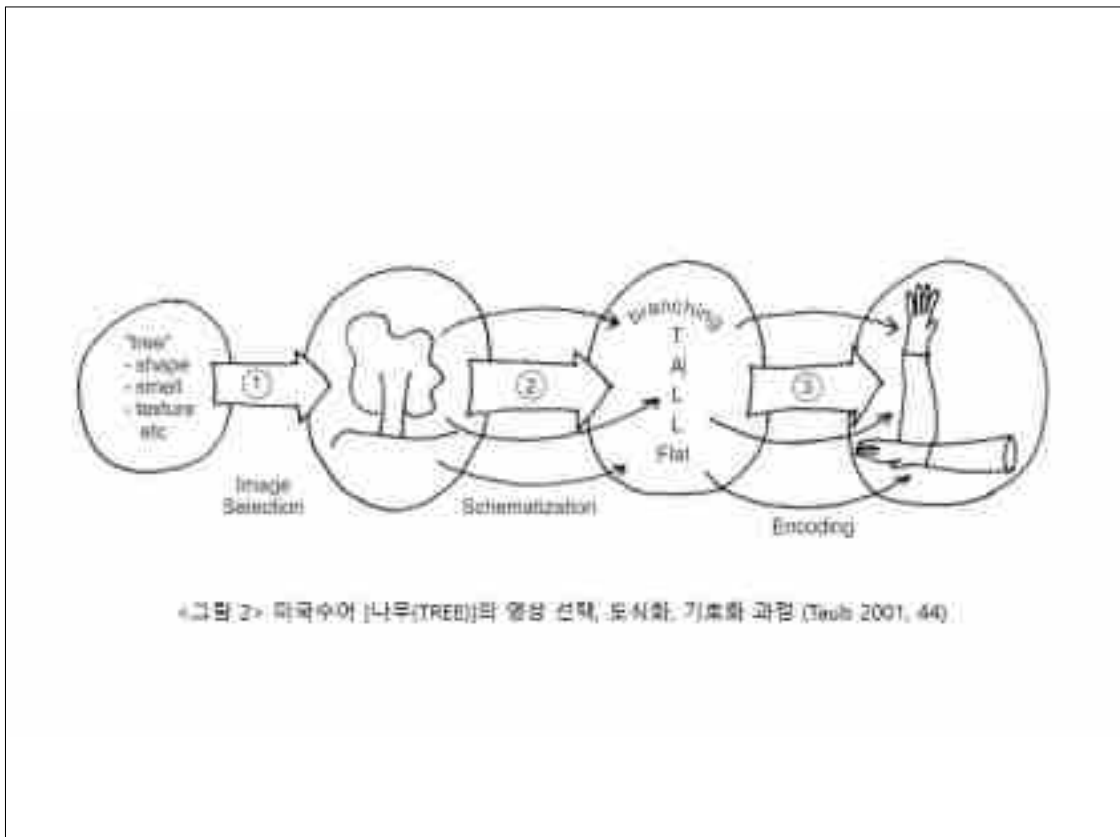
사물의 모양이나 움직임을 본떠서 개념화

→ 개념의 신체화에 대한 강력한 증거

- Wilcox(2000), Taub(2001), Wilcox, Wilcox and Jarque(2004):
미국수어의 은유적 혹은 환유적 의미 확장에 관하여 논의
- 한국수어: 인지의미론적 입장에서 의미 확장을 논의한 예시
 - 명지용·송현주(2015), 석수영(2016), 최영주(2017), 석수영·김기혁(2017),
임지용(2018)
 - 한국수어 혹은 중국수어의 도상성(iconicity)과 수어형성의 동기화
(motivation), 은유 및 환유적 의미 확장에 관하여 논의
- 그러나 그러한 의미 확장이 신체화 과정과 어떠한 연관성을 갖는지에 대한 논의가 없음

연구목표

- 한국수어 중 [봄], [여름], [가을], [겨울], [밤], [낮]과 같은 시간을 나타내는 수어가 도상적으로 그리고 환유적으로 어떻게 사상되는가를 분석하여 개념의 신체화에 대한 하나의 사례로 제시하고자 한다.





<그림 10> 한국수어 [학위(DEGREE)]의 영상선박(A), 도식화(B), 거호화(C) 과정
(Taub 2001, 102)



<그림 11> [수업] 수업의 도출형 사형 (Taub 2004: 102)

Iconicity Mapping of DRILL

TABLE S.4. Iconic Mapping for DRILL

ARTICULATORS	SOURCE
Dominant L	Long, thin object with handle (in particular, a drill)
Nondominant B	Flat surface
L inserted between fingers of B	Penetration of surface

<그린 드> [드릴] ⇨ 2차원 평면형 사형 (Tajiri 2004: 102)



<그린 드> THINK-PENETRATE(이해시키다)

TABLE 6.3. Iconic Mapping for THINK-PENETRATE

ARTICULATORS	SOURCE
1→	An object
Forehead	Head
1→ touches forehead	Object located in head
1→ moves toward locus of addressee	Sending an object to someone
Nondominant B	Barrier to object
1→ inserted between fingers of B	Penetration of barrier
Signor's locus	Sender
Addressee's locus	Receiver

<그림 7> [이해한다] 수어인 도상적 사형 (Taub 2004: 101)

TABLE 6.4. Double Mapping for THINK-PENETRATE

ARTICULATORS	METAPHORICAL MAPPING	
	SOURCE	TARGET
1→	An object	An idea
Forehead	Head	Mind, locus of thought
1→ touches forehead	Object located in head	Idea understood by signator
1→ moves toward locus of addressee	Sending an object to someone	Communicating idea to someone
Nondominant B	Barrier to object	Difficult in communication
1→ inserted between fingers of B	Penetration of barrier	Success in communication despite difficulty
Signor's locus	Sender	Originator of idea
Addressee's locus	Receiver	Person intended to learn idea

<그림 8> [이해한다] 수어의 이중 사형 (Taub 2004: 103)

[뵤], [어툼], [가을], [거울], [뵤], [낫] 수어 분석

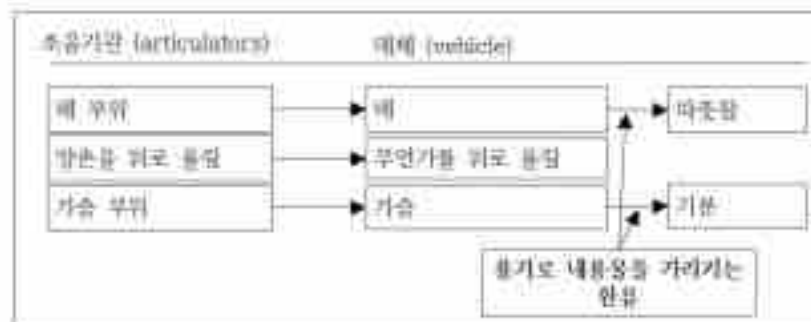
[뵤] 수어



<그림 10> [뵤/뵤리다], [뵤]

[뵤], [어툼], [가을], [거울], [뵤], [낫] 수어 분석

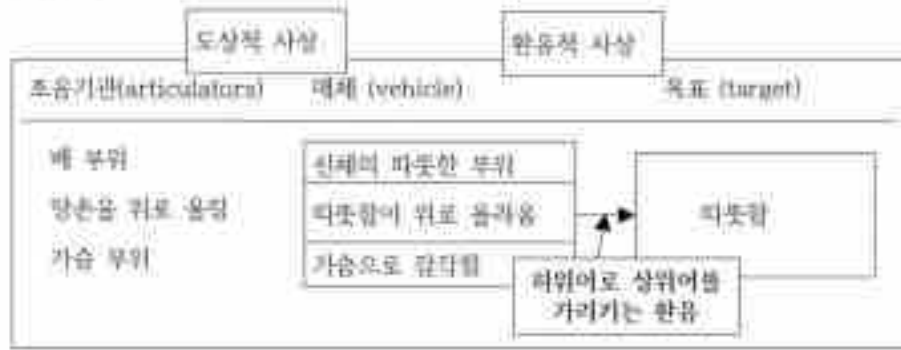
[뵤] 수어



<그림 11> [뵤/뵤리다] 수어의 도상적 사냥

[뵈], [아름], [가을], [거울], [뵈], [낯] 수어 분석

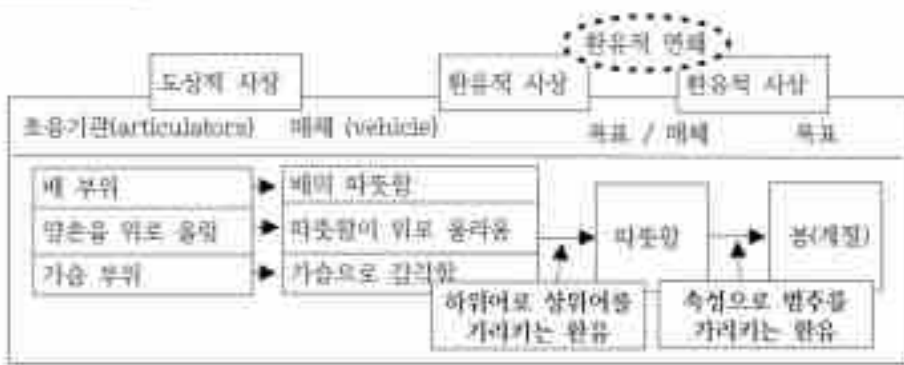
[뵈] 수어



<그림 12> [뵈(아름)] 수어의 도상적 사상 및 환유적 사상

[뵈], [아름], [가을], [거울], [뵈], [낯] 수어 분석

[뵈] 수어



<그림 13> [뵈(아름)] 수어의 [뵈(계절)]으로의 환유적 의미 확장

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

[여름] 수어



<그림 14> [말다], [이름]

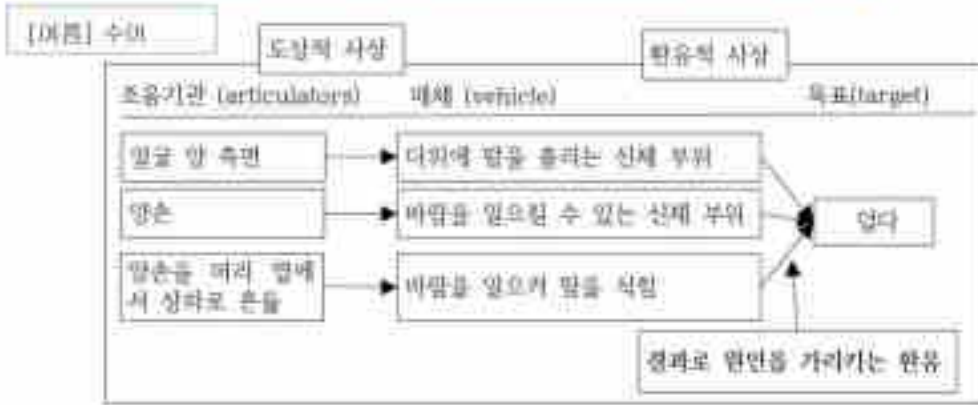
[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

[여름] 수어



<그림 15> [말다] 수어의 도상적 사할

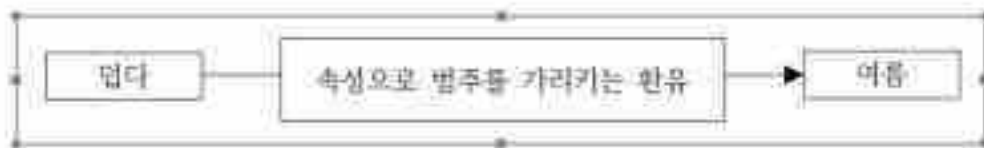
[뵈], [어름], [가을], [거울], [뱀], [낮] 수어 분석



<그림 16> [입다] 수어의 도상적 사상 및 환유적 사상

[뵈], [어름], [가을], [거울], [뱀], [낮] 수어 분석

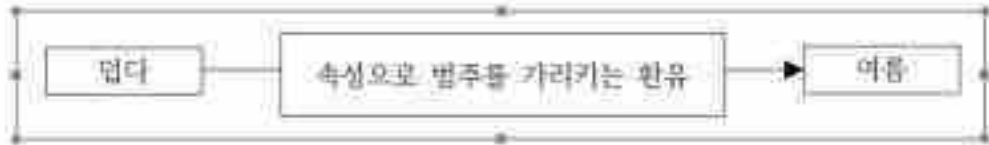
[어름] 수어



<그림 16> [어름] 수어의 환유적 사상

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

[여름] 수어



<그림 10> [여름] 수어의 환유적 구성

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

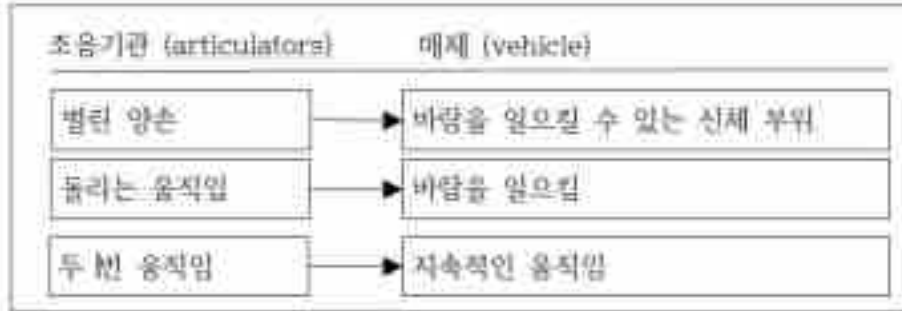
[가을] 수어



<그림 11> [바람(바다)], [가을]

[봄], [아름], [가을], [겨울], [밤], [낮] 수어 분석

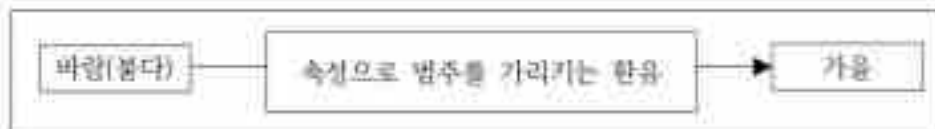
[가을] 수어



<그림 10> [바람(풍다)] 수어의 도상적 사실

[봄], [아름], [가을], [겨울], [밤], [낮] 수어 분석

[가을] 수어



<그림 11> [가을] 수어의 환유적 사실

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

[겨울] 수어



<그림 21> [겨울] 수어의 총유적 사상

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

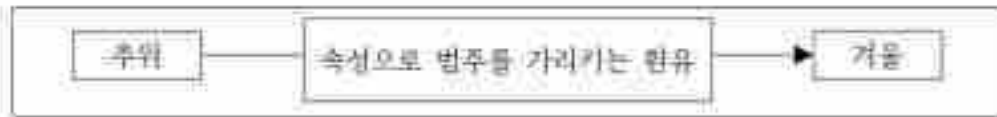
[겨울] 수어



<그림 22> [출처] 수어의 도상적 사상 및 환유적 사상

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

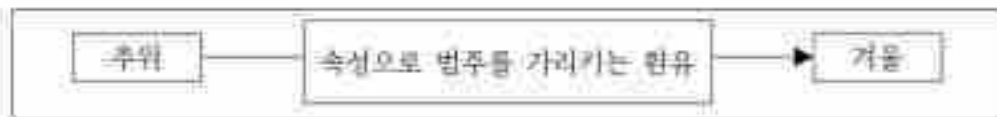
[겨울] 수어



<그림 23> [겨울] 수어의 환유적(상상)

[봄], [여름], [가을], [겨울], [밤], [낮] 수어 분석

[겨울] 수어



<그림 23> [겨울] 수어의 환유적(상상)

[뽕], [어름], [가을], [거울], [땀], [낮] 수어 분석

[뽕], [낮] 수

어



<그림 24> [어름다]



<그림 25> [뽕다]

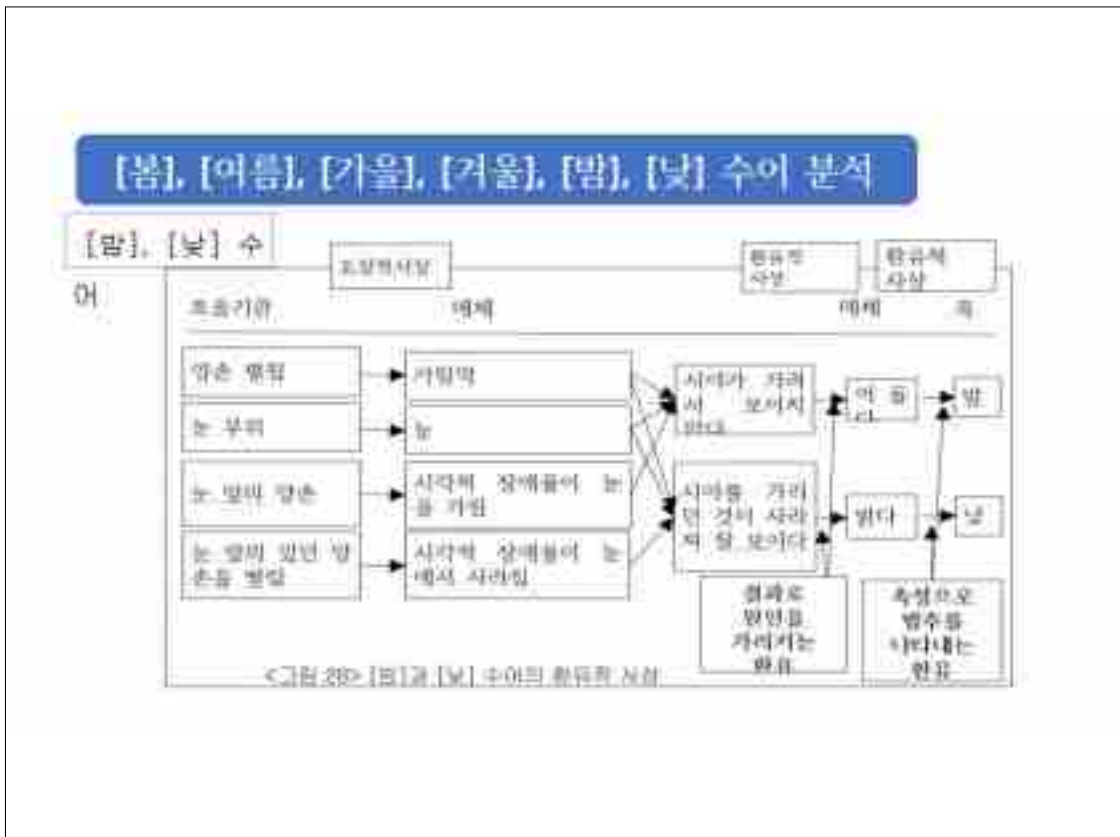
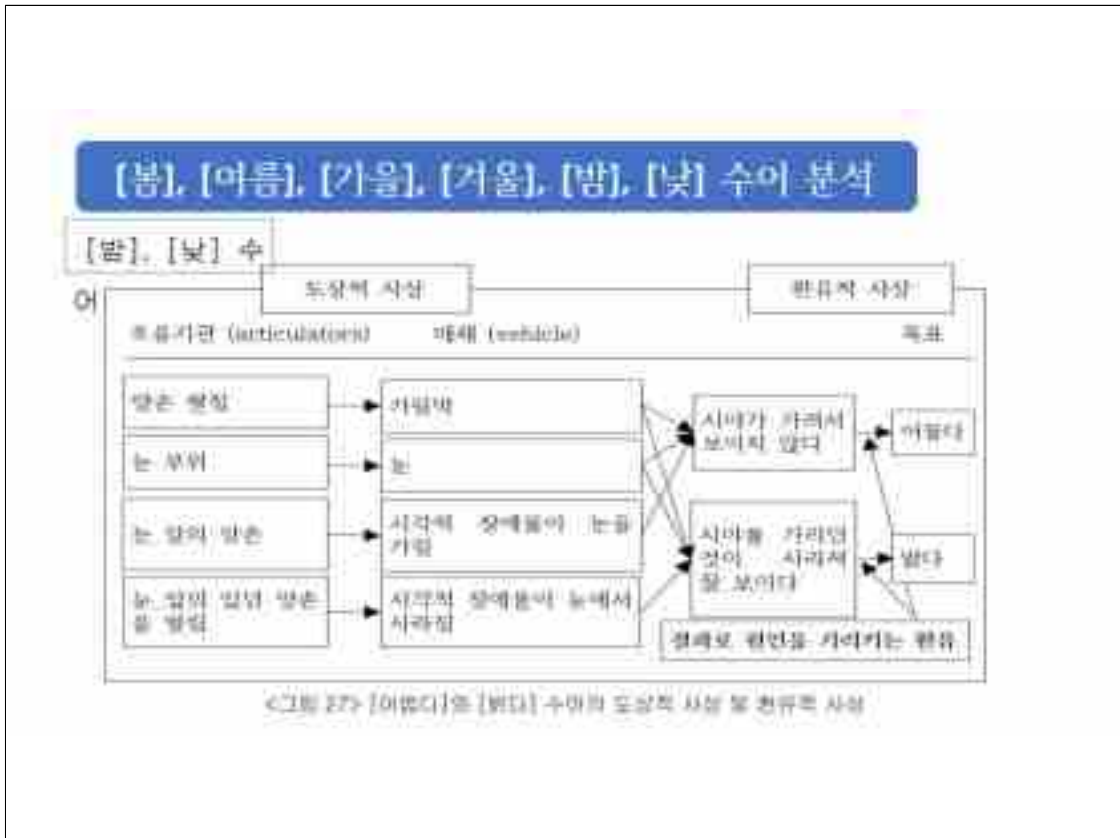
[뽕], [어름], [가을], [거울], [땀], [낮] 수어 분석

[뽕], [낮] 수

어

조음기관 (articulators)	매체 (vehicle)
양손 펼침	가림막
눈 부위	눈
눈 앞의 양손	가림막이 눈을 가림
눈 앞의 있던 양손을 벌	눈을 가리던 가림막 제거

<그림 26> [어름다]와 [뽕다] 수어의 도상적 사상



종합결과

수어	개념명	개념명 -> 행위	행위 -> 개념
몸	손가락 바탕잡음 가르코코 안무장	손가락, 손바닥으로, 정렬하는 '바탕잡음' 일러, 바탕잡음 기타지	바탕잡음 -> 몸
		바탕잡음, 바탕잡음, 가져가는 행위	바탕잡음, 바탕잡음, 가져가는 행위
아름	정렬이 정음 손으로 수어장면 어, 지칭	아름 이 다루는 것인, 정음으로 그, 정음인 수어를 가져감	정음 -> 아름
		정음과 정음을 가져가는 행위	정음으로, 정음을, 가져가는 행위
가을	손으로 바탕잡음 정음지	손가락, 손바닥으로, 정렬하는 바탕잡음, 바탕잡음, 바탕, 지칭 바탕, 바탕, 바탕잡음, '바탕' 정음을 가져감	바탕잡음 -> 가을
		바탕잡음, 바탕잡음, 가져가는 행위	바탕잡음, 바탕잡음, 가져가는 행위
아름	수어를, 정 음으로, 정음, 정	아름 이 다루는 것인, 정음으로 그, 정음인 수어를 가져감	정음 -> 아름
		정음과 정음을 가져가는 행위	정음으로, 정음을, 가져가는 행위
빛, 낮	손으로 바탕잡음 지칭 수어, 정음 정음, 정음 정음	바탕잡음, 바탕잡음, '바탕' 정음을, 정음, 정음으로, 바탕잡음, 지칭지칭, 수어를 가져감, 정음, 정음지칭을, 정 음지칭, 정음, 정음으로, 정음을 가져감	바탕잡음, 정음 -> 빛, 낮
		정음과 정음을 가져가는 행위	정음으로, 정음을, 가져가는 행위

결론

- 인지의미론의 기본 가정 중의 하나인 개념의 신체화에 대하여 많은 논의가 있었지만 대부분 정화어를 기반으로 한 것으로 수어에 관하여 이루어진 연구는 매우 드물다.
- 수어는 신체를 사용하여 의미를 전달하기 때문에 개념의 신체화가 수어에서 훨씬 더 풍부하게 나타난 것으로 예상되지만 수어에 대한 언어학적 접근, 특히 인지언어학적 접근이 미약하였기 때문에 개념의 신체화에 대한 수어 연구가 많지 않다.
- 본 연구는 한국수어의 '몸', '아름', '가을', '겨울'과 '빛', '낮'을 나타내는 수어를 분석하여 개념의 신체화에 대한 증거를 제시하였는데 그 의의를 갖는다.

Thank you!!

코 퍼 스

A Corpus-based Study of Specialized Vocabulary Lists for Maritime English

19th August, 2021

Lin Xu & Se-Eun Jhang
(Korea Maritime and Ocean University)

2021 KALS Online Summer Conference
The Korean Association of Language Sciences
(www.alakorea.or.kr)

Contents

1. Introduction
 2. Previous Studies
 3. Data and Method
 4. Results and Discussion
 5. Conclusions
- References

INTRODUCTION

Background & Purpose & Research Questions

This paper is part of Lin Xu's PhD dissertation titled "Developing vocabulary lists in specialized Maritime English corpora," which will be published in August, 2021.

1. Introduction: Research Background

❑ Vocabulary learning is important to language learning.

Wilkins (1972, p. 111) argues that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed". Vocabulary is essential in learning a second language since it carries "the main information load" (Read, 2004, p. 153) and "a particularly heavy functional load" of meanings (Jones, 1998, p. 144).

❑ Corpus-assisted vocabulary studies lead to the development of vocabulary/word lists.

Landscape vocabulary lists keep popping up, for example, *General Service List (GSL)*, *University Word List (UWL)*, *Academic Word List (AWL)*, *New General Service List (new-GSL)*, *Academic Vocabulary List (AVL)*, etc.

❑ Discipline-specific vocabulary lists are needed.

"It is wise to direct vocabulary learning to more specialized areas when learners have mastered the 2,000–3,000 words of general usefulness in English" (Nation, 2001, p. 187).

1. Introduction: Research Background

❑ Maritime English is important to ESP and EOP education, and communication in workplaces.

Maritime English can be seen as an effective and important communication tool to deliver useful information and reach a desirable instructing result in the maritime field.

❑ The study of maritime English vocabulary needs to be further developed.

Previous studies focus on maritime English teaching, literature translation, and stylistic features of different genres.

Vocabulary studies on maritime English are mainly in

- vocabulary size,
- discipline-specific vocabulary lists.

18

1. Introduction: Research Purposes

❑ Maritime Academic Word List (MAWL)

To meet the needs of discipline-specific academic vocabulary in Maritime English academic contexts.

❑ Maritime Technical Word List (MTWL)

To bridge the vocabulary gap between general vocabulary and discipline-specific vocabulary.

19

1. Introduction: Research Questions

Question 1. Which lexical items occur frequently in maritime academic texts?

Question 2. Which are the commonly-used technical words in a maritime English corpus?

Previous Studies

Vocabulary Studies & Maritime English

2. Previous Studies: Definitions

□ Academic vocabulary

- Academic vocabulary refers to the lexical items that are relatively frequent across a wide range of academic texts but are infrequent in other genres (Culford & Nation, 2001).

□ Technical vocabulary

- Technical vocabulary has become an established term to refer to a specific category of specialized words different from academic vocabulary, another group of specialized words (Nation 2013, Nation and Cheadle 2012).
- The broader view holds that technical words may occur in any subject (e.g. art, science, or engineering) and they include words of various types, ranging from those that are used almost exclusively in a subject area (e.g. *ballast*) to those that have a high-frequency in general language but are used with a subject-specific meaning (e.g., *port*) or are important concepts in a subject without a separate subject-specific meaning (e.g., *shipping*). In short, "technical vocabulary is subject related, occurs in a specialist domain, and is part of a system of subject knowledge" (Chung and Nation 2004: 253).

2. Previous Studies: Vocabulary Lists

Some influential general and academic vocabulary lists

Vocabulary types	Vocabulary lists	Brief description
General vocabulary	General Service List, GSL (West, 1953)	2,000 word families
	New General Service List, New-GSL (Brezina & Gollman, 2015)	2,494 lexemes
Academic vocabulary	Academic Word List, AWL (Culford, 2000)	570 word families
	Academic Vocabulary List, AVL (Gardner & Davis, 2014)	3,015 lexemes
	Academic Spoken Word List, ASWL (Dang et al., 2017)	1,741 word families

2. Previous Studies: Vocabulary Lists

Some previous studies on discipline-specific vocabulary lists

Academic word lists	Technical word lists
<ul style="list-style-type: none"> • Medicine: Lee & Liu 2018; Wang et al. 2000; Choi & Ge 2007; • Chinese Traditional Medicine: Hsu 2018 • Finance: Li & Qian 2010 • Engineering: Ward 2009 • Environment: Liu & Han 2015; • Agriculture: Nasa, 2012; Martinez et al. 2009 • Chemistry: Valpiani & Nanni 2015; • Nursing: Yang 2012 • Linguistics: Vengpravitth et al. 2009 • Textbooks: Coetzel et al. 2010; Nourzad et al. 2017 	<ul style="list-style-type: none"> • Business English: Hsu 2011 • Plumbing: Coetzel & Dunscheff 2018 • Engineering: Wilson-Teal 2017; Mahara 2010; • Environmental Science: Liu & Han 2015 • Finance: Tsagkos-Papavasiliou 2014 • Medicine: Coetzel and Quiro 2012 • Music: Wang & Picard 2013

11

2. Previous Studies: Key issues in developing vocabulary lists

Counting units

□ Types

- Each different word is counted as a type in a corpus.
- Types are suggested to word lists for beginners (spelling, recognizing words) (Carroll et al., 1971; Zeno et al. 1995).
- Word types are suggested in technical word lists (Nation, 2016).

□ Word families

- A word family is a base word with its inflectional forms and derivational forms.
- Bauer and Nation (1993) Word family levels.
- Learners have difficulties in derivational forms (Level 6 word families) (Schmitt & Zimmerman, 2002; Nagy & Tomazek, 2012; McLean, 2018).

12

2. Previous Studies: Key issues in developing vocabulary lists

□ Lemmas

- A lemma is a base word with part of speech and its form with inflectional suffixes.
- Lemmas (and above Level 2 word families) are suggested for receptive knowledge (reading and listening) (Nation, 2016:).
 - ✓ Display the important information for vocabulary learning—the part of speech.
 - ✓ Allow learners to clearly focus on the vocabulary items in order to ease the burden. (Gardner & Davies, 2014; Lei & Lin, 2016)

□ Flemmas

- The word “flemma” can be divided into “f” and “lemma”; “f” in flemma stands for family (Pinchbeck, 2014).
- A flemma is a base word and inflectional forms, regardless of the part of speech.
- Flemmas are better than word families, but it is not proved that flemmas are superior to lemmas (McLean, 2018; Stoeckel et al., 2018; Brown et al., 2020).

75

2. Previous Studies: Key issues in developing vocabulary lists

Stop words and overlaps

□ The exclusion of general words from specific word lists (e.g., AWL).

- The rationale behind such a practice is the assumption that language learners generally grasp high-frequency words before they learn academic words and low-frequency words.
- There are studies (e.g., Cobb, 2010; Gardner & Davies, 2014) show that some AWL items were among the most frequent words in BNC and COCA.

□ Vocabularies overlap in general, academic and technical word lists.

- General words have special or different meanings in the academic context and some high-frequency general words actually have a much higher frequency in academic English than in general English (Gardner and Davies's study, 2014).
- As reported in Nation (2013), there is 20.6% technical vocabularies in an academic textbook. Of the 20.6% technical vocabularies, 9.2% are shared in the GSE, 6.2% are in the AWL, and only 5.2% are not from the above two groups.

76

2. Previous Studies: Maritime English vocabulary

Studies on maritime English vocabulary

- ❑ multi-words compounds (Lee, 2016, PhD dissertation, KMOU).
- ❑ collocation analysis on the sea-related words (Yi, 2018, PhD dissertation, KMOU).
- ❑ near-synonyms in maritime English (Li, 2017, PhD dissertation, KMOU).
- ❑ lexical features of maritime English (Byun, 2011, MA thesis, KMOU).
- ❑ maritime safety words (Chang, 2007).
- ❑ vocabulary in marine news (Gao, 2014).
- ❑ conjunctions in marine English (Lu & Guo, 2009), and discourse markers in maritime English (Wu, 2014).

13

2. Previous Studies: Maritime English vocabulary

Maritime English corpora

- ❑ Hong and Jiang (2010) developed the Maritime English Corpus (MECC) (1 million) for the purpose of serving EFL teaching.
- ❑ Lee (2016) compiled a 4 million word maritime English corpus (MEC) consisting of academy, news, laws, and textbooks.
- ❑ Yi (2018) extracted his data from a 1.4 million-word corpus including safety at sea, shipping news, navigational and marine engineering technology, laws, rules and regulations and documents in all the related areas of maritime transportation.
- ❑ Li (2017) compiled a 2.5 million word specialized maritime English Corpus.
- ❑ Byun (2011) used a 1 million maritime English corpus, containing six kinds of maritime written texts such as journals, maritime law, media text, sea language, textbooks and other documents.
- ❑ And there are some other specialized corpora (e.g. Qi's corpus of marine accident investigation reports (2019), Liu's corpus of MARPOL (2020), Xu and Jiang's corpus of charter parties (2020)).

14

DATA & METHOD

Data & Instruments & Procedures

3. Data and Method: Corpora

Study corpora: Maritime English Corpus (MEC)

Descriptive statistics of MEC

Statistical items	Numbers
texts	1,030
tokens (counting words) in texts	8,896,348
tokens used for word list	8,320,812
types (distinct words)	129,077
type-token ratio (TTR)	1.29%
standardized TTR	33.56%
aid. div.	67.10

Descriptive data of sub-corpora in MEC

Sub-corpora	Texts	Tokens	Percentage (%)
Committee Resolutions (CRes)	246	1,782,382	18.111
Conventions and Regulations (CReg)	10	771,210	7.793
Journal Articles (JA)	744	5,825,872	58.429
Media Texts (MT)	264	1,786,042	18.347
Special Guides (SG)	249	1,842,376	18.618
Textbooks (TB)	20	1,880,611	19.003
Total	1,030	8,896,348	100

3. Data and Method: Corpora

Study corpora: Maritime English Corpus (MEC)

Sub-corpus	Brief Description
Committee Resolutions	IMO committee resolutions: 19 texts LEG-Regal committee resolutions: 3 texts MEPC (Marine Environment Protection Committee): 14 texts MSC (Maritime Safety Committee): 82 texts PALS (Pollution Committee) resolutions: 5 texts
Conventions and Regulations	Convention on the International Regulations for Prevention Collisions at Sea-1972; International Convention for the Control and Management of Ships' Ballast Water and Sediments-2004; Ship Recycling Convention; MAJPCOL; SOLAS; The Maritime Labour Convention; United Nations Conventions on the Law of the Sea, etc.
Journal Articles	From year 2010-2020, 244 texts from 13 academic journals in the maritime field.
Media Texts	From year 2010 to 2020. Mainly from maritime reports, IMO alert news, Maritime index news, World marine news, and some other maritime media websites.
Special Genres	Maritime academic arrangement reports, from 2010 to 2020, 83 texts Charter parties, 136 texts
Textbooks	26 texts including subjects of navigation, marine engineering, transportation and logistics, shipbuilding and design, maritime law, maritime economics, and management etc.

18

3. Data and Method: Corpora

Study corpora: Maritime Journal Article Corpus (MJAC)

Descriptive statistics of MJAC

Statistical item	Numbers
texts	478
tokens (running words) in text	3,548,028
tokens used for word list	3,204,532
types (distinct words)	54,670
type/token ratio (TTR)	1.71%
standardized TTR	30.35%
mi.doc	63.67

19

3. Data and Method: Corpora

Study corpora: Maritime Journal Article Corpus (MJAC)

List of specialized areas and journals in MJAC

Specialist areas	Journals
Policy and Management	<i>Marine Policy</i> , <i>Maritime Policy and Management</i> , <i>Oceans and Coastal Management</i>
Maritime Law	<i>Journal of Maritime Law and Commerce</i> , <i>The Tulane Maritime Law Journal</i> , <i>Loyola Maritime Law Journal</i>
Economics and Business	<i>Maritime Business Review</i> , <i>Maritime Economics and Logistics</i>
Transportation and Logistics	<i>Transportation Research</i> , <i>Journal of Transport Geography</i>
Marine Structures and Engineering	<i>Marine Structures</i> , <i>Offshore Engineering</i>
Navigation	<i>The Journal of Navigation</i>

21

3. Data and Method: Corpora

Study corpora: Maritime Journal Article Corpus (MJAC)

Descriptive statistics of sub-corpora in MJAC

sub-corpora	texts	tokens in text	tokens in word list	types	TTR	RTTR	std.dev.
Policy and Management	33	578,439	272,210	21,457	3.37	39.24	31.49
Maritime Law	23	388,477	148,143	17,871	3.23	36.72	33.28
Economics and Business	52	398,987	142,298	18,233	3.38	38.98	32.33
Transportation and Logistics	79	248,288	142,228	18,387	3.4	36.77	32.81
Marine Structures and Engineering	78	348,987	117,488	11,203	2.95	34.18	31.32
Navigation	101	228,488	100,225	18,172	3.27	34.78	34.81

22

3. Data and Method: Corpora

Reference corpus: BNC

Descriptive statistics of the subsets in BNC

	BNC (written)	BNC non-academic	BNC academic
tokens in text	88,639,272	72,672,904	15,966,370
tokens used for word list	87,190,400	71,572,616	15,017,783
types	329,391	302,484	120,485
TTR	0.39	0.42	0.57
STTR	43.25	43.97	38.95
std.dev.	57.50	55.54	58.73

23

3. Data and Method: Instruments

Stanford CoreNLP

- Stanford CoreNLP is a natural language processing toolkit.
- POS-tag the words in MJAC, and lemmatize the tokens in the word list.

WordSmith Tools 8.0

- A software designed for looking at vocabulary behaviour in texts.
- WordList and KeyWords tools were used to extract high-frequency words and keywords.

Technical words checklist

- A rating scale (Chung & Nation, 2004) to evaluate the degree of technical words.
- A revised version with three levels (Tongprasoo/Pattanasom, 2018) was used.

No.	Words	Level 1	Level 2	Level 3
1	history			
2	business			
3	length			
...	...			

24

3. Data and Method: Procedures

Extraction of the maritime academic words

- ❑ Stanford CoreNLP was used to POS-tag and lemmatization of words in the raw texts.
- ❑ Lemmas are used as the counting unit in the list.
- ❑ Selection criteria: a mix of AWL's and AVL's with some modifications.
 - Minimum frequency: 28.57 times per million (Conhead, 2000), or least 95 times in the present study.
 - Dispersion: DP (Orris, 2006), below 0.7.
 - Range ratio: 70% (Conhead, 2000).
 - Discipline measure: a lemma should occur less than three times the expected frequency in any of the six specialist areas (Gardner & Davies, 2014).
 - Frequency ratio: 1.5 (Gardner & Davies, 2014), which means a lemma must appear at least in the academic corpus 50% more times than in the non-academic corpus.
- ❑ Proper nouns, and abbreviations and acronyms were excluded.

25

3. Data and Method: Procedures

Extraction of the maritime technical words

- ❑ WordSmith Tools was used to identify keywords based on frequency (Scott, 1997) and dispersion (Egbert & Hiber, 2019).
 - **Frequency-based keywords:** a word with unusual frequency compared with its frequency in the reference corpus.
 - **Dispersion-based keywords:** a word disperses unusually more or less in the texts of the target corpus than it does in the reference corpus.



26

3. Data and Method: Procedures

Extraction of the maritime technical words



Examples of frequency and sex dispersion keyword lists

27

3. Data and Method: Procedures

Extraction of the maritime technical words

- ❑ BIC scores were used to assess keyness. BIC scores range from 0 to 10: below 0 not trustworthy; 0-2 only worth a bare mention; 2-6 positive evidence; 6-10 strong; more than 10 very strong (Scott, 2010). Only keywords with BIC scores above 2, indicating the positive keyness values, can be included into the keyword lists.
- ❑ The results of the two keyword lists were then synthesized; only common shared keywords were kept in the preliminary technical word list to guarantee both frequency and dispersion keyness criteria.
- ❑ The preliminary technical word list was ruled by five Maritime English experts using the revised technical words checklist.
- ❑ Words that ruled as in Level 4 are believed not related to the maritime field and then were excluded from the MITWL.
- ❑ To better serve the pedagogical purpose, the technical word list was divided into two sub-lists: the MITWL and a supplementary list of proper names, and acronyms.

28

4. Results and Discussion : MAWL

- The percentage of CEFL 2,000 word families in MAWL and AVL are different, but the number of word families are similar. This result indicates that there are high-frequency general words appearing frequently in academic texts of different domains, including maritime disciplines.
- There are fewer AVL word families in MAWL than in AVL. This result indicates that maritime academic texts bear its own features in vocabulary. Not all general academic words are frequently used in the maritime academic context, which makes MAWL more targeted.
- Including various disciplines, AVL incorporates more CEFL lists vocabularies than MAWL, which indicates that MAWL are more targeted to specialized learners and users. Thus, MAWL are more effective than general academic word lists from pedagogical perspective.

Range analysis on MAWL, AVL, and AVL

Base word lists	MAWL		AVL		AVL	
	word families	word families (%)	word families	(%)	word families	(%)
pe_1st_1000	441	27.38	310	14.73	-	-
pe_2nd_1000	114	13	155	6.2	-	-
pe_3rd_1000	389	33.89	497	20.42	178	1.88

31

4. Results and Discussion : MTWL

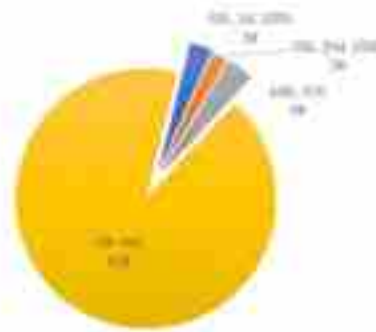
Maritime Technical Word List

- In order to better serve the pedagogical purpose, then the technical word list was divided into two sub-lists: the MTWL and a supplementary list of proper names, abbreviations, and acronyms.
- MTWL contains 527 word types, and the supplementary list contains 124 acronyms, abbreviations and proper names.
- The MTWL (527 words) covers 3.64% tokens in the MEC. If the supplementary lists of acronyms and proper names are added, the lexical coverage of the MTWL in the MEC increases to 4.23%.

32

4. Results and Discussion: MTWL

- The biggest proportion of the MTWL falls out of the GSL, high-frequency word lists and the AWL.
- There are overlapping vocabularies among technical vocabulary, general vocabulary and academic vocabulary.



Largest composition of GSL and AWL in MTWL

23

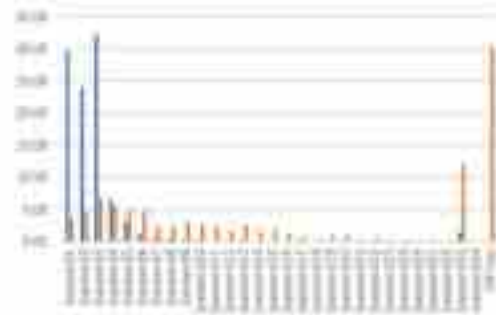
4. Results and Discussion : MTWL

- While these shared high-frequency technical words are the ones most commonly used in the Maritime English context, for example, *vessel, fishing ships, coast, bays, officer, float, loading, unloading, engineer, harbour, voyage*.
- One thing should be noted is that some of the MTWL and GSL shared words have different connotations in the maritime context and in the general English context, for example, *flag* in *admiral flag, answering flag, code flag, convenient flag ship, flag chest, flag locker, etc.* In most cases, these shared words keep their frequently used meanings in general English contexts, but the combination with other words makes them particular in the maritime technical context.

24

4. Results and Discussion: MAWL vs. MTWL

- Comparing with the MAWL, the MTWL has a bigger vocabulary size, ranging from the base word list 1 to 27, plus some compound words
- MAWL includes more high-frequency general words (top 2,000)
- These results are possibly due to the totally different extraction methods: MAWL includes high-frequency general words in it, while keyword analysis based on frequency and dispersion helps to exclude most of the shared high-frequency words in both general and technical contexts.

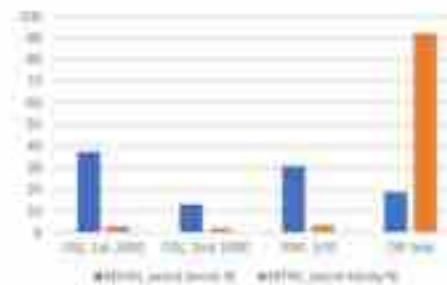


Comparison of vocabulary size between MAWL and MTWL

28

4. Results and Discussion: MAWL vs. MTWL

- The MAWL contains more general vocabulary, as well as general academic vocabulary, which allows the MAWL greater vocabulary coverage (about 23.89%) in the MEC compared with the coverage of the MTWL in the MEC (3.64%).
- In contrast, the vast majority of lexical items in the MTWL do not belong to high-frequency general vocabulary and are distinct from specialized vocabulary in other subject areas. This comparison reflects the features of technical vocabulary: that technical vocabulary is used less frequently and has a narrower coverage than academic vocabulary.



Comparison of MAWL and MTWL composition

29

4. Results and Discussion : MAWL vs. MTWL

- About 22% of the shared words fall into the group of the AWL 570 word families, followed by the GSL first 1,000 (about 19%) and in the GSL second 1,000 (about 10%).
- The biggest proportion of the shared words are off the lists of GSL and AWL.
- 55 shared words (word families) make 4.59 % in MAWL and 11.39% in MTWL.

GSL and MAWL and MTWL shared words

bullet	page	control	interface
boating	release	effort	radio
boat	transmit	board	average
charge	scene	propeller	weight
configuration	service	wheel	multimedia
community	engine	engine	traffic
fuel	warning	clutter	surface

MAWL and MTWL shared words in GSL and AWL

GSL 1st 1000	GSL 2nd 1000	AWL 570
can	bat	submarine
cast	engine	super
effort	surface	radio
fuel	bat	weight
in	engine	interface
in	weight	super
in		radio
interface		weight
power		interface
top		radio
weight		radio

31

CONCLUSIONS

Major Findings & Implications & Limitations

32

5. Conclusions

Answers to research questions

1. Which lexical items occur frequently in maritime academic texts?

Based on frequency, the maritime academic word list of 1,684 lemmas was generated. And it covers about 34% vocabulary in maritime academic texts. Compared with general academic word lists, the MAWL has a better coverage in maritime academic texts, and the vocabulary items were presented in lemmas, which is believed to better help learners in maritime English vocabulary learning, especially for lower-proficiency learners.

2. Which are the commonly-used technical words in a maritime English corpus?

With the hybrid method of keyword analyses (frequency-based & text-dispersion-based) and technical words checklist, the MTWL of 527 words and a supplementary list of 124 acronyms, and proper names were generated, which has about 4% lexical coverage in MEC. It could bridge the gap between general vocabulary and discipline-specific vocabulary in maritime English.

38

5. Conclusions

Pedagogical Implications

- ❑ The maritime English word lists can help learners, teachers, and syllabus designers focus their attention and energy effectively on the most useful vocabulary items.
- ❑ The MAWL and MTWL provide teacher and syllabus designers with objective evidence on which words are most likely to be encountered by learners in the maritime written context.
- ❑ Combining with general high-frequency word lists, the MAWL and MTWL can be used as base word lists to evaluate textbooks and other teaching materials.
- ❑ The word lists could also be used in language tests to assess learners' vocabulary competence in the maritime field.

40

5. Conclusions

Limitations of the present research

- **Overlap in academic and technical word lists**
It is hard to tell which types of vocabulary the overlaps should be comprised in.
- **Homonymy**
Both the meaning units of lemmas and types could not distinguish homonyms.
- **Technical words rates**
Only two maritime English experts were involved.
- **Multisyllable words**
It was found during the research that many of the technical terms are multisyllable words that are made up of general words.
- **Spoken data**
Due to the limited volume of spoken data, it is difficult to balance the written data and spoken data. Therefore, although a small part of spoken data was collected, they were not included in the MEC.

Suggestions for future research

- Although most agreed that there are overlaps in academic, technical and general vocabulary, some new criteria are needed to simplify and specify the vocabulary lists.
- More advanced research methods are needed to working out this problem to further elaborate on vocabulary lists.
- More participants from different institutions specialized areas would possibly improve the validity of the results.
- A list of technical multisyllable words could possibly facilitate the vocabulary learning of maritime English. Therefore, further studies with more focus on technical multisyllable words are suggested.
- In the future, more spoken data of maritime English are urgently needed to add to the compilation of maritime English corpora.

41

REFERENCES

- Barr, L., & Sains, F. (1995). *Vocabulary*. International Journal of Lexicography, 18, 223-239.
- Bond, V., & Dhillon, S. (2012). Is there a core general vocabulary? Describing the core general lexicon. *Applied Linguistics*, 33(1), 1-22.
- Brown, D., & Fraser, S. (2006). The most appropriate word set for L2 vocabulary research and pedagogy: A test of core of the lexicon. *Applied Linguistics*, 27(4), 691-709. <https://doi.org/10.1017/S0022268906000000>
- Chen, B. J. (2017). *Maritime English: a corpus-based lexical analysis*. (Linguistic Series, Korea Maritime and Ocean University).
- Conrad, J. B., Brown, P., & Richards, B. (1971). *The American seaman's word*. (Newark: New York: Harper & Row, Inc.). (Series: American Heritage).
- Chen, M. (2017). A corpus-based lexical analysis of maritime English in IMU documents: Application of vocabulary analysis in Maritime English corpora [Hainan thesis]. (Hainan Normal University).
- Georg, T., & Sains, F. (2008). Handling related vocabulary. *Lexicon*, 21, 211-243.
- Giles, T. (2002). Learning third language with business film content program. *Reading in a Foreign Language*, 21, 181-204.
- Conrad, S. (2005). *Learn another word list*. (2005). (January). 34, 233-236.
- Conrad, S., & Vance, F. (2001). The specialized vocabulary of English for academic purposes. In T. Flower, & J. Flower (Eds.), *Research perspectives on English for academic purposes* (pp. 172-207). Cambridge: Cambridge University Press.
- Conrad, S., Brown, J., & Tait, J. (2010). My weight is under: word methods to address context? *Web Journal of English as Applied Linguistics*, 2, 17-21.
- Conrad, S., & Quinn, B. (2012). Investigating a criteria vocabulary list to increase medical students' IELTS scores. *IELTS Academic Journal*, 2(1), 23-41.
- Conrad, S., & Dorschner, M. (2013). Investigating the technical vocabulary of pharmacy. *English for Specific Purposes*, 31, 64-78.
- Green, E., & Pech, A. (2010). Academic vocabulary in IELTS studies papers: A corpus based study. *Journal of English for Academic Purposes*, 21, 103-124.
- Dang, T. N. T., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 31, 68-76.
- Dang, T. N. T. (2016a). The best source spoken word list. *IELT—International Journal of Applied Linguistics*, 7, 44-71.
- Dang, T. N. T. (2016b). The status of vocabulary in academic spoken word list and its influence. *English for Specific Purposes*, 37, 36-42.
- Dang, T. N. T., Conrad, S., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 4, 293-307.

42

REFERENCES

- Elyot, J., & Bhat, J. (2018). Assessing the discriminability of word pairs: Corpus-based English learning language processing and linguistics. *PLoS*, 13(12), 1-14.
- Gu, S. (2011). A corpus-based study on lexicology extraction of Chinese English (Master thesis, Dalian Maritime University).
- Gutierrez, D., & Davis, H. (2018). A new machine readability test. *Applied Linguistics*, 39, 303-321.
- Guo, S. D. (2006). Disposition and cultural differences in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-421.
- Guo, Y. (2016). A corpus-based study on the lexical features of Chinese English (Master thesis, Dalian Maritime University).
- Hag, S. C., & Song, S. K. (2019). The readability of a Korean English corpus for EFL learners. *Asian Journal of English Language and Linguistics*, 18, 860-891.
- Hag, S. K., Park, H. M., Dwyer, T., & Lee, S. (2020). Design and readability of global academic journals using a lexical semantic network. *International Journal of Machine-Aided Foreign Language Studies*, 17(2).
- Lee, S. H. (2008). Lexical features of Chinese English corpus with multi-word compound: A lexical semantic network-based semantic network. *International Journal of Machine-Aided Foreign Language Studies*.
- Lee, J., & Lee, D. (2016). A new lexical analysis tool for a corpus-based study: multi-lexical methodology. *English for Academic Purposes*, 55, 42-57.
- Lee, J., & Park, J. (2017). A corpus-based environmental analysis tool for building and its readability test. *English for Specific Purposes*, 58(2), 1-11.
- McLean, S. (2018). Evidence for the adoption of the British as an appropriate word reading test. *Applied Linguistics*, 39(6), 825-840.
- Milroy, Y. L. (2015). The readability of English texts: population-specific studies in English. *English for Specific Purposes*, 36, 28-44.
- Nelson, P. (2000). *How to use a readability formula for writing and editing*. The Creative Studies Language Series, 4(1), 1-4.
- Nelson, P. (2011). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
- Nelson, P. (2016). *Adding and using word lists for language learning and testing: What Researcher's Platform*. <https://doi.org/10.1017/9781107301308>.
- Nelson, P., Davidson, A., Cheng, T. H., & Davis, S. (2018). *Spoken word lists: A R Package for building and using word lists for language learning and testing* (pp. 101-110). In *A Researcher's Platform*. <https://doi.org/10.1017/9781107301308>.
- Nelson, P., & Pared, E. (2018). *How to use and put them to use*. In T. Nelson (Ed.), *Adding and using word lists for language learning and testing* (pp. 61-74). *A Researcher's Platform*. <https://doi.org/10.1017/9781107301308>.

42

REFERENCES

- Nelson, P., & Pared, E. (2018). *How to use and put them to use*. In T. Nelson (Ed.), *Adding and using word lists for language learning and testing* (pp. 61-74). *A Researcher's Platform*. <https://doi.org/10.1017/9781107301308>.
- Paulsen, D. D. (2014). *Lexical frequency: profiling of a large corpus of Canadian high school English exam responses*. *Reading: 1-1* and 1-2. *Canadian Journal of Applied Linguistics*. Montreal, QC, USA.
- Paulsen, D. (2016). *Research in reading instruction*. *Journal Review of Applied Linguistics*, 33, 146-158.
- Paulsen, D. (2017). *Wordlist Tools to create a list*. *Research Library*. <https://www.researchlibrary.com/wordlist-tools/>.
- Reynolds, T., Jahn, T., & Reppert, F. (2019). Is the lexicon more representative than the lexicon in a word reading test? *Applied Linguistics*, 40(6), 881-894. <https://doi.org/10.1017/S0272261919000022>.
- Song, J., Song, S., & Gu, S. (2018). English learners' lexical word list. *English for Specific Purposes*, 57(5), 402-419.
- Song, S., & Pared, E. (2017). A word list for the readability of English: A corpus-based study. *Asian Journal of English Language Studies*, 11, 113-121.
- Wang, J. (2018). A new measuring English word list for less proficient students measuring vocabulary. *English for Specific Purposes*, 58(2), 175-182.
- West, M. (2011). *A general lexicon list of English words*. London, UK: Longman.
- Wilkins, D. A. (1972). *Lexicology in language teaching*. Cambridge: CUP Press.
- Yu, Q. (2014). A corpus-based study on discourse markers in the conversation of Chinese English (Master thesis, Dalian Maritime University).
- Yu, L., & Song, S. S. (2020). Keyword analysis of English learner papers. *English for Specific Purposes*, 71(2), 267-288.
- Xiao, G., & Nelson, P. (1984). A readability word list. *Language Learning and Communication*, 1, 213-228.
- Yang, M. (2015). A corpus analysis word list. *English for Specific Purposes*, 37, 27-38.
- Yu, Y. (2015). A corpus-based study on lexical features of Chinese academic writing papers (Master thesis, Dalian Maritime University).
- Yu, Y. (2018). *A corpus-based study on the lexical features of Chinese academic writing papers*. *International Journal of Machine-Aided Foreign Language Studies*.
- Zou, X. H., Zhou, X. H., Nishida, T., & Dorman, S. (1993). *The lexicon: a word frequency guide*. Routledge Applied Foreign Language Series.

43



THANK YOU.



Functional Analysis of Lexical bundles in rhetorical move of research article abstracts in Applied Linguistics

Xiaolei LIANG & Se-Eun Jhang
(Korea Maritime and Ocean University)

2021 Online Winter Conference
The Korean Association of Language Sciences
19th of August, 2021



Contents

1. Introduction
- Purpose of the Study /Research Background /Research Questions
 2. Literature Review
 3. Corpus and Methodology
 4. Results and Discussion
 5. Conclusion
- References



1. Introduction

- Purpose of the Study

- This comparative study attempts to explore how lexical bundle functions help to structure rhetorical moves in research article(RA) abstracts published by professional writers in international journals and Chinese Journals;
- in order to establish possible connections between lexical bundles and research article(RA) rhetorical moves in applied linguistic abstracts by L1 and L2 writers.
- This paper also examines the frequent used lexical bundles in rhetorical moves of RA abstracts.



- Research Background

- Genre analysis has been a popular research method in ESP and EAP. Many studies have been conducted under the genre analysis framework. Multiword unit has been examined under the different terminologies in different genre or registers for the past decades.
- Biber *et al.* (1999) defined lexical bundles as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status," simple sequences of words that commonly co-occur in natural language use.
- We have seen a great number of studies on this topic. Although studies have revealed different features in RA research, multiword units in each RA abstract move have been less explored.



- Research questions

- This present study aims to explore relationship between rhetorical moves in RA abstracts lexical bundle functions in each rhetorical move in RA abstracts by L1 and L2 writers.

So this present study will address the following research questions:

- **Q1:** What are the **frequent lexical bundles** and **lexical bundle functions** can be identified in association to each move in applied linguistic RA abstracts?
- **Q2:** can we find any **different preferences of lexical bundle functional categories** in the RA rhetorical moves of applied linguistic RA abstracts between L1 and L2 corpus?



2. Literature Review

- Identification of lexical bundles

- Four-word bundles have been an established criterion in bundle research for structural and functional analysis because they are representative and relatively manageable in manual intensive reading and coding (Cortes 2004; Hyland 2008a)
- Frequency thresholds for bundles have changed over time, from 10 per million words (pmw) (Biber et al. 1999) to 20 or 40 occurrences pmw (Biber, Conrad and Cortes 2004; Cortes 2004)
- Cut-off point for dispersion is that that bundles should occur in at least five different texts (Biber et al. 1999) or 10% of texts (Hyland 2008a)



- Lexical bundle studies

- Studies have been done on bundle variation, esp. in structures and functions between academic prose and other registers
- In recent years, bundle studies have been directed to the exploration of typical bundle structures and functions in different disciplines to help learners with their academic writing in a specific field.
- These studies have discovered that bundle use might differ across disciplines, especially in bundle numbers, structures, and functions (Cortes 2004; Hyland 2008b).
- Most recently, Hong (2020) compares the bundle-move connection in introductions of published RAs produced by L1-English writers with those of term papers produced by same Korean EAP students at two different academic levels.
- However, literature shows that there is need to explore the connection between lexical bundles features and rhetorical moves in RA's **abstracts** in Applied Linguistics, which is the main reason for this present study.



3. Corpus and Methodology

3.1 Corpus

- A RA abstract corpus was built out of RA abstracts by L1 and L2 professionals. See Table 3.1 for corpora information;
- L1 and L2 corpus contains 604 and 647 RA abstracts randomly selected from privileged international journals and Chinese Journals;
- The RAs were published from 2005 to 2015,
- L1 and L2 have 105,238 and 82,279 words, respectively
- L1 and L2 have 2,399 and 2,089 moves in total.

Table 3.1 Corpus information

Corpus	Text	Words	Move
L2	647	82,279	2,089
L1	604	105,238	2,399
Total	1,251	187,517	4,488

3.2 Analysis framework and data process procedure



- Swales (1990) and Santos (1996)'s five-move scheme will be used as move analysis framework in the present study.
- A typical five-move model divides an abstract into 'situating the research', 'presenting the research', 'describing the methodology', 'summarizing the results' and 'discussing the research'; see Table 3.2
- Pan, Reppen & Biber (2016, 2020) and Hyland (2008a)'s bundle functional classification will be employed for bundle function analysis in the present study.
- The scheme classifies 4-word lexical bundles into three functional categories: Research-oriented, Text-oriented, and Stance-oriented bundles
- See Table 3.3 Bundle functional classification scheme



Table 3.2. Move model for tagging

Move	Tag	Functions
Background	<STR>	Situating the research
Purpose	<PTR>	Presenting the research
Method	<DTM>	Describing methodology
Results	<STF>	Summarizing the findings
Conclusion	<DTR>	Discussing the findings



Bundle selection criteria and functions

- Bundle selection standard followed Biber et al. (1999, 2004) and Chen & Baker (2010) classification.
- AntConc 3.5.0 (Anthony, 2017) was used to retrieve all the 4-word lexical bundles in each sub-corpus respectively; Wordsmith 8.0 was also used to analysis bundle data.
- The range of N-grams sets at 5, which means each bundle should appear at least in 5 different texts.
- Those 4-word bundles that appeared at least 5 times in each corpus were taken out for further analysis.
- All the bundles were classified using bundle functional classification scheme (Pan, Reppen & Biber, 2016, 2020; Hyland, 2008a, 2008b); see Table 3.3

Table 3.3 Bundle functional classification scheme



TAGs:	
Research-oriented:	activities and experiences
R1 Location:	of the time of
R2 Procedure:	were randomized to receive
R3 Quantification:	in a ratio to
Text-oriented:	
T1 Relational signals:	these findings suggest that
T2 Structuring signals:	little is known about
T3 Framing signals:	on the basis of
Stance-oriented:	
S1 Epistemic modality:	were more likely to
S2 Deontic modality:	research is needed to (degree of obligation)

Adopted from Pan, Reppen & Biber (2016, 2020); Hyland (2008a, 2008b)

4. Results and Discussion



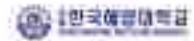
Research Question 1: frequently occurred bundle and its functional distribution in each move

The most frequently occurred 4-word bundles in L1 and L2 corpus were: *in this paper we* (L1 only), *this paper reports on*, *the results of the*, *the results suggest that*, etc.

In each move, we saw (Figure 1 & Figure 2):

- M1: L1 used all three functional categories (R:28%,S:44%,T:28%), L2 used only research-oriented (23%) and stance-oriented bundles(67%);
 - Frequently occurred lexical bundles: *an important role in*, *little is known about*, *research has shown that*, *the extent to which* ;
- M2: both L1 and L2 used more research-oriented bundles(90%) ; No Stance-oriented ones found in this move;
 - Frequently occurred lexical bundles: *in this paper we*, *this paper reports on*, *this study was to*, *this study investigates the* .

Research Question 1 (Cont'd)

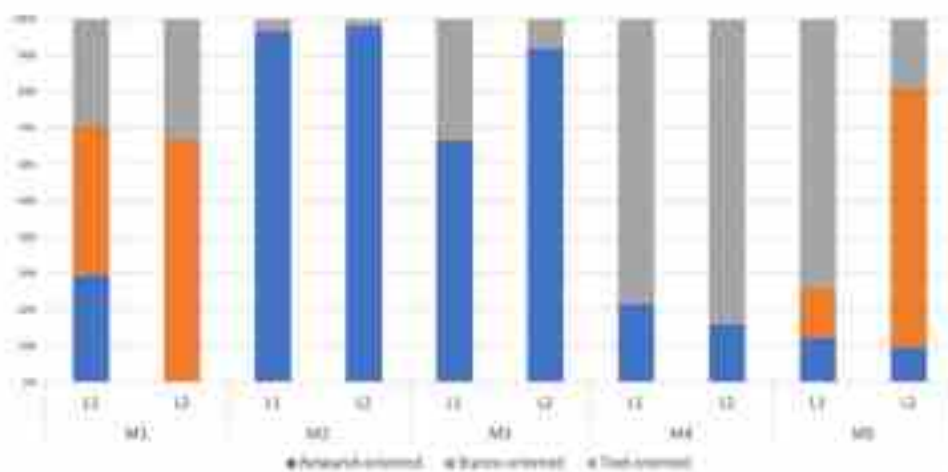


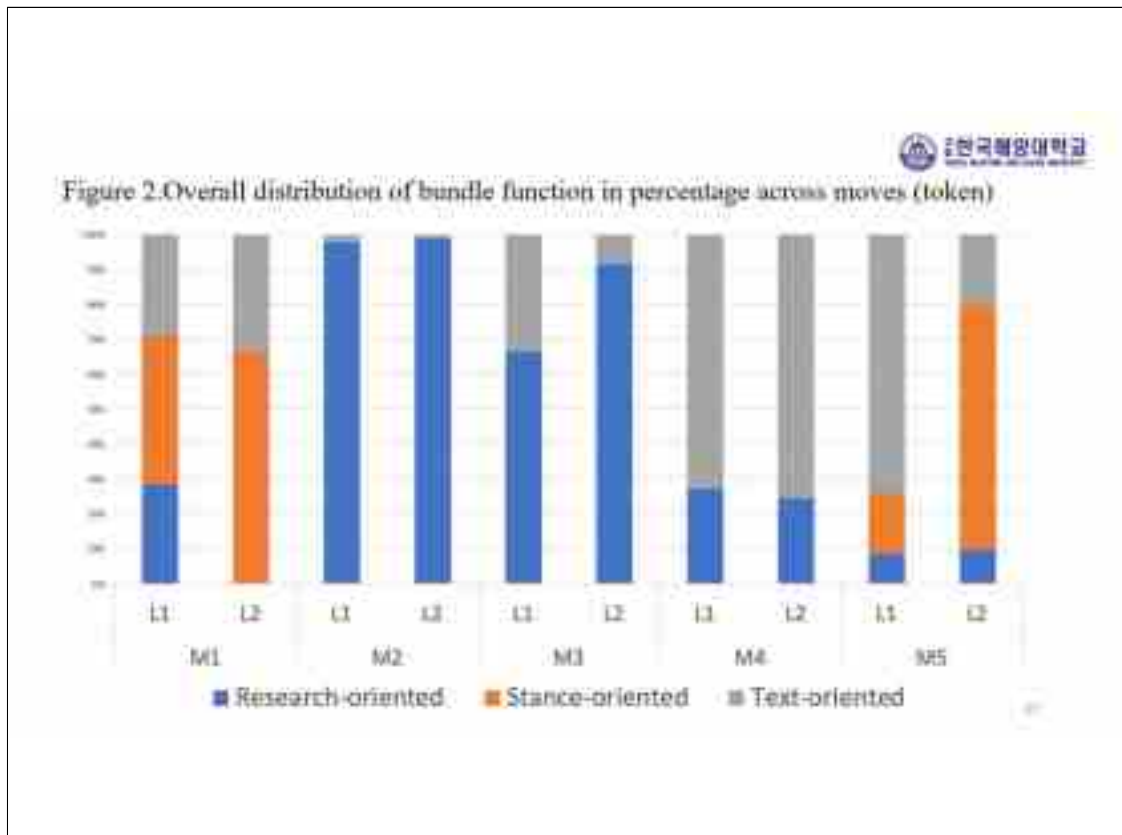
- M3: both writer groups prefer more research-oriented (L1:67%, L2:92%)
 - Frequently occurred lexical bundles: *analysis is based on, and the use of, in terms of the*


- M4: saw more text-oriented bundles both in L1 (72%) and L2(76%)
 - Frequently occurred lexical bundles: *the findings show that, the results of the, it was found that, the results suggest that*

- M5: L1 favors more Text-oriented (73%) while L2 interested more in Stance-oriented bundles (70%)
 - Frequently occurred lexical bundles: *these findings suggest that, implications of these findings, discussed in relation to*

Figure 1. Overall distribution of bundle function in percentage across moves (type)





Research Question 2: Comparison of bundle functional differences between L1 & L2  한국영양대학교

Move 1 is to give the background of the study or research.

- statistically significant differences in bundle token, indicating that L2 group used less research-oriented bundles than L1 did; see Figure 4.
- No statistical differences were found in bundle type; see Figure 3
- L2 writer group used less bundles than L1 writers in terms of total number and variety.

Move 2 describes the objective /purpose/aim of the research.

- No significant differences in bundle function were found in type & token
- Research-oriented bundles were heavily used by L1 and L2 writers

Figure 3. Chi-square test for functional distribution in each move (types)



Category	Type	Move	M1	M2	M3	M4	M5	F	p	
Reason-orientated	L1	Observed	2	49	8	8	2	65	134.00	p = 0.004 df=4 z-squared 7.66
		Expected	0.65	48.34	5.55	0.47	0.98			
		Residuals	1.67	0.09	0.19	-1.13	1.03			
	L2	Observed	0.00	99.00	11.00	23.00	1.00			
		Expected	1.35	99.66	11.45	19.55	2.02			
		Residuals	-1.16	-0.07	-0.13	0.79	-0.72			
Stance-orientated	L1	Observed	3	0	0	0	4	7	p = 0.506	
		Expected	2.19	0.00	0.00	0.00	-0.81			
		Residuals	0.55	NaN	NaN	NaN	-0.37			
	L2	Observed	2	0	0	0	7	9		
		Expected	2.61	0.00	0.00	0.00	0.19			
		Residuals	-0.40	NaN	NaN	NaN	0.33			
Text-orientated	L1	Observed	2	2	3	16	17	40	p = 0.0000	
		Expected	1.03	1.03	1.58	30.00	6.55			
		Residuals	0.95	0.95	1.58	-2.56	-4.08			
	L2	Observed	1	1	1	71	2	76		
		Expected	1.97	1.97	2.62	57.00	12.45			
		Residuals	-0.09	-0.69	-1.00	1.85	-2.90***			

*** p < 0.001

Figure 4. Chi-square test for functional distribution in each move (types)



Category	token	Move	M1	M2	M3	M4	M5	F	p
Reason-orientated	L1	Observed	10	267	24	24	14	657	p = 0.0000
		Expected	2.93	241.64	23.46	67.69	5.28		
		Residuals	8.13	1.63	0.11	-5.34	1.80		
	L2	Observed	0	357	56	200	4	617	
		Expected	7.07	382.36	56.54	158.31	12.72		
		Residuals	-7.07	-32.36	-0.04	41.69	-8.72		
Stance-orientated	L1	Observed	14	0	0	0	18	32	p = 0.11
		Expected	10.83	0.00	0.00	0.00	19.17		
		Residuals	0.96	NaN	NaN	NaN	-0.72		
	L2	Observed	12	0	0	0	30	42	
		Expected	15.17	0.00	0.00	0.00	26.83		
		Residuals	-0.91	NaN	NaN	NaN	0.61		
Text-orientated	L1	Observed	10	9	12	87	85	203	p = 0.0000
		Expected	4.30	5.37	4.56	105.80	24.97		
		Residuals	2.75	1.57	3.40	-6.00	12.01		
	L2	Observed	0	11	3	523	8	545	
		Expected	11.70	14.63	12.44	446.20	68.03		
		Residuals	-11.70	-3.63	-9.44	76.80	-60.03		

*** p < 0.001

Research Question 2 (Cont'd)



Move 3 is to describe research procedure, including study design, methods, participants, outcome measurement;

- L2 used significantly less text-oriented bundle in token compared with L1 corpus.
- No statistical significance found in other two functional categories in type & token.

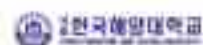
Move 4 is to summarize the findings of the study.

- Compared with L1, L2 group used significant more text-oriented bundles and research-oriented bundles in token;
- No statistical significances found in bundle type;

Move 5 makes the conclusion of the study.

- Compared with L1, L2 group used significant less text-oriented bundles(type & token)
- L2 also used less research-oriented bundles(type only)
- No

5. Conclusion



Q1: Frequently used lexical bundle functions are :

- M1 : *an important role in, little is known about, research has shown that, the extent to which*
- M2: *in this paper we, this paper reports on, this study was to, this study investigates the*
- M3: *analysis is based on, and the use of, in terms of the*
- M4: *the findings show that, the results of the, it was found that, the results suggest that*
- M5: *these findings suggest that, implications of these findings, discussed in relation to*

How each move was realized through bundles

Data analysis has shown :

- M1, L1 writers used all three functional categories, while L2 used only research-oriented and stance-oriented bundles;
- M2: both L1 and L2 used more research-oriented bundles(90%); No Stance-oriented ones found in this move;
- M3: both writer groups prefer more research-oriented;
- M4: saw more text-oriented bundles both in L1 and L2;
- M5: L1 favors more Text-oriented bundles while L2 interested more in Stance-oriented bundles

19

Q2: Bundle functional differences in Moves by L1 & L2


- In general, L1 writers used more bundles than L2 writers both in terms of total number and variety.
- For bundle token, L2 writers demonstrated certain tendency to **overuse** **research-oriented** and **text-oriented** bundles in **Move 4** and **underuse** them in **Move 5**
- Data analysis also confirmed that L2 writer group used **less** **text-oriented** in **Move 3**;
- For bundle type, L2 writers show an tendency to **overuse** **text-oriented** bundles in **M3** and to **underuse** them in **M5**.
- The difference may indicate that L2 writers use only certain bundles when they structured their RA abstract moves, showing a lack of variety in the move rhetorical realization.

20

References



- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers*, 26, 181-190.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Hong, J. (2020). Connecting lexical bundles and moves by Korean EAP students at different academic levels. *Korean Journal of Applied Linguistics*, 36(1), 3-31.
- Hyland, K. (2008a). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Pan, F., Reppen, R., & Biber, D. (2016) Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21:60-71.
- Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics. *Text-Interdisciplinary Journal for the Study of Discourse*, 16(4), 481-500.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.




A comparison of maritime-related characteristics in the 19th century English journal log-book corpus of East Asian regions through keyword analyses

코퍼스 키워드 분석을 통한 19세기 동아시아 항해일지 영어코퍼스에 나타난 한중일 문화 특성 비교

19th August, 2021
Se-Eun Jhang, Songhwa Lee & Hamin Park
(장세은, 이성화, 박하민)
Korea Maritime and Ocean University
(한국해양대학교)

2021 KALS-Online Summer Conference (한국언어과학회 온라인 여름학술회의)
hosted by The Korean Association of Language Sciences



Contents

1. Introduction
 - 1.1 Purpose of the Present Study
 - 1.2 Research Questions
2. Previous Studies
3. Data and Methodology
 - 3.1 Corpora
 - 3.2 Methods
4. Results and Discussion
5. Conclusions

References



1. Introduction

1.1 Purpose of the present study

- The purpose of the study is to investigate lexical characteristics of Maritime English used in the 19th centuries through the two approaches to keyword analyses (corpus frequency based keywords analysis and text dispersion based keywords analysis) of an English journal log-book corpus of East Asian regions consisting of 517,608 words and 233 texts (average 2,222 words per a text) divided from six expedition journals composed of Broughton(1804), Hall(1818), Göttsche(1834), Belcher(1848), King(1861) and John(1880), which were written in English from the late 18th century to the late 19th century.
- We focus on a **multilateral comparison** of maritime-related important words used in the present study corpus with those used in the BE06 corpus as contemporary general British English as well as those used in the contemporary maritime English corpus (CMEC).



1.2 Research Questions

- (1) What are **lexical characteristics of Maritime English** used from the early to the late 19th century? If keywords carry significant implications of a text, facilitate understanding of the main point of the text, and reflect important themes that characterize what the text is about, which approach to keyword analysis shows 'a most welcome keyword list' to access the textual features which are domain-specific?
- (2) How similarly and differently did these 19th century maritime-related words change over time, multilaterally compared with the BE06 corpus as a reference corpus, the contemporary maritime English corpus, and online corpora such as COHA, and COCA?




2. Previous studies on keyword analysis

2.1 traditional approach: Corpus Frequency Keywords

- Scott (1997; 2020) : A key word may be defined as a word which occurs with unusual frequency in a given text. This does not mean high frequency but unusual frequency, by comparison with a reference corpus of some kind.
- Baker (2004) : Keywords carry significant implications of a text or culture. In quantitative analysis they appear to be more frequent or infrequent in a particular text or corpus than in a reference.
- Scott and Tribble (2006): Keywords are derived from comparative quantitative corpus analysis. Keyness is a term used in linguistics to describe the quality a word or phrase has of being "key" in its context.
- Bondi (2010) : Keywords may facilitate understanding of the main point of a text, constituting chains of repetition in the text or a corpus.
- Stubbs (2016) : Content words directly indicate the propositional content of texts, while function words are the most frequent ones in the language and they are also the most frequent players in phrase structures.
- Jhang and Lee (2013) : Visualization of collocational networks using maritime English keywords.

2.2 New Approaches : Text Dispersion Keywords and Hybrid of Frequency and Text Dispersion Analysis

- Egber and Dibeer (2019) : It has been proved that frequency-based calculation is probably not ideal, since dispersion is not taken into consideration. They proposed a text dispersion keyword analysis. The effectiveness of a keyword analysis can be evaluated concerning two criteria: **content-distinctiveness** and **content-generalisability**. **Content-distinctiveness** requires keywords that are truly typical in a specific domain, and point out the textual features which are distinguished from the other domains. **Content-generalisability** needs keywords that have the quality to be generalized in the other texts in a similar domain and "offer insight into the actual content-'aboutness' of those texts."
- Xu and Jhang (2020) : They suggested a combination of the frequency-based method and the dispersion-based method for analyzing a specialized corpus to develop a consolidated keyword list.
- Gries (2021) : Text dispersion keyword analysis proposed by Egber and Biber (2019) can still be profitably extended by utilizing both frequency and dispersion for keyness computations. He proposes a new two-dimensional approach to keyness and exemplifying it on the basis of the *Clinton-Trump Corpus* and the *British National Corpus*.
- Lee and Jhang (2021) : Text dispersion keyword analysis is better than traditional corpus keywords analysis through cross-comparison of the two keyword analyses of the 19th century English journal log-book corpus focusing on Vancouver and Broughton expedition journals.



3. Data and Methodology

3.1 Study Corpus: Six expedition journal logbooks

Broughton(1804), Hall(1818), Gützlaff(1834), Belcher(1848), King(1861) and John(1880)

- Broughton(1804) *The coast of Korea in "A Voyage of Discovery to the South Pacific Ocean ... Performed in His Majesty's Ship Providence, and her smaller *Porpoise* (1791), in the years 1791, 1796, 1797, 1798", by William Robert Broughton, London: T. Cadell and W. Davies.*
- Hall(1818) *Account of voyages of discovery to the west coast of Corea, and the Great Leo-Choo Island by Captain Basil Hall, London: John Murray.*
- Gützlaff(1834) *Journal of three voyages along the coast of China in 1831, 1832, & 1833, with notices of Siam, Corea, and the Leo-Choo Islands by Charles Gützlaff, London: Frederick Westley and A. H. Davis.*
- Belcher(1848) *Notes from a journal of research into the natural history of the countries visited during the voyage of H.M.S. Samarang under the command of Captain Sir Edward Belcher, C.B., F.R.S., by Arthur Adams, R.N., F.L.S., London: Murray, Baskin and Baskin.*
- King(1861) *The China pilot; the coasts of China, Korea, and Tartary; The Sea of Japan, Gulf of Tartary and Amur, and Sea of Okhotsk, and the Bering, Bashi, Formosa, Malacca-Sin, Lu-Chu, Lufones, Bonin, Japan, Sughala, and Kien Islands by John W. King, Master, R.N. (Third Edition), London: J.D. Potter.*
- John(1880) *Notes and sketches from the wild coasts of Nippon with chapters on visiting after pirates in Chinese waters by Captain H. C. St. John, R.N. Edinburgh: David Douglas.*



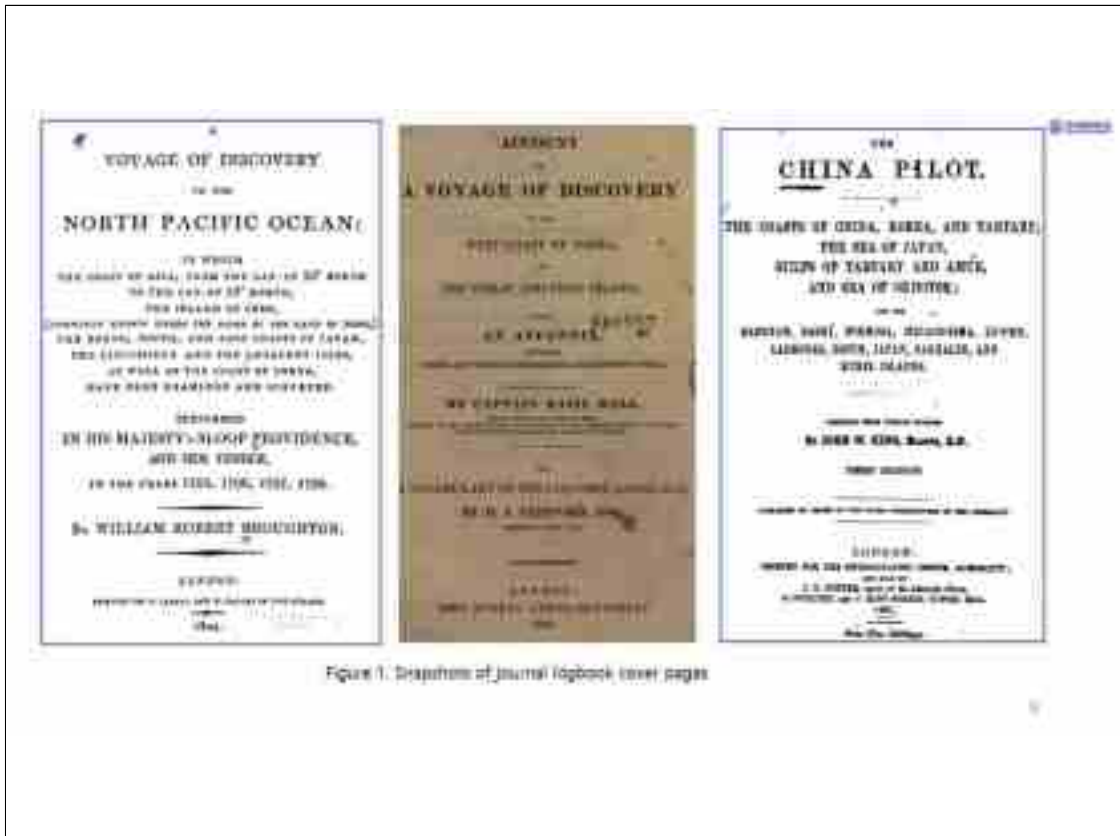
Text No	tokens (running words) in text	type (distinct words)
Overall	517,808	2,233
Text 001_Broughton(1804)_Coast of Korea	2,083	478
Text 002_Broughton(1804)_Coast of Korea	2,178	388
	4,261	866
Text 003_Hall(1818)_Chapter 01 of 01.txt	2,371	770
Text 007_Hall(1818)_Chapter 02 of 02.txt	2,164	411
	4,535	1,181
Text 004_Gützlaff(1834)_China, Korea, Chapter 01 of 01.txt	2,177	387
Text 005_Gützlaff(1834)_China, Korea, Chapter 02 of 02.txt	2,255	393
	4,432	780
Text 006_Belcher(1848)_Samarang, 001, Chapter 01 of 01.txt	1,778	304
Text 007_Belcher(1848)_Samarang, 002, Chapter 02 of 02.txt	1,883	774
	3,661	1,078
Text 008_King(1861)_The China pilot, Chapter 01 of 01.txt	1,314	368
Text 009_King(1861)_The China pilot, Chapter 02 of 02.txt	1,228	387
	2,542	755
Text 010_John(1880)_Wild Coasts of Nippon, Chapter 01 of 01.txt	1,282	714
Text 233_1880_1881 Coasts of Nippon, Chapter 18 of 28.txt	2,447	760
	27,947	6,082


Table 1. Basic information of study corpus.

The study corpus consists of **517,808 words** and **233 texts** (average 2,222 words per a text) according to 2,000 words standard per a text in each chapter of a journal logbook.

Corpus size


- King(1861) **92 texts** >
- John(1880) **39 texts** >
- Belcher(1848) **38 texts** >
- Gützlaff(1834) **37 texts** >
- Hall(1818) **23 texts** >
- Broughton(1804) **4 texts**





3.2 Reference Corpora

- Reference corpus: The BE06 Corpus
 - A **one million word** corpus of published **general written British English**.
 - The same sampling frame as the LOB and FLOB corpora.
 - Consisting of **500 files of 2000 word samples** taken from 15 genres of writing.
- Other corpora (when necessary to be compared)
 - Contemporary Maritime English Corpus (CMSEC) comprised of **four million words**
 - Online corpora: COHA (475 million), and COCA (1 billion) (<https://www.english-corpora.org/>)



Reference Corpus

The BE06 Corpus

Building the BE06 Corpus:
Collecting files from the web

- o 500 files (2,000 words each)
- o Texts collected randomly where possible
- o 10 minutes to locate a text sample, copy it and log it =
- o 5,000 minutes or 83 hours or 12 working days

Table 2: Genre and sources of The BE06 Corpus

Genre	Source(s)
A Press Reports	Newsweek
B Press Editorial	Letters
C Press Reviews	
D Religion	Blackwell and Cambridge publications, Faithful magazine
E Self, Travel and History	Watership 10 magazine, The Times, The Guardian, The Sunday Telegraph etc.
F Popular Fiction	HarperCollins, Orion, and other general pulp publishers, The Guardian
G Letters, ethics, Biographies, Essays	HarperCollins, Orion, The Economist, Atlantic, Collins, Blackwell, etc.
H Miscellaneous (the other industrial reports etc)	Philips encyclopedia, Penguin, Random House, Longman, etc.
J Academic prose	HarperCollins, etc.
K General Fiction	Random House and Orion
L Mystery and Detective Fiction	Random House and Orion
M Science Fiction	Random House and Orion
N Adventure and Western	Random House, Orion
P Science and Technology	Random House, Orion
R History	HarperCollins, Orion

3.2 Methods

- **Cross-comparison of the two keyword lists:**
WordSmith Tools 8.0 for keywords extraction using the two keyword analyses, traditional approach to corpus frequency keyword analysis and new approach to text dispersion keyword analysis
- **Multilateral comparison of maritime-related keywords** using the BE06 corpus as a reference corpus and the contemporary maritime English corpus (CMEC) as well as online corpora such as BNC, CGHA, and COCA when necessary to be compared.



Figure 3. A snapshot of WordSmith Tools 8.0

3.2 Methods (Cont'd)

- **Keyness computation:**

- To consider a significant level, we used a log-likelihood test (Dunning, 1993), because it gives a better estimate of keyness, especially when contrasting long texts or a whole genre against a reference corpus (Scott, 2012). Thus we set a relatively low p-value threshold such as 0.000000000001 (1 in one hundred trillion) to produce more reliable data.
- Log Likelihood test which gives a better estimate of keyness, especially when contrasting long texts or a whole genre against a reference corpus.
- Dunning (1993: 65): The likelihood ratio could lead to much better statistical results for small corpora.
- Bertels and Spermann (2013: 24): Log-likelihood could be an efficient statistical measure for both small and large corpora. This can be supported by other keyword related journal articles.
- Previous ESP research prefers to log-likelihood (Bertels and Spermann, 2013; Ginzburg, 2015; Phuong Le and Hartmann, 2015; Wilkinson, 2015).

4. Results and Discussion

4.1 Comparison of two keyword lists

Text-Dispersion (TD) keywords by employing the text dispersion function

- A total of 185 keywords were extracted.

Corpus Frequency (CF) keywords by not employing the text dispersion function

- A total of 378 keywords were extracted.
- The top 185 keywords were compared with the keywords with the dispersion function.

◀ There appeared 135 shared keywords

➔

So 50 unshared keywords are compared.

Table 3. 135 Shared keywords extracted by the two keyword analyses

Rank	TD Ranking	Shared Keywords	CF Ranking	Rank	TD Ranking	Shared Keywords	CF Ranking
1	71	ARMIST	122	26	98	SEE	122
2	17	AVIATION	64	25	27	STRANGE	51
3	3	ANCHORAGE	38	22	15	SYSTEMS	116
4	54	HIGHBALL	32	37	32	EXTEND	32
5	49	APPARANCE	105	34	108	STYVES	74
6	157	APPEAR	104	39	1	LOVING	8
7	122	SHINE	107	49	92	HOT	37
8	22	SHIP	14	47	34	CONCRETE	170
9	24	SHARPS	52	51	44	FOOT	49
10	35	SHARK	63	43	80	INCIDENTS	143
11	46	SOFT	66	44	74	UNUSUALLY	34
12	20	SOFTS	74	47	117	UNUSED	171
13	120	BOTTOM	102	46	21	HARVEST	24
14	61	CABLE	35	42	84	HILL	29
15	128	CABLES	103	48	33	HILLS	175
16	132	CAMPAIGN	111	49	64	SHOCKS	124
17	100	CAYMAN	110	50	27	APPARENTLY	124
18	107	CHAMBER	19	51	14	SLAND	1
19	57	CHINA	62	52	9	SLANDS	74
20	15	CHINESE	34	53	13	SOFT	37
21	74	COAST	23	54	42	TRUTH	46
22	191	CONSIDERABLE	172	70	3	IT'S	7
23	146	CONCRETE	106	56	107	WORK	140
24	30	CATAWISS	108	62	114	WORK	443
25	148	DEPTH	96	63	21	STAYS	12
26	23	DECEMBER	18	58	71	LAND	33
27	60	DEGREE	64	60	34	LAT	74
28	144	INSTANT	104	61	27	LATITUDE	124
29	127	MANUAL	171	62	147	DAYTON	144
30	175	LAT	22	63	88	LINE	38
31	22	NATURAL	144	64	102	W	120
32	33	NATURAL	43	65	171	MADISON	112
33	18	NATURAL	10	66	114	MADISON	37

No.	TD Ranking	Keyword	CF Ranking	No.	TD Ranking	Keyword	CF Ranking
47	43	WELL	25	102	178	WITNESS	37
48	40	WELL	4	103	179	WITNESS	87
49	39	WINDSOON	39	104	180	WITNESS	103
50	117	WINDWARD	134	105	181	WITNESS	112
51	37	WELL	187	106	182	WITNESS	14
52	109	WELL	5	107	183	WITNESS	178
53	144	WIND	104	108	184	WITNESS	18
54	50	WINDWARD	135	109	185	WITNESS	165
55	77	WINDWARD	48	110	186	WINDWARD	83
56	33	WIND	49	111	187	WINDWARD	88
57	144	WIND	35	112	188	WINDWARD	48
58	50	WINDWARD	105	113	189	WINDWARD	102
59	140	WINDWARD	117	114	190	WIND	37
60	14	WINDWARD	28	115	191	WIND	136
61	76	WINDWARD	37	116	192	WINDWARD	151
62	5	WIND	7	117	193	WINDWARD	119
63	113	WIND	130	118	194	WIND	9
64	36	WINDWARD	36	119	195	WIND	167
65	54	WINDWARD	42	120	196	WIND	46
66	125	WIND	34	121	197	WIND	75
67	110	WINDWARD	146	122	198	WIND	88
68	38	WIND	38	123	199	WIND	168
69	70	WINDWARD	146	124	200	WIND	71
70	30	WIND	42	125	201	WIND	25
71	38	WIND	37	126	202	WIND	33
72	30	WINDWARD	171	127	203	WIND	108
73	31	WIND	129	128	204	WIND	178
74	34	WIND	33	129	205	WIND	5
75	77	WIND	39	130	206	WIND	12
76	8	WIND	21	131	207	WINDWARD	96
77	22	WIND	39	132	208	WINDWARD	37
78	20	WIND	4	133	209	WINDWARD	64
79	20	WIND	139	134	210	WIND	66
80	26	WIND	39	135	211	WIND	39
81	22	WIND	34				

Table 4. 50 Unshared Keywords extracted by the two keyword analyses

Text Dispersion Keywords				Corpus Frequency Keywords				
No.	Ranking	TD Keyword	No.	Ranking	CF Keyword	No.	Ranking	CF Keyword
1	136	APPROX	26	176	WINDWARD	38	41	WIND
2	361	WINDWARD	27	180	WINDWARD	37	75	WIND
3	170	WINDWARD	28	177	WINDWARD	38	47	WIND
4	179	APPROACHED	29	179	WINDWARD	39	183	WINDWARD
5	171	APPROXIMATE	30	182	WINDWARD	40	140	WINDWARD
6	174	APPROXIMATE	31	183	WINDWARD	41	176	WINDWARD
7	180	WINDWARD	32	187	WINDWARD	42	180	WINDWARD
8	120	WIND	33	187	WINDWARD	43	125	WINDWARD
9	170	WINDWARD	34	188	WINDWARD	44	18	WINDWARD
10	176	WINDWARD	35	194	WINDWARD	45	179	WINDWARD
11	104	WINDWARD	36	194	WINDWARD	46	140	WINDWARD
12	44	WINDWARD	37	195	WINDWARD	47	127	WINDWARD
13	102	WINDWARD	38	196	WINDWARD	48	17	WINDWARD
14	104	WINDWARD	39	196	WINDWARD	49	174	WINDWARD
15	178	WINDWARD	40	197	WINDWARD	50	17	WINDWARD
16	126	WINDWARD	41	198	WINDWARD	51	13	WINDWARD
17	101	WINDWARD	42	198	WINDWARD	52	12	WINDWARD
18	140	WINDWARD	43	199	WINDWARD	53	140	WINDWARD
19	141	WINDWARD	44	200	WINDWARD	54	100	WINDWARD
20	80	WINDWARD	45	200	WINDWARD	55	191	WINDWARD
21	34	WINDWARD	46	201	WINDWARD	56	144	WINDWARD
22	178	WINDWARD	47	201	WINDWARD	57	171	WINDWARD
23	101	WINDWARD	48	202	WINDWARD	58	17	WINDWARD
24	140	WINDWARD	49	202	WINDWARD	59	12	WINDWARD
25	401	WINDWARD	50	217	WINDWARD	60	120	WINDWARD

4.2 Why is text dispersion keyword analysis better than corpus frequency keyword analysis

Apparently the text dispersion based keywords in the present study corpus are more meaningful words.

A total of 185 keywords employing the dispersion function were analyzed as follows:

1. Function words
2. Maritime-related words
3. Theme specific content words
4. General content words

Table 5. Comparison of the two keyword lists according to keyword classification

No	Keyword Classification	Shared keywords	Unshared Keywords		TOTAL (TD Keywords)
			Text Dispersion	Corpus Frequency	
1	Function words	4	0	3	4
2	Maritime-related words	80	23	28	105
3	Theme specific content words	10	5	7	22
4	General content words	10	22	17	54
	TOTAL	108	50	60	185

Comparison of maritime-related keywords between text dispersion keyword analysis (TD keywords) and corpus frequency keyword analysis (CF keywords) out of 50 unshared keywords

Table 6. Comparison of unshared maritime-related keywords between TD keyword analysis and CF keyword analysis

No.	Ranking TD Keywords	Ranking CF Keywords
1	274 APPRAISED	61 BAKE
2	133 APPROXIMATE	171 BELONG
3	144 ARCHDELAINE	113 BRACE
4	121 BATS	179 BURDEN
5	119 BEARS	44 CAPS
6	134 BASKETS	81 CHAL
7	44 BERTHOUS	181 CLAD
8	108 BUSHNET	182 COB
9	85 BATHER	162 CREAK
10	69 BOB	131 DEEP
11	64 BODENBACH	10 2
12	160 BONDING	137 FISHING
13	178 GALE	161 GULF
14	114 HITCHER	617 HAY
15	114 HOOKSANDW	173 HANG
16	103 HOOKSAND	16 HINT
17	91 QUARTER	140 KAN
18	98 HULLING	119 SOGA
19	144 JERRY	11 HITCH
20	85 HOOKHOLE	99 STRACK
21	179 GUNLOCK	161 TA
22	108 WHEEL	161 TAIL
23	91 HULLING	11 STIFF

• **Content-distinctiveness** requires keywords that are truly typical in a specific domain and that point out the textual features, which are distinguished from the other domains.

• **Content-generalisability** needs keywords that have the quality to be generalized to the other texts in a similar domain and that "offer insight into the actual content: 'aboutness' of those texts" (Egbert & Biber, 2019: p. 79).

• Table 6 displays the different results of the unshared maritime-related keyword lists between text-dispersion keywords and corpus frequency keywords generated by WordSmith Tools 8.0.



4.3.2 Maritime-related English Keywords

• A total of 105 types are categorized into Maritime-related words and grouped into 7 themes

- 1) Geographical features
- 2) Direction and/or location
- 3) Types of vessels
- 4) Navigation-associated words
- 5) Weather
- 6) Anchor-associated words
- 7) Measurements in maritime

39

Table 7. Seven Categories Classification for 105 keywords of maritime-related words

Categories	Maritime-related Keywords
Geographical features (26)	ARCHIPELAGO, BANKS, BAY, BAYS, BEACH, CHINA, COAST, COASTS, EXTREME, EXTREMITY, FORT, HARBOUR, ISLAND, ISLANDS, ISLET, ISLETS, JAPAN, MUD, PATCHES, PENINSULA, PORT, REEF, REEFS, RIVER, ROCK, ROCKS, ROCKY, SAND, SANDY, SEA, SHORL, SHOALS, SHORE, SHORES, STREAM, WATER
Direction and/or location (24)	DIRECTION, DIRECTIONS, EAST, EASTERLY, EASTERN, EASTWARD, FARTHER, LAT, N, NORTH, NORTHERLY, NORTHERN, NORTHWARD, QUATER, QUARTER, QUARTERS, S, SOUTHERLY, SOUTHERN, SOUTHWARD, W, WESTERLY, WESTERN, WESTWARD
Types of vessels (8)	BOAT, BOATS, JUNK, JUNKS, SHIP, SHIPS, VESSEL, VESSELS
Navigation-associated words (22)	APPROACHED, APPROACHING, BEARING, BEARS, CAPTAIN, CHYNNEL, DRAUGHT, EBB, FISH, FISHERMEN, FISHING, GROUND, NARROW, PASSAGE, PROCEEDING, PROCEEDS, SHALLOW, STEER, SLIPPER, SWELL, TIDE, TIDES
Weather (5)	GALE, MONSOON, WIND, WINDS, WEATHER
Anchor-associated words (5)	ANCHOR, ANCHORAGE, ANCHORED, DEPTH, SOUNDINGS
Measurements in maritime (5)	NET, PATHWAYS, RICHES, MILE, MILES

40

4.3.2 Maritime-related English Keywords (cont'd)



1) Geographical features

- 36 types
- ARCHIPELAGO, BANKS, BAY, BAYS, BEACH, CHINA, COAST, COASTS, EXTREME, EXTREMITY, FORT, HARBOUR, ISLAND, ISLANDS, ISLET, ISLETS, JAPAN, MUD, PATCHES, PENINSULA, PORT, REEF, REEFS, RIVER, ROCK, ROCKS, ROCKY SAND, SANDY, SEA, SHOAL, SHOALS, SHORE, SHORES, STREAM, WATER

2) Direction and/or location

- 24 types
- NORTH, NORTHERLY, NORTHERN, NORTHWARD, OUTER, QUARTER, QUARTERS, S, SOUTHERLY, SOUTHERN, SOUTHWARD, W, WESTERLY, WESTERN, WESTWARD

→ X-erly, X-ward; x = north, south, west, (east)



4.3.2 Maritime-related English Keywords (cont'd)



3) Types of vessels

- 8 types
- BOAT, BOATS, JUNK, JUNKS, SHIP, SHIPS, VESSEL, VESSELS



Figure 3. Junk
 • Chinese ship with
 a flat bottom and square sails
 中国帆船

4) Navigation-associated words

- 22 types
- APPROACHED, APPROACHING, BEARING, BEARS, CAPTAIN, CHANNEL, DRAUGHT, EBB, FISH, FISHERMEN, FISHING, GROUND, NARROW, PASSAGE, PROCEEDING, PROCURED, SHALLOW, STEER, SUNKEN, SWELL, TIDE, TIDES



4.3.2 Maritime-related English Keywords (cont'd)

v) Weather

- 5 types
- GALE, MONSOON, WIND, WINDS, WEATHER

vi) Anchor-associated words

- 5 types
- ANCHOR, ANCHORAGE, ANCHORED, DEPTH, SOUNDINGS

vii) Measurements

- 5 types
- FEET, FATHOMS, INCHES, MILE, MILES
- Fathoms (물의 깊이 측정 단위. 6피트 또는 1.8미터에 해당)
- cf. leagues (3마일)

4.3.2 Maritime-related English Keywords (cont'd)

4.3.2.1 Geographical features extremity

- Oxford dictionary definition: "the furthest point, especially from the centre"
- 128 tokens were found in the study corpus
- No tokens were found in BE06



Figure 4. Concurrency of extremity

4.3.2 Maritime-related English Keywords (cont'd)



4.3.2.2 Direction: eastward, northward, southward, westward (cont'd)

The BE06 Corpus

- No southward, westward
 - One northward, two eastwards found
- (3) a. They set off **northward**. [adverb]
- b. In Colchester the population grew, and houses sprawled far beyond the old town walls - into the countryside to the south, and **eastward** as far as the River Colne [adverb]
- c. (...) the expansion **eastward** and the remaking of Europe's relation with Washington [noun]



4.3.2 Maritime-related English Keywords (cont'd)



4.3.2.2 Direction: eastward, northward, southward, westward (cont'd)

Contemporary Maritime English Corpus

15 eastward, 6 westward, 6 northward, 2 southward

- (4) a. While there can be no denying the gradual **eastward** shift in shipping's overall centre of gravity, it is (...) [adjective]
- b. (...) constituted by the 30°N parallel from Florida **eastward** to 77°30' W meridian, [noun]
- c. (...) that wind particles moving toward the equator would come from a region of lower **eastward** velocity and enter a region of higher **eastward** velocity [adjective]
- ❖ 'X-ward' in the 18th and 19th centuries tend to be used as a noun; in the contemporary English, more diverse parts of speech (noun, adverb, adjective) seem to be used.



4.3.2 Maritime-related English Keywords (cont'd)
 4.3.2.3 Navigation-associated words : **procured**

Procured in BE06 and CMEC

- One token in BE06
- In the text of 'Christine Borland: Reconnecting anatomy to a sense of humanity'

(5) She had learned that a long established trade in human skeletal material for medical purposes was about to end and, with the assistance of the firm involved, she **procured** the last two human skulls in the trade ...

- No tokens in CMEC

4.3.2 Maritime-related English Keywords (cont'd)
 4.3.2.3 Navigation-associated words : **procured** (cont'd)

'Procured' in COHA (1820-2010) and COCA (2010-2019)

	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	2010-2019	2019-2019
FREQ	210	334	287	300	254	211	136	207	157	140	135	64	61	54	42	35	28	26	28	30	57	57
PER MIL	30.06	24.38	18.79	22.28	15	18.55	9.87	10.13	7.14	8.06	9.25	2.31	2.23	1.88	1.44	1.21	0.87	0.76	0.83	0.45	0.67	0.67

PER MIL

- ◆ 'Procure(d)' has a negative connotation in contemporary British English and its usage has dramatically decreased since the early 20th century in American English.
- ◆ 'Procure(d)' seems to have been neutrally used as a meaning of simple purchase in the 18th and 19th centuries.

4.3.2 Maritime-related English Keywords (cont'd)

4.3.2.4 Anchor-associated: soundings (수심)



162 token (in 75 texts) in the study corpus (40th keyword ranking)

Figure 9. Concordance of soundings



4.3.2 Maritime-related English Keywords (cont'd)

4.3.2.4 Anchor-associated: soundings (수심)




CMEC and BE06

• 6 tokens in CMEC

- (6) a. (...) on failure to sight land, or a navigation mark or to obtain soundings by the expected time.
- b. (...) if, unexpectedly, land or a navigation mark is sighted or a change in soundings occurs.


• No tokens in BE06


❖ 'Soundings' seems to be used exclusively in the maritime domain.




References

- Baker, Paul. (2006). Querying keywords: Questions of definition, frequency, and issues in keyword analysis. *Journal of English Linguistics*, 32, 249-258.
- Baker, M. (2007). Perspectives on keywords and keyness: An introduction. In M. Baker and M. Duff (eds.), *Keywords in use: Annotated*. John Benjamins Publishing Company.
- Biber, T., & Myles, D. (1999). Incorporating text frequency into keyword analysis. *Corpus: Corpus-based language learning, language assessment and linguistics*, 24, 117-146.
- Biber, D., & Conrad, S. (2012). A new approach to (text) keyword analysis: Using frequency, and not the frequency. *Research in English Linguistics*, 1(1), 1-11.
- Biber, D., & Lee, A. M. (2012). Visualization of multidimensional keyword analysis: English keywords. *Language Awareness*, 23(5), 181-192.
- Biber, D. P., & Conrad, S. (2011). The vocabulary of academic English: Keyword analysis of the English homepage of post-graduate journal for world. *Journal of English Language and Linguistics*, 2(2), 189-193.
- Biber, D. P., & Conrad, S. (2010). The construction of a 19th-century English corpus for ESP studies. *Journal of English Language and Linguistics*, 2(2), 189-193.
- Lee, S. R., and Biber, D. P. (2011). Keyword analysis of the 19th-century English journal log-book corpus. Vancouver and Singapore, a manuscript presented at 2011 KALLI (Spring Conference on English Linguistics) on 14 of June, 2011, hosted by The Korean Association for the Study of English Language and Linguistics.
- Baker, M. (1997). The analysis of text: Text and key words. *Journal of English Linguistics*, 23(2), 179-191.
- Baker, M., & Conrad, S. (2008). *Keyword analysis: Key words and corpus analysis in language education*. Amsterdam, The Netherlands: John Benjamins.
- Baker, M. (2002). *Wordbank: Text analysis: A practical approach*. Available from <http://www.leeds.ac.uk/leeds/leeds/leeds/leeds/>
- Biber, D. (1997). *Text analysis: A practical approach*. In M. Baker and M. Duff (eds.), *Keywords in use: Annotated*. John Benjamins Publishing Company.
- Ba, L., & Biber, D. (2010). Keyword analysis of English-chinese papers. *Linguistic Research*, 47(2), 267-284.








Thank You Very Much For Listening.

Any Questions?





A Hybrid Approach to Keywords Analysis of Aviation Accident Investigation Reports Corpus

19th August, 2021

Yang Qi, Se-Eun Jhang, & Homin Park

(Dalian Maritime University & Korea Maritime and Ocean University)

2021 KALS Online Summer Conference
hosted by The Korean Association of Language Sciences
(<http://www.alskorea.or.kr>)



Contents

1. Introduction
 2. Literature Review on keywords analysis
 3. Data and methods
 4. Results and Discussion
 5. Conclusion
- References



1. Introduction

1.1 Purpose of the present study

1.2 Research Questions

1.3 Aviation Accident Investigation Reports



1.1 Purpose of the present study

- The purpose of the study is to compare three different ways of keyword lists of an Aviation Accident Reports Investigation Corpus (AAIRC) as a study corpus through three different approaches to keywords analysis: corpus frequency-based keywords analysis (Scott 2010), text dispersion based keywords analysis (Egbert and Hiber 2019; Scott 2020) and Gries' (2021) hybrid approach.
- We focus on the cross-comparison of aviation-related keywords out of the top 100 keywords extracted by using three different methods and then provide heuristic explanations in order to investigate which method is the best.
- We also illustrate two/three dimensional diagrams for keywords of words in AAIRC as a study corpus and the COCA Academic 2011-2012 corpus and a reference corpus as general English, as positive evidence for a hybrid approach to keywords analysis as the best way.





1.2 Research Questions

- (1) First, how do we extract a list of keywords by using a hybrid approach to keywords analysis proposed by Gries (2021)?
- (2) Second, what are significant differences between Scott's (2016) frequency based keywords analysis, Egbert and Biber's (2019) text dispersion based keywords analysis, and a hybrid approach of DKL-frequency and DKL-dispersion proposed by Gries (2021)?



1.3 Aviation Accident Investigation Reports

- Accident Reports are one of the main products of an National Transportation Safety Board (NTSB) investigation.
- NTSB is an independent U.S. government investigative agency responsible for civil transportation accident investigation.
- Reports provide details about the accident, analysis of the factual data, conclusions and the probable cause of the accident, and the related safety recommendations.
- Most reports focus on a single accident, though the NTSB also produces reports addressing issues common to a set of similar accidents.
- Accident Investigation Reports include Aviation Accident Reports, Hazardous Materials Accident Reports, Highway Accident Reports, Marine Accident Reports, Pipeline Accident Reports and Railroad Accident Reports.





2. Literature Review on Keywords Analysis

2.1 Traditional approach: Corpus Frequency Keywords

2.2 New Approach: Text Dispersion Keywords (Egbert and Biber 2019)

2.3 A hybrid Approach (Gries 2021)



2.1 Traditional Approach: Corpus Frequency Keywords

- Scott (2016, 2020) : a keyword is a very important word and the frequency of its first occurrence in a given text is very unusual, but it doesn't mean that its high frequency is abnormally high.
- Baker (2004) : the keywords themselves carry some important background information, or the implied information of the article, as well as the corresponding cultural and knowledge content.
- Scott (1997); Baker (2004); Scott & Tribble (2006) : as a quantitative perspective, keywords are those that have statistical significance in terms of the frequency of occurrence in the text or corpus compared with the standard set by the reference corpus.





2.2 New approach: Text Dispersion Keywords by Egbert and Biber (2019)

- According to Egbert and Biber (2019), corpus frequency keywords are frequently used in a corpus.
- However, they are not dispersed broadly through each text of the corpus. That is to say, such corpus frequency keywords cannot really reflect the characteristics of the discourse in this field, nor can they describe the characteristics of this corpus in a scientific and objective way.



2.2.1 Text dispersion keyness for keyword analysis

- Egbert & Biber (2019) use log-likelihood or G2 to evaluate these figures arithmetic.
- This formula for G2 is:

$$G^2 = 2 \sum_i O_i \ln \left(\frac{O_i}{E_i} \right)$$

	Corpus frequency keywords	Text dispersion keywords
Definition	Words that are statistically more frequent in a target corpus when compared with a reference corpus.	Words that occur in statistically more texts in a target corpus when compared with a reference corpus.
Variable	Frequency in corpus	Dispersion point ratio
Formula	Log-likelihood O = observed word frequency E = expected word frequency	Log-likelihood O = observed word dispersion (number of texts) E = expected word dispersion (number of texts)
Requirements	- At least one text in target and reference corpora. - Software (e.g., AntConc, WordSmith) or spreadsheet program.	- Many texts in target and reference corpora. - Specialized program.



2.2.2 Egbert and Biber (2019)

- The two goals of Egbert and Biber (2019) are to introduce text dispersion keyness and also tell us the importance of text dispersion.
- This method focuses on dispersion and generates key wordlist.
- There is **no** related word frequency, and this measure is a kind of range.
- Both the sizes and corpus parts are **not** considered in this method.



2.3 A hybrid approach (Gries 2021)

- Egbert and Biber's (2019) text dispersion analysis for keyword research can still be upgraded a little bit, and we can consider both frequency and text dispersion at the same time. Then we calculate this keyness value.

2.3.1 Overview of a hybrid approach to keywords analysis

- This study focuses on a hybrid approach to keywords analysis proposed by Gries (2021), providing **heuristic explanations** and top 100 keywords examples to show how we extract a keyword list by using this hybrid approach.
- A majority of the keywords analyses on educational and genre studies are a blend of the two ways. They are frequency lists and association statistics. Two steps are needed: First, we establish a frequency list of a study corpus and another frequency list of a reference corpus. Second, we can create a contingency table (2 plus 2 table) which is usually used in collocation statistics, as shown in Table 1.

Table 1. A Contingency Table

	target corpus	reference corpus	sum
target word	a	b	a+b
other words	c	d	c+d
sum	a+c	b+d	N

2.3.1 Overview of the hybrid approach to keywords analysis (Cont'd)

- The LLR log-likelihood ratio for this contingency table is computed accordingly.
- The expected frequencies and log-likelihood score are used in the followed equation.

$$(1) \theta_{expected} = \frac{(a+b)(a+c)}{n}$$

$$(2) LLR/G^2 = 2 \times \left(a \times \log \frac{a}{\theta_{expected}} + b \times \log \frac{b}{\theta_{expected}} + c \times \log \frac{c}{\theta_{expected}} + d \times \log \frac{d}{\theta_{expected}} \right)$$



2.3.2 DKL frequency: New frequency-based keyness measure by Gries (2021)

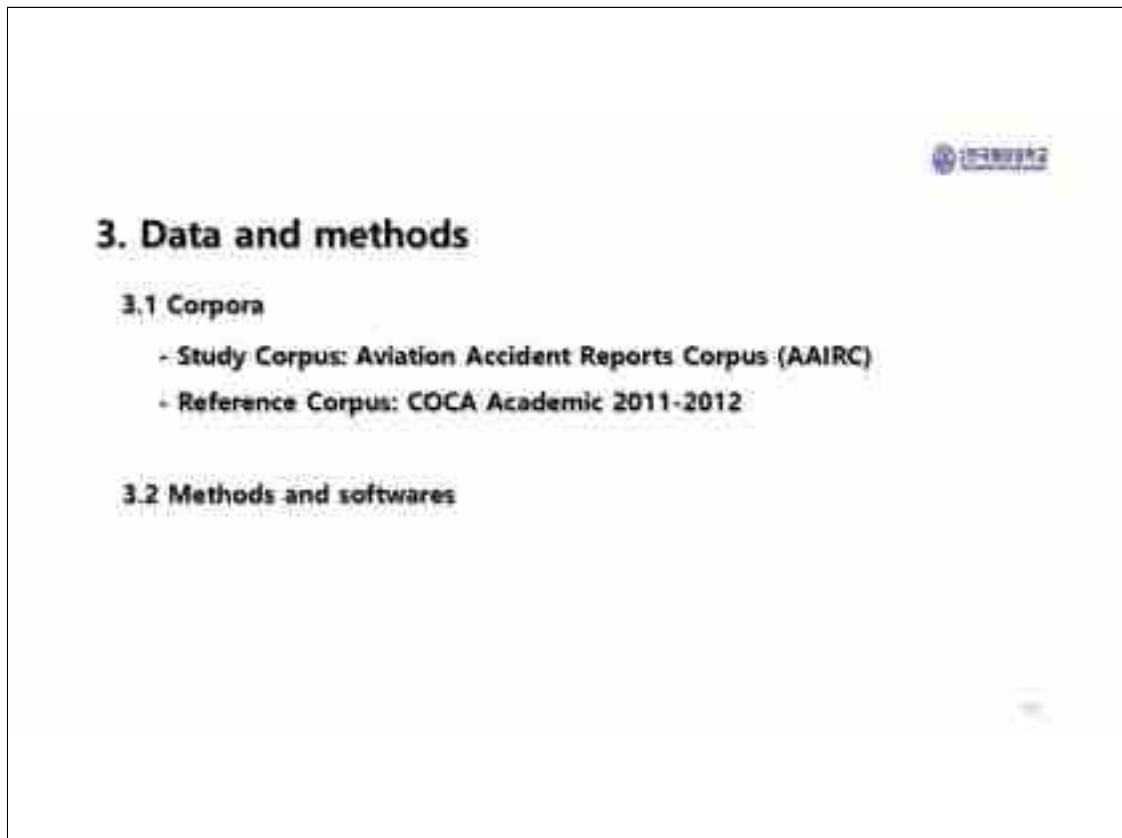
- A dispersion measure beyond range such as Deviation of Proportions (DP)
- Gries' (2021) first enhancement is a new keyness measure with frequency information.
- Gries' (2021) first method is less associated with frequency.
- This measure is called "Kullback-Leibler (KL) divergence".
- It is written as **DKL** (posterior/data/priority /ibency)" (Gries 2021, p.14)



2.3.3 DKL dispersion : Application of dispersion-based keyness measure by Gries (2021)

- Gries' (2021) second enhancement is a new keyness measure with frequency information and dispersion information.
- Gries' (2021) second method is to add dispersion to the keyword analysis, so that we can try to solve a task of which method is the best out of several keyness measures.







3.2 Methods and softwares

- WordSmith Tools 8.0 was used to extract corpus frequency keywords and text dispersion keywords.
- Python was used to process the data to get DKL frequency, DKL target corpus dispersion and DKL reference corpus dispersion for a hybrid keyword.
- DKL dispersion are calculated in both the AALBC as a study (-target) corpus and COCA Academic 2011-2012 as a reference corpus.
- After reading the book Corpus Data Processing using Python (with Professor Leilei and co-author Homin Park), we wrote some code that used Python programs to extract the corresponding keywords. First of all, in the aspect of data processing of corpora, we carried out sentence and word segmentation. Second, we carried out the magnetic permeability, extraction of keywords, extraction of words list, delete the stop words, and calculate the collocation strength of individual words.

78



4. Results and Discussion

- 4.1 Comparison of three keyword lists
- 4.2 Comparison of aviation-related keywords between K1, K2 and K3 keywords analyses
- 4.3. Gries' hybrid approach to keywords analysis

79



4.1 Comparison of three keyword lists

- **K1 keywords analysis:** Corpus Frequency (CF) keywords as a traditional analysis by not employing the text dispersion function in keywords setting of WordSmith Tools 8.0 (Scott 2016, 2020).
- **K2 keywords analysis:** Text Dispersion (TD) keywords as a new analysis by employing the text dispersion function in keywords setting of WordSmith Tools 8.0 (Scott 2020)
- **K3 keywords analysis:** keywords of DKL frequency and DKL dispersion as a hybrid approach to keywords analysis
- The top 100 keywords by using three different keywords analyses were compared
- There appeared 17 shared keywords.

Table 2. 17 shared keywords

Rank	Keyword	K1	K2	K3
1	aircraft	2	1	3
2	engine	5	2	2
3	engine	3	5	4
4	engine	27	11	10
5	engine	26	24	21
6	engine	11	5	2
7	air	6	12	10
8	engine	22	8	6
9	air	11	2	11
10	engine	10	24	10
11	air	7	12	11
12	air	7	5	3
13	engine	1	5	2
14	air	1	1	1
15	air	1	1	10
16	engine	8	12	8
17	engine	21	18	11

Table 3. 83 unshared keywords in K1


Rank	Keyword	K1	K2	K3
18	air	20	63	10
19	air	20	44	20
20	engine	20	41	10
21	engine	22	47	10
22	air	22	45	10
23	engine	24	42	10
24	engine	24	34	10
25	air	25	33	10
26	air	25	32	10
27	engine	26	30	10
28	air	26	24	10
29	engine	26	22	10
30	air	26	22	10
31	engine	26	22	10
32	air	26	22	10
33	air	26	22	10
34	air	26	22	10
35	air	26	22	10
36	air	26	22	10
37	air	26	22	10
38	air	26	22	10
39	air	26	22	10
40	air	26	22	10
41	air	26	22	10
42	air	26	22	10
43	air	26	22	10
44	air	26	22	10
45	air	26	22	10
46	air	26	22	10
47	air	26	22	10
48	air	26	22	10
49	air	26	22	10
50	air	26	22	10
51	air	26	22	10
52	air	26	22	10
53	air	26	22	10
54	air	26	22	10
55	air	26	22	10
56	air	26	22	10
57	air	26	22	10
58	air	26	22	10
59	air	26	22	10
60	air	26	22	10
61	air	26	22	10
62	air	26	22	10
63	air	26	22	10
64	air	26	22	10
65	air	26	22	10
66	air	26	22	10
67	air	26	22	10
68	air	26	22	10
69	air	26	22	10
70	air	26	22	10
71	air	26	22	10
72	air	26	22	10
73	air	26	22	10
74	air	26	22	10
75	air	26	22	10
76	air	26	22	10
77	air	26	22	10
78	air	26	22	10
79	air	26	22	10
80	air	26	22	10
81	air	26	22	10
82	air	26	22	10
83	air	26	22	10

Table 4. K1 unshared keywords in K2

No.	Keywords	Count	Percentage	Keywords	Count	Percentage
1	비행기	28	40	비행	25	35
2	비행사	25	35	비행기	25	35
3	비행기	25	35	비행기	25	35
4	비행기	25	35	비행기	25	35
5	비행기	25	35	비행기	25	35
6	비행기	25	35	비행기	25	35
7	비행기	25	35	비행기	25	35
8	비행기	25	35	비행기	25	35
9	비행기	25	35	비행기	25	35
10	비행기	25	35	비행기	25	35
11	비행기	25	35	비행기	25	35
12	비행기	25	35	비행기	25	35
13	비행기	25	35	비행기	25	35
14	비행기	25	35	비행기	25	35
15	비행기	25	35	비행기	25	35
16	비행기	25	35	비행기	25	35
17	비행기	25	35	비행기	25	35
18	비행기	25	35	비행기	25	35
19	비행기	25	35	비행기	25	35
20	비행기	25	35	비행기	25	35
21	비행기	25	35	비행기	25	35
22	비행기	25	35	비행기	25	35
23	비행기	25	35	비행기	25	35
24	비행기	25	35	비행기	25	35
25	비행기	25	35	비행기	25	35
26	비행기	25	35	비행기	25	35
27	비행기	25	35	비행기	25	35
28	비행기	25	35	비행기	25	35
29	비행기	25	35	비행기	25	35
30	비행기	25	35	비행기	25	35
31	비행기	25	35	비행기	25	35
32	비행기	25	35	비행기	25	35
33	비행기	25	35	비행기	25	35
34	비행기	25	35	비행기	25	35
35	비행기	25	35	비행기	25	35
36	비행기	25	35	비행기	25	35
37	비행기	25	35	비행기	25	35
38	비행기	25	35	비행기	25	35
39	비행기	25	35	비행기	25	35
40	비행기	25	35	비행기	25	35
41	비행기	25	35	비행기	25	35
42	비행기	25	35	비행기	25	35
43	비행기	25	35	비행기	25	35
44	비행기	25	35	비행기	25	35
45	비행기	25	35	비행기	25	35
46	비행기	25	35	비행기	25	35
47	비행기	25	35	비행기	25	35
48	비행기	25	35	비행기	25	35
49	비행기	25	35	비행기	25	35
50	비행기	25	35	비행기	25	35

Table 5. K3 unshared keywords in K2

No.	Keywords	Count	Percentage	Keywords	Count	Percentage
1	비행기	25	35	비행기	25	35
2	비행기	25	35	비행기	25	35
3	비행기	25	35	비행기	25	35
4	비행기	25	35	비행기	25	35
5	비행기	25	35	비행기	25	35
6	비행기	25	35	비행기	25	35
7	비행기	25	35	비행기	25	35
8	비행기	25	35	비행기	25	35
9	비행기	25	35	비행기	25	35
10	비행기	25	35	비행기	25	35
11	비행기	25	35	비행기	25	35
12	비행기	25	35	비행기	25	35
13	비행기	25	35	비행기	25	35
14	비행기	25	35	비행기	25	35
15	비행기	25	35	비행기	25	35
16	비행기	25	35	비행기	25	35
17	비행기	25	35	비행기	25	35
18	비행기	25	35	비행기	25	35
19	비행기	25	35	비행기	25	35
20	비행기	25	35	비행기	25	35
21	비행기	25	35	비행기	25	35
22	비행기	25	35	비행기	25	35
23	비행기	25	35	비행기	25	35
24	비행기	25	35	비행기	25	35
25	비행기	25	35	비행기	25	35
26	비행기	25	35	비행기	25	35
27	비행기	25	35	비행기	25	35
28	비행기	25	35	비행기	25	35
29	비행기	25	35	비행기	25	35
30	비행기	25	35	비행기	25	35
31	비행기	25	35	비행기	25	35
32	비행기	25	35	비행기	25	35
33	비행기	25	35	비행기	25	35
34	비행기	25	35	비행기	25	35
35	비행기	25	35	비행기	25	35
36	비행기	25	35	비행기	25	35
37	비행기	25	35	비행기	25	35
38	비행기	25	35	비행기	25	35
39	비행기	25	35	비행기	25	35
40	비행기	25	35	비행기	25	35
41	비행기	25	35	비행기	25	35
42	비행기	25	35	비행기	25	35
43	비행기	25	35	비행기	25	35
44	비행기	25	35	비행기	25	35
45	비행기	25	35	비행기	25	35
46	비행기	25	35	비행기	25	35
47	비행기	25	35	비행기	25	35
48	비행기	25	35	비행기	25	35
49	비행기	25	35	비행기	25	35
50	비행기	25	35	비행기	25	35



4.2 Comparison of aviation-related keywords between K1, K2 and K3 keywords analyses

- More meaningful words are found in K3 of the present target corpus based on the observation below:
- The AAIRC usually contains information on the following details: Type of accident, accident location and time, airplane information (name of the airplane, official number, owner, nationality, damage, etc.), summary of the accident, pre-accident condition, the process of the accident, measures taken during the accident, results, recommendations and so on.




Table 6. A total of 100 keywords belonging to three categories of keywords, as follows (Lee and Jiang 2021):


1. Function words
2. Aviation-related words
3. Theme specific content words
4. General content words

No.	KEYWORDS CLASSIFICATION	SUMMED KEYWORDS
1	Function words	0
2	Aviation-related words	10
3	Theme specific content words	4
4	General content words	2
	Total	17

Table 7. 17 shared keywords in three keywords analyses

KEYWORD	AAIRIC	COCA	COCA ACADEMIC	KEYWORD CLASSIFICATION
accident	1	1	1	General content word
investigation	1	1	1	Aviation-related word
report	1	1	1	Aviation-related word
findings	1	1	1	Aviation-related word
conclusion	1	1	1	Aviation-related word
cause	1	1	1	Aviation-related word
error	1	1	1	Aviation-related word
human factors	1	1	1	Aviation-related word
weather	1	1	1	Aviation-related word
air traffic control	1	1	1	Aviation-related word
communication	1	1	1	Aviation-related word
procedures	1	1	1	Aviation-related word
training	1	1	1	Aviation-related word
fatigue	1	1	1	Aviation-related word
cockpit	1	1	1	Aviation-related word
instrument	1	1	1	Aviation-related word

⇒ A heuristic explanation: Most of the keywords generated by KS analysis have higher ranking than other analyses.



4.3. Gries' hybrid approach to keywords analysis

- To study a hybrid approach to keywords analysis mixed with frequency and dispersion, we investigated some plots in 2-dimensions and 3-dimensions.
- The x-axis represent the words' frequency or DKL-frequency or DKL-dispersion. For example, in case that values of (-1,1) appear on the plot, the range (-1,0) represents words whose frequency or DKL-frequency or DKL-dispersion distribution decides them to be AAIRC keywords.
- The more deviates, the stronger a word's frequency preference for either AAIRC, or COCA academic. That is to say, the strongest AAIRC words in terms of frequency or DKL-frequency or DKL-dispersion will be far on the left. At the same time, the strongest COCA Academic words in terms of frequency or DKL-frequency or DKL-dispersion will be far on the right.

4.3.1 Three dimensional diagram illustrating DKL-frequency and DKL-dispersion



Three dimensional keyness of words diagram illustrates both the relationship between LLR, DKL-Frequency, and DKL-Dispersion, as shown in Figure 2.

Figure 2. Three dimensional keyness of words in AALRC and COCA Academic Corpus

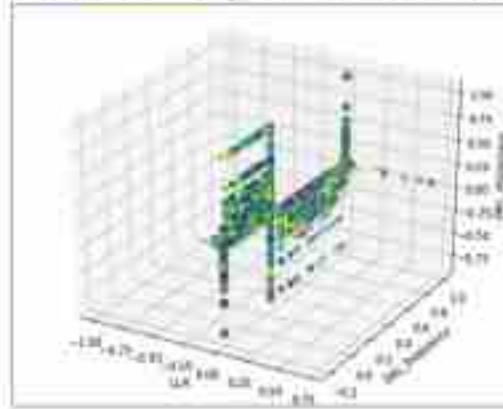
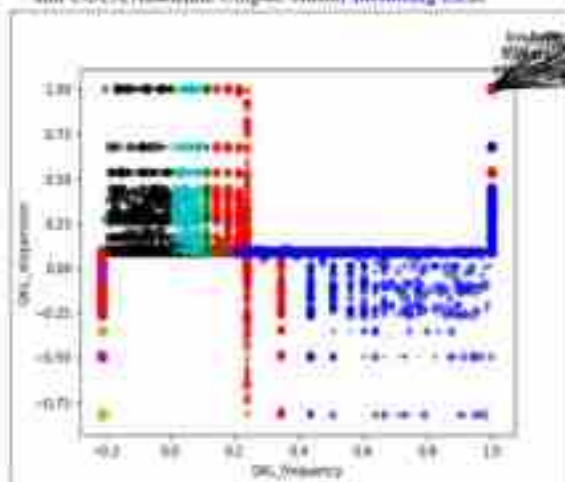
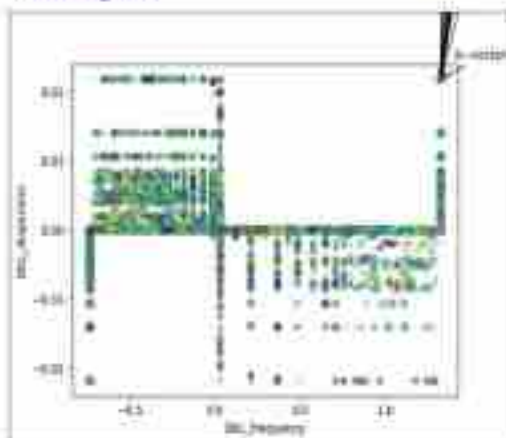


Figure 3. Two-dimensional keyness of words in AALRC and COCA Academic Corpus cluster including LLR



- The two-dimensional keyness of words diagram illustrates both the relationship between DKL-Frequency and DKL-Dispersion.
- The cluster-cluster diagram is actually a 2D scatter representation of a three-dimensional graph. The x axis is DKL-Frequency, and the y axis is DKL-Dispersion. This information is reflected in color. We cluster all the values of LLR into 7 categories, and then use a color for each category, so that we can clearly see that LLR is a category of data in which DKL-Frequency and DKL-Dispersion are not aggregated, which indicates that LLR is not a representative indicator for finding keywords and key keywords.
- So it's not very accurate to judge by LLR in keywords analysis, compared with DKL-Frequency and DKL-Dispersion, as is illustrated in Ertes' example in Chien's Study, just like the graph of choosing the Omi CHI I mentioned in his paper.

Figure 4. Two-dimensional keyness of words in AAIRC and COCA Academic Corpus-cluster not including LLR



- This diagram does not consider and add the information of LLR on the basis of cluster diagram, and it is separately taken out for LLR.
- The x-axis is DKL-frequency, and the y-axis is DKL-dispersion, and then we have a graph that looks like this.

Table 8. Part of the top N wordlist

Word	Distance
seem—stout	1.4142135623730951
wing—blue	1.4142135623730951
wing	1.4142135623730951
line—here	1.4142135623730951
gemfibrozil	1.4142135623730951
lystocytosis	1.4142135623730951
evenright	1.4142135623730951
conventual	1.4142135623730951
company-related	1.4142135623730951
subsets	1.4142135623730951
radius	1.4142135623730951
efficiency	1.4142135623730951
unproven	1.4142135623730951
male-terminator	1.4142135623730951
passenger-type	1.4142135623730951
stress-free	1.4142135623730951
medium-range	1.4142135623730951
tachystrophia	1.4142135623730951
aw-stand	1.4142135623730951

- Top-n actually corresponds to the distance between all points in Figure 3 and the far point.
- The further the distance is, the higher the key degree is, the more critical the word is.
- It is deduced that this word is a relatively key word in the AAIRC corpus. So, the distance between the left and right words from the origin is calculated here.
- That's the distance between all the points in the legend, and then we sort it from top to bottom, and then we open up the graph top-n, and we can see, on the left is the word, and on the right is the value of an index, and that's the sorting result as keywords.



Table 9. Part of the top N-ed wordlist

```

W1421 000011 100001 00001 00001
rate 1.4142135623730951
fruit 1.4142135623730951
inclusion 1.4142135623730951
starfish 1.4142135623730951
approach-to-staff 1.4142135623730951
ship 1.4142135623730951
oil 1.4142135623730951
shy 1.4142135623730951
step 1.4142135623730951
stairway-to-st 1.4142135623730951
rory 1.4142135623730951
go-amount 1.4142135623730951
var 1.4142135623730951
te 1.4142135623730951
sustain 1.4142135623730951
q-amount 1.4142135623730951
c 1.4142135623730951
please 1.4142135623730951
youth 1.4142135623730951
column 1.4142135623730951
    
```

- Top-n-ed graph is just cut the upper right corner, the values greater than 0, and then sort a keyword. Then the corresponding result is the figure 4 diagram.
- It essentially calculates the order of all the words or points it.

Figure 4; one distance from the origin, and doesn't take into account the values that are less than zero, so the results are slightly different between the two.




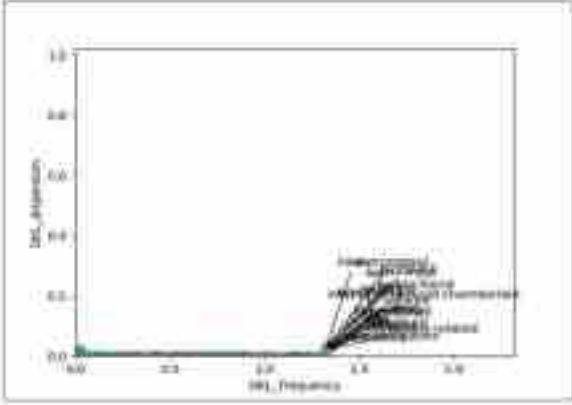



Figure 5. Two-dimensional keyness of words in AAJC and COCA Academic Corpus



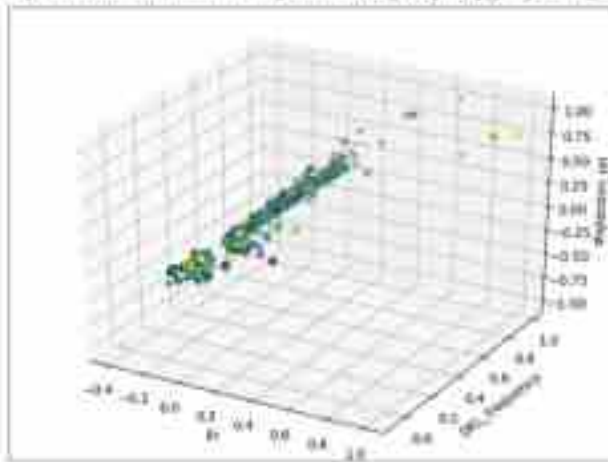
The ZOOM diagram in Figure 5 is a magnified version of the upper right corner of Figure 3, which is also the same as diagrammed in Gries (2021), the upper right corner: commonly used in his paper is enlarged.



4.3.2 Results and the best out of three keywords analyses

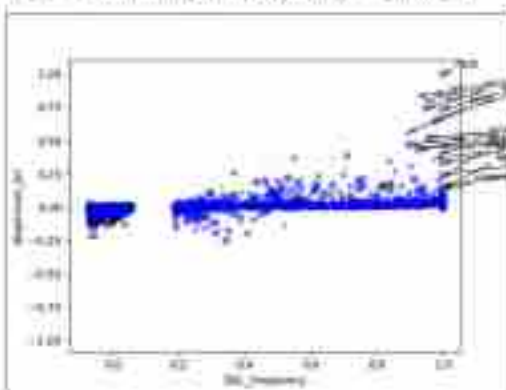


Figure 6. Three dimensional diagram of DKL, Frequency Dispersion, Ori, LLR



- The 3D DKL, Frequency Dispersion, Ori, LLR diagram between -1 and 1 based on the five values we calculated with a normalization.
- The 3D has three dimensions, so we took those of the five results and drew this picture. The purpose is to show the relationship between them.
- DKL, Frequency is one dimension, the dispersion is one dimension, and LLR is one dimension. Thus a three dimensional diagram was drawn.
- This is the result of nesting the words, using the combination of the two methods, Geiss (2021) and Egbert and Bibee (2019).
- It means that we need to calculate the first and second values separately, Then we calculate the square and the square root of the two values, and we can find the relationships in AAIRC, as shown in Figure 6.

Figure 7. 2D-Cluster_DKL, Frequency Dispersion, Ori



- Cluster DKL, Frequency Dispersion, Ori is the cluster diagram. The ordinate is the dispersion value calculated by Egbert and Bibee's (2019) method.
- The color is also divided into 7 classes according to the logarithmic likelihood ratio.
- This is a clustering diagram, which is also used by three parts, LLR, DKL-frequency (Geiss 2021), dispersion Ori (Egbert and Bibee 2019).
- LLR is used for clustering analysis.
- We found the left lower part and upper right part are blue dots, and they represent different colors.
- LLR cannot distinguish good and different words, and they are also unable to extract keywords.
- DKL-frequency is good to differentiate.
- In the dimension of DKL-frequency, the dots to the right are more important.





Table 10. A list of keywords in KJ-analysis

Keyword	Value
rtab	1.4142135621730951
flight	1.342092807865094
aircraft	1.240999155294017
fee	1.2338787180260004
accident	1.2041809760003248
airplane	1.1749074084486355
aviation	1.0907256205526714
cockpit	1.0880567795779112
runway	1.0728668825099044
pilot	1.0517530175456888
postaccident	1.052313389844002
takeoff	1.0411841976478942
airspeed	1.0249233573632879
car	1.021767097588568
airport	1.021450763488201
city	1.0152243512745809
crew	1.0121719734282223

- The best way out of three keyword analyses is the combination of Egbert and Bibac (2019) and Gries (2021).
- Gries (2021) proposes an improved version of Egbert and Bibac (2019) that is DKJ-dispersion, which has no better effect than Egbert's method on the AADRC-CECA Academic dataset.
- Gries (2021) alone is not generally applicable to all the texts because of its limitations.
- As shown in Figure 10, top 17 words are obtained by using the combination of the two methods of Egbert and Bibac (2019) and Gries (2021).
- It means that we need to calculate the first and second values separately. Then we calculate the square and the square root of the two values.





5. Conclusion

5.1 Summary

5.2 Pedagogical Implications



5. Conclusion

5.1 Summary

- We have investigated which keywords analysis can generate more appropriate keywords of AAIRC as a study corpus through **heuristic explanations** of multilateral comparison of top 100 keywords using three different approaches to keywords analysis as well as **two/three dimensional diagrams** for keyness of words in AAIRC and COCA Academic Corpus.
- The best way out of three keyword analyses is the combination of Tugheer and Biber (2019) and Gries (2021), which is K3 analysis, because this analysis generated more significant and meaningful words in AAIRC such as type of accident, accident location and time, airplane information (name of the airplane, official number, owner, nationality, damage, etc.), summary of the accident, pre-accident condition, the process of the accident, measures taken during the accident, results, recommendations and so on.

5.2 Pedagogical Implications

- The appropriate use of English for aviation can save human lives and avoid accidents and incidents in an aviation field.
- Understanding important words of English for aviation can make L2 writers more efficient to improve ESP writing such as aviation accident investigation reports. To do this, one of the best guides for the teaching and learning of Aviation English writing in an aviation field is likely to provide L2 writers with a list of more appropriate keywords extracted from a corpus of aviation accident investigation reports.

References



- Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keyword analysis. *Journal of English Linguistics*, 31, 346-359.
- Egbert, J. and Bibot, D. (2019). Incorporating text dispersion into keyword analysis. *Corpora*, 16(1), 77-104.
- Giles, S. T. (2021). A new approach to (key) keywords analysis: Using frequency, and now also dispersion. *Research in Corpus Linguistics*, 8(2), 1-33.
- Lee, S.H. and Jiang, S.E. (2021) Keyword analysis of the 19th century English journal log-book corpus: Vancouver and Houghton, a manuscript presented at 2021 EASLFL Spring Conference on English Linguistics on 5th of June, 2021, hosted by The Korean Association for the Study of English Language and Linguistics.
- Scott, M. (2016). *WordSmith Tools version 7*. Birmingham: Lexical Analysis Software.
- Scott, M. (1997). PC analysis of key words – and key key words. *System*, 25(2), 233-245.
- Scott, M. (2020). *WordSmith Tools version 8*. Birmingham: Lexical Analysis Software.
- Scott, M. and Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam, The Netherlands: John Benjamins.

99



Thank You Very Much For Listening.

Any Questions?

100

A Corpus-based Analysis of Syntactic Features on Headlines of Maritime News

August 19th, 2021

Melhan Wang
(Korea Maritime and Ocean University)

2021 KALS Online Summer Conference

CONTENTS

- 0 Abstract and Key Words
 - 1 Introduction
 - 2 Previous Studies
 - 3 Data and Methods
 - 4 Results and Discussion
 - 5 Conclusions
- References

0. Abstract

With the development of the shipping industry, more and more people pay close attention to maritime news. Reading maritime news is a relatively convenient and efficient way for maritime English learners and shipping industry operators to learn about the latest shipping trends and related knowledge. Reading maritime news headlines also is a quicker way to grasp the main points. The **Maritime News Headline Corpus (MNHCC)** used in this paper is selected from the news headlines of the world's major maritime organization classification societies. The **General News Headlines Corpus (GNHC)** is collected from the world's influential newspapers. **The purpose of this study explores the syntactic features of maritime news headlines from four aspects: sentence structure patterns, sentence forms, sentence voices and omissions** with the help of SPSS, WordSmith Tools and GLADSH. The research results and findings of this paper have certain referential significance for Maritime English learners, writers of maritime news headline and operators in the shipping field.

Key words: maritime news headlines; corpus; syntactic features; structure patterns; tenses; voices; omission

1 Introduction

1.1 Background and Significance of the Research

With the recovery and revival of shipping industry, a great number of people pay their attention to maritime news. As the enrichment and essence of the news texts, the news headlines play an important role in satisfying the need for efficient and convenient access to information for the readers.

Compared with reading whole texts of news, reading news headlines that are regarded as the eye of news is a comparatively faster method to grasp main ideas. As the gist of the news, the language of news headlines is compressed and condensed with the unique rules and regulations (Reah, 1998, p.14).

1 Introduction

1.1 Background and Significance of the Research

Fairclough (1989) claims that headlines possess special characteristics in the syntactical level, making them a grammatical particularity (p.21). Maritime English, as a category of English for Special Purpose, has different language characteristics (including news language) compared with other genres.

Therefore, the research on maritime news has never stopped. A large number of studies on the language of maritime news, including maritime news headlines, have focused on the vocabulary and stylistic features in a qualitative way. Few scholars discuss the characteristics of maritime news headlines from the syntactic level.

1 Introduction

1.2 Research Questions

Q1: What syntactic features do maritime news headlines employ regarding sentence structures, tenses & aspects, voices, and omission?

Q2: What are similarities and differences between these syntactic features of maritime news headlines and those of general news headlines?

2 Previous studies

2.1 Theories about Syntactic Features

2.1.1 Sentence Structures

According to Yule (2014), "The word 'syntax' came originally from Greek and literally meant 'a setting out together or 'arrangement' (p.100)".

Quirk et al. (1985) puts forward the classification of sentences: "Sentences are either simple or multiple. Specifically speaking, a simple sentence is composed of only a clause. A multiple sentence consists of two or more clauses and it can be further divided into compound sentence and complex sentence. A compound sentence refers to a sentence consists of two or more coordinate clauses and a complex sentence composes of one or more independent clauses.

2 Previous studies

2.1 Theories about Syntactic Features

2.1.1 Sentence Structures (Cont'd)

Generally speaking, though the classifications of sentence can be different based on different perspectives, their classification are all on the base of the number of independent clauses and dependent clauses a sentence contains. In this thesis, the classification of the sentence is based on **the Quirk's theory**, namely **the simple sentence and multiple sentence** and the later can be further divided into compound sentence and complex sentence.

2 Previous studies

2.1 Theories about Syntactic Features

2.1.1 Sentence Structures (Cont'd)

Table 2.1 Seven Patterns of Simple Sentence

Pattern of Simple Sentences	
SV	Subject+Verb
SVO	Subject +Verb + Object
SVC	Subject +Verb+ Subject Complement
SVA	Subject +verb + Adverbial
SVOC	Subject +Verb + Object +Complement
SVOO	Subject + Verb + Indirect Object + Direct Object
SVOA	Subject +verb + Object +Adverbial

Quirk's theory (1985)

2 Previous studies

2.1 Theories about Syntactic Features

2.1.1 Sentence Structures (Cont'd)

- Compound sentence refers to a sentence that contains at least two independent clauses, which is often joined by coordinating conjunction such as **or, but, for, and**.
- According to Zhou (2012), "The complex sentence is combined with **one main clause** and **one or more than one subordinating clause**."

2 Previous studies

2.1 Theories about Syntactic Features

2.1.1 Sentence Structures (Cont'd)

According to Zhang (2004), the conjunctions are divided into subordinating conjunctions, interrogative adverbs, relative pronouns, relative adverbs and interrogative pronouns. The specific examples of these conjunctions are listed in the following Table 2.2.

Table 2.2 Examples of Conjunctions in Complex Sentence

Types	Examples
Subordinating Conjunctions	where, when, after, because, so that, etc.
Interrogative Adverbs	when, where, why, how, etc.
Relative Pronouns	who, which, whom, whose, that, etc.
Relative Adverbs	when, why, where, etc.
Interrogative Pronouns	who, which, whom, whose, what, etc.

2 Previous studies

2.1 Theories about Syntactic Features

2.1.2 Sentence Tenses

Curme (1931) points out that the tenses are divided into **absolute tense** (present, past, present perfect and future) and **relative tense** (past perfect and future perfect). The absolute tense means expressing the moment of speaking, while the relative tense describes time relatively to the preceding absolute tense.

2 Previous studies

2.1 Theories about Syntactic Features

2.1.2 Sentence Tenses

Zhang (1995) proposes, "time has four parts: present, past, future and past future and action has four types: the indefinite, the progressive, the perfect and the perfect progressive. The combination of these time concepts and actions results in sixteen tenses (p.356)."

Table 2.3 Distribution of Sixteen-Tense System

Simple Present	Present Progressive	Present Perfect	Present Perfect Progressive
Simple Past	Past Progressive	Past Perfect	Present Perfect Progressive
Simple Future	Future Progressive	Future Perfect	Future Perfect Progressive
Simple Past Future	Past Future Progressive	Past Future Perfect	Past Future Perfect Progressive

11

2 Previous studies

2.1 Theories about Syntactic Features

2.1.3 Sentence Voices

According to Quirk (1985), as a grammatical category, voice makes it possible to perceive the action of a sentence in two ways (active and passive) without change in the facts reported. Generally speaking, the active voice usually emphasizes on the active role of the subject, namely the agent of the action; while the passive voice usually underlines the passive role of the subject-recipient of the action.

12

2 Previous studies

2.2 Previous Studies on English News Headlines

- Bell (1991) claims that the headline has the power to draw reader's attention and provoke them to read the whole story.
- Garrison (1996) insists that a news headline not only offers information guidance to audiences, moreover, it also orients readers the way how to sense the news.
- Jarovici and Amel (1989) consider that news headlines possess a "double function" simultaneously, which are semantic function and pragmatic function. The semantic function means the headlines deal with the referential text; the pragmatic function is concerned with the readers to whom the text is addressed.

2 Previous studies

2.2 Previous Studies on English News Headlines (Cont'd)

- Van Dijk (1998), as the leader in the field of western news language study, he finds some ways to combine linguistic theory and discourse analysis with news headlines studies in his book *News Analysis*. Through analyzing some samples, he argues that the most prominent feature the headlines had in news was to present the cardinal information and to give readers the topics before preceding the whole discourse for the macro structural position.
- Luo (2007) makes a study on stylistic features of news headlines based on corpus approach, in which the data is collected from yahoo news website and the study finds that the SVO/Ad patterns take the first position and SV/Ad pattern takes the second position and the third position is accounted by SVed/Adj.

2 Previous studies

2.3 Previous Studies on News Headlines of Maritime English (Cont'd)

- Guo (2014) presents a study on Lexical features;
- Li (2013) carries out a research on lexical cohesion;
- Liu (2014) gives a study of lexical bundles of maritime English news
- *A Corpus-Based Analysis of Stylistic of Headlines of Maritime News* was carried out by Liu Yang (2015), who presents his research on stylistic features of maritime news headlines in the four levels: graphological level, lexical level, and rhetorical level and syntactical level.

03 Data and Method

3.1 Corpus-Based Approach

- ◆ with the combination of qualitative and quantitative ways, the corpus-based approach was employed to analyze the syntactical features of maritime news headlines.
- ◆ Leech (1991) defines that "a corpus is a database that is a collection of massive information used in a real situation".
- ◆ Biber, Conrad, Reppen (1998) claim that "it is a large and principled collection of natural texts. According to Sinclair (1999), as the stored electronically texts, the corpus is a massive and principled collection with naturally-occurring texts and it is used for summarizing and characterizing a state or variety of a language.

03 Data and Method

3.2 Corpora Used in the Study

3.2.1 Description of MNHC and GNHC

Maritime News Headlines Corpus (short for MNHC) is a newly self-built corpus, making up the main research data in this present study.

The websites related to maritime news were selected as the sources of corpus collection. The Maritime News Headlines Corpus consists of 86,315 words and 10029 items of maritime news headlines are downloaded from the websites.

1. IMO (<http://www.imo.org/>)
2. ITI seafarers (<http://www.itiseafarers.org/index.cfm>)
3. BIMCO (<https://www.bimco.org/About/Press.aspx>)
4. World Maritime News (<http://worldmaritimeweb.com>)
5. Marine insight (<http://www.marineinsight.com/>)

03 Data and Method

3.2 Corpora Used in the Study

3.2.1 Description of MNHC and GNHC (Cont'd)

General News Headlines Corpus (short for GNHC) as a reference corpus is also newly self-built. The data was downloaded from the websites as follows:

1. <https://www.thomsonreuters.com/en.html>
2. <https://www.un.org/en>
3. <https://www.foxnews.com/>
4. <https://www.ap.org/en-gb/>
5. <https://wall-street.com>

03 Data and Method

3.2 Corpora Used in the Study

3.2.1 Description of MNHC and GNHC (Cont'd)

Table 3.1 Descriptive Statistics of MNHC and GNHC

Corpus	Words	Headlines
MNHC	86313	10029
GNHC	126795	10000

03 Data and Method

3.2 Corpora Used in the Study

3.2.2 Description of S-MNHC and S-GNHC

- 1,000 headlines will be extracted randomly from MNHC and GNHC respectively to be used for analyzing sentence length and sentence structures. They are named as S-MNHC and S-GNHC respectively.
- Because the major study is focused on the syntactical features and sentences serve as the basic language unit, the phrase headlines need to be deleted manually.

Table 3.2 Descriptive Statistics of S-MNHC and S-GNHC

Sample	Words	Phrase Headlines	Sentence Headlines
S-MNHC	8610	38	962
S-GNHC	12981	29	971

03 Data and Method

3.3 Tools for Data Processing:

3.3.1 SPSS

- SPSS (the Statistical Package for Social Science) is a software package widely used for statistical analysis, ranging from the field of market researching, survey companies to education researching, data mining. SPSS provides lots of functions of statistical analysis with various charts. The descriptive statistics includes mean, mode, median, range, t-test, ANOVA, Correlation, standard error and Nonparametric tests. The non-parametric test known as the Chi-square test is especially useful for which can be used for checking whether there is significant statistical difference in distributions of certain linguistic features.
- In this paper, the Chi-square test is conducted to test whether there is significant difference in the use of sentence structures between maritime news headlines and general news headlines.

03 Data and Method

3.3 Tools for Data Processing:

3.3.1 SPSS

The null hypothesis and alternative hypothesis are as follows:

- H₀: There is no significance in distribution of sentence structures between SMNHIC and S-GNHIC.
- H₁: There is a significance in distribution of sentence structures between SMNHIC and S-GNHIC.

The data are obtained through running the SPSS as shown in Table 3.3.

Table 3.3 Chi-Square Test

	Observed N ^a	df	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Exact Sig. (1-sided)	Exact Sig. (2-sided)
Chi-Square Test Statistics	388.967 ^a	2	<.001	<.001		<.001
Continuity Correction	447.402	2	<.001	<.001		<.001
Fisher's Exact Test	492.168			<.001		<.001
Linear-by-Linear Association	332.736 ^a	1	<.001	<.001	<.001	<.001
N of Valid Cases	1800					

a. If there is a 2x2 table, continuity correction is used. The continuity correction is used if there is a 2x2 table.

b. The standard error is 14.776.

03 Data and Method

3.3 Tools for Data Processing

3.3.2 CLAWS

CLAWS4 Part-of-speech (POS) tagging, also called grammatical tagging, is the commonest form of corpus annotation, and was the first form of annotation to be developed by University Centre for Computer Corpus Research on Language (UCREL) at Lancaster.

The latest version of the tagger, CLAWS4, was used to POS tag c.100 million words of the British National Corpus (BNC). CLAWS4 has consistently achieved 96-97% accuracy (the precise degree of accuracy varying according to the type of text).

Because a standalone version of CLAWS4 was paid software, the free web tagging service of CLAWS4 (Free-CLAWS-web-tagger.lancaster.ac.uk/), which was accessible to the latest version of the tagger, CLAWS4, is used in this thesis.

03 Data and Method

3.3 Tools for Data Processing

3.3.2 CLAWS

After tagging these data, the frequency distribution of different parts of speech in MNHC was realized through WordSmith 8.0. For example, the frequency of verbs used in MNHC can be calculated by searching for each corresponding code in the tagset related to verbs (POS codes related to verbs in the C7 tag set are as follows).

03 Data and Method

3.3 Tools for Data Processing

3.3.2 CLAWS (Cont'd)

Table 3.4 Codes Related To Verb in Accordance With C7 Tagset

VOC	to, from, have	V1F3	infinitive form
VOC	are	V1F2	forming
VOC	are	V1F1	verb, auxiliary
VOC	being	V1F0	infinitive participle
VOC	is, are, was, were	V1F4	is
VOC	is	V1F5	finite auxiliary
VOC	was	V1F6	finite auxiliary (copula, verb)
VOC	am	V1F7	finite form of main verb
VOC	is	V1F8	participle of main verb
VOC	to have been doing	V1F9	infinitive of main verb inf. partic. forming
VOC	is	V1F10	infinitive auxiliary forming inf. partic. forming
VOC	was	V1F11	finite
VOC	is, are, was, were	V1F12	infinitive of main verb
VOC	is	V1F13	participle of main verb
VOC	am	V1F14	finite of main verb (inf. partic. forming)
VOC	to have been doing	V1F15	

03 Data and Method

3.3 Tools for Data Processing

3.3.2 CLAWS (Cont'd)

Table 3.5 Sample of Tagged Corpus

operations NN2 Bureau NP1 starts NP2 launches VV2 Digital JJ And CC
 independent JJ Tool NN1 To JJ Support NN1 Design NN1 Verification NN1 Of IO
 floating JJ Units NN2 TEN MC ; Market VV0 Recovery NN1 Is V5Z Already RR
 Upon JJ JJ NP1 Clean NN1 Updates VVZ GuideLines NN2 For JJ Remote JJ
 Survey NN2 Cruise VV0 Ship NN1 Interiors NN2 Expo NP1 Europe NP1 Releases VVZ
 Agenda NN1 Blueprint NP1 Releases VVZ New JJ Guide NN1 To X CH NN1 In JJ
 Braker NN1 Mending NN1 B CC Discharge VV0 K.ZZ) Line NN1 Commences V5Z
 Proof NN1 Of IO Concept NN1 With JW Nokia NP1 In JJ The_AT Field NN1 Of IO
 DR NP1 For JJ Ships NN2 Port NN1 of IO Birmingham NP1 hydrogen NN1 supply NN1
 bid NN1 submitted VV0 Watch NN1 ; 80 MC Years NN2 Old JJ USS NP1
 Johnson NP1 ; Worlds NN2 Deepest JJ Shipwreck NN1 Explored VV0
 DeepOcean JJ to TO sport VV1 UK NP1 entry NN1 and CC create VV1 two MC new JJ
 divisions NN2 Takara NN1 closes VV2 Titan NP1 deal VV0 Greek JJ
 Shipowners NN2 Back RR Ship NN1 Power NN1 limitation NN1 Proposal NN1 also NN1
 specific JJ young JJ workers NN1 get VV0 connected JJ AET TO Wire VVZ SRS NP1
 Ship NN1 Owner NN1 Of IO The_AT Year NN1 At JJ Singapore NP1 International JJ
 Machine JJ Awards NP1 USC NFA Gayane NP1 crewman NN1 gets VV2 S+ FO yes NN6JJ
 in JJ proof NN1 for Fdrug NN1 smuggling VV0 Mitsubishi NP1 slashes VVZ

03 Data and Method

3.3 Tools for Data Processing

3.3.3 WordSmith Tools

WordSmith Tools was developed by the British linguist Mike Scott at the University of Liverpool. The core areas of the software package include three modules:

- (1) **Concord** is used to create concordances, so all the hits from a search within a previously defined body text.
- (2) **WordList** lists all the words or on word forms that are included in the selected corpus and statistical data are different from the text corpus.
- (3) **KeyWord** creates a list of all those words and word forms according to certain statistical criteria in the text corpus significantly occur rarely or frequently.

In the present study, the WordSmith 6.0 is employed to search the frequency of sentence tenses and sentence voices. The following table presents an example of searching the simple present tense in MNHC.

03 Data and Method

3.3 Tools for Data Processing

3.3.3 WordSmith Tools

Table 3.6 Illustration of Sentence Patterns of Simple Sentences of MNHC

1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

04 Results and Discussion

4.1 Sentence Structures

Table 4.1 Descriptive Statistics of Sentence Structures in S-MNHC and S-GNHC

Structure	S-MNHC		G-MNHC	
	Number	Percent	Number	Percent
Simple Sentence	956	99.37%	647	66.63%
Compound Sentence	2	0.21%	113	11.63%
Complex	4	0.42%	211	21.73%
Total	962	100.00%	971	100.00%

04 Results and Discussion

4.1 Sentence Structures (Cont'd)

Figure 4.1 Descriptive Statistics of Sentence Structures in S-MNHC and S-GNHC



04 Results and Discussion

4.1 Sentence Structures (Cont'd)

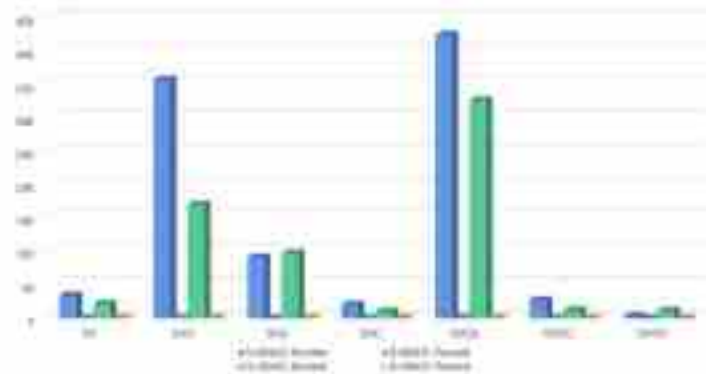
Table 4.2 Seven Basic Patterns of Simple Sentence in S-MNHC and S-GNHC

Sentence Pattern	S-MNHC		S-GNHC	
	Number	Percent	Number	Percent
SV	38	3.50%	23	3.25%
SVO	357	37.34%	170	26.27%
SVA	91	9.52%	37	14.99%
SVC	20	2.09%	10	1.54%
SVOA	424	44.35%	326	50.39%
SVDC	27	2.82%	12	1.85%
SVOD	3	0.03%	11	1.70%
Total	956	100.00%	647	100.00%

04 Results and Discussion

4.1 Sentence Structures (Cont'd)

Figure 4.2 Seven Basic Patterns of Simple Sentence in S-MNHC and S-GNHC



04 Results and Discussion

4.2 Sentence Tenses

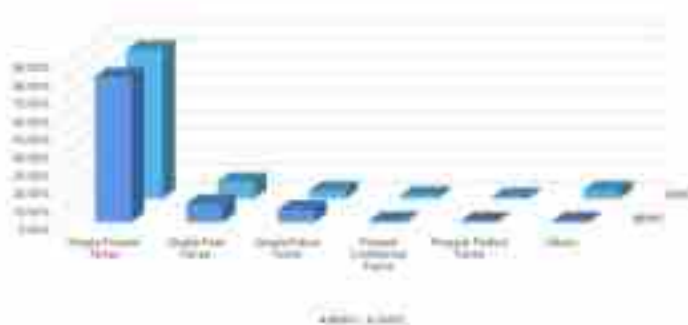
Table 4.3 Distribution of Tense Used in the Corpora

Tense	MNHC		GNHC	
	Number	Ratio	Number	Ratio
Simple Present Tense	8067	80.74%	8256	82.90%
Simple Past Tense	1027	10.28%	931	9.34%
Simple Future Tense	702	7.03%	458	4.60%
Present Continuous Tense	79	0.79%	181	1.82%
Present Perfect Tense	55	0.55%	87	0.87%
Others	61	0.61%	47	0.47%
Total	9991	100%	9971	100%

04 Results and Discussion

4.2 Sentence Tenses (Cont'd)

Figure 4.3 Bar Chart of Frequency of the Tense Used in the Corpora



04 Results and Discussion

4.2 Sentence Tenses (Cont'd)

According to the Table 4.3 and Figure 4.3 above, it can be seen that the simple present tense is generally the first in both MNHC and GNHC; this finding is consistent with the previous research on news headlines introduced in Chapter 2. Zhang Jian (1990) indicated that in order to enhance the freshness and immediacy of news reports, the present tense is often used instead of the past tense. Xu Xiaoping (2009) claimed that using the present tense can show the timeliness of news reports. The simple past tense ranks second, with 1027 occurrences in MNHC and 931 occurrences in GNHC; simple future tense ranks third in MNHC with a small proportion.

In general, the distribution of tenses in the two corpora follows a similar pattern, which supports previous studies on the tenses of news headlines.

04 Results and Discussion

4.2 Sentence Tenses (Cont'd)

4.2.1 Simple Present Tense

In the most cases, the events of maritime news headlines occurred in the past time. The simple past tense is supposed to be used in the writing. However, the journalists prefer using the simple present tense in most cases to reflect the timeliness of the news, which is mentioned in the introduction to characteristics of the news in Chapter 2. Here are the examples:

- ITF **assists** crews Sudan and Lebanon first time
- Safety at Sea and BIMCO **publish** cyber security white paper.
- Neste **Brings** New IMO 2020-Compliant Fuel to Germany

04 Results and Discussion

4.2 Sentence Tenses (Cont'd)

4.2.2 Simple Past Tense

In many cases, news events occurred in the past, so it is natural to use the past tense in maritime news headlines, especially when the events happened at a relatively early time. In addition, when the purpose of the news is to state facts, the past tense can be used.

Examples:

- Petronas **wrapped up** its first LNG bunkering in Port Klang
- New Generation Antenna Platform To Future-Proof Vessel Connectivity In An Uncertain World **Launched**
- Many seafarers **could quit** the industry over mishandling of COVID-19 crew change crisis

04 Results and Discussion

4.2 Sentence Tenses

4.2.3 Future Tense

The basic structures of the future tense include "be + to do", "will + verb", "be going to", and "be about to". Due to space limitations, "be going to" and "be about to" are used the least in news headlines. However, there is another form of expressing "be to do" in the future tense. In addition, the verb "be" is usually omitted from news headlines. Generally speaking, infinitive forms of verbs (such as to go, to do, etc.) are used more frequently in maritime news headlines than the verb will. The reason why infinitives are used more is that the expression of infinity "to" is made up of two letters, while the word "will" consists of 4 letters. The length of the word "to" is shorter than the length of the word "will". In fact, these two words have no practical meaning. Shorter words are likely to be used in news headlines, so infinitives are often employed in news headlines.

04 Results and Discussion

4.2 Sentence Tenses

4.2.3 Future Tense (Cont'd)

Table 4.4 Distribution of Two Forms of the Future Tense in MNHC

	Frequency	Ratio
will + do	57	8.12%
To do	645	91.88%
Total	702	100%

Figure 4.4 Distribution of Two Forms of the Future Tense in MNHC



04 Results and Discussion

4.2 Sentence Tenses

4.2.3 Future Tense (Cont'd)

For example:

Carnival Corporation Brands To Resume Cruising In US, Caribbean And Europe

Wisdom Marine to Expand Fleet with Two More Bulkers

Rockhopper to resolve Ombrina Mare arbitration in July

04 Results and Discussion

4.3 Sentence Voices

- In the field of news headlines, a great number of studies have remarked that the active voice is more preferred than the passive voice.
- The active voice is considered more colorful and fantastic in the view of rhetoric, and more avant-garde and powerful in expression. (Zhuang, 2006). However, if the author's purpose is to emphasize the action or the object, the passive voice is usually used and the verb "be" is omitted.
- Zu (2005) also found that the active voice has priority in news language, for the sense of directness can be achieved.

04 Results and Discussion

4.3 Sentence Voices (Cont'd)

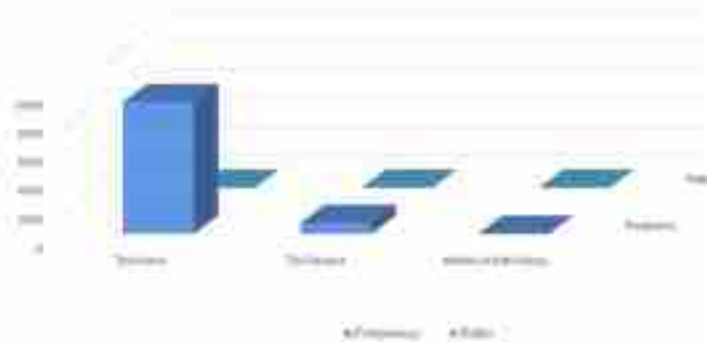
Table 4.5 Distribution of the Active Voice and the Passive Voice in MNHC

	The Active	The Passive	Mixture of both Voices	Total
Frequency	9229	760	2	9991
Ratio	92.37%	7.60%	0.20%	100%

04 Results and Discussion

4.3 Sentence Voices (Cont'd)

Figure 4.5 Distribution of the Active Voice and the Passive Voice in MNHC



04 Results and Discussion

4.4 Sentence Omissions

As is known to all, simplicity and accuracy are the most important characteristics of news headlines. Some words need to be omitted with the purpose of keeping the title as short as possible at times.

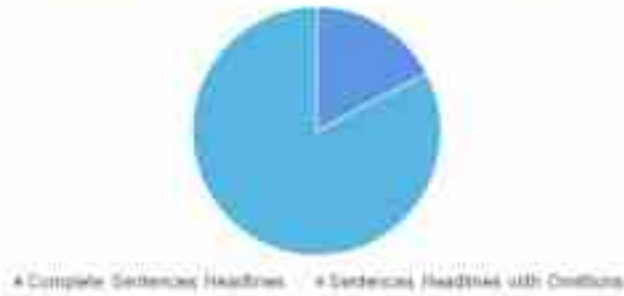
Table 4.6 Distribution of Sentence Headlines and Sentence Headlines with Omissions

Headlines	Number	Ratio
Complete Sentences Headlines	167	17.36%
Sentences Headlines with Omissions	795	82.64%
Total	962	100%

04 Results and Discussion

4.4 Sentence Omissions (Cont'd)

Figure 4.6 Distribution of Sentence Headlines and Sentence Headlines with Omissions



04 Results and Discussion

4.4 Sentence Omissions (Cont'd)

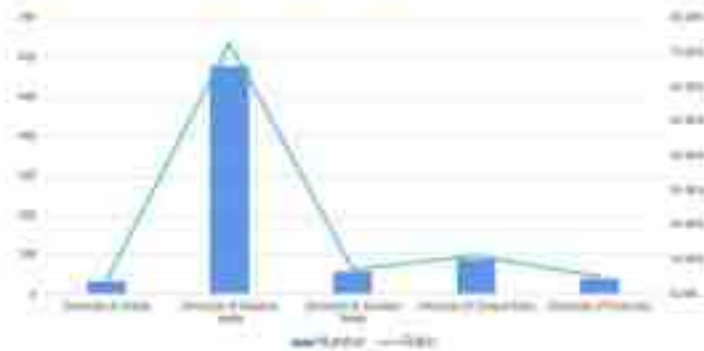
Table 4.7 Distribution of Sentence Headlines with Omissions in S-MNHC

S-MNHC	Omission of Article	Omission of Relative verbs	Omission of Auxiliary verbs	Omission of Conjunctions	Omission of Pronouns	Total
Number	32	577	57	89	40	795
Ratio	4.03%	72.58%	7.17%	11.19%	5.03%	100%

04 Results and Discussion

4.4 Sentence Omissions (Cont'd)

Figure 4.7 Distribution of Sentence Headlines with Omissions in S-MNHC



05 Conclusion

5.1 Major Findings: Summary

This study explored the syntactic features of maritime news headlines from four aspects: sentence structures, sentence patterns, sentence voices and omissions in sentences by establishing MNHC and GNHC corpora. The language of maritime news headlines belongs to the genre of maritime English and also serves as a sort of news language. Therefore, the syntactic features of maritime news headlines are similar to general news headlines, but there are also certain differences.

05 Conclusion

5.1 Major Findings: Firstly,

● Sentence Structures

Simple sentences are used much more frequently in maritime news headlines than general news headlines. Among the seven basic modes of simple sentences, **SVD**, **SVOA** and **SVA** are the most frequently used modes in maritime news headlines and general news headlines. In terms of multiple sentences, the frequency of use of maritime news headlines is lower than that of general news headlines.

In addition to considering conciseness and space limitations, maritime news headlines are different from general news headlines in that they are related to the shipping industry and contain some necessary technical terms and jargon. Therefore, using a large number of short sentences plays a critical role in reducing the reading difficulty and attract the attention of readers.

05 Conclusion

5.1 Major Findings: Secondly,

● Sentence Tenses

The **simple present tense** is extensively used to represent past events in both MNHC and GNHC. The finding is consistent with previous research, which found that the simple present tense is commonly used to replace the past tense, because the present tense can express the timeless of the news. For the sake of space, when the **future tense** is needed, the **infinitive form "to do"** is preferred instead of the "will do" form.

05 Conclusion

5.1 Major Findings: Thirdly,

- **Sentence Voices**

In terms of sentence voice, the **active voice** is used more frequently in maritime news headlines and general news headlines.

On the one hand, the use of active voice is appreciable in emphasizing the active role of the agent, helping readers to read the title and master the main ideas more effectively. On the other hand, the active voice is deemed to have more advantages in attracting readers' attention. Nevertheless, if the receiver of the action needs to be emphasized in the news headline, the passive voice should be used.

05 Conclusion

5.1 Major Findings: Fourthly,

- **Omission in Sentences**

The omission of **articles**, **relative verbs**, **auxiliary verbs**, **conjunctions**, and **some pronouns** can be found both in maritime news headlines and general news headlines, and most of the omitted words only carry low information. In this study, we focus on the omission of the S-MNHC. Omission of **relative verbs** occurs most frequently and accounts for 72.58%.

05 Conclusion

5.2 Implications

- Firstly, maritime news headlines play a significant role in conveying shipping industry news.
- Secondly, for editors and reporters of news headlines, especially writing of maritime news headlines, mastering the syntactic features can help them create correct, vivid, and concise headlines.
- Thirdly, the translators, who need to translate maritime news headlines, should understand and appreciate and translate headlines at the same time. This study provides some help for them to distinguish sentence-structures, tenses, voices and omissions based on the study of the syntactic level of maritime news headlines. Only when translators have an accurate understanding of the headlines, they can translate them accurately and fluently.

References

- Bell, A. (1991). *The language of News Media* (pp. 84-85). Oxford: Blackwell.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
- Curme, G. D. & Hurst, H. (1933). *A Grammar of the English Language in Three Volumes* (Vol. 3). Heath.
- Fairclough (1989). *Language and Power*. Longman, London and New York.
- Garrison, B. (1998). *Computer-Assisted Reporting*. Psychology Press.
- Guo, M.Y. (2014). *A Corpus-based Study on the Lexical Features of Maritime English News*. M.A. Thesis. Dalian: Dalian Maritime University.
- Larovic, Edith, Amel, Rodica. (1989). *The Strategy of the Headline*. *Semiotica* 77(4), 441-459.
- Leech, G. (1991). *Corpora and Theories of Linguistic Performance*. Berlin: Mouton de Gruyter, 125-148.
- Leech, G. (1992). *Corpora and Theories of Linguistic Performance*. *Directions in Corpus Linguistics*, 105-122.
- Lees, R. B. & Chomsky, N. (1957). *Syntactic Structures*. *Language*, 33(3 Part 1), 375-408.
- Li, F. F. (2013). *A Study on Lexical Cohesion in Maritime News*. M.A. Thesis. Dalian: Dalian Maritime University.
- Liu, Y. P. (2014). *A Corpus-based Study on Lexical Bundles in Maritime News*. M.A. Thesis. Dalian: Dalian Maritime University.
- Quirk, R., Greenbaum, S., Leech, G & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

References

- Leech, G. (1991). *The Language of Newspapers*. London: Routledge.
- Sinclair, J. (1991). In J. Sinclair and R. Carter, Editors, *Corpus, Concordance, Collocation*. Oxford: University Press.
- Sinclair, J. (1999). *The Computer, the Corpus and the Theory of Language*. *Lingua*, 7(99), 24-32.
- Van Dijk, T.A. (1998). *News Analysis: Case Studies of International and National News in the Press*. New Jersey: Erlbaum Associates.
- Yule, G. (2000). *The Study of Language*. Beijing: Foreign Language Teaching and Research Press.
- Zhou, Z. Y. (2012). *On the Sentence Structure Features of China English-A Comparative Analysis of International Political News in China Daily and The New York Times*. M.A. Thesis. Shanghai: Shanghai Normal University.
- 胡壮麟(Hu Zhuanglin). (2000). 《理论文体学》. 北京: 外语教学与研究出版社.
- 罗天法(Luo Tianfa). 从自建语料库看英语新闻标题的文体特色. 广西广播电视大学学报, 2007, (3): 70-73
- 张健(Zhang Jian)(2004). 新闻英语文体与范文评析. 上海: 上海外语教育出版社, 2004.
- 章振邦(Zhang Zhenbang). 张月村(Zhang Yuecun). 新编英语语法教程. 上海: 上海外语教育出版社, 1995.



Thank You Very Much For Listening to My Presentation.

Any Questions?

제4분과

응용/언어교육





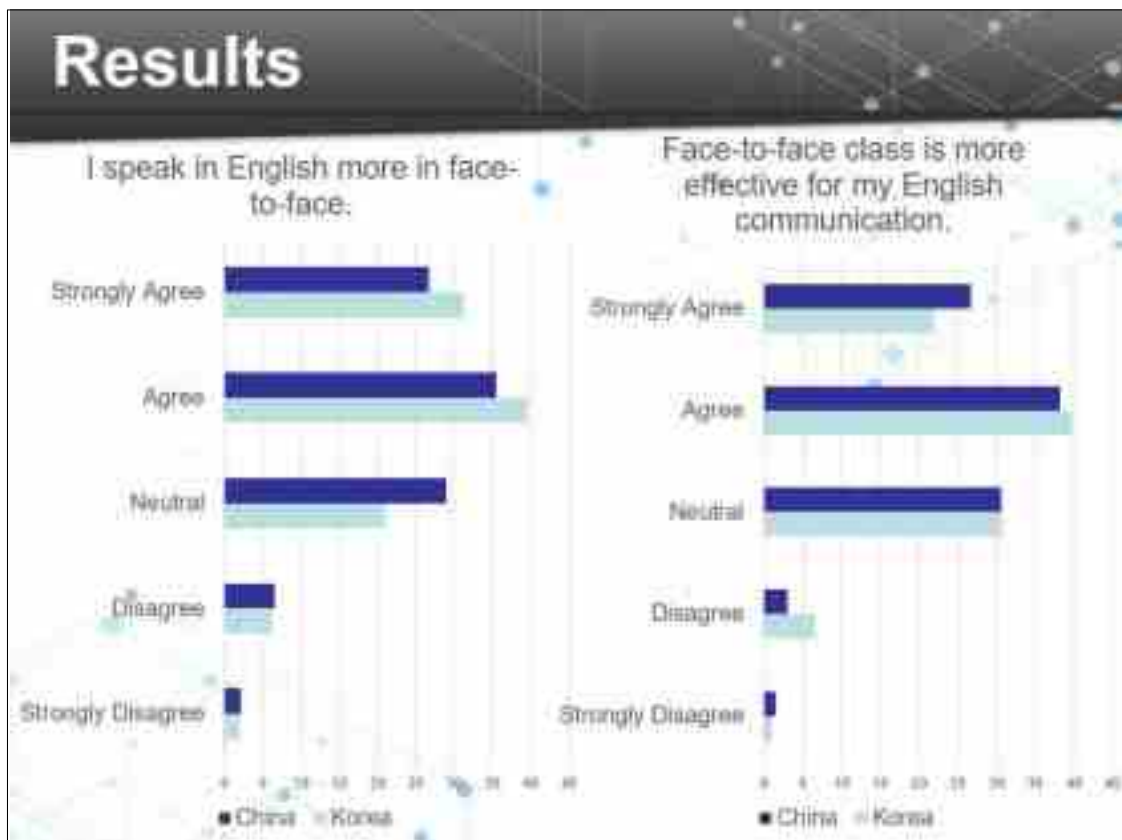
2. Academic Significance

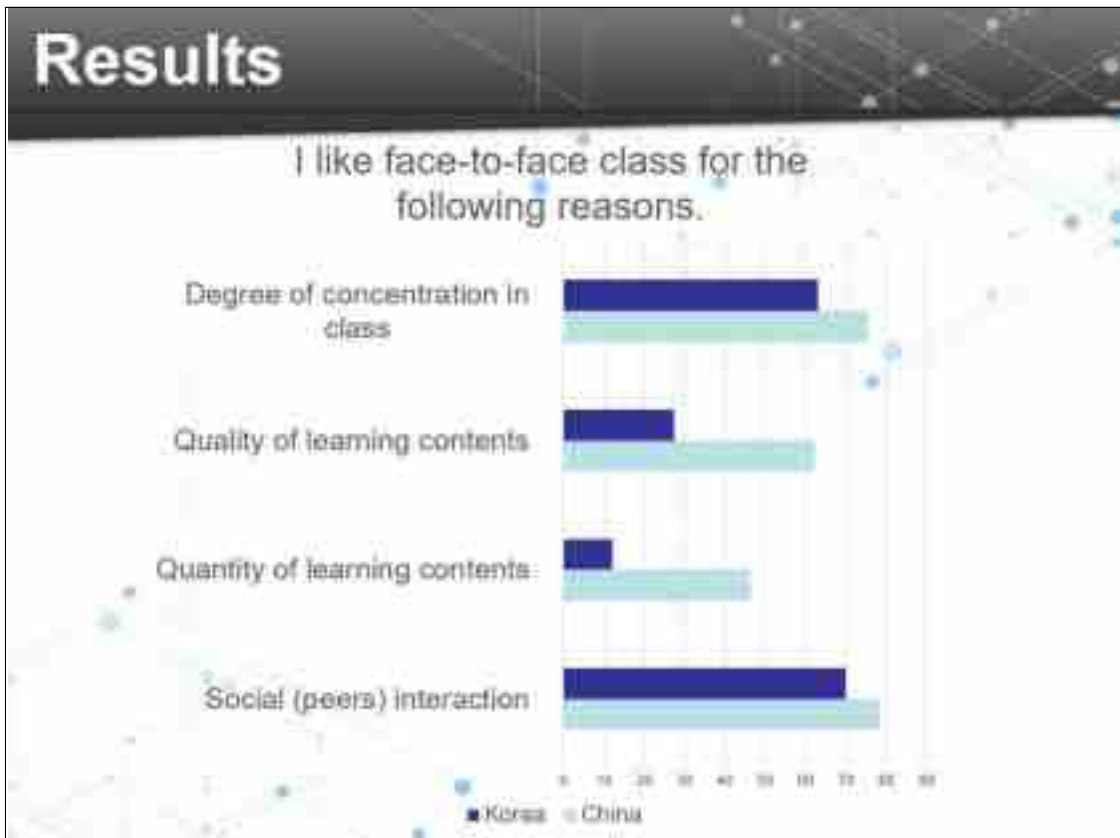
Academic Significance

Online Education	Students' Perception	Online English Education
<p>Quality, Flexibility, Sensitivity (Alvandar & Gola, 2007; Blac et al., 2007)</p> <p>Factor Analysis: Learners, Instructors, System, and Learning (Daryaning, Farhan, and Khrmaz, 2020; Kim, Shin, & Jung, 2020)</p>	<p>Mixed Results on Students' Satisfaction in Psychology, Sociology, Chemistry, Biology, Computing Classes (Dinnit, Benoit, Miao, and Hansen, 2020; Lester & Kline, 2008)</p>	<p>Students' Preference of Face-to-Face Learning (Moqadam-Tabrizi, 2018; Tratik, Um & Jereb, 2019; Sitwchai, 2019)</p>

3. Present Research







Qualitative Result

Teacher Pedagogy

(I want teacher to) leave some questions for students to think for themselves when teaching online.

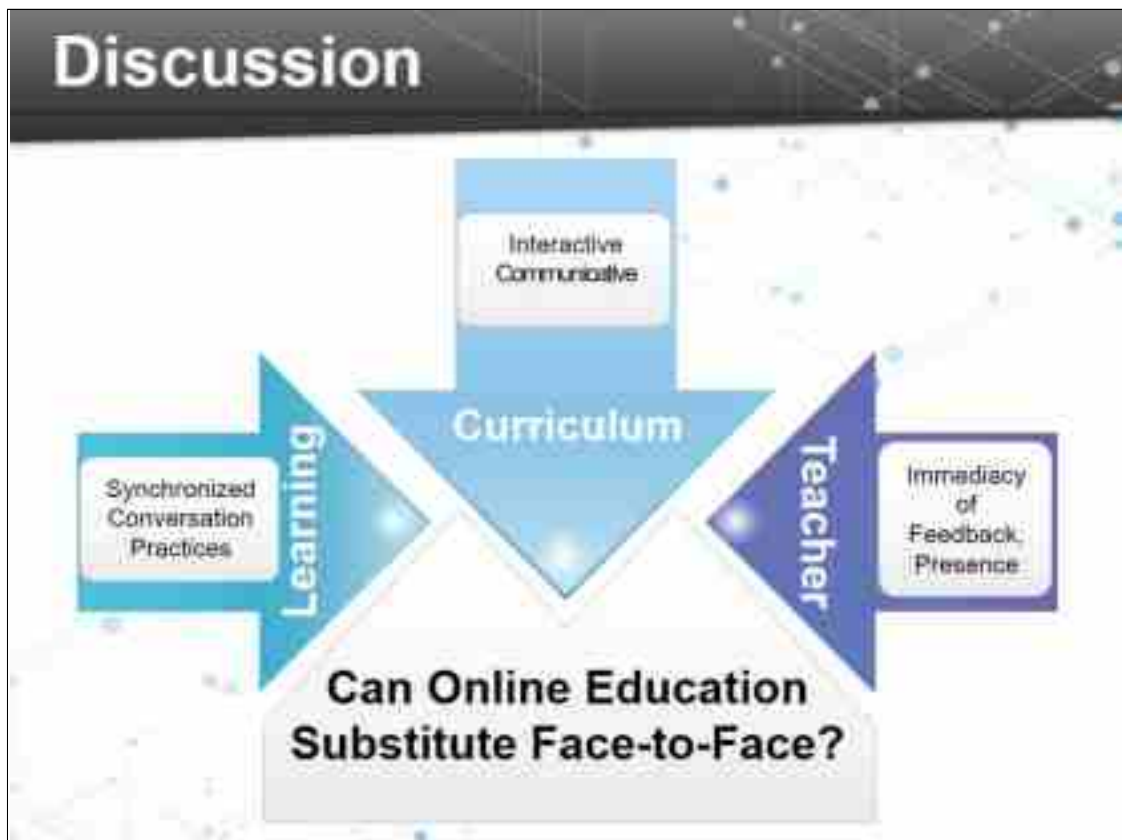
I hope that in face-to-face instruction, teachers can increase interaction, rather than simply ask people to answer questions.

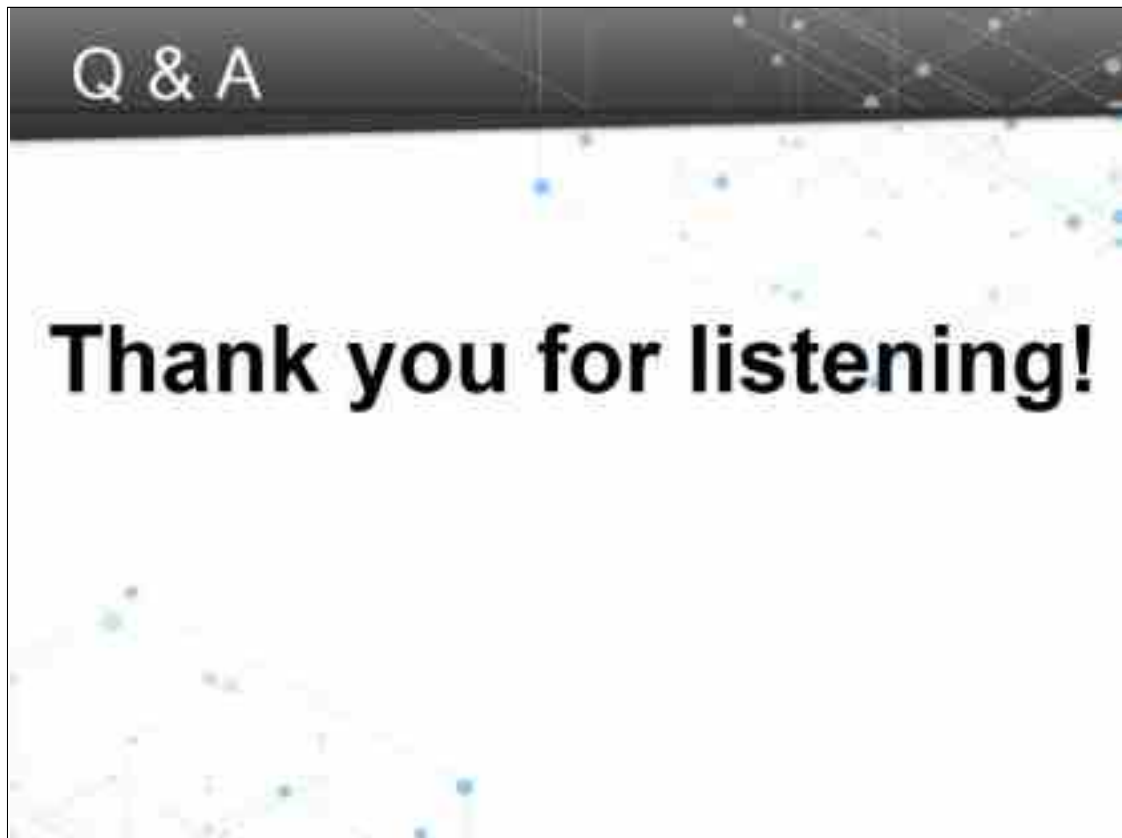
Face-to-face vs. Online

I prefer face to face.

I learn more knowledge, and more focused in face-to-face class.

The combination of face-to-face and online teaching methods can be adopted.





L2 Learners' Perceptions of Online English Learning in the Post-Corona Era: Focusing on Affective Factors and Language Proficiency

Youusn Shin

(Suncheon National University)

I. INTRODUCTION

Over the past decades, online classes were considered to be an alternative teaching method to traditional classroom classes (Artino, 2020; Larreamendy-Joerns & Leinhardt, 2006; Tallent-Runnels et al., 2006;), and now online classes are being almost exclusively used as one of the main teaching methods particularly over the last two years. The unprecedented coronavirus pandemic has led to a huge change in the way instructors deliver classes to students, but due to the time limitations, they have focused on how learners have access to content rather than worrying about how well their teaching materials are designed and delivered (Egbert, 2020). Instead of classes held in the classroom, students were required to complete assignments and learn L2 at home which eventually led to a paradigm shift in teaching and learning. Therefore, we need to be more aware of students' perception of online classes, and related research is also required to improve the quality of online classes we have currently conducted in order to deal with learning descriptions and losses caused by the current coronavirus.

With so many changes already underway in the educational environment, the shift to the online class paradigm will not be able to revert to educational practice in the pre-coronavirus period. In this sense, we need to know how online L2 classes, which are not most commonly utilized under the pandemic context, are accepted by L2 prospective teachers. This study was expected to tell intriguing results in that the roles of participating students can be defined as learners now and as teachers who will be teaching students after graduation. It is very timely to know how positive or negative feelings, L2 proficiency, values for their achievements are interacting with the perceptions of online L2 classes. Additionally, it is necessary to respond well to educational changes affected by the emergence of a new variable called coronavirus, and to get prospective teachers prepared for providing better educational resources to students. The

main purpose of the current study is, thus, to examine the correlations between the perceptions of online L2 classes, affective factors, and L2 proficiency. By doing so, this study will be able to provide researchers and instructors with better understanding of the effective factors and language ability which might affect online learning.

II. LITERATURE REVIEW

The contextual features formulating the study consist of students' affective factors such as enjoyment or anxiety, the value of learning activities, L2 proficiency, and perceptions toward online L2 learning which are presumed to link to important academic outcomes, such as satisfaction or academic achievement. Thus, for this study, four components were used to assess learners' enjoyment and anxiety, subjective task value, L2 proficiency, and perceptions of online learning during the COVID-19 pandemic in Korea.

1. Perceptions of Online L2 learning

Some previous research focusing on different types of teaching contexts, either transitional classroom or online learning environment, pointed out that online approach not only makes language learning more flexible and authentic but also it may play a role in increasing learners' attendance and task engagement (Gacs, Goertler., & Spasova, 2020; Felix, 2008). On the other hand, other studies showed that positive effects of online learning can be limited by technical and individual issues which include both learners and teachers' low self-efficacy for online learning or computer use (Artino, 2010), and learners' low motivation or low active participation (Hampel, 2003).

Recent research conducted by Islam, Kim., & Kwon (2020) showed that learners' preferences during the COVID-19 pandemic leaned toward learning materials which can be found on online platforms, and in which learners are able to meet their own needs. Apart from learning materials, being able to effectively manage class time can be part of essential conditions for successful online learning (Johnson, Veletsianos., & Seaman, 2020).

2. L2 learners' Affective Factors toward Language Learning

In language learning context, anxiety has been described as one of the strongest predictors in terms of success or failures (MacIntyre, 1999). MacIntyre and Gardner (1994, 1994) showed that anxiety can play a disruptive role to acquire and produce language during every stages of the input, processing and output process. Previous studies have consistently shown that foreign language anxiety

negatively affects communication and learning itself and potential sources of foreign language anxiety include harsh error correction done by teachers or peers (Gregersen, 2003; Young, 1991), competitiveness among peers (Bailey, 1983), concerns about public speaking (Cohen & Norst, 1989), tolerance for ambiguity in L2 (Dewaele & Shan, 2013). While a majority of research has been concentrated on negative emotions of language learners, Arnold and Brown (1999) conducted the study about positive foreign language emotion and concluded that

Task value as one of the motivational beliefs can be defined as learners' judgement of how meaningful, useful, and interesting a program is to learners (Eccles & Wigfield, 2002). The result of the previous research suggested that task value belief positively predicted learners' academic achievements including choice of future learning activities or professions (Eccles & Wigfield, 2002), and school achievement (Artino & Stephens, 2009; Zusho, Pintrich, & Coppola, 2003).

Research questions of this current study were formulated as follows:

- (1) Are there any significance among L2 learners' emotions toward language learning (the Foreign Language Enjoyment questionnaire & the Foreign Language Classroom Anxiety), subjective task values, L2 language proficiency (QPT & self-perceived L2 proficiency), and perceptions of online L2 learning during the COVID-19 pandemic crisis by grade levels and gender? (one-way ANOVA)
- (2) What are the relationships among the variables – emotions toward language learning, subjective task values, L2 language proficiency and the perceptions of online L2 learning? (correlational study)
- (3) To what extent do L2 learners' emotions, subjective task values, L2 language proficiency predict participants' perceptions of online L2 learning during the COVID-19 pandemic crisis? (multiple regression)

The main purpose of the present study is thus to identify the independent variables linked to the perceptions of L2 online learning that can help instructors create the optimal learning environment to make up for academic losses due to the unprecedented COVID-19 pandemic crisis since 2020.

III. METHODS

1. Participants

A total of 67 participants (Female: 45, Male: 22) completed all sets of online questionnaires. Participants's age ranged from 18 to 30 ($M=22.34$, $SD=2.29$). Participants were Koreans and had Korean as their first language. They have majored in English Education at a university located in Jeolla province and had studied English as a foreign language at least since 3rd grade at elementary school

for 10 years at the time of the study. The number of participants by grade is as follows: 1st year (N=15), 2nd year (N=15), 3rd year (N=26), 4th year (N=11).

2. Instruments

The instruments used in the study were as follows:

(1) The Foreign Language Enjoyment questionnaire (hereinafter referred to as FLE) & The Foreign Language Classroom Anxiety Scale (FLCA): FLE and FLCA were used and adapted from the original forms in translated forms. The FLE has 19 items and FLCA has 8 items, with options on a 5-point Likert scale, from 1 (Strongly Disagree) to 5 (Strongly Agree). The FLE was developed to assess learners' enjoyment toward foreign language learning (i.e., favorable learning atmosphere, positive reactions to foreign language learning, teacher's support or encouragement, etc) while the FLCA measures foreign language anxiety (i.e., anxiety, fear, confusion, discomfort, etc). Cronbach Alpha of FLE is .881 (19 items) while Cronbach Alpha of FLCA is relatively low with .799 (8 items), which can be considered as reliable.

(2) Quick Placement Test (QPT): this test was designed to measure L2 learners' English language proficiency, which was developed by Oxford University Press and Cambridge ESOL to provide teachers with reliable but accessible results for their level of English proficiency. The test consists of 60 multiple choice items (maximum possible score: 60). Of the two available versions (computer-based & a paper and pen), a paper and pen version was used in this study. The majority of the participants took this test when they were in their 1st or 2nd year of university. The results of the QPT were converted into band descriptors and then compared to the results of self-perceived language proficiency based on the Common European Framework of Reference for Languages.

TABLE 1
Score Reports of QPT and Equivalent Level Descriptors of Council of Europe Level

QPT scores	Level & Descriptors	Council of Europe Level & Descriptors
0-10	0.1 Beginner	--
11-17	0.2 Breakthrough	A1 Basic User
18-29	1 Elementary	A2 Basic User
30-39	2 Lower Intermediate	B1 Independent User
40-47	3 Upper Intermediate	B2 Independent User
48-54	4 Advanced	C1 Proficient User
55-60	5 Very advanced	C2 Proficient User

(3) Self-Perceived Language Proficiency: The participants were asked to identify their self-perceived English proficiency based upon the translated form of the Common European Framework of Reference for Languages (CEFR). CEFR consists of five major communicative functions (Listening, Reading, Spoken Interaction, Spoken Production, and Writing), with six levels of language proficiency (A1 & A2 Basic user, B1 & B2 Independent user, C1 & C2 Proficient user). In each grid, 'can-do' statements describe language learners' ability to perform with language at each proficiency level. Cronbach Alpha for self-perceived language proficiency is 0.851.

(4) Subjective Task Value questionnaire: This questionnaire consists of four main categories (Efficacy for English, Interest Value, Utility Value and Attainment Value) with 20 items, which were adapted from the original form and were translated into Korean. The items are intended to measure the participants' intrinsic task value of English learning. Cronbach Alpha is very high with .901.

(5) Perceptions of L2 Online Learning questionnaire: The questionnaire comprises 20 items with the five major categories (Perceived usefulness, Attitude toward using, Intention to use, Computer anxiety, Enjoyment, Self-efficacy). It is intended to identify L2 learners' perceptions of the general online learning environment. The reliability index, Cronbach Alpha, shows that it is very high with .952. All the items in the questionnaire were measured on 5-Likert scale and higher scores mean higher levels of perception on each item. In the case of negative statements in Likert scales, reverse coding was performed and calculated for reliability and other statistics.

3. Procedures and Analysis

The participants were asked to complete the Quick Placement Test at the beginning of the year, 2018, 2019, 2020, and 2021. The rest of the questionnaires (FLE, FLCA, self-perceived language proficiency, Subjective Task Value) were collected during the 2021 spring semester and entered into the Excel program for further analysis. SPSS 26 was used to analyze the collected data and to obtain the validity of the instruments in this study.

IV. RESULTS

The participants' QPT mean scores were 35.42 (Possible Maximum Score: 60) (SD=5.889) while the total mean of self-perceived language proficiency was 3.34 (SD=.771). Of the four questionnaires, the mean scores of the FLCA were the lowest (M=2.71, SD=.691) while the mean scores of the Subjective Task Value were the highest (M=4.14, SD=.475), followed by FLE mean scores (SD=4.03, SD=.456).

TABLE 2
Descriptive Analysis of Language Proficiency, Emotions, and Perceptions of Online Learning

Variables	M	SD	Min	Max	왜도	첨도
QPT	35.42	5.889	19.00	48.00	-.060	-.116
Subjective Task Value	4.14	.475	2.90	5.00	-.396	-.205
FLE	4.03	.456	2.94	4.89	-.109	-.435
FLCA	2.71	.691	1.25	4.25	-.127	-.316
Listening	3.24	.971	1	6	.727	1.867
Reading	3.70	.835	1	6	-.509	1.504
Spoken Interaction	2.82	.920	1	5	.369	-.074
Spoken Production	3.04	1.036	1	6	.161	.430
Writing	3.90	1.089	1	6	-.150	-.207
Language Proficiency (Total Mean)	3.34	.771	1.00	5.20	-.014	1.190
Perceptions of L2 Online Learning	3.91	.605	2.14	4.86	-.317	-.090

QPT mean scores of proficiency levels of the participants and self-assessed language proficiency (total mean) using CEFR were 35.42(SD=5.889) and 3.34(SD=.771), respectively. When switching these means to a band descriptor shown in Table 1, both means fell within the ‘upper intermediate’ level, implying both the test and self-assessment could serve as legitimate tools in assessing participants’ language proficiency.

1. Result of Grade Levels and Gender Effect

The first research question focuses on the effects of grade levels and genders on six dependent variables (QPT, Task Value, FLE, FLCA, Self-perceived language proficiency, perceptions of online L2 learning). For our analysis, two types of comparison groups were created depending on grade levels (1–2 years vs. 3–4 years) and gender (female vs male). The descriptive analysis in Table 3 shows that there was only one significant mean difference between the grade levels by university years on QPT ($F_{(1,65)}=21.887$, $p<.000$) scores. The overall score tendency suggested that 3–4th year students yielded better scores across all the variables even though the mean scores of the instruments were not high enough to show significance except QPT.

TABLE 3
Descriptive Analysis of all the Variables and ANOVA Results by Grade Levels

Dependent Variables	1-2 year(N=30) Mean (SD)	3-4 year(N=37) Mean (SD)	F	Sig.
QPT	32.04(5.332)	38.18(4.831)	21.887**	0.000
Subjective Task Value	4.10(0.565)	4.17(0.392)	0.379	0.540
FLE	3.98(0.498)	4.07(0.421)	0.675	0.414
FLCA	2.74(0.678)	2.68(0.711)	0.106	0.746
Language Proficiency (Total Mean)	3.34(0.880)	3.34(0.682)	0.000	0.998
Perceptions of L2 Online Learning	3.84(0.574)	3.97(0.631)	0.858	0.358

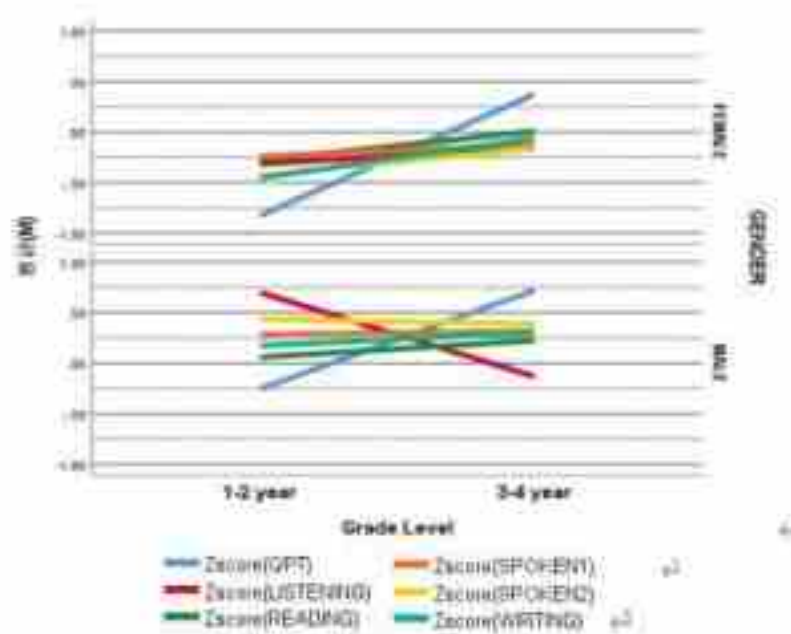
Table 4 showed the descriptive statistics of all the variables and ANOVA results with respect to the effect of gender. Male students obtained higher mean scores across all the instruments except the perceptions of online L2 learning. Unlike the effect of grade level, significant differences were found in FLE ($F_{(1,65)}=8.184$, $p=.006$), FLCA ($F_{(1,65)}=6.633$, $p=.012$), self-perceived language proficiency ($F_{(1,65)}=5.472$, $p=0.022$).

TABLE 4
Descriptive Analysis of all the Variables and ANOVA Results by Gender

GENDER	FEMALE (N=45) Mean (SD)	MALE (N=22) Mean (SD)	F	Sig.
QPT	34.90 (6.007)	36.38(5.678)	0.864	0.356
Subjective Task Value	4.06(0.463)	4.29(0.473)	3.372	0.071
FLE	3.929(0.425)	4.25(0.450)	8.184*	0.006
FLCA	2.56(0.708)	3.01(0.561)	6.633*	0.012
Language Proficiency (Total Mean)	3.19(0.754)	3.65(0.730)	5.472*	0.022
Perceptions of L2 Online Learning	3.96(0.553)	3.82(0.705)	0.802	0.374

Figure 1 delineates the score patterns of QPT and self-assessment language proficiency of the participants by grade levels and gender. For the comparisons among the scores, z-transformed scores were used in order to compare two different types of raw scores (*as part of a standard score, z-scores allow to compare different kinds of variables by standardizing the distributions).

FIGURE 1.
Score Patterns of QPT and Self-perceived Language Proficiency



Regardless of gender, the higher the grade, the higher the score on the QPT test. On the other hand, when comparing individual language skills by gender, male students generally outperformed female students in QPT and all language skills. The graph also demonstrated that the low-grade male group assessed their listening ability the highest among all the language skills, but the listening ability was rated the lowest for the high-grade male group. It should not be overlooked the possibility, however, that female students' self-assessment has been more stringent than male students. The reason may be a stricter propensity for self-assessment done by female students, or it could be because of the better language ability that male students have developed in the program.

Given the fact that most participants took QPT within a year after they got admitted into the program, the gap between their English ability before and after COVID-19 pandemic crisis becomes clear. Obviously, their ability difference cannot be attributed entirely to COVID-19, but it is assumed that it had some effect. In further research, it will be necessary to examine the impact of COVID-19 pandemic crisis and the efforts to recoup subsequent academic losses from it.

2. The Relationships among the Variables

Research question 2 addressed the interrelations among the variables. The correlations regarding the variables were tested using the Pearson correlation shown in Table 5. The perceptions of L2 online learning are positively and significantly correlated with foreign language enjoyment and subjective task value. The results

of correlation showed that the higher foreign language enjoyment and the subjective task value, the higher the participants' perception levels toward online L2 learning.

TABLE 5
Correlation Analysis among All the Variables (N=67)

Variables	1	2	3	4	5
1. QPT	1				
2. Language Proficiency (total mean)	.288*	1			
3. TV	.154	.428**	1		
4. FLE	.275*	.322**	.716**	1	
5. FLCA	.133	.443**	.384**	.169	1
6. Perceptions of L2 Online Learning	-.057	.066	.435**	.532**	-.013

* $p < .05$, ** $p < .001$

3. Critical Factors for the Perceptions of Online L2 learning

Starting from the correlation analysis, multiple regression analysis was conducted to test if the seven independent variables significantly predicted participants' perceptions of online L2 learning. The result revealed that only two predictor variables were found to be significant. The two predictor variables were gender ($t = -2.330$, $p < .024$) and foreign language enjoyment ($t = 3.320$, $p < .002$).

TABLE 6
Results of Regression Analysis for Variables Predicting Perceptions of Online L2 Learning

Dependent variable	Independent variables	B	β	t	P	VIF
	(Constant)	1.444		2.148	.036	
	Grade Level	.087	.071	.534	.596	1.565
	GENDER	-.371	-.291	-2.330*	.024	1.365
Perceptions of Online L2 Learning	QPT	-.025	-.237	-1.719	.092	1.672
	TV	.174	.138	.801	.427	2.591
	FLE	.744	.565	3.320*	.002	2.534
	FLCA	-.015	-.017	-.137	.891	1.386
	Language Proficiency	.012	.015	.114	.910	1.499
		$F_{(5, 61)} = 4.278$	$p = .002$	Adjusted $R^2 = .199$ $R^2 = .260$		

As shown in Table 6, the total amount of variance that was predictable from the two predictors was 26.0% and the adjusted R^2 change of 19.9%. The overall regression model was successful in explaining approximately 19.9% of the adjusted variance in the perceptions of online L2 learning. In sum, gender and foreign language enjoyment were found to be significantly associated with the perceptions of online L2 learning at $p < .05$. The result indicated that the perceptions of online L2 learning could be predicted with the degree of foreign language enjoyment and gender. The rest of the variables entered did not significantly contribute to predicting participants' perceptions of online learning.

V. DISCUSSIONS AND CONCLUSION

The current study intended to examine the participants' language proficiency, language learning enjoyment/anxiety, and subjective task value on the perceptions of online L2 learning during COVID-19 pandemic crisis. The first research question focused on the effects of grade levels and gender effect on different types of dependent variables. The ANOVA results showed that the grade levels of the participants only lead to significant differences in QPT, indicating that the participants who got admitted into the program before and after COVID-19 pandemic demonstrated different levels of language proficiency. Meanwhile, there was significance found in foreign language enjoyment, foreign language class anxiety, and self-assessed language proficiency with respect to gender. Male students obtained higher scores in all the above-mentioned variables compared to female students.

For the second research question, we tried to examine the relationships among all the variables used in this study which are assumed to have an impact on the perceptions of online L2 learning. The result of correlation analysis suggested that there were positive and significant correlations between foreign language enjoyment, subjective task value, and the perceptions of online L2 learning. The last research question was to identify crucial factors affecting participants' perceptions of online L2 learning during COVID period. The regression analysis was successful in explaining approximately 19.9% of the adjusted variance in the perceptions of online L2 learning. The finding showed that foreign language enjoyment and gender effect had contributed to the significant contribution to explain the perceptions of online learning.

REFERENCES

- Artino, A. R. (2010). Online or face-to-face learning? Exploring the personal factors that predict students' choice of instructional format. *The Internet and Higher Education*, 13, 272-276.
- Artino, A. R. (2012). Emotions in online learning environments: Introduction to the special issue. *The Internet and Higher Education*, 15, 137-140.

- Artino, A. R., & Stephens, J. M. (2009). Beyond grades in online learning: Adaptive profiles of academic self-regulation among Naval Academy undergraduates. *Journal of Advanced Academics*, 20, 568–601.
- Boudreau, C., MacIntyre, P., & Dewaele, J.-M. (2018). Enjoyment and anxiety in second language communication: An idiodynamic approach. *Studies in Second Language Learning and Teaching*, 8, 149–170. doi:10.14746/ssl.2018.8.1.7
- Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, 4, 237–274. doi:10.14746/ssl.2014.4.2.5
- Dewaele, J.-M., Franco Magdalena, A., & Saito, K. (2019). The effect of perception of teacher characteristics on Spanish EFL learners' anxiety and enjoyment. *The Modern Language Journal*, 103, 412–427. doi:10.1111/modl.12555
- Dewaele, J.-M., & MacIntyre, P. D. (2019). The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety. In M. Sato, & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of instructed second language acquisition studies* (pp. 263–286). Routledge. doi:10.4324/97811351190558-12
- Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety in the FL classroom: The effect of teacher and learner variables. *Language Teaching Research*, 22, 676–697.
- Eccles, J. (2005). Subjective Task Value and the Eccles et al. Model of Achievement-Related Choices. *Handbook of Competence and Motivation; Guilford Publications* (pp. 105–121).
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Egbert, J. (2020). The new normal?: A pandemic of task engagement in language learning. *Foreign Language. Annals*, 53, 314–319.
- Felix, U. (2008). The unreasonable effectiveness of CALL: What have we learned in two decades of research? *ReCALL*, 20, 141–161.
- Gacs, A., Goertler, S., & Spasova, S. (2020). Planned online language education versus crisis-prompted online language teaching: Lessons for the future. *Foreign Language. Annals*, 53, 380–392.
- Gregersen, T. (2020). Dynamic properties of language anxiety. *Studies in Second Language Learning and Teaching*, 10, 67–87.
- Hampel, R. (2003). Theoretical perspectives and new practices in audio-graphic conferencing for language learning. *ReCALL*, 15, 21–36.
- Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, 70, 125–132. doi:10.1111/j.1540-4781.1986.tb05256.x
- Islam, M., Kim, D.-A., & Kwon, M. (2020). A Comparison of Two Forms of Instruction: Pre-Recorded Video Lectures vs. Live ZOOM Lectures for Education in the Business Management Field. *Sustainability*, 12, 8149.
- White, C.J. (2017). Distance Language Teaching with Technology. *The Handbook of Technology and Second Language Teaching and Learning* (pp. 134–148), Wiley.
- Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education*, 25, 1081–1094.

글쓰기 과업에서 언어형태의 복잡성이 주목과 학습에 미치는 영향

조윤경

(부경대학교)

1. 서론

외국어 학습에 있어서 입력(input)의 중요성은 오랜 전부터 강조되었지만, 출력의 역할은 입력을 처리하여 생긴 언어지식의 결과물로서 다소 한정적으로 다루어져 왔다. 하지만, Swain(1985)은 출력가설(output hypothesis)을 제안하여 입력만큼이나 출력 또한 언어학습 과정에 중요한 역할을 한다고 강조하였다. 구체적으로 Swain은 출력이 학습과정에서 주목 유도 기능(noticing/triggering function), 가설 검증 기능(hypothesis-testing function) 및 상위 언어적 기능(metalinguistic function)을 가지고 있다고 설명하였다. 이 중에서 본 연구에서 다루고자 하는 주목 유도 기능은 출력 과정을 통해 학습자들이 모르고 있는 요소를 깨닫게 되며 이로 인해 이후의 입력자료에서 그와 관련된 요소에 더 주목하게 됨을 의미한다.

학습자들의 목표언어형태에 대한 주목은 학습과정에 매우 중요하다. 이는 특히 Schmidt(1990)의 주목가설(noticing hypothesis)을 통해 더욱 강조되어왔다. 주목가설에 따르면, 학습자들이 어떤 요소를 학습하기 위해서는 해당 요소에 의식적으로 주목해야 한다. Schmidt(1990, 1995)의 주목가설에 근거를 두어 Swain(1985)이 주장한 출력의 주목 유도 기능을 검증하고자 시도한 선행연구들은 각기 다른 연구결과를 보고하였다. Izumi와 그의 동료들(Izumi and Bigelow 2000; Izumi et al. 1999)은 출력 중심의 활동과 입력 중심의 활동에서 학습자들의 주목과 학습이 유사한 수준으로 이루어졌음을 보고하였다. 반면에, 국내에서 실시된 몇몇 연구들은 출력활동이 목표언어형태에 대한 학습자들의 주목 및 학습에 긍정적인 영향을 미치는 것으로 보고하였다.

한편, Uggen(2012)은 출력활동에서 목표언어형태의 복잡성 수준이 학습자들의 주목 및 학습에 미치는 영향을 분석하였다. 가정법의 과거형 구문(present hypothetical-conditional structure)과 과거완료형 구문(past hypothetical-conditional structure)을 각각 복잡성 수준이 낮고 높은 목표언어형태로 가정하고, 각각의 목표언어형태와 관련된 에세이 쓰기(essay writing) 활동이 학습자들의 주목 및 학습에 미치는 영향을 비교하였다. 분석 결과에 따르면, 학습자의 목표언어형태에의 주목에 있어서는 목표언어형태의 복잡성 수준에 따른 차이는 발견되지 않았다. 반면에, 사후 작문시험에서는 복잡성 수준이 높은 목표언어형태와 관련된 출력활동을 한 집단이 입력활동을 한 통제집단과 복잡성 수준이 상대적으로 낮은 목표언어형태와 관련된 출력활동을 한 집단보다 더 높은 점수를 받았다.

본 연구는 Uggen(2012)의 연구를 부분 반복연구(partial replication study)를 실시하여, Uggen(2012)의 연구와 동일한 목표언어형태를 대상으로 목표언어형태의 복잡성 수준에 따

라 출력활동이 학습자들의 주목 및 학습에 미치는 영향이 달리 나타나는지를 분석하고자 한다. 하지만, 출력활동으로 에세이 쓰기 활동을 활용한 Uggen(2012)과는 달리 본 연구에서는 그림보고 문장쓰기(picture-cued sentence writing) 활동을 사용하고자 한다. 다양한 출력양상이 나올 수 있는 에세이 쓰기 활동과는 달리 그림보고 문장쓰기 활동은 주어진 그림에 의거하여 문장을 작성함으로써 학습자들의 출력향상을 다소 통제할 수 있다. 이는 곧 출력활동을 통해 발현되는 학습자들의 출력 결과물과 출력과업 이후에 주어지는 예시글(model writing) 즉 입력자료와의 높은 유사성으로 인해, 학습자들이 예시글에 좀 더 주목할 가능성이 높음을 의미한다. 또한, 본 연구는 주목을 측정하기 위해 좀 더 엄격한 방법을 사용하고자 한다. 주목을 측정하기 위하여 예시글에의 밑줄 굵기를 분석한 Uggen(2012)과는 달리, 본 연구에서는 Philp(2003)의 연구에서 사용한 주목 측정 방법을 활용하고자 한다. 즉 첫 번째 출력활동에서 목표언어형태와 관련된 오류가 후속적으로 제시된 피드백을 받거나 입력 자료를 읽은 후에 실행한 동일한 출력활동에서 수정된다면 해당 요소와 관련하여 학습자들의 주목이 일어난 것으로 간주하고자 한다. 이러한 측정 방법은 학습자들의 모든 주목을 측정하기에는 한계가 있지만, 의식적인 인식(conscious attention)을 반드시 요구하지는 않는 밑줄 굵기에 비교하여 좀 더 정확하게 학습자들의 주목을 측정할 수 있다는 점에서 의미있다.

실험설계상에서 이러한 두 가지 측면을 변형하여, 본 연구는 한국 고등학교에 재학 중인 영어학습자들을 대상으로 Uggen(2012)의 부분 반복연구를 실시하고자 한다. 즉 본 연구는 출력활동이 학습자들의 주목 및 학습에 미치는 영향이 목표언어형태의 복잡성 수준에 따라 다르게 나타나는지를 분석하고자 한다. 또한 목표언어형태에 대한 학습자의 주목과 학습간의 관계가 목표언어형태의 복잡성 수준에 따라 달라지는지를 살펴보고자 한다. 구체적인 연구과제는 다음과 같다.

- (1) 출력활동에서 학습자들의 목표언어형태에의 주목이 목표언어형태의 복잡성 수준에 따라 다르게 나타나는가?
- (2) 출력활동에서 학습자들의 목표언어형태 학습이 목표언어형태의 복잡성 수준에 따라 다르게 나타나는가?

2. 연구 방법

본 연구에는 2016년 2학기에 부산시 소재 고등학교 1학년 네 학급의 여학생 108명이 참여하였다. 연구에 참여한 학교는 인문계 고등학교이며, 해당 학교에서 연구 참여자들에게 정규 영어교과를 가르치는 교사가 본 연구의 모든 실험과정을 진행하였다. 실험에 참여할 네 학급은 두 실험집단과 두 통제집단으로 무선할당되었다. 두 실험집단은 복잡성 수준이 다른 두 목표언어형태에 대한 실험처치를 각각 받았다. 연구 참여자들은 상대적으로 복잡성 수준이 낮은 목표언어형태(즉 가정법 과거형 구문)에 대해 실험처치를 받은 실험집단 A반 27명, 복잡성 수준이 높은 목표언어형태(즉 가정법 과거완료형 구문)에 대해 실험처치를 받은 실험집단 B반 27명, 가정법 과거형 구문과 과거완료형 구문 각각에 대한 사전·사후 시험만을 수행한 통제집단 A반 27명과 통제집단 B반 27명으로 이루어졌다.

본 연구에서 활용된 연구 도구는 실험처치 전·후에 목표언어형태에 대한 학습자들의 지식 정도를 평가하기 위한 사전·사후시험이다. 우선, 사전·사후시험은 본 연구의 목표언어

형태인 가정법 과거형 구문과 과거완료형 구문에 대하여 각각 24문항 그리고 목표언어형태와 관련 없는 10문항으로 구성되었다. 사전·사후시험에서 목표언어형태와 관련된 총 24개의 문항들은 이해능력과 표현능력을 평가하는 문항들로 각각 50%씩 구성되었다. 이 중 이해능력 영역은 주어진 문장이 올바르게 표현된 것인지를 판단하는 문법성 판단 유형으로 이루어졌다. 표현능력 영역은 주어진 동사를 사용하여 해당 문장의 빈 칸을 완성하는 빈 칸 채우기 유형으로 구성되었다. Cronbach α 계수를 통해 분석된 사전·사후시험의 신뢰도는 각각 .83과 .86으로 분석되었다.

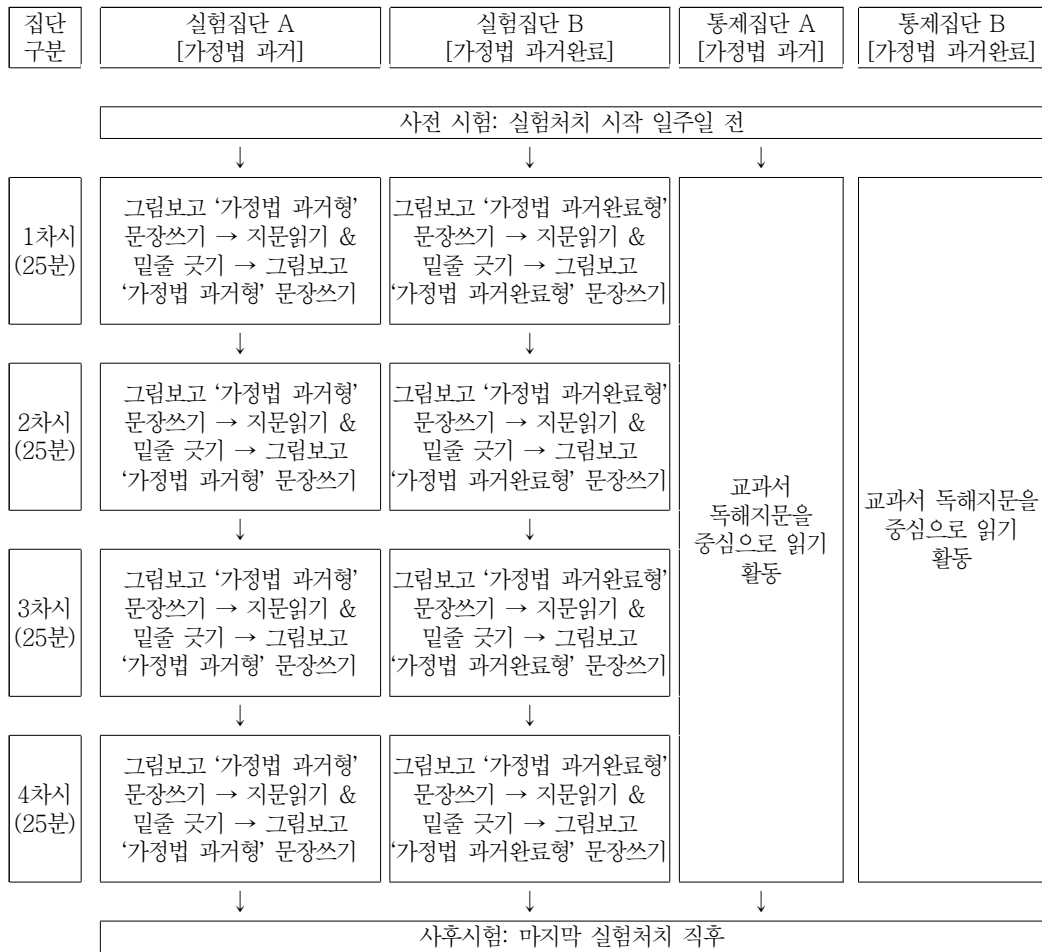
실험처치가 실시되기 전에 사전시험을 통해 목표언어형태에 대한 네 집단의 사전지식 수준을 비교하였다. 실험처치는 정규 영어교과 수업시간에 총 4차시 걸쳐 실시되었다. 각 차시는 제시되는 그림과 사용되어야 하는 표현만 다를 뿐 동일한 방법으로 실시되었다. 우선 두 실험집단의 학습자들은 3-4개의 문장으로 이루어지고 이야기의 배경이 되는 짧은 문단을 읽었다. 이 후에 학습자들은 각각 원인과 결과를 표현하는 2개의 그림을 총 4쌍을 받고, 각 쌍의 그림들이 의미하는 문장을 if-절을 사용하여 표현하였다. 즉 각 차시당 총 4문장을 작성하였다.

Chan-ho applied for an internship in an American company. The two-month internship includes travelling to the company's headquarters in the US and some of its branches in Europe. Yesterday, he went to the company for an English interview, but he failed. He regrets the following things:



문장쓰기 활동을 완료하고 활동지를 제출한 후에 실험집단의 학습자들은 문장쓰기 활동의 예시글이 될 수 있는 입력자료를 제공받았다. 해당 입력자료에 대해 읽기 활동을 마치고 예시글을 다시 교사에게 제출한 후에 학습자들은 앞서 수행했던 그림보고 문장쓰기를 다시 실시하였다. 각 차시의 활동은 약 25분이 소요되었다. 4차시의 활동이 완료된 직후에 실험집단들은 사후시험을 수행하였다. 한편, 두 통제집단은 실험집단에서 관찰될 수 있는 학습의 효과가 오로지 실험처치에 의한 것임을 보여주기 위하여 그리고 성숙효과를 배제시키기 위해 설정되었다.

두 통제집단은 사전·사후시험을 일주일 간격으로 수행하였다. 그리고 그 일주일동안 통제집단은 정규 영어 교과서의 독해 지문을 중심으로 교사의 강의 중심의 읽기 활동에 참여하였다. 구체적으로, 통제집단은 교과서의 독해 지문에서 제시된 어려운 어휘에 대한 설명을 듣고, 해당 지문의 주제문장을 찾고 세부 정보를 찾는 활동에 참여하였다. 이 때 독해 지문에서 목표언어형태는 제시되지 않았다.



<그림 1> 실험절차

3. 결과 및 논의

1) 출력활동에서 목표언어형태의 복잡성 수준이 목표언어형태에의 주목에 미치는 영향

본 절에서는 출력활동의 주목 유도 기능이 목표언어형태의 복잡성 수준에 따라 다른 양상을 보이는지를 분석하기 위하여 출력활동에 참여한 두 실험집단의 출력활동 결과를 분석하고자 한다. 이를 위해, 각 차시에서 첫 번째 출력활동에서 작성한 작문에서 목표언어형태와 관련된 오류 개수를 세고, 후속적으로 제시된 예시글을 읽은 후에 다시 작성한 작문에서 해당 오류가 수정된 개수 및 비율을 구하였다. 표 1은 4차시에 걸쳐 출력활동에서 나타난 오류 개수와 수정 개수 및 비율을 제시한다. 앞서 설명하였듯이, 가정법 과거형 구문과 관련된 출력활동에서 나타날 수 있는 오류 개수는 총 48개이며, 가정법 과거완료 구문의 경우에는 총 80개이다.

우선 가정법 과거형 구문과 관련된 출력활동에서 4차시에 걸쳐 한 학생당 평균적으로 약 41개(약 86%)의 오류가 나타났으며, 후속적으로 제시된 예시글을 읽은 후에 참여한 동일한 활동에서 해당 실험집단의 학습자들은 약 23개의 오류를 수정함으로써 처음 오류의 약 56%가 2차 작문에서 수정한 것으로 분석되었다. 2번의 동일 출력활동을 한 후에 최종적으로

로 나타난 오류의 개수는 4차시 합계 총 약 18개로 전체 가능한 오류개수의 약 38%로 나타났다. 한편, 가정법 과거완료 구문과 관련된 첫 번째 출력활동에서는 가능한 오류 총 80개에서 약 92%에 해당되는 약 74개의 오류가 관찰되었으며, 후속 예시글을 읽은 후에 다시 실행한 출력활동에서는 약 34개의 오류가 수정되어 처음 오류의 약 47%가 수정된 것으로 나타났다. 두 번째 출력활동에서 최종적으로 남은 오류의 개수는 전체 4차시 합계 약 39개로 나타났으며, 이는 전체 가능한 오류 개수 80개에서 약 49%에 해당된다.

<표 1> 출력활동에서 나타난 평균 오류 개수와 수정 개수 및 비율

	최초 오류		오류 수정		최종 오류	
	개수	비율 (최초오류/가능한 오류)	개수	비율 (수정/최초 오류)	개수	비율 (최종오류/가능한 오류)
과거형 실험집단 (가능한 오류 총 48개)	41.3	86.04%	23.21	56.20%	18.09	37.69%
과거완료 실험집단 (가능한 오류 총 80개)	73.84	92.30%	34.41	46.60%	39.43	49.29%

2) 출력활동에서 목표언어형태의 복잡성 수준이 목표언어형태 학습에 미치는 영향

본 절에서는 목표언어형태의 복잡성 수준에 따라 출력활동이 목표언어형태 학습에 미치는 영향이 다른지를 살펴보고자 한다. 우선, 사후시험 평균점수(24점 만점)는 가정법 과거형(이하 ‘과거형’) 실험집단과 가정법 과거완료(이하 ‘과거완료’) 실험집단이 각각 17.98점과 16.91점으로, 과거형 통제집단(M=8.96)과 과거완료형 통제집단(M=8.48)보다 높게 나타났다. 사후시험의 세부영역인 이해영역 및 표현영역에서도 유사한 양상이 발견된다. 이해영역(12점 만점)에서 과거형 실험집단과 과거완료형 실험집단의 평균 점수는 각각 8.65점과 7.54점으로 나타난 반면에, 과거형 통제집단과 과거완료형 통제집단의 평균점수는 각각 5.76점과 5.65점으로 실험집단보다 낮게 나타났다. 표현영역(12점 만점)에서도 과거형 실험집단(M=9.33)과 과거완료형 실험집단(M=7.87)의 평균 점수가 과거형 통제집단(M=3.20)과 과거완료형 통제집단(M=2.83)보다 높았다.

<표 2> 목표언어형태의 복잡성 수준 및 활동유형에 따른 집단별 사후시험 총점 및 세부영역 소합계에 대한 일원배치분산분석 및 다변량 분산분석 결과

	제곱합	df	평균제곱	F	유의확률	사후분석
총점	2072.787	3	690.929	48.308	.000	과거형 실험집단 = 과거완료형 실험집단 > 과거형 통제집단 = 과거완료형 통제집단
이해영역	170.917	3	56.972	16.438	.000	과거형 실험집단 > 과거완료형 실험집단 > 과거형 통제집단 = 과거완료형 통제집단
표현영역	872.433	3	290.811	47.898	.000	

이러한 평균점수의 차이가 통계적 유의성을 지니는지를 살펴보기 위해 사후시험 총점 및 세부영역 소합계에 대해 일원배치분산분석 및 다변량 분산분석을 각각 실시하였다. 표 2에서도 알 수 있듯이, 우선 사후시험 총점에서 목표언어형태의 복잡성 수준 및 활동유형에 따른 집단 간에 유의미한 차이가 나타났다. 사후분석 결과는 두 실험집단의 총점이 두 통제집단의 총점보다 유의미하게 높음을 보여준다. 또한, 이해 및 표현영역의 소합계에서도 목표언어형태의 복잡성 수준 및 활동유형에 따른 네 집단 간에 유의미한 차이가 나타났다. 사후분석 결과에 따르면, 이해 및 표현영역 모두에서 두 실험집단의 소합계는 두 통제집단의 소합계보다 유의미하게 더 높았으며, 과거형 실험집단이 과거완료형 실험집단보다 통계적으로 유의미하게 더 높은 점수를 받았다.

세부 평가영역에서 과거형 실험집단이 과거완료형 실험집단보다 유의미하게 높은 점수를 받았음을 보여주는 분석 결과는 정보처리 이론(VanPatten 1996, 2004)을 지지한다. 이 이론에 따르면, 외국어 학습자들의 정보처리 능력은 한정적이기 때문에, 목표언어형태와 의미를 동시에 처리하기에는 많은 어려움이 따른다. 특히 복잡한 목표언어형태를 수반하는 과업은 더 많은 정보처리 과정이 필요하기 때문에, 제한된 정보처리 능력을 가진 외국어 학습자들이 복잡한 목표언어형태에 집중하는 것이 어려울 수 있다. 이러한 인지적 부담은 학습에 부정적인 영향을 미칠 수 있다(McLaughlin and Heredia 1996; McLaughlin et al. 1983). 반면에, 상대적으로 복잡성 수준이 낮은 목표언어형태는 좀 더 수월한 정보처리를 수반하므로, 학습자들이 해당 목표언어형태에 좀 더 집중할 수 있게 되며 나아가 긍정적인 학습 결과를 가질 수 있다.

4. 결론

본 연구의 분석 결과를 요약하면 다음과 같다. 우선, 출력활동에 참여한 두 실험집단은 교과서 중심의 읽기 수업을 받은 통제집단보다 사후시험 점수에서 유의미하게 높은 점수를 받았다. 사후시험의 이해영역 및 표현영역 소합계에 대해서는 두 실험집단 중에서 상대적으로 복잡성 수준이 낮은 목표언어형태와 관련된 활동을 한 집단이 유의미하게 높은 점수를 받았다. 또한, 복잡성 수준이 상대적으로 낮은 목표언어형태와 관련된 출력활동을 한 집단에서 좀 더 많은 오류수정이 일어났다. 마지막으로, 복잡성 수준이 높은 목표언어형태를 학습한 집단보다는 복잡성 수준이 상대적으로 낮은 목표언어형태를 학습한 집단에서 오류수정 비율과 목표언어형태의 학습간의 상관관계가 더 강한 것으로 나타났다.

하지만, 이러한 연구결과는 다음과 같은 연구방법상의 제한점으로 인해 한정적으로 해석되어야 한다. 첫째, 본 연구에서는 첫 번째 출력활동에서의 오류가 관련 예시글을 읽은 후에 다시 실행한 두 번째 출력활동에서 수정되었다면 학습자는 입력자료에서 해당 목표언어형태를 주목하였다고 가정하였다. 이러한 주목 측정 방법은 목표언어형태와 관련하여 학습자들이 인지한 것을 모두 측정하였다고 간주하기 어렵다. 후속 연구에서는 주목을 측정하기 위해 좀 더 다각적으로 측정할 필요가 있다. 둘째, 본 연구에서는 학습자들이 목표언어형태를 적절하게 사용하는지를 측정하기 위하여 빈 칸 채우기 유형을 사용하여 학습자들의 표현 능력을 측정하였다. 이러한 방법은 한 문장에 국한하여 목표언어형태의 사용을 측정한 것이므로, 후속연구에서는 실험처치 후 학습자들의 학습 정도를 측정하기 위하여 자유작문(free writing) 또는 면담 등의 방법을 사용하여 학습자들이 좀 더 자연스러운 상황에서 목표언어형태를 정확하게 사용할 수 있는지를 평가할 필요가 있다.

참고문헌

- Shinichi, I., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly*, 34(2), 239-278.
- Shinichi, I., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition." *Studies in Second Language Acquisition*, 21(3), 421-452.
- Barry, M., & Heredia, R. (1996). Information processing approaches to research on second language acquisition and use. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 213-228). New York: Academic Press.
- Barry, M., Rossmann, T., & McLeod, B. (1983). Second language learning: An information processing perspective. *Language Learning*, 33(2), 135-157.
- Jennifer, P. (2003). Constraints of noticing the gap: Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25(1), 99-126.
- Richard, S. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Richard, S. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning." In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu, HI: University of Hawaii Press.
- Merrill, S. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & M. Carolyn (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Uggen, M. S. (2012). Reinvestigating the noticing function of output. *Language Learning*, 62(2), 506-540.
- VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.
- VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research and commentary* (pp. 5-32). Mahwah, NJ: Erlbaum.

한국 자폐스펙트럼장애 유아의 직접 화행 능력의 선별적 장애

이명민 (University of Hawaii at Manoa), 이주미 (University of Arizona), 정계영 (Columbia University), 이리숙 (연세대학교), 조석진 (세종대학교), 조영준 (서울대학교)

2021.8.19 한국언어학회 제9회 연례대회

Overview

- I. 뇌론
- II. 자폐스펙트럼장애 (ASD) 아동의 화행 능력
- III. 방법론
- IV. 결과 및 논의

I. 서론

자폐스펙트럼장애와 화형

- 자폐스펙트럼장애(Autism spectrum disorder, ASD)
 - 1943년 Leo Kanner의 연구 결과 (Kanner, 1943; Kasari & Paulson, 2018; de Vries et al., 2004)
 - 1980년 DSM-III 판본을 통해 공식적으로 인정받게 되지만, 최근에는 인지, 정서, 사회적 상호작용에 대한 다양한 양상과 범용성에 초점을 맞춤
- 화형(Speech act)
 - 화형이란, 화형의 의도나 청각적으로 지칭된 사물(대상)과 화형의 내용(의사)을 일치시키려는 시도를 통해 이루어지는 언어 사용의 과정이다.
 - "이제 그만 하자"
 - 화형의 목적은 화형의 내용과 화형의 의도를 일치시키는 것인데, 화형의 목적과 화형의 내용이 일치하는 것은 화형의 성공을 의미한다.
 - "이제 그만 하자"

마음이론 능력(Theory of Mind)

- 자신뿐만 아니라 타인의 마음과 행동을 이해하는 능력
- 아동 능력은 마음이론(theory of mind) 능력의 발달과 더불어 30개월 전후에 크게 상승함
- ASD 아동은 타인의 사고, 느낌, 의도를 감응에 따라 적절한 이해와 반응을 보이지 않음
 - 타인 눈치 보지 않기
 - 타인의 감정 상황에 알맞지는 적절한 반응을 보이지 않음

(Baron-Cohen, 2000; Tager-Flusberg, 2000)

8

II. 자폐스펙트럼장애(ASD) 아동의 화행능력

9

한국 ASD 아동의 화행능력 연구 (cont.)

- 최유원·최경은·홍민아(2017)
 - 초등학교 1학년에서 내한성 고기능자폐장애 아동 22명과 거동 과잉/과소 아동 22명의 능력 수준 연구
 - 특히 13세까지 유지된 경우, ASD 아동의 언어 화행 능력에 대한

- 조희숙 외(2007) 연구, 외관 다른 연구들은 언어 자료가 모두 비구조적 환경 상황에서 수집 → 피험자들의 화행 능력과 실제 기회가 균등하지 않음
- 간접 화행의 능력에 대한 연구는 아직 근거가 부족 → 보다 정밀한 분석 필요

⇒ 한국 자폐스펙트럼장애 (autism spectrum disorder)(ASD) 아동의 직접 화행(direct speech act) 능력 연구 필요

12

III. 방법론

13

자료 분석

총 46개 화행 코드, 7개 범주

1. 지시/반응 (Directives & Responses)
2. 발화유도/반응 (Speech Elicitation & Responses)
3. 약속/반응 (Commitments & Responses)
4. 고백/반응 (Declarations & Responses)
5. 표시/반응 (Markings & Responses)
6. 진술/반응 (Statements & Responses)
7. 질문/반응 (Questions & Responses)
8. 평가 (Evaluation)
9. 발성 (Vocalization)

17

자료 분석 (cont.)

- 연구자 1인→리서널 1인→무사별1인 < 총 3인 0(2인)도구법
- 총 3회 분기 - 1회기 자료 1회 분석 > 분석하지 않을 경우 0점 기록 가능
- 최종 1회기 분석 이후 나머지 2회기 자료도 2차 분석 > 100% 확인할 때 0점도 분석 가능
- SPSS 및 통계치 해석법 (SPSS One-way ANOVA)과 여러 변수의 상관성 (levene test for equality of variances)

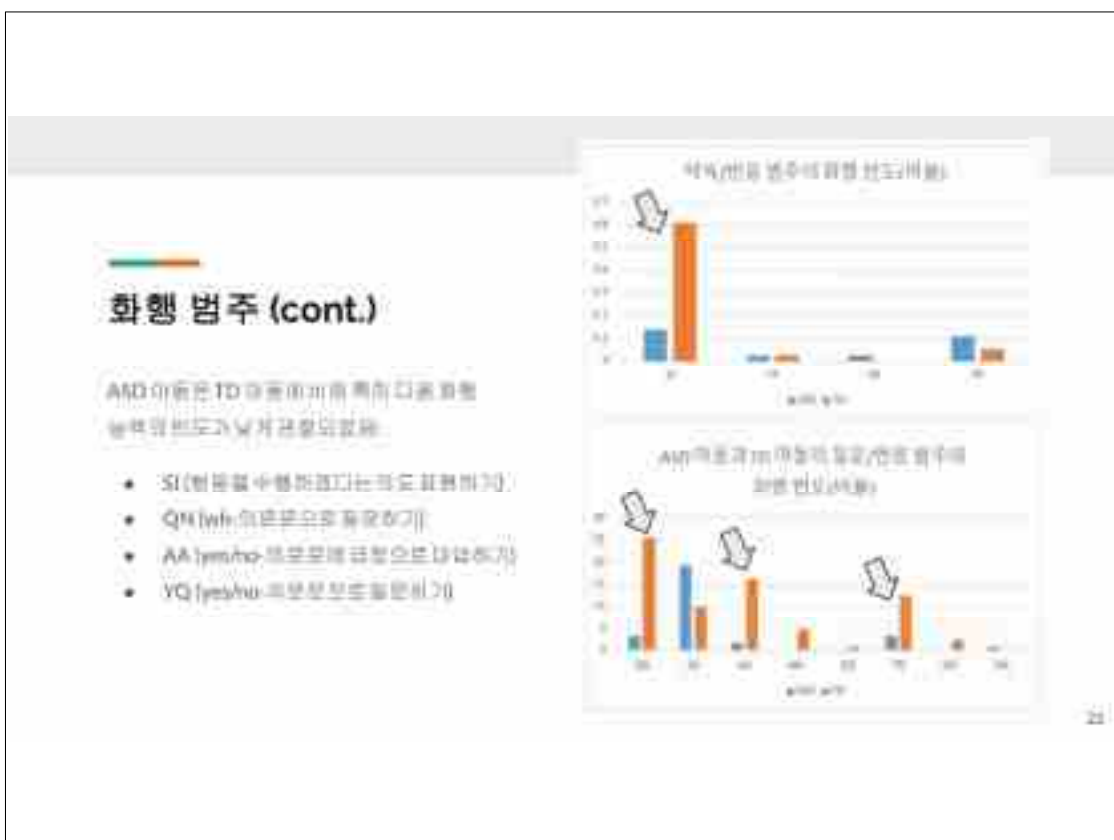
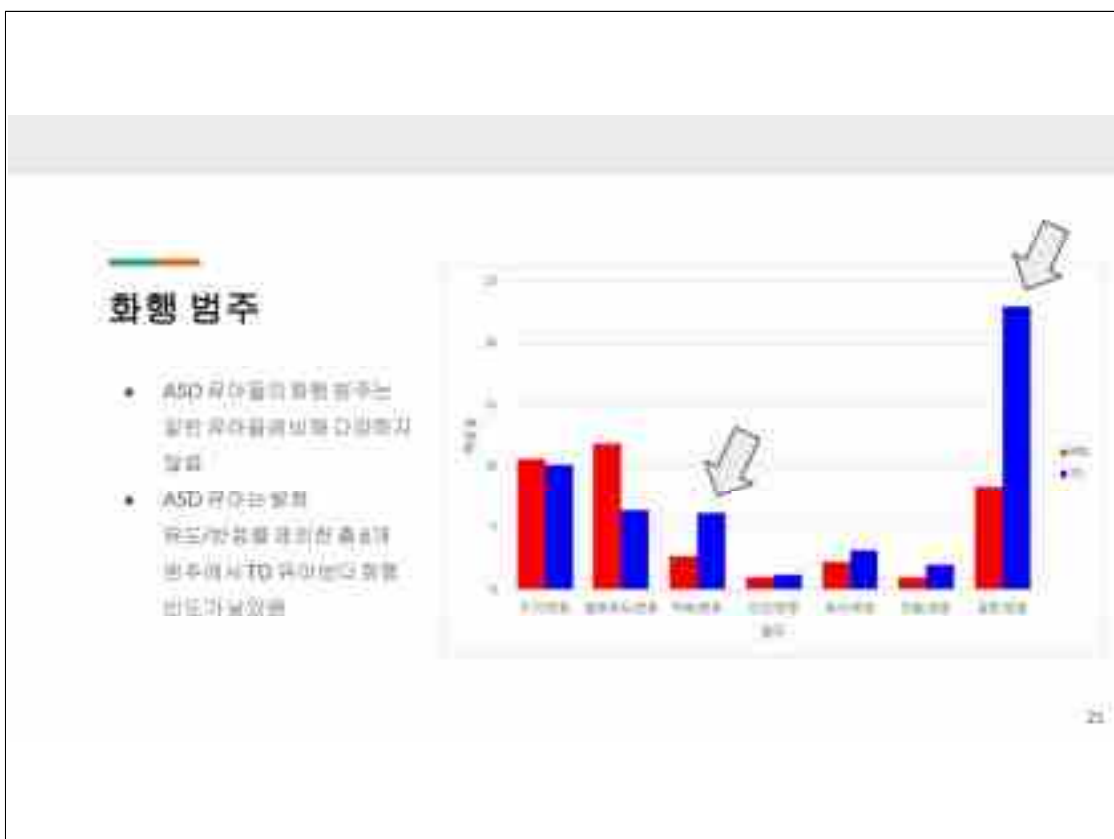
18

IV. 결과 및 논의

총 발화 빈도

- ASD 유아 > TD 유아
- 두 집단 발화 빈도의 차이는 통계적으로 유의미
- 과반(Lowrie)의 영보상 실험 결과
 $F=6.089, \eta^2=0.096, df=5, p=0.027$

ASD 유아		TD 유아	
아동	빈도수	아동	빈도수
H	28	Y	33
I	66	C	44
W	22	K	32
G	74	D	224
합계	210 (38.67%)		333 (#1,33%)



화행 범주 (cont.)

- ASD 유아들이 TD 유아들과 비교 높은 언어 발달을 보이는 경우
- CLIP에 있는 것만 보면 높은 언어 능력
- 하지만 사회적 상호작용 부족(ASD 특징)
- ASD라는 사회적 제한한 범주를 거부하지
- SA(언어만)나 그외의 사회적 능력 평가
- SA/WashU-21은 사회적 상호작용

23

개의 및 결론

ASD 유아는 TD 유아보다 높은 언어 수준이 발달함에 따라 사회적 능력 면에서는 부족한 모습을 보임

- 화행 범주가 일반 유아들보다 더 다양하지만
- 화행의 질도 높지만 높은 정도가 일반적으로 낮음
- 특정 화행 범주에서 특히 부족한 모습을 보임 → ASD 유아들의 화행 능력은 특정 화행 범주에 종류에 국한된 선역적(selective) 어휘 발달을 시사
- 학습(만성)과 화행(언어)은 두 범주도 화행 능력이 특히 부족함(ASD 유아 요인의 미묘한) 능력에 TD 유아들보다 저조하므로 발달이 더 어려움 가능

24

참고문헌

- 김희정(2012). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 19(1), 1-11.
- 김희정(2016). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 23(1), 1-11.
- 김희정(2017). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 24(1), 1-11.
- 김희정(2018). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 25(1), 1-11.
- 김희정(2019). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 26(1), 1-11.
- 김희정(2020). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 27(1), 1-11.
- 김희정(2021). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 28(1), 1-11.
- 김희정(2022). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 29(1), 1-11.
- 김희정(2023). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 30(1), 1-11.
- 김희정(2024). 자폐스펙트럼장애 유아의 언어적 의사소통 능력의 선별적 장애. *한국언어학회지*, 31(1), 1-11.

11

감사합니다

12

부록: 화행 코드 체계와 화행 빈도

화행 코드	영문 코드	의미	빈도	
			ADD 빈도	TS 빈도
7) 의사행위 Greetings and Responses	PP	Receive/Propose action for hearing (수용하다/제안할 요구/제정함)	13	28
	CA	Call attention to hearing by name or by substitute (eye/ear/body) 이름이나 다른 표현 사용 및 언급 주의를 끌기	14	3
	EE	Signal to start performing act (act do not run or not a fall) 행위를 수행하고 시작함/오류 신호	1	1
	AD	Agree to carry out act requested by other (응답/수용이 가능한 행위를 수행하는 것/응답함)	3	2
	RD	Refuse to carry out act requested by other (다른 사람이 요구한 행위를 거부함)	12	5
	DA	Disagree to carry out act proposed by other (다른 사람의 행위를 수행하지 않음)	2	4
	DO	Disagree to carry out act requested or proposed by other (다른 사람이 요구한 행위를 수행하지 않음)	0	2
	CS	Counter-suggestion, as indirect refusal (반박/거부/거절함)	0	2
	GR	Give reason justify a request to accept, refuse, prohibit, etc. 이유(들)	1	0

27

부록: 화행 코드 체계와 화행 빈도 (cont.)

화행 코드	영문 코드	의미	빈도	
			ADD 빈도	TS 빈도
8) 발음과도 관련 Speech Elicitation and Responses	E	Elicit revision of word or sentence by vocal comment (음성적 피드백/단어나 문장의 수정 유도함)	1	4
	EC	Elicit completion of word or sentence (어휘나 문장의 완성 유도함)	0	1
	EX	Express ex- surprise, or enthusiasm, express enthusiasm for the speaker's level performance (놀라움/의사/흥분/칭찬/수용/찬양/행복/칭찬/감동/칭찬)	4	2
	EA	Elicit rewording of utterance made by another, and do so (다른 사람이 한 말을 수정함)	14	5
	ES	Repeat the child's own utterance (자신의 말을 반복)	4	13
	EL	Self talk (자기 자신에게 하는 말)	0	0
	ET	Repeat/imitate other's utterance (다른 사람의 말을 흉내/모방함)	14	5

28

부록: 화행 코드 체계와 화행 빈도 (cont.)

화행 유형	화행 코드	의미	빈도	
			ADD 빈도	TD 빈도
(3) 의견/의견 Comments is and Responses	SC	State intent to carry out act by speaker. 화자가 행위를 수행하고자 하는 의도를 표명하기	8	22
	PC	Ask for permission to carry out act by speaker. 화자가 화자(상대방)의 동의를 위해 화자 권한을 묻기	1	1
	PR	Requester needs to perform an act and 화자가 화자(상대방)의 협력을 요청하기	1	4
	PP	Prohibitional intent to perform an act 화자가 화자의 행위를 금지하기	4	2
(4) 선언/선언 Declarative is and Responses	DC	Declare and create a new state of affairs by declaration 선언형으로 새로운 사실을 창조하기	2	1
	DP	Declare in certain state before verify by declaration 선언형으로 화자의 상태를 증명하기	3	1
	ND	Disagree to a declaration 선언에 반대하기	0	1
	YD	Agree to a declaration 선언에 동의하기	0	2

28

부록: 화행 코드 체계와 화행 빈도 (cont.)

화행 유형	화행 코드	의미	빈도	
			ADD 빈도	TD 빈도
(5) 표현/표현 Wishes and Responses	MM	Mark occurrence of event (i.e., mark, greet, apologize, congratulate, etc.) 화자 발제를 '표현'하는 것(예를 들어, 인사하기, 축하하기, 축하하기 등)	8	8
	EM	Express in distress, pain, disappointment, and frustration 슬픔, 고통, 낙담, 그리고 실망의 표현	3	0
	DN	Express positive emotion 긍정적 감정 표현하기	1	3
	XX	Exhibit attentiveness to listener 화자의 주의를 끌기	0	1
(6) 선언/선언 Statements and Responses	ST	State or make a declarative statement 선언으로 진술하기	42	56
	AS	Agree with the other's suggestion 제안에 동의함	1	1
	AP	Agree with proposition expressed by previous speaker 화자 제안에 동의 요청을 받아들임	0	2
	DW	Disagree with proposition expressed by previous speaker 화자 제안에 반대 요청을 받아들임	0	2

29

부록: 화행 코드 체계와 화행 빈도 (cont.)

화행 범주	화행 코드	정의	빈도	
			ADD 빈도	TD 빈도
	VE	Express a wish (소원을 표현하기)	1	0
	CV	Could 수 세기	2	0
(7) 질문과 답변 Questions and Responses	QI	Wh-question or ask a present question (wh-이문으로 묻기)	4	25
	SA	Answer a wh-question with a statement (Yes/No-이문으로 답하기)	25	15
	AA	Answer in its affirmative to yes/no question (Yes/No-이문으로 긍정으로 답하기)	2	21
	AN	Answer in its negative to yes/no question (Yes/No-이문으로 부정으로 답하기)	3	6
	EG	Elaborating question tag (확정적 화행과 같은 조건으로 질문하기)	0	1
	YQ	Ask a yes/no question (Yes/No-이문으로 묻기)	4	16

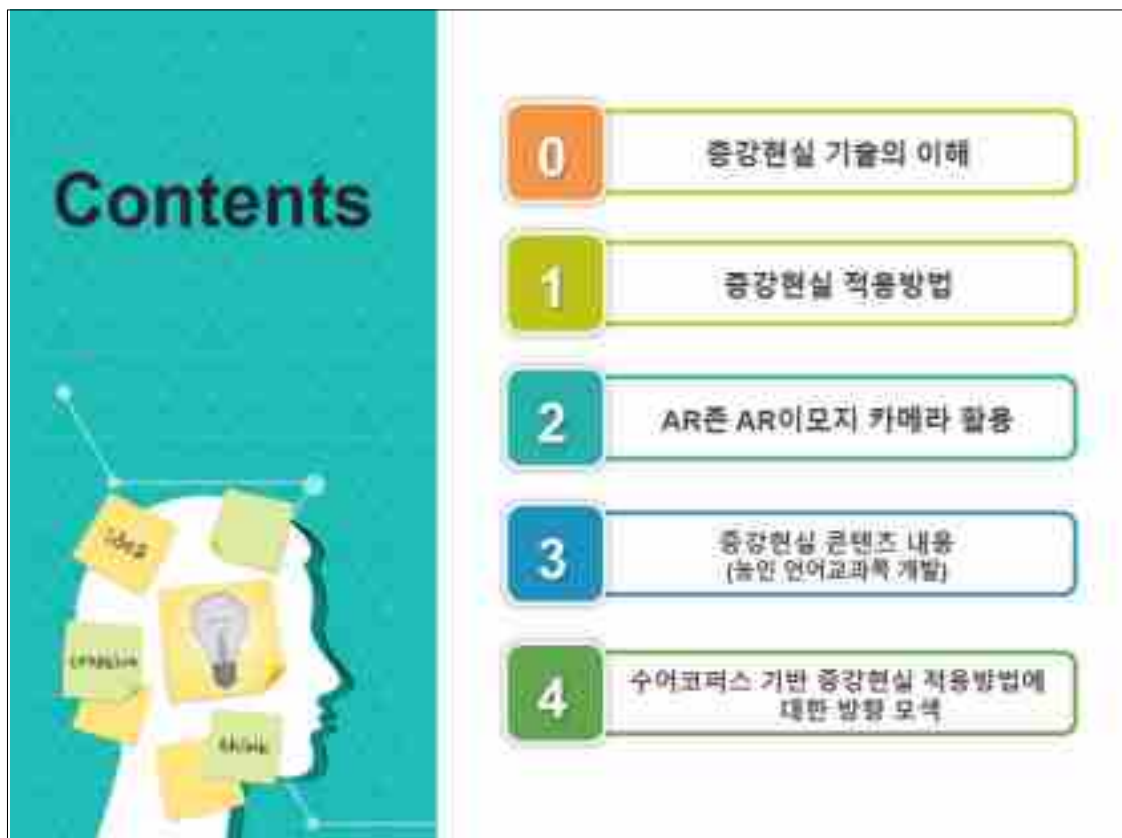
21

부록: 화행 코드 체계와 화행 빈도 (cont.)

화행 범주	화행 코드	정의	빈도	
			ADD 빈도	TD 빈도
	QA	Answer a question with a wh-question (wh-이문으로 질문을 답하기)	0	3
	NA	Non-satisfying answer to question (질문과 같은 화행으로 답하기)	1	0
(8) 평가 Evaluations	AE	Express positive evaluation of hearer's or speaker's acts (표준 화행에 대한 긍정적 평가를 표현하기)	1	0
	DE	Express negative evaluation of hearer's or speaker's acts (표준 화행에 대한 부정적 평가를 표현하기)	1	0
(9) 평가 Utterances	CO	Unrelated utterances or evaluations (일관되지 않는 화행 또는 평가)	1	0
	VV	Utterances without a clear function (화행의 기능이 불분명한 화행)	12	30
합계			210	335

(참고: 화행 코드는 Hays & Stein (1986)과 오카모토 (2007)의 코드를 바탕으로 하되, 수평 선이 들어간 화행 코드는 이 연구에서 추가된 것이다.)

22



0 증강현실 기술의 이해

1. 증강현실(Augmented Reality, AR) 기술

- **증강현실이란?**
 - 현실의 정보에 가상의 정보를 합성해 사물이나 이미지의 정보를 **증강**시키는 것
 - 현실 세계의 가상 세계라는 두 차원에 겹치는 위치에서 증강현실이 차이한다(Co-located, collocated, overlaid)
 - **"현실 세계 위에 3차원의 정보를 입혀"** 보여주는 기술
 - 증가된 Augmented **실제**현실 Reality
 - **현실과 가상의 조합 / 실시간 상호작용 / 3D에 기반**




출처: TED <http://www.ted.com> "Met Milk: Image recognition that triggers augmented reality" (박지민 발표)

2. 가상현실

- **가상현실(Virtual Reality, VR)이란?**
 - 존재하는 현실의 사물과 공간을 **가상**으로 재구성
 - 최초의 가상현실: 오펜하이머(Othmer) 기기(Othmer Helio, 1961) (optical type)
 - 최초의 가상현실 헤드셋: 데이브 데이(Dave Davies, 1961) (가상 / 헤드드 디스플레이(Head-mounted display))
 - **"현실 세계를 간단한 완벽한 디지털 환경"**을 구축하여 가상의 세계에서 활동하는 것






Virtual Reality & Augmented Reality

출처: 안호진(가리다)를 'VR'이라고 불러줘야 할까? (안호진 발표)

3. 혼합현실

- 혼합현실(Mixed Reality, MR)이란?
 - 가상의 형상을 현실에 겹쳐 더욱 발전된 가상세계를 구현하는 기술
 - 가상의 비의지적·구분된 공간에 현실 정보를 덧붙이거나, 현실 정보와는 무관하게 가상 이미지를 삽입
 - 「가상현실 + 증강현실」의 장점등 결합한 형태



www.BAND.CAM.com

혼합현실...가상 공간 + 실제 공간

<출처: 유튜브 채널 '2019 한국콘텐츠진흥원' 교육자료 영상첩>

3.1 증강현실 기술의 이해와 실례

혼합현실
Mixed Reality(MR)

현실세계
Real Environment

증강현실
Augmented Reality (AR)

증강가상
Augmented Virtuality (AV)

가상현실
Virtual Environment

현실과 가상세계의 연속체 Milgram, Kishino, 1994

현실 세계



증강현실 세계



가상현실 세계



실제 현실



증강현실



가상현실



← 실제 현실 증강현실 가상현실 →

<출처: 송신영 저, 『Augmented Reality: 현실과 가상의 연속체』, 2019년 10월 20일 교육자료 영상첩>



1 증강현실 적용 방법

1.1 증강현실(Augmented Reality, AR)

- 2015 개정 교육과정
 - 교육부(2018) : **일의용업별 인재 양립 목표**
 - 최고형사교육 향상을 위해 노력
 - 교육부(2018) : "증강현실 기술의 실질적 콘텐츠를 제공하는 '디지털 교과서 제작'"
 - 사회, 과학, 영어 교과목 디지털 교과서 보급
 - 초등학교 3~4학년, 중학교 1학년
 - 교육부(2019) : 실시간 대응태크 콘텐츠 제공하기 **교실현장 구축**
 - 학교교육혁신사업
 - 미래교육의 지원 방안(www.koel.go.kr)

Three images illustrating AR applications: 1. A tablet displaying a globe with a graduation cap and a pencil. 2. A news anchor on a screen with a globe background, with a caption '모뎀 GO' 증강현실, 교과서에 적용. 3. A classroom scene with students at desks, some wearing AR headsets.

출처: 교육과 기술의 융합, 최유리(2019) [E-learning] / YTN 모바일에서 인용함

1.1.1 증강현실로 구현할 수 있는 관련 내용

① 동물관찰(사이언스레벨업 실감형 콘텐츠, 한국과학창의재단)_App : 31종류의 동물카드 (2020)

<출처> AR 콘텐츠 개발 콘텐츠 개발하기 동영상 <<https://www.youtube.com/watch?v=4346CA3P>>

② 증강현실(마커와 동사카드) (부산대 해재간용범서양언, 한국연구재단)_App : 68종류의 동사카드 (2015-2020)

2.2.1 AR존 이모지 만들기

절차: AR존 -> AR이모지 카메라 -> 시퀀스촬영 -> 스타필링액 -> 이모지 서타카생성
 -> 이모지로드 완료



The image shows a sequence of four screenshots from a mobile application. The first shows a grid of AR zone icons. The second shows a selection of AR zone characters with handwritten labels. The third shows a grid of AR zone characters with handwritten labels. The fourth shows a grid of AR zone characters with handwritten labels.

3 증강현실 콘텐츠 내용

3.1 농인 언어교과목 교재 개발 절차



The flowchart illustrates the development process for AR content. It starts with '비이티구독' (Bite) and '교재 검토가이드' (Textbook Review Guide). The process involves '동영상 수집과 전사' (Video Collection and Transcription) and '코퍼스 구축' (Corpus Construction). The final output is 'AR이모지' (AR Emoji) and '증강현실' (AR).

<출처> 2021 울산대학교 언어생활연구소 학술세미나 및 중등어학연구포럼사업 (4차년도) -> 2020년도 수어교과부서 부처특성 융합연구 지원금 인용됨.

2020년도 선정 「언어사회연구소지원사업」

3.2 참여연구원

구분	연구자				역할
	성명	소속	직위	전공	
연구책임자	최영주	조선대학교	교수	언어학	한국수어자료 수집 및 분석, 발동치 구축 및 교재 개발에 관련된 연구 총괄
일반공동연구원	장세은	한국해양대학교	교수	코퍼스언어학	수어자료 발동치 구축 및 한국수어 연구
	윤병천	한국나사렛대학교	교수	특수교육	수어자료 공유 및 한국수어 교재 개발 분과총괄
	조창훈	서남대학교 (중국중경)	부교수	언어학	영어교과과정 분석 영어교재 개발 분과 총괄
	이재기	조선대학교	교수	국어교육	국어교과과정 분석 및 국어교재 개발 분과 총괄
	박형준	조선대학교	교수	산업공학 (증강현실)	증강현실 교과목 설계제작
	권순화	부산대학교	교수	언어치료학	증강현실 교과목 설계제작

3.3 능인 언어교과목 개발

(1) 3차원 영상 정보를 보여줌으로써 청각장애학생의 수어(sign language)교과목에 효율적 적용

- 발동치(corpus) 구축 및 수어 문장과 개별 어휘/다어휘(관용표현)로 작성
 - 발동치 분석 후 학년에 맞는 문장으로 분석한 후 문장으로 구현하는 방법
 - 단어의 단음과 화문성(대학 4-5 어휘)
해당하는 화교에서 공부할 한다.
- 초등학생(1, 2, ... 6학년) : 계획서 상 한 단원에 1문장의 적용(초등학교: **샘플 10개**를 제작)
 - 1학년 : 3학년 1학기(총 10개 문장)
 - 국어(5개 단원) 4개 권(총 5개 문장)
 - 계획방법(문단: **문단과 교과**를 제시하고 용어: **문단**을 제시, **문단**을 제시, **문단**을 제시)
 - 영어(5개 단원) 5개 권(총 5개 문장) : (1차로는 10개 제작)

기 간 (추진년월)	증강현실 방법과 모재의 구현 방법으로 중4까지를 모재함	비 고
1차년도	1. 증강현실로 제시 2. 글자모 제시 3. 음절으로 제시 4. 동인의 수어 발음으로 제시	

4 수어코퍼스 기반 증강현실 적용방법 모색

4.1 코퍼스 전사 데이터 수집 후 빈도수를 고려한 문장 선정

내용 및 방법

- 코퍼스 빅 데이터를 바탕으로 초등학생의 각 단원별 문장 선정
 - 초등학교(학년별) 코퍼스 전사 내용과 기준으로 문장 선정: 10개
 - 2학년 4학기
 - 국어 문장(5개)
 - ①번, ②번, ③번, ④번, ⑤번
 - 영어 문장(5개)
 - ①번, ②번, ③번, ④번, ⑤번

코퍼스 전사 데이터 수집 후 선정

교과목	학년	학기	단원	문장
국어	2학년	4학기	1단원	①번
영어	2학년	4학기	1단원	①번

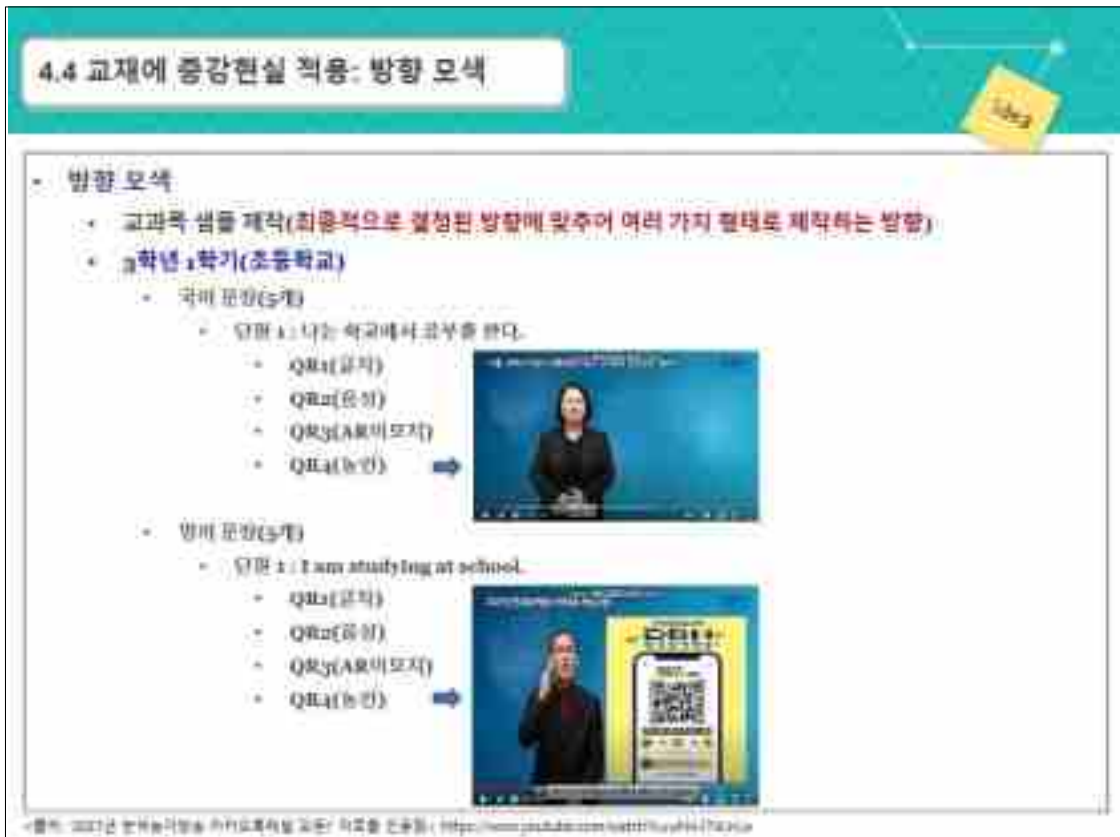
4학년도 전사 문장 현황

* 출처: 2021 초등학교 언어용집현수교과본(국어, 영어) 및 중학교 언어용집(국어, 영어) 1학년부, 2학년부, 3학년부 국어, 영어 교과서 내용

4.2 초등학교 교재에 증강현실 적용(국어, 영어문장)

증강현실 방법적 모색(문자(x), 음성(y), 아바타 or 3AR이모지(z))

- 2학년 4학기(초등학교)
 - 국어 문장(5개)
 - 단문 x: 나는 학교에서 공부를 한다.
 - QR1(문자) → 나는 학교에서 공부를 한다.
 - QR2(음성) →
 - QR3(AR이모지) →
 - QR4(노션) →
 - 단문 y, z, 4, 5
 - 영어 문장(5개)
 - 단문 x: I am studying at school.
 - QR1(문자) → I am studying at school.
 - QR2(음성) →
 - QR3(AR이모지) →
 - QR4(노션) →
 - 단문 y, z, 4, 5



4.4.1 교재에 증강현실 적용

교과서 형식(발달)

- 3학년 1학기(초등학교)
 - 국어 문장(단원 1)
 - 단원 1: 나는 학교에서 공부해 본다.
 - 단원 목표: 학습-몰입감, 집중도, 흥미, 주의집중, 학습태도에 긍정적 영향을 준다.
 - 콘텐츠 유형: 초등형
 - 콘텐츠 내용: 우리 기쁜 곳으로 카메라 화면에 QR코드를 찍으면 자신의 미래의 글자, 음성, 손짓, 사람이 나타난다.
 - QR1 (AR이모지)
 - QR2 (글자)
 - QR3 (음성)
 - QR4 (손짓)

4.4.2 교재에 증강현실 적용(국어, 영어)

3학년 1학기(초등학교)

- 국어 문장(단원 1)
 - 단원 1: 나는 학교에서 공부해 본다.
- 3학년 1학기(초등학교)
 - 영어 문장(단원 1)
 - 단원 1: I am studying at school.

