

2021년 겨울학술대회

K^A_LS 한국언어과학회

빅 데이터와 언어학: 현재와 미래

Big Data and Linguistics: Present and Future

2021년 2월 18(목) **일시**

온라인(Zoom) **장소**

주관

부산대학교 언어정보학과

주최

K^A_LS 한국언어과학회

[한국언어과학회 2021 겨울학술대회] Zoom 접속 주소

①분과(Zoom 1): 개회식, 어휘/의미 발표, 초청특강, 연구윤리교육 및 총회

<https://pusan.zoom.us/j/83772894412?pwd=UnozVzILYzEzRDF1TTdoc0laWHZxQT09>

회의 ID: 837 7289 4412

암호: 123456

②분과(Zoom 2): 통사/화용

<https://pusan.zoom.us/j/85949609966?pwd=d01XVFNrYjI3aHdOQ1pRYnV3VUVYdz09>

회의 ID: 859 4960 9966

암호: 123456

③분과(Zoom 3): 코퍼스/담화/텍스트

<https://pusan.zoom.us/j/9709769748?pwd=TU5ucGxKcFBVVkpWYkpSOXk5Sm5HUT09>

회의 ID: 970 976 9748

암호: 123456

④분과(Zoom 4): 언어 교육

<https://pusan.zoom.us/j/82875282876?pwd=WkVSSWxaRTJOTUU2Z2tRVGhQTyt1Zz09>

회의 ID: 828 7528 2876

암호: 123456

한국언어과학회 2021 겨울학술대회

“빅 데이터와 언어학: 현재와 미래”

일시: 2021년 2월 18일(목)

장소: 온라인(Zoom)

주관: 부산대학교 언어정보학과

주최: 한국언어과학회



12:50 ~ 13:00		① 개회식 회장 권연진(부산대)			
시간	분류	①어휘/의미 좌장: 박기성(부산대)	②통사/화용 좌장: 권기양(영산대)	③코퍼스/담화/텍스트 좌장: 김동국(영산대)	④언어 교육 좌장: 우길주(부산교대)
사회		이원빈(대전대)	홍신철(부산외대)	권미분(신라대)	전지현(부산대)
13:00 ~ 13:30	제목	어휘유형론, 공어휘화, 의미의 기본요소에 관한 소고	Notes on Freezing Effect and Anti-Freezing Effect	키워드 네트워크 분석을 통한 글로벌 해양 저널의 주요 주제와 경향 분석	The Underuse of English Progressive Aspect of Korean ESL learners: A Collostruational Analysis
	발표	김은일(부경대)	권기양(영산대)	장세은(한국해양대)	노강산 & 송상현 (고려대)
	토론	박기성(부산대)	황규홍(동아대)	이성화(한국해양대)	김양희(안동대)
13:30 ~ 14:00	제목	형용사 hard의 의미범주와 한국어 대응표현	Labelability of Elliptical Properties: Evidence from Fragmentary Questions in Korean	한국어 영화 대사와 영어 자막 번역에 나타난 감정 분석 비교: 건축학개론을 중심으로	A Corpus-based Study on Chinese EFL Learners' Use of Amplifier Collocations
	발표	정희란(부경대)	전혜원(부경대)	이성화 & 박호민 (한국해양대)	Li Shaojie(한국해양대)
	토론	김은일(부경대)	정현경(신라대)	하명호(신라대)	장세은(한국해양대)
14:00 ~ 14:30	제목	현대 일본어와 한국어의 동사 조건형의 어휘화에 관한 대조 연구 -시각활동을 나타내는 동사 「みる」와 「보다」의 경우-	Assessing the Linguistic Ability of BERT on Co-Indexation of English Reflexives	Exploring Lexical Bundles in Rhetorical Move of Research Article Abstracts in Applied Linguistics	디지털 미디어를 통한 한국인 영어 학습자들의 교실 밖 영어 학습 현상과 정의적 요인과 관계에 관한 연구
	발표	하재필(부산대)	김현수 & 송상현 (고려대)	Liang Xiaolei (한국해양대)	이향미(부산대)
	토론	이소영(부산대)	신유선(부경대)	이성화(한국해양대)	정영경(부경대)
사회		황규홍(동아대)	김수태(신라대)	김대익(영산대)	남정미(신라대)
14:30 ~ 15:00	제목	한국어와 일본어의 한자음 음절 대응 관계	CNPC Effects of Why-Stripping in Korean	영어참고서에 나타난 번역 양상 연구-중학교 학년별 중심으로	실무용 한국어 교재 개발 방안 연구
	발표	허지애(부산대)	이두원(한국교통대)	신순란(부산대)	원해영(부산외대)
	토론	서민정(부산대)	이성용(고신대)	우길주(부산교대)	이은진(부산대)
15:00 ~ 15:30	제목	몽골어와 한국어 사이의 공통어휘로 보이는 명사들	The Split-VP Hypothesis Revisited	A Corpus-based Study of GET-Adjective Patterns Used by Chinese English Majors	Seeking for the potential of using Social Network Analysis via NVivo R 1
	발표	박재익(고신대)	장경철(부산대)	Wang Meihan (한국해양대)	박종원(부경대)
	토론	강은지(부산대)	심재영(부산교대)	하명호(신라대)	전지현(부산대)
15:30 ~ 16:00	제목	A New Approach to Keyword Analysis: focusing on Old and New Testament	Two types of node-sprouting	온라인 영화 댓글의 특성: 영화<기생충>의 한국어 댓글과 영어 댓글 비교	비대면 온라인 영어수업에 대한 영어과 학생 교사들의 인식조사
	발표	하명호(신라대)	최재훈(대구대)	백다운(부산대)	김양희(안동대), 신유선 & 정영경(부경대)
	토론	장세은(한국해양대)	장경철(부산대)	김소영(동명대)	황희정(중원대)

16:00 ~ 17:00	초청 특강	① 정덕길 (동의대): 인공지능 시대에 언어학의 발전 방안 사회: 김동국(영산대), 토론: 장세은(한국해양대)
17:00 ~ 17:30		① 연구 윤 리 교 육 진행: 편집위원장
17:30 ~ 18:00		① 학 술 상 시상식 ① 입 시 총 회 진행: 총무이사

[한국언어과학회 2021 겨울학술대회] Zoom 접속 주소

①분과(Zoom 1): 개회식, 어휘/의미 발표, 초청특강, 연구윤리교육 및 총회

<https://pusan.zoom.us/j/83772894412?pwd=UnozVzILYzEzRDF1TTdoc0laWHZxQT09>

회의 ID: 837 7289 4412

암호: 123456

②분과(Zoom 2): 통사/화용

<https://pusan.zoom.us/j/85949609966?pwd=d01XVFNrYjI3aHd0Q1pRYnV3VUVYdz09>

회의 ID: 859 4960 9966

암호: 123456

③분과(Zoom 3): 코퍼스/담화/텍스트

<https://pusan.zoom.us/j/9709769748?pwd=TU5ucGxKcFBVWkpWYkpSOXk5Sm5HUT09>

회의 ID: 970 976 9748

암호: 123456

④분과(Zoom 4): 언어 교육

<https://pusan.zoom.us/j/82875282876?pwd=WkVSSWxaRTJOTUU2Z2tRVGhQTyt1Zz09>

회의 ID: 828 7528 2876

암호: 123456

Contents

Zoom 1

- 어휘유형론, 공어휘화, 의미의 기본요소에 관한 소고 3
/ 김은일
- 형용사 hard의 의미범주와 한국어 대응표현 9
/ 정희란
- 현대 일본어와 한국어의 동사 조건형의 어휘화에 관한 대조 연구 17
-시각활동을 나타내는 동사 「みる」와 「보다」의 경우-
/ 하재필
- 한국어와 일본어의 한자음 음절 대응 관계 29
/ 허지애
- 몽골어와 한국어 사이의 공통어휘로 보이는 명사들 37
/ 박재익
- A New Approach to Keyword Analysis: focusing on Old and New Testament 49
/ 하명호

Zoom 2

- Notes on Freezing Effect and Anti-Freezing Effect 63
/ 권기양
- Labelability of Elliptical Properties: Evidence from Fragmentary Questions in Korean 71
/ 전혜원
- Assessing the Linguistic Ability of BERT on Co-Indexation of English Reflexives 81
/ 김현수, 송상현
- CNPC Effects of Why-Stripping in Korean 91
/ 이두원
- The Split-VP Hypothesis Revisited 107
/ 장경철
- Two types of node-sprouting 109
/ 최재훈

Zoom 3

- 키워드 네트워크 분석을 통한 글로벌 해양 저널의 주요 주제와 경향 분석 115
/ 장세은
- 한국어 영화 대사와 영어 자막 번역에 나타난 감정 분석 비교: 건축학개론을 중심으로 133
/ 이성화, 박호민
- Exploring Lexical Bundles in Rhetorical Move of Research Article Abstracts in Applied Linguistics 151
/ Liang Xiaolei
- 영어참고서에 나타난 번역 양상 연구-중학교 학년별 중심으로 161
/ 신순란
- A Corpus-based Study of GET-Adjective Patterns Used by Chinese English Majors 189
/ Wang Meihan
- 온라인 영화 댓글의 특성: 영화<기생충>의 한국어 댓글과 영어 댓글 비교 203
/ 백다운

Zoom 4

- The Underuse of English Progressive Aspect of Korean ESL learners: 211
A Collostructional Analysis
/ 노강산 & 송상현
- A Corpus-based Study on Chinese EFL Learners'Use of Amplifier Collocations 219
/ Li Shaojie
- 디지털 미디어를 통한 한국인 영어 학습자들의 교실 밖 영어 학습 현상과 정의적 요인과 관계에 관한 연구 233
/ 이향미
- 실무용 한국어 교재 개발 방안 연구 245
/ 원해영
- Seeking for the potential of using Social Network Analysis via NVivo R 1 259
/ 박종원
- 비대면 온라인 영어수업에 대한 영어과 학생 교사들의 인식조사 279
/ 김양희, 신유선, 정영경

[초청특강] Zoom 1

- 인공지능 시대에 언어학의 발전 방안 289
/ 정덕길

Zoom 1

개회식, 어휘/의미 발표, 초청특강, 연구윤리교육 및 총회

<https://pusan.zoom.us/j/83772894412?pwd=UnozVzILYzEzRDF1TTdoc0laWHZxQT09>

회의 ID: 837 7289 4412

암호: 123456

어휘유형론, 공어휘화, 의미의 기본요소에 관한 소고

김은일
(부경대학교)

1. 서론

- 다른 언어들의 유의어 사이의 유사점과 차이점

- (1)a. He caught the butterfly.
b. 그는 나비를 잡았다.
- (2)a. He caught the bouncing ball.
b. 그는 튀어 오르는 공을 잡았다.
- (3)a. He caught a taxi.
b. 그는 택시를 잡았다.
- (4)a. He caught the train/bus/flight.
b. *그는 기차/버스/비행기를 잡았다.
- (5)a. He caught a cold.
b. 그는 감기를 *잡았다/감기에 걸렸다.
- (6)a. 우리는 결혼날짜를 잡았다.
b. We *caught/fixed the date for the wedding.
- (7)a. 나는 그가 화장실에서 담배 피우는 것을 잡았다.
b. I caught him smoking in the bathroom.
- (8)a. 정부는 투기 열기를 잡았다.
b. You can still catch the fever of animation.

2. 어휘유형론(Lexical Typology): François(2008)

- 목표:
다양한 언어에서 어휘의 의미장(semantic field)이 어떻게 구성되는지를 다룬다.

• 이론적 근거:

언어마다 다른 방식으로 의미영역을 어휘화하지만 주어진 환경 속에서 유추(analogy)할 수 있는 인간의 능력 때문에 다양한 언어의 어휘에서 의미적 전이(semantic shift)가 비교적 제한된 범위에 한정되고 보편성(universals)과 유형론적 경향성(typological tendency)을 나타낸다. cf. Haspelmath (2003)

- 연구절차: ① 의의(sense) 구분하기; ② 의의들 간 연결하기, ③ 어휘의미지도 그리기

A. 의의구분하기

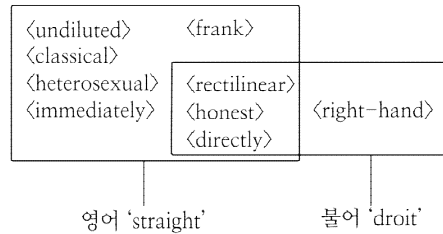
- 의의구분이 중요한 이유?: 영어 'straight' vs. 불어 'droit'
 - 단어의 수준에서 두 단어를 비교하면 그 차이점을 발견하기 힘들

(9) 영어 형용사 'straight' 의 의의

<rectilinear> (a straight line)	<heterosexual> (gay or straight)
<frank> (straight talking)	<undiluted> (straight whisky)
<honest> (a straight guy)	<directly> (straight to the point)
<classical> (a straight play)	<immediately> (straight away)

(10) 불어 형용사 'droit' 의 의의

<rectilinear> (un trait droit)	<directly> (aller droit au but)
<honest> (un type droit)	<right-hand> (le côté droit)



<그림 1> 영어 'straight' 와 불어 'droit' 의 의의 겹침

• 의의(sense)란?

- 의의 자체를 밝히는 것이 목적이 아니기 느슨한 정의 사용
- Fillmore and Akins(2000)의 용법(uses)
- François(2008)의 단어보다 더 작은 의미단위 cf. 단어(분자) / 의의(원자)
- 다양한 사전에서 제시한 의의들이나 코퍼스에서 발견된 예를 활용

• 의의구분 방법: 김은일(2016) 'open' vs. '열다'

- 연구자의 직관력(intuition)
- 'He opened the tumbler/beer bottle.' cf. 액체를 담은 용기(container)

-실증적 관찰(empirical observation): 공어휘화(colexification)
 ‘그는 텀블러/*맥주병을 열었다.’ cf. 맥주병을 따다 (일회성)

☞ 공어휘화(colexification):

- 기능적으로 구분되는 의미들을 동일한 어휘형태로 연관지을 수 있을 때 공어휘화한다고 한다. cf. <그림 1>
- 타동사의 경우는 목적어의 특성과 밀접한 관련; 목적어 언어와 관련된 공어휘화 ☞

B. 의미들 간 연결하기

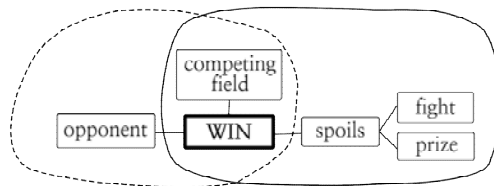
- 구분된 의미들을 의미적으로 가까운 의미들끼리 그룹으로 묶고,
- 가까운 순서대로 연결선으로 연결하여 다의망(polysemous network)을 구성

- 의미구분 방법: 김은일(2017) ‘win’ vs. ‘이기다’
- 연구자의 직관력(intuition)
- 실증적 관찰(empirical observation): 공어휘화(colexification)

- (11) a. He won the right/freedom/medal/prize money. <spoils(전리품)>
 b. 그는 권리/자유를 *이겼다/*땀다/얻었다. <right>
 c. 그는 메달/상금을 *이겼다/*얻었다/땀다. <prize>

C. 어휘의미지도(lexical semantic map) 그리기: 김은일(2017) ‘win’ vs. ‘이기다’

- 언어외적 기준연결망(etic grid) 과
- 언어내적 범주화(emic categorization)로 구성



<그림 2> ‘WIN’의 어휘의미지도

3. 의미의 기본요소(basic component) ☞

-어휘의 기본요소가 그 어휘 의미의 일부가 된다. Leech(1990)
 eg.) needle의 기본요소: ‘thin, sharp, steel, instrument’
 ‘needle leaf’ : needle=thin

- ‘잡다’의 원형의미와 중심의의

<seize with hands>: ‘손으로 움키고 놓지 않다’

-<make easy use of>: 손에 움켜진 것은 용이하게 활용할 수 있음

-<fix>: 움직이는 대상을 손에 움켜지면 그 대상을 고정할 수 있음

-<control>: 특정한 속성을 가진 것을 손으로 움켜지면 그 속성을 통제할 수 있음

- ‘catch’ 의 원형의미와 중심의의

<chase and capture>: ‘쫓아가서 잡다’ cf. 도구를 이용

-<reach in time>: 움직이는 대상을 쫓아가서 잡으려면 시간에 맞춰 대상이 있는 곳에 도달해야 함

-<capture quickly>: 움직이는 대상을 잡기 위해서는 재빨리 잡아야 함

4. ‘잡다’ vs. ‘catch’

- 한 <seize with hands> ∩ 영 <chase and capture>

(1)a. He caught the butterfly.

b. 그는 나비를 잡았다.

(2)a. He caught the bouncing ball.

b. 그는 튀어 오르는 공을 잡았다.

- 한 <make easy/exclusive use of> vs. 영 <reach in time>

(3)a. He caught a taxi.

b. 그는 택시를 잡았다.

(4)a. He caught the train/bus/flight.

b. *그는 기차/버스/비행기를 잡았다.

cf. 관광버스를 한 대 잡았다.

- 영 <reach in time>: <meet> vs. 한 ∅

(5)a. He caught a cold.

b. 그는 감기를 *잡았다/감기에 걸렸다.

cf. The cold caught him. <be infected; reflect; push along; become alight>

cf. <control>

- 한 <fix> vs. 영 ∅

(6)a. 우리는 결혼날짜를 잡았다.

b. We *caught/fixed the date for the wedding.

- 한 <seize with hands> vs. 영 <capture quickly>

(7)a. 나는 그가 화장실에서 담배 피우는 것을 잡았다.

b. I caught him smoking in the bathroom.

cf. He caught a nap.

- 한 <control> vs. 영 <reach in time>: <overtake>
 - (8)a. 정부는 투기 열기를 잡았다.
 - b. You can still catch the fever of animation.
 - cf. I was caught in the rain.

5. 결론

참고문헌

- 「국립국어원 표준국어대사전」(The National Institute of the Korean Language's Korean Standard Unabridged Dictionary). <http://stdweb2.korean.go.kr/main.jsp>
- 김은일(Kim, E.). 2016. 다의어 동사 ‘open’ 과 ‘열다’ 에 대한 어휘유형론적 접근 (A Lexical Typological Approach to the English Verb ‘open’ and the Korean verb ‘yeolda’). 「현대문법연구」 (Studies in Modern Grammar) 91, 167–187. <http://dx.doi.org/10.14342/smog.2016.91.167>
- 김은일(Kim, E.). 2017. He Won his Wife in a Card Game: WIN의 의미지도 (He Won his Wife in a Card Game: Semantic Maps of WIN). 「새한영어영문학」 (The New Korean Journal of English Language and Literature) 59.4, 173–196. DOI: 10.25151/nkje.2017.59.4.010
- 이기동(Lee, K.). 2015. 「인지문법에서 본 영어동사사전」 (English Verb Dictionary: A Cognitive Grammar View). 서울: 한국문화사 (Seoul: Hankuk Publishing Company).
- Fillmore, C. J. and B. T. S. Akins. 2000. Describing Polysemy: The Case of ‘Crawl’ . In Ravin, Y. and C. Leacock (eds.), *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press. 91–110.
- François, A. 2008. Semantic Maps and the Typology of Colexification. In Vanhove, M. (ed.), *From Polysemy to Semantic Change*. Amsterdam: John Benjamins Publishing Company. 163–215.
- Haspelmath, M. 2003. The Geometry of Grammatical Meaning: Semantic Maps and Cross-linguistic Comparison. In Tomasello, M. (ed.), *The New Psychology of Language* (Vol. II). Mahwah, NJ: Lawrence Erlbaum Associates. 211–243.
- Leech, G. N. 1990. *Semantics: The Study of Meaning* (Second Edition). Harmondsworth, Middlesex: Penguin Books.

형용사 hard의 의미범주와 한국어 대응표현

정희란

(부경대학교)

1. 서론

영어에는 많은 다의어가 존재하고, 단어들의 다의성에 대한 연구는 주로 기본 의미에서 의미가 어떻게 확장되는가하는 과정에 집중되어 있다. 하지만, 학습자들이 영어 어휘를 학습할 때, 우선은 영어 동사에 대한 모국어의 의미를 먼저 확인하게 된다. 이 때, 대응되는 표현이 한 두 개 정도면 상관없지만, 번역되는 표현이 많아지면, 학습자들의 어휘학습에 혼란을 겪을 수 있다. 예를 들어, 영어 형용사 hard의 경우는 한국어의 의미를 한국인 학습자들은 거의 ‘단단한’이나 ‘어려운’의 의미 정도를 알고 있다. 하지만 영한사전¹⁾에 나타난 hard의 한국어 의미를 살펴보면 다음과 같다.

(동아출판 프라임 영한사전)²⁾

① 굳은, 단단한, 견고한 ② 딱딱한 표지의 ③ 곤란한, 이해하기 어려운, 노력을 요하는, 하기힘든, 힘겨운 ④ 시대 등이 건디기 어려운, 쓰라린, 괴로운 ⑤ 운명 등이 가혹한, 힘든 ⑥ 날씨가 거친, 험악한, 사나운 ⑦ 일 등에 열심인, 열중한, 부지런한 ⑧ 강렬한, 언동·감정 등이, 신랄한, 적의가 있는 ⑨ 기질·성격·행동 등이 격렬한, 엄한; 매물찬, 양갈진, 쌀쌀맞은, 냉혹한, 난폭한 ⑩ 관찰 등이 날카로운, 면밀한 ⑪ 사실·증거 등이 엄연한, 확실한, 신뢰성이 있는, 구체성이 있는 ⑫ 빗갈·윤곽 등이 너무 짙은, 너무 강한, 명암 대조가 확실한 ⑬ 소리 등이 불쾌한, 딱딱한, [음성] 경음(硬音)의 ⑭ 사고방식 등이 냉정한, 현실적인 ⑮ 매매 등이 조건이 까다로운, 양보하지 않는 ⑯ 시가(時價) 등이 강세의 ⑰ 상품 등이 내구성이 강한, 오래가는 ⑱ 음료가 알코올 성분이 많은, 술이 독한 ⑲ 묘사가 노골적인

사전에서 hard는 아주 다양한 한국어로 설명되어 있다. 하지만, 학습자들은 hard를 ‘단단한’이나 ‘어려운’ 등의 의미로만 사용하는 경향이 있어, 다의어에 대한 효율적인 어휘 학습이 이루어지지 않는다. 또한, 위에 예시로든 사전의 의미 기술을 살펴보면, 비슷한 한국어 표현이 중복되어 나타나 hard의 중심 의미를 이해하는 데 어려움을 줄 수 있다. 어휘 학습의 경제성을 위해서 단순한 의미의 나열이 아니라 동일한 의미를 하나로 묶어 범주화 시켜 정리하면 학습자들의 어휘 학습에 상당한 도움을 줄 수 있을 것이다. 어휘의 의미 범주와 함께 한국어로 어떻게 대응되는지 살펴볼 필요성이 있다. 영어 형용사 hard는 한국어로 대응될 경우는 ‘단단한’이나 ‘어려운’이라는 표현 이외에도 많은 표현으로 대응되어 나타난다. 예문을 통해 다른 대응 표현을 살펴보자.

- 1) 본 연구에서는 네이버 영한사전(<https://en.dict.naver.com>)을 참고하였으며, 네이버 영한사전은 옥스퍼드 영한사전, 동아출판 프라임 영한사전, YBM 올인올, 교학사 및 슈프림 등 모두 다섯 개의 영한사전으로 구성되어 있다.
- 2) 사전에서는 형용사 hard의 의미를 26개로 정의하고 있지만, 본 연구에서는 설명에 필요한 부분만을 정리하여 표시하였다.

- (1) a. hard wood
b. 단단한 목재(나무)
- (2) a. a hard fact
b. 확실한 사실
- (3) a. a hard stare
b. 냉혹한 시선

예문 (1)처럼, 형용사 ‘hard’의 한국어 대응으로 자연스럽게 사용할 수 있는 대응 표현은 ‘단단한’이지만, hard가 수식하는 대상에 따라, 예문 (2)와 (3)처럼 ‘확실한’이나, ‘냉혹’ 등의 표현으로 대응될 수 있다. (2)에서 ‘사실’이라는 명사와 연어를 이룰 경우는 ‘확실한’으로 표현되는 것이 자연스럽고, 바라보는 ‘시선’이 우호적이 않다면 ‘냉혹한’이라는 한국어 표현으로 대응되는 것이 적절하다. 이와 같이 영어에서는 하나의 형용사 hard로 나타나지만 한국어에서는 다른 한국어 표현으로 사용되는 형용사 hard를 가지고, 명사 연어와 결합할 때, 어떠한 한국어로 표현되는지를 알아보고자 한다.

본 연구는 다의성을 가진 형용사 hard의 의미 범주를 분류하고, 범주에 따라 명사 연어와 결합할 때 한국어에서는 어떠한 대응어로 나타나는지를 살펴보고자 한다. 먼저, 범주를 정하기 위해서 코퍼스에서 형용사와 연어를 이루는 명사를 추출하고, 영한사전을 참고하고 한국어 대응 예문을 분석하여 유사한 의미끼리 묶어 범주를 나눈다. 범주 분류에 활용한 자료는 영한사전과 코퍼스 자료에서 형용사와 명사 연어를 이루는 예문을 뽑아 분석하였다.

2. 이론적 배경 및 연구 방법

2.1. 연어와 코퍼스

연어(collocation)의 사전³⁾적 정의는 “두 개 이상의 단어가 결합하여 의미적으로 하나의 단위를 이루는 말”이다. 기본적으로 연어는 아주 긴밀하게 결합되어, 오랫동안 함께 나타나는 두 개 이상의 어휘의 결합이라고 할 수 있다. 이러한 어휘의 긴밀한 결합을 연어 관계(collocational relation)라고 하고, 구체적 내용을 가리킬 때에는 연어 구성(collocational structure)이라고 한다. 또한 연어의 결합은 여섯 가지로 분류할 수 있다. 첫째, 동사+명사(submit a report), 둘째, 형용사+명사(a difficult decision), 셋째, 명사+동사(the fog closed in), 넷째, 명사+명사(radio station), 다섯 번째, 부사+형용사(extremely inconvenient), 마지막으로 동사+부사(appreciated sincerely) 등의 형태로 결합이 있고, 이를 어휘적 연어라고 한다.

본 연구에서는 어휘적 연어에서 두 번째에 해당하는 형용사+명사의 결합 관계를 통해 형용사 hard와 명사 연어를 분석하여 형용사의 의미 범주를 정할 것이다. 다음으로 코퍼스에 대해 살펴보자.

코퍼스(corpus)는 라틴어로 ‘몸’을 뜻하는 말에서 유래하였는데, 언어학에서 코퍼스란, 언어 연구를 위해 모아놓은 텍스트의 집합인 언어자료를 말하며, ‘말뭉치’로 쓰인다. 컴퓨터를 활용하여 언어 자료 수집이 용이해지면서, 전산화된 코퍼스를 통해 얻은 객관적인 데이터를 바탕으로 한 코퍼스 기반의 언어 연구가 활발히 이루어지고 있다. 코퍼스를 통해서 다양한

3)네이버 국어사전(<https://ko.dict.naver.com>)

통계 등 언어에 대한 실증적 자료를 얻을 수 있다는 점에서 언어 연구에 많은 도움을 받을 수 있다. 두 언어에 대한 병렬 언어 코퍼스의 경우는 번역에도 상당한 도움을 줄 수 있다. 아울러, 외국어 학습자를 대상으로 한 학습자 코퍼스 등은 외국어 학습 시 생기는 오류에 대해 파악할 수 있어, 학습자들에게 유용하게 활용된다. 이처럼 본 연구의 자료 분석도 코퍼스를 기반으로 이루어질 수 있다. 본 연구에서도 형용사 hard의 의미 범주를 분석하기 위해 코퍼스의 자료를 활용하고자 한다.

2.2. 연구 방법

본 연구의 진행과정은 다음과 같다. 먼저, 코퍼스 자료인 COCA(Corpus of Contemporary American)를 이용하여, 영어 형용사 hard와 연어를 이루는 명사를 추출하여, 형용사가 어떤 의미를 가지고 있는지 분석하여 의미 범주를 분류한다. 형용사와 결합하는 명사 연어를 알아보기 위하여, COCA에서 앞, 뒤 나타나는 단어의 수를 앞에는 0개로 지정하고, 뒤에 이어지는 어휘의 수는 1개로 하여 뒤쪽에 명사 연어가 오게 설정하였다. 또한 COCA의 문어자료에서, MI(mutual information)수치⁴⁾ 0이상의 빈도수 상위 100개의 명사 목적어를 추출하여 분석하였다. 형용사의 의미를 비슷한 의미끼리 서로 묶어 범주화 시키고, 범주 내에서 코퍼스를 통해 나타난 명사 연어를 바탕으로 형용사가 명사와 함께 사용되었을 때, hard의 한국어 대응 표현에 대해 살펴본다.

3. 본론

3.1 hard의 의미 범주

hard의 의미 범주는 크게 ‘not weak’, ‘not easy’, 그리고, ‘not soft’ 등 세 가지로 나눌 수 있다. 영어 형용사 hard의 기본적인 의미는 수식하는 대상이 약하지 않고 단단하여 강하다는 것을 의미한다. 어떤 구체적인 대상이 단단하다는 것은 의미는 추상적인 대상의 경우, 단단하게 굳어 다루기에 쉽지 않다는 의미로 확대 되어 쓰인다. 즉 처리하기 어렵고 힘든 상황을 나타내거나, 사람의 태도나 대상의 표면이 부드럽지 않다는 의미로 확대되어 쓰인다. 다음, <표 1>는 hard의 의미 범주와 연어를 이루는 명사의 수와 빈도를 나타낸 것이다.

<표1> 형용사 hard의 의미 범주별 명사 연어 수 및 빈도수

의미 범주	not weak	not easy	not soft
명사 연어 수 (총 77개)	45개	21개	11개
빈도수(100%)	58.3%	27.3%	14.3%

4) MI(Mutual Information score for the collocates)수치는 연어로 된 두 단어가 코퍼스자료 전체에서 나온 빈도수와 두 단어가 일정 범위 내에서 근접하여 나타난 빈도수를 비교해서 나온 수치를 말하며, MI 수치가 0이상이면 두 언어의 연어 결합은 유의미하며, 수치가 높을수록 결합된 두 언어의 정보공유 정도와 결합 선호도가 높다.

코퍼스 자료에서 hard와 연어를 이루는 명사 어휘 추출하여 분석한 결과⁵⁾, 위의 표에서 볼 수 있듯이, 빈도수가 가장 높은 의미 범주는 ‘not weak’로 나타났으며, 전체에서 58.3%의 높은 빈도를 보여주고 있어, hard는 대상이 약하지 않고, 단단한 의미로 가장 많이 쓰는 것으로 나타났다. 다음으로는 ‘not easy’ 그리고, ‘not soft’의 의미 순으로 나타났다.

3.2 의미 범주 별 한국어 대응 표현

1) not weak: 약하지 않음

not weak의 범주에 속하는 hard는 다양한 한국어 대응표현으로 나타난다. 다음의 <표2>는 {not weak} 범주에서 hard의 한국어 대응 표현과 그에 따른 사전적 의미, 그리고 대표적인 명사 연어의 예를 나타낸 것이다.

<표2> ‘not weak’의 한국어 대응 표현과 사전적 의미 및 명사 연어의 예

한국어 대응 표현과 사전적 의미		명사 연어 유형
단단한	1. 어떤 힘을 받아도 쉽게 그 모양이 변하거나 부서지지 아니하는 상태에 있다. 2. 험겁거나 느슨하지 아니하고 튼튼하다.	rock, knot
확실한	실제 사실과 꼭 맞아 틀림없다.	evidence, fact, data
독한	사물의 맛이나 냄새 등이 지나치게 자극적이고 심하다.	liquor, drug blow, line, rain
강한	1. 물리적인 힘이 세다. 2. 수준이나 정도가 높다	
경(硬)	명사 앞에 붙어, ‘딱딱한’이나 ‘어려운’ 또는 ‘격식적인’의 뜻을 더하는 말.	currency, science

- (4) a. a hard rock/knot
b. 단단한 바위/매듭
- (5) a. hard evidence/fact/data
b. 확실한/명백한/엄연한 증거/사실/데이터
- (6) a. hard liquor/drug/blow//line/rain
b. 독한/강한/센 술/타격/노선
- (7) a. hard currency
b. 경통화

5) 코퍼스에서 hard와 결합하는 명사 연어 100개에서, 단, 복수형이 함께 나타난 경우 하나는 포함시키지 않았으며, ‘hard bob’과 같은 고유명사 등은 제외하고, 모두 77개의 명사 연어를 추출하였다.

not weak의 범주에서 hard는 ‘바위’와 같이 내구성이 강하여 어떤 충격에도 변하지 않는 대상을 꾸며 줄 때는 (4b)와 같이 ‘단단한’으로 대응되어 쓰인다. 또한, ‘매듭’을 꾸며 주는 경우에도 ‘단단한’으로 표현된다. 한국어 사전의 정의처럼 험겍거나 느슨하지 않게 하는 경우에는 한국어로 ‘단단한’으로 표현할 수 있기 때문이다. 하지만, (5)에서, 어떤 증거나 사실, 데이터 등 정보와 관련 있는 어휘와 함께 쓰일 때, 한국어로는 ‘확실한’으로 대응하거나, 때로는 ‘명백한’ ‘엄연한’ 등으로 표현된다. 정보의 정도가 강하다는 것은 정보의 내용이 구체적이고 확고하여 약하지 않다는 의미로 볼 수 있다. (6)의 한국어에서 ‘술’이나 ‘약물’의 맛이나 냄새의 정도가 심한 경우에 hard는 ‘독한’으로 대응하든지, ‘강한’이나 ‘센’으로 대응되고 치거나 차는 등의 공격적인 행위인, ‘타격’이나 정치적 노선이나, 표현의 수위가 세거나 내리는 비의 정도가 약하지 않은 경우에 hard는 ‘강한’, ‘센’ 등의 한국어로 대응될 수 있다. 또한, 주로 격식적인 표현을 사용할 때는 (8)처럼 단단함을 의미하는 한자어 ‘경(硬)’을 붙여 표현할 수 있다.

2) not easy : 쉽지 않음

not easy의 범주에 속하는 hard는 주로 ‘어려운’, ‘힘든’, ‘고된’ 그리고 ‘곤란한’으로 대응된다. <표 3>를 통해서 hard의 한국어 대응어와 사전적 의미 그리고 대표적인 명사의 종류를 살펴보자.

<표3> ‘not easy’의 한국어 대응 표현과 사전적 의미 및 명사 언어의 예

한국어 대응 표현과 사전적 의미		명사 언어
어려운	1. 하기가 까다로워 힘에 겹다. 2. 겪게 되는 곤란이나 시련이 많다.	time, life decision, choice, question work
힘든	1. 힘이 쓰이는 면이 있다. 2. 어렵거나 곤란하다.	
고된	1. 일이나 훈련이 어렵고 힘들다. 2. 사람이 생활이나 일이 어렵고 힘들다. 3. 사람이 어찌하기가 어렵고 힘들다.	
곤란한	사람이 어떤 일이나 형편이 몹시 딱하고 어렵다.	

- (8) a. time/life/work
b. 어려운/힘든/고된/곤란한 시간/생활/일
- (9) a. decision/choice/question
b. 어려운/힘든/곤란한 결정/선택/질문
- (10) a. a hard worker
b. 열심히 일하는 근로자

not easy에서 hard는 수식하는 대상이 감당하기 어렵거나 처리하기 힘든 것을 나타낸다. 일반적으로 가장 많이 대응되는 표현은 ‘어려운’이지만, 육체적인 노력이나 힘이 드는 경우에는 ‘힘든’으로 표현될 수 있고, 살아가는 일의 어려움과 형편이 쉽지 않다고 판단되는 경우는 ‘고된’이나, ‘곤란한’ 등 표현

으로 대응될 수 있다. (8)과 (9)처럼, 주로 시간이나, 삶 그리고 일 등의 명사를 수식할 때나, 무언가를 결정하거나, 선택할 때, 그리고 대답하기 어려운 질문 등의 명사와 어울릴 때 **hard**는 ‘어려운’, ‘힘든’, 그리고 ‘곤란한’으로 나타난다. **not easy**의 범주에서 **hard**는 하나의 표현으로 한정되지 않고, 위에 설명된 한국어 표현이 두루 대응되어 쓸 수 있다. 하지만, **not easy**의 범주에 포함되는 어휘 중, **work**에서 일을 하는 사람인, **worker**와 함께 쓰일 경우, 영어에서는 ‘일’과 ‘일하는 사람’ 둘 다를 **hard**로 수식할 수 있지만, 한국어에서는 ‘일’과 함께 쓰일 때, **hard**는 위에 언급한 ‘어려운’이나 ‘힘든’ 또는 ‘고된’등으로 대응되는 반면에 예문(10)에서 ‘일하는 사람’을 수식할 때는 아주 다른 한국어로 대응되어 나타난다. 즉, ‘일’, 또는 ‘공부’하는 사람과 함께 쓰일 경우에는 한국어에서는 ‘열심히 하는’이나 ‘근면한’으로 표현된다..

3) **not soft**: 부드럽지 않음

not soft의 범주에 속하는 **hard**는 ‘딱딱한’, ‘냉혹한’, 또는 ‘혹독한’ 등의 한국어 표현으로 대응되어 나타났다. ‘딱딱한’으로 대응되는 경우는 대상의 표현이 부드럽지 않는 경우나, 태도 말씨 등이 상냥하지 않은 경우에 사용하며, ‘냉혹한’이나 ‘혹독한’의 경우도, 감정이나 상황들이 나쁠 경우에 쓸 수 있다.

<표4> ‘not soft’의 한국어 대응 표현과 사전적 의미 및 명사 언어의 예

한국어 대응 표현과 사전적 의미		명사 언어
딱딱한	1. 몹시 굳고 단단하다. 2. 태도, 말씨, 분위기 따위가 부드러운 맛이 없이 엄격하다.	mattress, shell feeling, stare winter
냉혹한	태도나 상황 따위가 인정이 없이 몹시 가혹하다.	
혹독한	1. 무엇이 정도가 몹시 심하다. 2. 마음씨나 하는 짓이 매우 모질고 나쁘다.	

(11) a. a **hard mattress/shell**

b 딱딱한 침대/껍질

(12) a **hard feeling/winter**

b.냉혹한/혹독한 감정/겨울

(11)처럼, 침대 매트리스가 폭신하지 않을 경우, 영어에서는 **hard**로 대상의 표면적 질감이나 재질이 부드럽지 않은 것을 나타낼 수 있지만, 한국어에서는 표면이 부드럽지 않는 것은 주로 ‘딱딱한’으로 표현된다⁶⁾. 영어에서 **hard**는 성격이나 태도, 감정 그리고 바라보는 시선 등의 어휘를 수식할 때, 친절하지 않는 것을 의미할 수 있다. 하지만, 한국어에서는 사전의 설명처럼, 사람의 태도와 감정, 그리고 겨울의 날씨처럼 가혹하거나 모진 것을 표현할 때, 예문 (12)와 같이 ‘냉혹한’이나 ‘혹독한’등으로 대응되어 나타난다.

6) 한국어 ‘딱딱한’과 ‘단단한’은 의미 면에서 비슷하지만 약간 차이가 있다. ‘딱딱한’은 ‘단단한’과 마찬가지로 대상이 ‘굳고 단단한’ 것을 의미하지만, ‘뽕’이나 ‘떡’처럼 질감이 처음에는 부드럽다가 시간이 지나면서 굳어진 것 표현하거나, ‘의자’의 표면의 폭신하지 않는 경우에도 쓸 수 있다.

4. 결론

본 연구는 형용사 hard와 연어를 이루는 명사의 결합을 통해 hard의 의미를 분석하여 범주 별로 분류하고, 이를 바탕으로 영어 hard가 한국어에서 어떠한 표현으로 대응되어 나타나는지를 살펴보았다. hard의 의미 범주를 분류하기 위한 절차는 먼저, 코퍼스를 통해서 형용사가 수식하는 명사를 수식할 때, hard가 어떤 의미를 가지고 있는지 구분하였다. 또한, 의미 구분을 위해 영한사전을 참고 하였으며, 가까운 의미끼리 묶어 의미범주를 나누었다. 이러한 의미 범주에 따라, 형용사 hard의 한국어 대응표현을 알아보았다.

코퍼스와 사전의 자료에서 통해서 영어 형용사 hard의 의미 범주는 ‘not weak’, ‘not easy’, ‘not soft’ 등의 세 가지로 구분된다. 범주의 빈도수는 not weak(58.3%) > not easy(27.3%) > not soft(14.3%)의 순으로 나타났다. 다음은 의미 범주에 따른 한국어 대응표현을 <표 5>로 정리하였다.

<표 5> hard의 의미범주와 명사 연어 유형에 따른 한국어 대응표현

의미 범주	명사 연어 유형	한국어 대응표현
not weak	단단한 물체 /매듭	단단한 경(硬)(격식적인 표현)
	사실, 증거, 자료	확실한
	술, 약물	독한, 강한, 센
	표현/치거나 차는 행동/비	강한, 센
not easy	시기, 삶, 일	어려운, 힘든, 고된, 곤란한
	결정, 선택, 질문	
	일하는 사람	열심히 하는
not soft	표면, 질감	딱딱한
	태도, 감정, 눈길	냉혹한, 혹독한
	계절	

지금까지 여러 개의 의미로 나열된 영어 형용사 hard의 의미를 같은 의미로 묶어 범주 별로 분류하고, 범주에 따라 hard의 한국어 대응표현에 대해 코퍼스 자료를 통해 살펴보았다. 연구를 통해서, 영어는 세 가지의 범주에서 하나의 어휘 hard로 표현하는 반면에 한국어는 명사 연어의 유형에 따라 아주 다양한 한국어 표현으로 대응되어 나타났다. 어휘를 학습할 때 학습자들이 많이 의지하는 텍스트가 영한사전인데, 주로 사전에서 제일 먼저 나타나는 한국어 표현이 그 어휘의 주된 표현인줄 알고 학습하는 경우가 많이 있다. 하지만, 이렇게 의미를 범주로 나누어 이해한다면 효율적인 어휘 학습이 이루어 질 수 있고, 여러 가지 한국어 대응표현을 적용하여 어휘를 이해한다면 영어 어휘 학습에 유용하게 활용 될 수 있을 것이다. 본 연구가 다의성을 가진 외국어 어휘 학습에 도움이 되기를 바라며 이 글을 마무리 한다.

참고문헌

- 김은일(Kim, E. I). 2017a. 다의어 동사 ‘open’과 ‘열다’에 대한 어휘유형론적 접근 (A Lexical Typological Approach to the English Verb ‘open’ and the Korean Verb ‘yeolda’). 「현대문법연구」 (Studies in Modern Grammar) 91, 167-187.
- 김은일(Kim, E. I). 2017b. He Won his Wife in a Card Game: WIN의 의미지도 (He Won his Wife in a Card Game: Semantic Maps of WIN), 「새한영어영문학」 (The New Korean Journal of English Language and Literature) 59.4, 173-196.
- 박진호(Park, J. H). 2015. 언어유형론이 한국어 문법 연구에 제시하는 것들 (What Linguistic Typology Reveals to the Study of Korean Grammar). 「일어일문학연구」 (The Japanese Language and Literature Association of Korea) 94.1, 3-27.
- 엄영임(Aum, Y. N). 2016. 영어 형용사 hard, heavy, strong 사용에 관한 코퍼스 기반 연구 (A Corpus-Based Study of the Use of the English Adjectives hard, heavy, and strong). 부산대학교 석사학위 논문 (M.A. Thesis. Busan National University).
- 이기동(Lee, K. D). 2015. 이기동의 영어 형용사 연구(BASIC) (English Adjectives) 경기, 교문사.
- 정연창, 김은일, 박정훈(Jung, Y. C., E. I. Kim and J. H. Park). 2016. 영어 형용사 ‘difficult’와 ‘hard’의 사용에 대한 코퍼스 기반 연구와 시사점 (A Corpus-Based Study of the Use of the English Adjectives difficult and hard and Its Implications: The Lexical Approach). 「현대문법연구」 (Studies in Modern Grammar) 88, 19-45.
- Geeraerts, D. 1993. Vagueness’s Puzzles, Polysemy’s Vagaries, *Cognitive Linguistics* 4, 223-272.
- Haspelmath, M. 2003. The Geometry of Grammatical Meaning: Semantic Maps and Cross-linguistic Comparison. In M. Tomasello (ed.), *The New Psychology of Language* 2. New York: Lawrence Erlbaum Associates Publishers: 211-243.
- Langacker, R. W. 1991. *Foundations of Cognitive Grammar, Vol II: Descriptive Application*. California: Stanford University Press.
- Nerlich, B., Z. Todd, V. Herman and D. Clarke. (eds). 2003. *Polysemy: Flexible Patterns of Meaning in Mind and Language*. Berlin: Mouton de Gruyter.
- Riemer, N. 2005. *The Semantics of Polysemy: Reading Meaning in English and Warlpiri*. Berlin: Mouton de Gruyter.

현대 일본어와 한국어의 동사 조건형의 문법화에 관한 대조연구 -시각활동을 나타내는 동사 「みる」「보다」의 경우-

하재필*
(부산대학교)

< 목 차 >	
I. 서론 1. 들어가기 2. 선행연구 3. 연구대상 및 연구방법	II. 본론 1. 「みれば」와 「보면」의 본동사 용법 2. 문법화한 용법의 대조 III. 이후의 과제

키워드 : 동사 조건형, 문법화, 후치사, 접속사, 접속조사

I . 서론

1 . 들어가기

본 연구에서는 현대 일본어와 현대 한국어의 동사 조건형을 대상으로, 양 언어의 동사 조건형이 동사성을 상실해서 구문적 기능을 수행하는 문법 형식으로 쓰이는 문법화 현상에 주목하여 양 언어의 공통점과 차이점을 고찰한다. 본 발표에서는 시각활동을 나타내는 동사 「みる」「보다」에 한정하여 논한다.

일본어와 한국어에서 종속문과 주문 사이에 조건관계가 성립하는 경우, 종속문의 서술어는 조건형을 취하는데, 시각활동을 나타내는 동사 「みる」「보다」는 예문 1, 2와 같이 「みれば」「보면」의 형태를 취해서 <조건>임을 나타낸다¹⁾.

- (1) 村人が狼を**みれば**、羊飼いの少年を助けにくるだろう。
- (2) 마을 사람들이 늑대를 **보면** 양치기 소년을 도우러 오겠지.

그리고 잘 알려진 바와 같이 「みれば」와 「보면」에는 예문 3, 4와 같은 파생용법이 있다.

- (3) 村人の立場から**みれば**、嘘つきの羊飼いの少年は問題児だ。
- (4) 마을 사람들의 입장에서 **보면** 거짓말쟁이 양치기 소년은 문제아다.

* 일어일문학과 조교수, 현대일본어문법

1) 조건형이라고 해도 일본어의 경우 「すれば」 외에 「すると」「したら」「するなら(ば)」「しても」 등의 형식이 있다. 한편 한국어의 경우 「하면」 외에 「하거든」「하더라도」 등의 형식이 있다. 본 발표에서는 지금까지의 연구결과에 바탕하여 일본어는 「すれば」「すると」 형식에, 한국어의 경우 「하면」 형식에 한정한다. 또한 문법화 현상에 있어서 「みれば」와 「みると」 사이에 거의 차이가 없기 때문에 본 발표에서는 편의상 「みれば」 형식의 예문을 사용한다.

예문 1, 2와 3, 4의 「みれば」 「보면」은 형식으로 보면 같아 보이지만 그 어휘·문법면에서의 성질은 다르다. 우선 어휘면을 보자면 예문 1, 2의 「みれば」 「보면」은 시각활동을 나타내는 어휘적 의미를 가지고 있다²⁾. 그러나 예문 3, 4의 경우 시각활동을 나타내는 의미는 추상화되어 사고활동을 나타내는 「かんがえれば」 「생각하면」에 가깝거나, 그 어휘적 의미가 희박해져 있어 「村人の立場からみれば」 「마을 사람들의 입장에서 보면」을 「村人にとっては」 「마을 사람들의 입장에서는」과 같이 다른 문법 형식으로 바꾸어 말할 수도 있다.

이어서 문법면을 보자. 예문 1, 2의 「みれば」 「보면」은 종속문의 서술어로 각각 주어인 「村人」 「마을 사람들」의 속성인 동작을 서술하고 있다. 그리고 서술어이기 때문에 상(aspect)나 극성(polarity) 등의 문법 범주에 따라 어형변화를 할 수 있다.

(5) 村人がじっとみていれば、狼は近づいてこないだろう。

(6) 마을 사람들이 **보고 있으면** 늑대는 다가오지 않겠지.

(7) 羊が嘔まれて死んでいるのを見なければ、村人は少年の言うことを信じないだろう。

(8) 양이 물려서 죽어 있는 것을 **보지 않으면** 마을 사람들은 소년이 하는 말을 믿지 않겠지.

그러나 예문 3, 4와 같은 용법이 되면 「みれば」와 「보면」은 동사성을 상실하여 종속문의 서술어로서 기능하지 않는다. 그렇기 때문에 대응하는 주어는 없으며 어형은 고정된다.

그런데 「みれば」 「보면」이 문법화를 하게 되는 경우에 관해서는 학습자를 위한 문형사전 등에서 사용빈도가 높은 표현을 다루고는 있다(松木·森田1989, グループ・ジャマシイ1998, 백봉자2006 등). 그러나 발표자의 조사에 한해서 말하자면 동사의 일반적인 성질에 바탕을 두어 일본어와 한국어의 동사 조건형의 문법화 현상을 고찰한 본격적인 연구는 없는 듯 하다.

이에 본 발표에서는 다음과 같은 세 가지 문제에 초점을 맞추어 시각활동을 나타내는 동사 「みる」 「보다」의 조건형인 「みれば」 「보면」를 대상으로 하여 대조 고찰을 한다

- ①본동사에 문법화가 일어날 때 수반되는 어휘·문법면의 일반적 변화
- ②문법화 용법의 변종과 파생경로에서 보이는 양 언어 간의 공통점과 상이점
- ③문법화 용법에서 보이는 양 언어 간의 공통된 또는 상이한 구문적 특징

2. 선행연구

2.1 품사로서의 동사와 동사 조건형의 형태론적 위치

일반적으로 말해 품사로서의 동사는 어휘적 의미, 구문적 기능, 형태론적 어형변화와 같은 기준을 통해 규정된다. 가장 전형적인 동사는 어휘적으로는 사람·사물의 운동(동작·변화)를 나타내고, 구문적으로는 사람·사물을 나타내는 주어의 속성을 서술하는 서술어로서 기능한다. 그리고 형태론적으로는 술어가 되는 동사는 화자와 현실과의 관계, 화자와 청자

2) 문법화를 일으키는 단어가 갖는 어휘적 측면의 경향으로서 그 어휘적 의미가 일반적이며 유의어 안에서 상위어에 속한다(Hopper and Traugott 2003:101). 그렇기 때문에 「みれば」 「보면」은 본동사 용법에 있어서 다른 유의어로 바꾸어 말할 수 있다. 예를 들어 「晴れの日、海岸から南の方をみれば、対馬が見える」 「맑은 날 해안가에서 남쪽을 보면, 대마도가 보인다」의 경우 「みれば」 「보면」은 각각 「眺めれば」 「바라보면」으로 바꾸어 말할 수 있다. 그러나 「みれば」 「보면」에 문법화가 일어나 버리면 이러한 유의어 관계는 성립하지 않는다.

와의 관계 등을 반영해서 시제(tense), 법(mood), 상(aspect) 등의 문법 범주에 따라 어형변화를 한다.

그리고 동사는 그 구문적 위치에 따라 <종지 終止>, <연용 連用>, <연체 連體>, <접속 接續>의 기능을 하며, 그 기능에 맞추어서 <종지형 終止形>, <연용형 連用形>, <연체형 連體形>, <접속형 接續形>의 체계를 이룬다.

본 발표에서 다루는 동사 조건형은 <접속형 接續形>에 속하며 일본어에는 「すれば」「すると」「したら」「するなら」「しても」「したのに」 등의 형식이 있다. 한편 한국어의 조건형에는 「하면」「하라면」「하거든」「하더라도」 등과 같은 형식이 있다.

2.2 동사가 동사성을 잃을 때의 변화

다카하시 高橋(2003)은 주로 동사의 구문적 기능, 동사의 문법 범주 및 그 형태론적 성격을 다룬 연구이다. 그 중에서 동사가 동사성을 잃을 때 어떠한 변화가 일어나는지를 지적하고 있다. 다카하시는 ①의미가 변한다, ②기능이 변한다, ③형태론적 범주의 체계가 변한다고 하였다(같은 논문: 261-262).

이와 같은 다카하시의 지적은 일본어를 대상으로 한 것이지만 해외의 문법화 연구에서도 자주 언급되는 바이기도 하며, 발표자의 소견으로는 한국어의 현상에도 적용할 수 있는 것으로 보인다.

3. 연구대상 및 연구방법

3.1 연구대상

본 발표에서는 어휘적 의미가 인간의 인식활동과 연관되어 문법화의 소재로 종종 거론되는 시각활동을 나타내는 동사 「みる」와 「보다」³⁾를 대상으로 한다. 그리고 일본어와 한국어의 조건형 중에서 <조건>을 나타내는 전형적인 형식이라 할 수 있는 「すれば」 형식과 「하면」 형식을 취한 경우를 대상으로 한다.

3.2 연구방법

본 발표에서는 실제로 사용에 있어서의 의미·용법을 기술하기 위해 소설, 수필과 같은 문학작품, 그리고 신문, 논설문에서 데이터를 수집하였다.

기본적인 분류는 동사 조건형의 구문적 기능과 의미에 주목하여 진행하였다. 우선 「みれば」「보면」가 종속문의 서술어로 기능하는 경우와 그렇지 않은 경우로 나눈다. 동사 조건형이 종속문의 서술어로 기능하는 경우, 주문의 사건이 성립하는 데에 관여하는 <조건>을 나타내는 용법(전형적인 용법)과 주문의 사건이 성립하는 데에 직접적으로 관여하지 않는 용법로 나눈다(주변적인 용법)⁴⁾.

본 발표에서는 「みれば」「보면」이 종속문의 서술어로 기능하지 않는 경우를 연구대상으로 하고 있지만, 이 경우 「みれば」「보면」은 기본적으로 다른 구문요소에 후접하는 패

3) 「돌아보다, 살펴보다, 바라보다」는 「보다」가 자립어가 아니라 어구성의 요소이므로 본 발표에서는 다루지 않는다.

4) 조건형의 주변적 용법에는 이른바 <발견의 조건>이 있다. 예를 들어 「개찰구를 나와서 오른쪽을 보면 커피숍이 있습니다」의 경우 동작 주체가 개찰구를 나와서 오른쪽을 보는 일이 성립하지 않아도 카페는 존재한다. 이러한 관점에서 보자면 전형적인 조건을 나타낸다고 보기는 어렵다. 그러나 동작주체가 개찰구를 나와서 오른쪽을 보는 일이 일어나지 않으면 카페의 존재를 확인할 수 없다. 이러한 경우를 스즈키 鈴木(1972: 353)에 따라 <발견의 조건>이라고 부른다.

턴으로 쓰인다. 데이터를 보면 「みれば」 「보면」의 의미·용법은 선행하는 요소에 따라 영향을 받으므로 본 발표에서는 양 형식의 구문적 위치와 더불어 선행 요소의 형식과 의미적 유형에 주목한다.

II. 본론

본론에서는 우선 간략하게 「みれば」 「보면」의 본동사 용법을 확인하면서 본동사로서의 어휘·문법적 특징을 확인한다. 이어서 동사성을 잃은 용법을 보면서 동사 조건형이 어떠한 용법으로 발달하는지를 비롯하여 각 용법에서 어떠한 구문적 특징이 나타나는지 고찰한다.

1. 「みれば」와 「보면」의 본동사 용법

본동사 용법의 경우 「みれば」와 「보면」은 종속문의 서술어로 기능하고, 종속문에 제시된 사건⁵⁾이 주문에 제시된 사건의 성립에 관여하는 <조건>임을 나타낸다. 그러나 단순히 <조건>이라 하여도 문장에 서술된 내용과 현실과의 관계에 따라 몇 가지 유형으로 나눌 수 있다. 대표적으로는 예문 9, 10과 같은 <가정적 조건> 외에, <반복·습관적 조건>, <일반적 조건> 등이 있다.

① 가정적 조건

- (9) 警察の連絡では岩に顔をぶつけたらしく、身許が不明だと聞かされたが……仲間なら分かるよな。どんなに会ってなくても、バンドラの小さな体を見れば皆がきつと思いでくれるはずだ。(高橋克彦『バンドラ・ケースよみがえる殺人』)
- (10) 이윽고 그는 급히 집으로 향했다. 한쪽 눈을 잃은 것을 보면 부모님께서 몹시 놀라실 것이다.(김성중『여명의 눈동자 5』)

② 반복·습관적 조건

- (11) このことは私が一つの煙を見、次にその煙の下に行ったことを示している。煙をみれば、必ずそこへ行ったのだった。(大岡昇平『野火』)
- (12) 학벌 콤플렉스는 무효가 된 지 오래건만, 나는 그때까지도 연필 끝으로 종이를 찢어 내는 버릇을 못 고치고 있었다. 선회는 내가 그 짓을 하고 있는 것만 보면 그림을 그리고 있다가도 다가와서 뒤로부터 팔을 잡으며 못 하게 했다.(박완서『그 산이 정말 거기 있었을까』)

③ 일반적인 조건

- (13) 古今東西、人を恋するのは人情、美人を見れば、誘惑にもかられる。(毎日新聞910924・総合面)
- (14) 순범은 눈앞에서 어른거리는 칼날에 현혹되면 안 된다는 생각이 본능적으로 떠올랐다./(눈을 보아야만 한다.)/칼끝을 보면 그 변화에 현혹되고 교란되어 반드시 실수를 하게 마련이다.(김진명『무궁화 꽃이 피었습니다 2』)

이와 같은 용법에서 「みれば」 「보면」은 종속문의 서술어이므로 앞서 예문 5~8에서 확인한 바와 같이 상, 극성 등에 따른 어형변화가 가능하다.

그러나 파생 용법에서 「みれば」 「보면」은 이러한 동사성을 상실하고 구문 요소와 요소

5) 기본적으로 눈으로 구체적인 사물과 사건을 인식하는 시각활동은 나타내지만 잘 알려진 바와 같이 동사 「みる」 「보다」와 조합하는 명사의 유형이나 구문적 환경에 따라 /읽다, 조사하다, 생각하다/ 등 다양한 의미를 실현한다.

를 연결짓거나 문장과 문장을 연결짓거나 하는 문법형식이 된다. 이 현상의 배경에는 동사 조건형이 종속문의 서술어=비종지(非終止)라는 구문적 위치에 의한 요인이 토대가 된다. 예를 들어 예문 15~18을 보면 알 수 있듯이 동사 조건형에는 비과거의 사건과 과거의 사건 사이에 형태론적 시제 대립이 없다.

- (15) 太郎は、犬をみれば、いつも逃げる。
- (16) 太郎は、犬をみれば、いつも逃げた。
- (17) 민호는 개를 보면 항상 도망친다.
- (18) 민호는 개를 보면 항상 도망쳤다.

이로부터 「みれば」「보면」의 문법화에는 동사의 어휘적 의미의 일반성과 더불어 문법형식의 비종지라는 구문적 특성과 이에 수반되는 형태론적 특성이 동사성을 상실케 하는 요인으로서 작용하고 있다 볼 수 있다.

2. 문법화한 용법의 대조

본절에서는 「みれば」와 「보면」이 동사성을 잃은 용법을 중심으로 서술하면서 일본어와 한국어에서 보이는 공통점과 상이점을 확인한다. 우선 일본어와 한국어의 「みれば」와 「보면」의 문법화에 수반되는 다양한 어휘·문법적 성질의 변화를 정리한다(2.1). 이어서 단어로서의 자립성을 잃은 「みれば」「보면」이 접속하는 선행요소의 품사와 그 형식을 중심으로 두 형식의 분포와 파생경로의 전체상을 확인한다(2.2). 마지막으로 문법화한 용법의 구체적인 사례를 확인하면서 양 언어 간의 구문적 특징에서 나타나는 공통점과 상이점을 확인한다(2.3).

2.1 문법화에 수반되는 어휘·문법면의 변화

「みれば」와 「보면」이 본동사로 쓰이는 경우와 문법화한 경우를 비교하였을 때 확인할 수 있는 어휘·문법면의 변화를 정리하면 <표1>과 같다.

<표1> 본동사의 용법과 문법화한 용법 사이의 어휘·문법면의 변화

	본동사 용법	문법화한 용법
문장 구조	복문이다	단문이다 (단문으로 이행한다)
주문의 구문적 특징	동사술어문이 중심 평서문이 중심	형용사·명사술어문이 중심 평서문에 한정
어휘적 의미	기본적으로 시각활동을 나타내고 파생된 의미를 갖는다	추상화되거나 희박해진다
형태론적 특징	어형활용이 가능	어형이 고정

또한 본동사 용법의 경우 「みれば」는 「ながめる」「のぞく」「にらむ」「のぞむ」등과 「보면」은 「바라보다」「들여다보다」「짜려보다」등과 유의어 관계에 있으나, 문법화가 일어나면 「みれば」「보면」의 어휘적 의미가 추상화가 되기 때문에 경우에 따라 「かんがえる」「생각하다」와 같은 사고활동을 나타내거나, 그 의미가 희박하게 되어 형식

만이 남게 된다.

<표1>의 「みれば」와 「보면」의 본동사 용법을 보면 개별 언어의 특성에 의해 주문에 쓰이는 문장의 전달적 유형-평서문, 명령문, 권유문 등-에 차이는 보이지만⁶⁾ 전체적으로는 문법화와 함께 일본어와 한국어에 일정한 변화가 일어난다고 볼 수 있다.

그런데 앞서 「みれば」와 「보면」에 문법화가 일어나면 단어로서의 자립성을 상실하고 다른 구문 요소에 후접하는 양상을 보인다고 하였다. 여기서 「みれば」와 「보면」에 선행하는 주요한 형식을 정리하면 <표2>와 같다.

<표2> 「みれば」 「보면」에 선행하는 형식의 분포

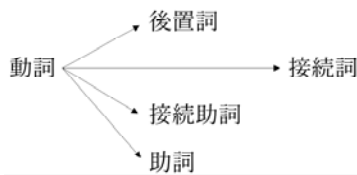
선행요소의 품사	「みれば」의 선행요소	「보면」의 선행요소
명사	~から ⁷⁾ 、~で、~を ~であって、~にして	~에서、~으로、~을/를 ~이고
동사	そして、して	그리고、~하다 ⁸⁾

<표2>의 내용으로부터 다음과 같이 말할 수 있다.

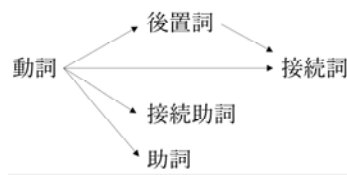
우선 「みれば」 「보면」 모두 명사와 동사에 접속한다. 이러한 사실로부터 「みれば」와 「보면」이 후치사와 접속사로 발달하는 것을 예상할 수 있다. 그리고 「보면」에 한해서 말하자면 「~하다」와 같은 동사 접속형에 후접하는 것으로 보아 복문에서 접속조사로 쓰이고 있을 것이 예상된다.

2.2 「みれば」와 「보면」의 문법화 경로의 전체상

이번에 수집한 데이터를 바탕으로 「みれば」와 「보면」 각각의 문법화한 용법을 고찰한 결과를 도식화하면 【그림 1】 【그림 2】와 같다.



【그림 1】 「みれば」의 문법화 경로



【그림 2】 「보면」의 문법화 경로

【그림 1】 과 【그림 2】로부터 다음과 같이 점을 지적할 수 있다.

「보면」의 경로에서 후치사로부터 접속사로 발달하는 경로가 보이는 것을 제외하면⁹⁾

6) 일본어에서 종속문의 조건형이 「すれば」 「すると」의 형식을 취하는 경우, 주문은 기본적으로 평서문에 한정된다.
 7) 시간부사인 「今、現在、後」가 「~から」와 같은 형태로 표시되고 「みれば」가 접속하는 경우가 있다. 이러한 경우 <과거의 사건을 다시 판단·평가하는 화자의 태도>를 나타내는 복합적인 진술부사(문장부사)로 발달하는 어휘화가 일어난다.
 8) 예를 들어 「철수는 만화책을 읽다(가) 숙제가 있었다는 것이 생각났다」와 같이 종속문의 서술어가 「~하다(가)」 형식을 취하는 경우를 말한다. 문법화한 「보면」이 접속하는 것은 「~하다」 형식에 한정된다.
 9) 다만 【그림 1】은 어디까지 이번에는 수집한 데이터에 바탕을 두고 있어 일본어의 경우에 있어서도 구문적 조건이 갖추어지면 후치사에서 접속사로 발달하는 형식을 상정할 수 있다. 그렇게 되면 적어도 「みれば」와

「みれば」와 「보면」의 문법화 경로는 유사하다고 할 수 있다. 이러한 유사성은 동사의 어휘적 의미가 갖는 구문적 능력에 의한 것이라 생각할 수 있다. 즉 <표2>에서 거론한 문법화에 쓰이는 형식은 모두가 연용적(連用的)인 형식으로 어떠한 연용적 형식과 연계되는가는 동사의 어휘적 의미가 갖는 구문적 능력에 의한다. 그러하다면 동사의 어휘적 의미가 갖는 유사성이 문법화한 용법과 문법화 경로의 유사성에 영향을 미치고 있다고 볼 수 있다.

그리고 위 그림에는 명시되지는 않았으나, 동사 조건형이 어떠한 문법적 기능을 하기 위한 형식으로 바뀌어 가도 동사 조건형의 본래의 성격 - 문장에 나타나지 않은 가능성으로서의 다른 사건을 함의하는 문법적 의미, 두 문장을 접속하는 기능-은 유지된다¹⁰⁾.

2.3 「みれば」와 「보면」이 문법화한 용법

본절에서는 「みれば」와 「보면」의 문법화한 용법과 그 형식을 확인하면서 양 언어 간의 공통점과 상이점을 고찰한다.

2.3.1 <후치사화>한 용법¹¹⁾

「みれば」와 「보면」은 명사에 후접하여 어떠한 사물·사건에 대한 평가·판단을 할 때의 <평가·판단의 근거>를 한정한다. <평가·판단의 근거>의 하위 분류로는 평가·판단의 <입장>, <재료>, <관점>이 있다. 이러한 용법에는 「~からみれば」 「~でみれば」 「~をみれば」 「~にしてみれば」 「~에서 보면」 「~(으)로 보면」 「~을/를 보면」와 같은 형식이 쓰인다.

이하 본동사 용법에 기반을 두면서 일본어와 한국어의 공통점과 상이점을 확인할 수 있는 형식을 중심으로 선행하는 명사(명사구) 형식과 의미적 특징을 확인한다.

① <평가·판단의 입장>

「みれば」 「보면」이 어떠한 사물·사건에 대한 평가·판단의 입장을 한정하는 경우가 있다. 이때 「みれば」는 「~からみれば」 「~でみれば」 「~にしてみれば」 형식으로 쓰인다. 한편 「보면」은 「~에서 보면」 「~(으)로 보면」의 형식으로 쓰인다.

우선 「~からみれば」와 「~에서 보면」의 경우를 보겠다. 일본어의 'から'격과 한국어의 '에서'격은 「展望台から星をみる」 「전망대에서 별을 보다」와 같이 동작이 이루어지는 장소를 나타내고 있어 (아라 荒1983 : 422, 趙1995 : 45) 'から'격과 '에서'격의 명사와 동사의 조합이 공통되고 있음을 확인할 수 있다. 이러한 조합이 중간단계를 거쳐¹²⁾ 후치사화한 형식으로 발달한다.

「~からみれば」 「~에서 보면」이 접속하는 명사의 유형을 보면 집단을 나타내는 도시명, 국가명 (예문 19, 20) 외에 추상화된 공간인 「~の立場, ~側」 「~입장, ~쪽」 (예문 21, 22)이 쓰인다.

「보면」의 문법화 경로는 일치한다고 보아도 좋을 것이다.
 10) 동사 조건형의 문법적 의미와 기능의 존속에 관해서는 河 (2019) 를 참고 바란다.
 11) 본고에서 말하는 <후치사>란 기본적으로 스즈키 鈴木 (1972:499) 의 규정을 따른다.
 「단독으로는 문장 성분이 될 수 없으며 명사의 격 형식과 조합하여 해당 명사가 다른 단어에 대한 관계를 나타내기 위해 발달한 보조적 단어이다」
 12) 위 예문 19를 가지고 간략히 설명하겠다. 예문 19와 같은 문장 구조는 「~を~からみれば~」 (札幌を東京からみれば, まだまだ田舎だ) 와 같은 중간단계에서 '을'격 명사가 「~は」로 표시되어 화제를 나타내게 되면서 「東京からみれば, 札幌はまだまだ田舎だ」와 같은 구조가 만들어진다.

- (19) 「札幌の人間が聞いたら、それは吹き出すよ。だが、そうは言っても、まあ東京からみればまだまだ田舎だからねえ」(三浦綾子『塩狩峠』)
- (20) 「일본에서 보면 한국 남자들 참 못났어요. 촌스럽고 통명스럽고.」(김진명『무궁화 꽃이 피었습니다 3』)
- (21) ポスト海部をうかがう竹下の立場からみれば、腹心が海部の幕引き役をするのは、なんとも奇妙な具合、作戦的には好ましくない。(毎日新聞910416・2面)
- (22) 생산수단의 공유와 평등은 소련사회에서 가장 중시되는 가치였다. 그러나 소련 인민들의 입장에서 보면 이것은 어디까지나 추상적인 차원의 문제였다.(유시민『부자의 경제학 빈민의 경제학』)

이와 같이 「~からみれば」 「~에서 보면」이 접속하는 명사의 유형에 공통점이 보이지만 「~からみれば」에는 사람명사도 쓰여 상이점도 확인할 수 있다.

- (23) お前のような男なら、向うの人から見れば、結婚の相手として不足はないんだからね。(石川達三『青春の蹉跎』)

또한 일본어의 경우 'から'격 외에 'で'격을 취하는 공간명사와의 조합도 동작이 이루어지는 장소를 나타내어 (「展望台で星をみれば…」¹³⁾, 이 용법에서 <평가·판단의 입장>을 나타내는 후치사로 발달한다.

- (24) 麻原被告が公判延期をもくろんで弁護人を解任したとしたら、悔しいけれど「やられた」という感じ。オウムはこれまでも常識では考えられない手を繰り返してきたが、また先手を取られた。だがサリン事件の被害者の立場で見れば、こんなふざけた話はない。(毎日新聞951026・社会面)

한편 한국어의 경우 「~(으)로」가 활동의 자격을 나타내는 용법 (「레지던트가 당직으로 환자를 보다」)에서 <평가·판단의 입장>을 나타내는 후치사로 발달한 것으로 보인다. 「~(으)로 보면」의 경우 집단·사람을 나타내는 명사가 쓰인다.

- (25) "아이고, 정군하고 임씨 애 먹었제?"/어머니는 명훈을 짓혀두고 정군과 임씨만을 반색하며 맞았다. 이웃 군에서 온 일곱 일꾼 중에서 마지막 남은 둘이었다. 그것도 탄사람과 함께 품삯지불이 나온 개간지로 옮겨갔다가 무엇인가 틀어져 되돌아온 것이지만 명훈네로 보면 고맙기 짝이 없는 사람들이었다.(이문열『변경 5』)

② <평가·판단의 재료>의 한정

「みれば」 「보면」가 어떤 사물·사건에 대한 평가·판단을 할 때의 재료를 한정하는 경우가 있다. 이때 「みれば」는 「~でみれば」 「~からみれば」 「~をみれば」과 같이 쓰인다. 한편 「보면」은 「~(으)로 보면」 「~을/를 보면」이 쓰인다.

우선 'で'격과 '(으)로'격 명사가 구체명사로 구체적인 동작을 나타내는 동사와 결합하는 경우를 살펴보면 「顕微鏡で細胞をみる」 「현미경으로 세포를 보다」와 같이 'で'격과 '(으)로'격 명사는 동작의 수단을 나타내는 공통점이 있다. 이를 바탕으로 두 형식을 비교한다.

- 13) 마찬가지로 동작이 이루어지는 장소라 하더라도 공간명사가 'から'격을 취하는 경우 동작의 방향성을 나타내는 뒤양스가 수반된다 (荒1983: 404).

「みる」「보다」가 조건형을 취하면서 추상명사의 '데'격과 '(으)로'격의 명사가 조합하면 「みる」「보다」의 어휘적 의미가 /판단하다, 생각하다/와 같은 사고활동을 나타내는 의미로 추상화된다.

(26) 内心ひやひやしていた。一人の生徒の偏差値が同校の推薦基準に達していないのだ。だが、この成績表には同じ生徒について二つの偏差値が記されている。県内の高校受験に使うものと、都内の高校用のもの。三の生徒の場合、県の偏差値で見れば基準に達していた。(毎日新聞910312・3面)

(27) 형색이 걸인처럼 볼품없고 깡마른 사람이 효녀의 집 앞에서 고개를 빼고 기웃거리는 게 보였다. 차림새로 보면 영락없는 걸인이었다. 해골처럼 뼈쩍 마른 형상이 더욱 그랬다. (구석봉 『한국사를 뒤흔든 여인들』)

이어서 잘 알려진 바와 같이 본동사인 「みる」「보다」는 그 시각활동의 대상으로서 '을'격과 '을/를'격의 구체명사를 취한다. 그러나 「みる」「보다」가 「みれば」「보면」의 형식을 취하고 '을'격, '을/를'격의 명사에 상황이나 모습을 나타내는 명사(명사구)가 쓰여서 <평가·판단의 재료>를 나타낼 수 있다.

(28) 現代のヤングの行動パターンを見れば、ヒットをねらえる雑誌はコミックでしょう。(毎日新聞910313・3面)

(29) 그가 가고난 뒤 여욕은 좀 당황한 기분이었다. 이렇게 점잖은 손님은 처음이었던 것이다. 경계심이 풀리면서 대신 호기심이 일었다. 군인들과 함께 온 걸 보면 군인인 것 같은데 여느 군인들과는 좀 다른 데가 있었다. (김성중 『어명의 눈동자 4』)

일본어의 경우 「~からみれば」의 용법에 <평가·판단의 재료>를 나타내는 용법이 있다. 이 경우 모습이나 상황을 나타내는 명사(명사구)가 'から'의 형식으로 표시된다.

(30) 栄子はもともと父の妾だった。前身は何であったか知らない。いつも濃い化粧をするところから見れば、どうせ何か水商売であったに違いない。(石川達三 『青春の蹉跎』)

③ <평가·판단의 관점>의 한정

어떠한 사물·사건을 판단할 때 평가·판단의 관점을 한정하는 경우가 있다. 얼핏 보면 평가·판단의 입장을 한정하는 경우와 유사하게 보인다. 명확한 경계선을 긋기는 어렵지만, 알기 쉽게 말하자면 입장을 한정하는 경우는 '누가 그 평가·판단을 하는가', 그 주체에 초점이 맞춰진다고 하면 관점을 한정하는 경우 '어떠한 잣대로 평가·판단을 하는가, 그 사고방식이나 가치관에 초점이 맞춰진다고 할 수 있다.

일본어의 경우 「~からみれば」 형식이 쓰이는데, 관점을 한정하는 용법은 「窓から空をみる」와 같이 동작의 경유점을 나타내는 용법에서 만들어진 것이 아닌가 여겨진다. 「~からみれば」가 쓰이는 명사를 보면 어떤 특정한 분야·사물에 대한 사고방식, 그리고 이들 명사에 「~目, ~観点」과 같은 명사의 조합이 쓰인다.

(31) 先の石山氏は、ヨーロッパ型の生活の質の維持をめざすE.Cの基準から見れば、狭い家に住んで長時間通勤・労働に耐える日本のサラリーマン社会は文明的な敵と映ると述べていたが…(毎日新聞911030・文化面)

(32) 今回の人的貢献は「平和の回復」という観点から見れば国連平和維持活動(P.K.O)の一環と言えなくもない。(毎日新聞910506・2面)

한편 한국어도 평가·판단의 관점을 한정하는 형식이 있으나 한국어의 경우 「보면」에 선행하는 명사의 형식은 도구·수단을 나타내는 「~(으)로」가 쓰인다. 「~(으)로 보면」의 형식으로 쓰이는 명사의 유형을 보면 일본어와 차이가 없어 보인다.

- (33) 하인리히 마르크스는 과격한 자유주의자는 아니었지만 그 시대의 기준으로 보면 매우 진보적인 사람이었다. (유시민 『부자의 경제학 빈민의 경제학』)
- (34) 물론 해외로 빼돌린 재산만 없다면 그 돈이 결국 우리나라 안에 있기 때문에 국부적인 관점으로 보면 큰 문제가 아닐 수도 있는 것이다. (김홍신 『인간시장 9』)

2.3.2 <접속조사화>한 경우

「みれば」와 「보면」이 접속조사로서 쓰이는 경우를 보면 「~であってみれば」와 「~하다 보면」의 형식이 쓰여 양 언어 간의 형식·의미상의 공통성은 보이지 않는다. 다만 복문에 있어서 전후의 문장을 연결짓는 구문적 기능이 공통된다. 본 발표에서는 각 용법을 간략히 소개만 하겠다.

① 「~であってみれば」

「~であってみれば」로 연결되는 전후 문장의 내용을 보면 두 문장을 인과관계로 연결짓고 있다. 종속문에는 사물·사건의 항상적인 특징이 제시되고 주문에는 그 특징이 원인이 되어 실현되는 사건이 제시된다.

- (35) 玉枝の叔母である佐木田もんというひき手婆は、本通りの弁天わきにある「かつらぎ楼」につとめていた。もとより、通行人をよびとめて、客ひきをするひき手婆がもんの仕事であってみれば、午すぎた時刻はまだつとめに出していない。(水上勉 『越前竹人形』)

② 「~하다 보면」

동사의 접속형인 「~하다(가)」에서 발달한 것으로 보이며 「~하다 보면」으로 고정된 형식이다. 「~하다 보면」에서 「보면」의 어휘적 의미는 표백되어 서술어로서는 기능하지 않고 그 전후의 문장을 연결짓는 접속조사로서 기능한다. 여기에서 「~하다 보면」은 종속문에 제시된 동작이나 활동이 원인이나 방법=<조건>이 되어 주문에 제시된 사건이 성립하는 것을 나타낸다(河2020: 174).

- (36) 내성적이란, 겉으로 나타내지 않고 혼자서만 속을 앓는다는 말이다. 이런 사람을 달래다 보면 이쪽이 지치고 만다.(전영우 『교양인의 화법』)
- (37) 현재 자신이 종사하고 있는 일이 자신의 마음에 들지 않고 희망하는 길과는 전혀 다른 분야라고 할지라도, 자신에게 주어진 일을 열심히 하다 보면 성공의 길에 이르는 수가 있습니다. (곽광택 『삶의 행복을 주는 114가지 지혜』)

2.3.3 <조사화>한 경우

여기서 예시하는 유형은 그 구문적 위치를 보자면 종속문이 아니라 연체수식을 받고 있는 단어에 「みれば」 「보면」이 붙어 있어 2.3.2의 경우와는 다르다.

형식의 구성을 보면 일본어의 경우 2.3.2의 접속조사화한 용법에서 서술한 「~であってみれば」와 같은 형식이다. 그러나 2.3.2의 「~であってみれば」가 서술어로 쓰이고 있지

만, 여기에서 말하는 「~であってみれば」는 서술어가 아니라 「~ならば」에 상당하는 조사로 쓰이고 있다. 한편 한국어의 경우 명사의 연용적 형식인 「~이고」에 「보면」이 후접하고 있다. 이 「~이고 보면」도 「~(이)라면」과 같이 전제를 제시하는 기능을 수행하고 있다(河2020: 175).

- (38) 修一郎は黙して答えず、四谷に帰って行った。懲役二年、執行猶予四の刑を受けた身であってみれば、学校を出ても就職はまず無理だった。 (立川正明『冬の旅』)
- (39) 맛선을 보고 서로가 호감을 가지게 되며 다시 만나게 된다.(중략) 더구나 결혼을 전제로 한 만남이고 보면 거기에는 기쁨과 즐거움이 생기게 되고, 애정이 싹트게 된다.(전영우 『교양인의 화법』)

2.3.4 <접속사화>한 용법

일본어와 한국어 모두 형식적으로는 지시표현을 수반한 「そうする」「그러하다」의 어형을 바꾼 「そうして」에 「みれば」가, 「그러고」에 「보면」이 후접한다. 다만 이 유형의 「そうして」「그러고」는 전형적인 지시표현의 경우와는 성격이 다르다.

예문 40, 41의 경우 「そうして」「그러고」는 선행하는 주체의 의지적 동작을 나타내고 「みれば」「보면」은 그 동작에 이어지는 시각활동을 나타낸다.

- (40) フロアシートにボールペンの跡がついてしまったら、液体ムヒできれいに取れます。ボールペンの跡に液体ムヒをぬった後ティッシュで拭きます。そして見れば、ボールペンの跡はきれいに取れているはずです。 (作例)
- (41) 장판에 불펜자국이 났을 때는 물파스로 말끔히 지울 수 있습니다. 불펜자국 위에 물파스를 살짝 바르고 나서 휴지로 닦습니다. 그리고 보면 불펜은 말끔히 지워져 있을 것입니다. (作例)

그러나 「みれば」「보면」이 동사성을 잃는 경우 「そうして」「그러고」는 의지적인 동작을 나타내지 않고 「みれば」「보면」의 어휘적 의미는 사고활동으로 추상화된다. 예를 보면 알 수 있듯이 「そうしてみれば」「그러고 보면」은 화자(작중인물)의 내리는 판단의 근거가 되는 문장과 그 근거로부터 내리는 판단을 서술하는 문장을 연결 짓는다.

- (42) これこそ楡基一郎独特の謀略、しかし彼にしてみればごく当然の、先見の智慧にあふれた決定なのであった。桃子には聖子の二の舞を踏ませてはならない。それにこの末娘は、どうもある方面が早熟で、しかもこれまでの一、二の事実が証明するようにけっこう男好きがするのではないか。そしてみれば基一郎が、世間一般の人間より遙かにすばやく反応する自分の指先を早手まわしにうごかしたのは当然である。 (北杜夫『楡家の人々』)
- (43) “위로 9대 이백 년을 살아온 집입니다.” / “그랬군요. 낡았지만 대단했어요. 요즘도 이처럼 산골인데-그 옛날에... 그리고 보면 인철씨네 꽤 명문 대가었던가 봐요.” (이문열 『변경 5』)

한편 한국어의 경우 형식적으로는 「~에서 보면」 형식으로 후치사화한 경우와 같지만, 선행하는 명사에 「이런」「그런」과 같은 지시표현이 수반되는 경우가 있다. 본 발표에서는 문장과 문장을 연결하는 점을 보고 접속사화한 경우로 분류한다.

- (44) 그러나 이러한 선거법 개정으로 중간계급의 대다수는 선거권을 가지게 되었지만 재산이 거의 없거나 적은 하층민들은 여전히 선거권이 없었다. 이런 면에서 보면 1832년 선거법 개정은 의회 개혁의 출발점에 불과했다.(안효상 『상식 밖의 세계사』)

(45) 그런데 이 세 가지 신화를 낳은 그리스인, 이탈리아인, 켈트인, 게르만인은 다 같이 인도유럽어 족에 속한다. 따라서 4천 년 전 중앙아시아에서 인도와 이란으로 이동한 아리아인과도 같은 뿌리를 지니고 있다. 그런 의미에서 보면 인도 신화와 이란의 페르시아 신화도 앞의 세 신화와 형제지간의 신화라고 할 수 있겠다. (강웅천 『바이킹전사들의 북유럽 신화여행』)

이렇게 보면 「みれば」 「보면」가 문장과 문장을 연결짓는 요인으로서의 우선 문장 레벨에서 두 사건을 연결짓는 구문적 기능에 바탕을 두고 있다고 볼 수 있다. 그리고 선행하는 문장을 나타내는 지시표현과의 조합이 텍스트 레벨에서 문장과 문장을 연결지을 수 있도록 도움을 주고 있는 것이 아닌가 보인다. 그렇다면 「みれば」와 「보면」의 접속화에는 문장 레벨에서 기능하는 형식이 텍스트레벨에서 기능하게 되는 기능적 확장이 수반된다.

Ⅲ. 이후의 과제

이후 조사대상으로 하는 어휘를 언어활동, 사고활동을 나타내는 「いう」「かんがえる」「말하다」「생각하다」로 확장시키고, 한편으로 어형도 여타 접속형으로 확장시켜 개별 언어에서 나타나는 문법화 양상을 확인할 필요가 있다.

<인용문헌>

- 백봉자 (2006) 『외국어로서의 한국어 문법사전』 하우기획출판
 —— (2019) 「脱動詞化にみられる文法的な意味と機能の存続—動詞条件形の脱動詞化を中心に—」 『日本研究』 51 中央大学校日本研究所 pp.113-131
 —— (2020) 「동사 ‘보다’의 조건형 ‘보면’의 문법화」 『언어와 정보 사회』 40 서강대학교 언어정보연구소 pp.149-181
 荒正子 (1983) 「から格の名詞と動詞とのくみあわせ」 言語学研究会編 (1983) 『日本語文法・連語論 (資料編)』 むぎ書房
 奥田靖雄 (2015) 「動詞論」 『奥田靖雄著作集4 言語学編3』 むぎ書房 pp.5-23
 グループジャマシイ (1998) 『教師と学習者のための日本語文型辞典』 くろしお出版
 言語学研究会編 (1983) 『日本語文法・連語論 (資料編)』 むぎ書房
 鈴木重幸 (1972) 『日本語文法・形態論』 むぎ書房
 高橋太郎 (2003) 『動詞九章』 ひつじ書房
 趙義成 (1995) 「現代朝鮮語の-에서格について」 『朝鮮學報』 150 朝鮮學會 pp.19-72
 陳真理子 (1996) 「現代朝鮮語の-로格について—單語結合論の観点から—」 『朝鮮學報』 160 朝鮮學會 pp.1-64
 松木重恵・森田良行 (1989) 『日本語表現文型—用例中心・複合辞の意味と用法』 アルク
 Hopper, P. and Trougott, E. C. 2003 *Grammaticalization 2nd Edition*. Cambridge University Press.



한국어와 일본어의 한자음 음절 대응관계

한국언어과학회 2021겨울 학술대회

2021.02.18. 목

허지애 (부산대)

목차

1. 연구 목적
2. 선행 연구
3. 연구 방법
4. 한국어와 일본어의 음절 구조
5. 한자음 음절 구조 대응
6. 한국어와 일본어의 한자음 음절표
7. 한자음의 음절 구조 제약
8. 한국어와 일본어의 한자음 대응의 기본 원리
9. 한국어와 일본어의 한자음 음절 대응

참고 문헌



1. 연구의 필요성 및 목적

- 한국어 학습자를 위한 한글 표기의 한자음 학습 전략 요구
- 현대 한국어 한자음과 현대 일본어 한자음의 음절 대응 관계 도출
- 한자음 대응 규칙을 활용한 한국어 한자음 및 한자어 교수법 개발 기대

2. 선행연구

샤우쥬나스-까찐스까스 마르띠나스(2011)

한자 4,316자

한자음의 자음을 대상으로 한중일 언어에서의 변화 양상 고찰

자음만으로 대상으로 하여 음절 대응 관계 도출에 한계

조광호(2013)

교육용 기초 한자 730자

한국어와 일본어의 한자음 대조 분석

자료의 한계

반충무(2016)

천자문의 1,00자

한중일 한자음을 공식적으로 비교 연구

나운연(2019)

한중일 상용 한자 4,262자

한중일 한자음을 통시적으로 정리하고 공식적으로 비교

중국어 한자음 기준으로 하여 한국어 한자음과 일본어 한자음의 직접적 대응관계 도출에 한계

3. 연구방법

자료: 郭錫良의 『漢字古音手冊(2014)』

13,000여자의 한자

중국 중고음을 기준으로 하여
현대 중국음, 한국음, 일본음을 IPA로 입력

한국어 한자음의 음절 기준으로 재 배열



4. 한국어와 일본어의 음절 구조



V=모음, C=자음, G=활음, ()=수의적



5. 한자음 음절 유형

현대한국어 한자음

음절 유형		보기	음절 유형		보기
1) V형	V	아[a]	3) VC형	VC	안[an]
	GV	요[jo]		GVC	용[jon], 엽[jɛp]
2) CV형	CV	고[ko]	4) CVC형	CVC	갈[kal]
	CGV	교[kjo]		CGVC	광[kwan]

현대한국어 한자음

음절 유형		보기	음절 유형		보기
1) V형	V	아[a]	3) VN형	VN	안[aN]
	GV	요[jo]		CVN	칸[kaN]
2) CV형	CV	고[ko]	4) CVN형	CGVN	칸[kjaN]
	CGV	교[kjo]		5) CV-CV형	CV-CV

6-1. 현대한국어 한자음 음절표

음절	ㅂ	ㅃ	ㅍ	ㅑ	ㅓ	ㅕ	ㄴ	ㄷ	ㄹ	ㅈ	ㅊ	ㅋ	ㆁ	ㅇ	개수
이	비	피	미		시	니	리	지	치	기	히	이			12
입		핍			십	닙	립	집	칩			입			7
일		빌	밀		실	빌	질	칠	길	힐		일			9
익		픽			식	닉	직	척	깅			익			7
임					심	림	짐	침	깅			임			7
인	빈	핀	민		신	린	진	침	깅			인			9
잉	빙						징	칭				잉			4
예					세		제	체	계			예			5
애	배	패	매	대	태	새	내	래	재	채	개	해	애		13
액	백	박	맥		택	색			채	객	핵	액			8
앵		퐁	멍		탱	생		랭	쟁	갱	행	앵			9
위					쉬				취	귀	휘	위			5
권											권				1
															9
															2
															2
															5
															5
우				득	득			죽	죽	죽	극	후			7
우					뵤			름	름	름	금	훈			7
우										름	금	훈			4
우				등	등	송	하	름	름	름	금	훈			9
우				터	터	섭		저	저	저	거	허			7
우	뵤				섭			접	접	접	겁	업			6

漢韓大辭典(2015, 단국대학교 동양학연구소)의 한자음 513개를
곽석량(2014)의 한자와 대조하고, 김무림 외(2017)을 참고하여
음절을 더하고 빼어 한자음 503개를 정리함.

6-2. 현대일본어 한자음 음절표

제1음절 \ 제2음절	o	い i	う u	ち chi	っ chu	き ki	く ku	ん N	개수
ア a	ア	アイ		アチ	アツ		アク	アン	6
カ ka	カ	カイ		カチ	カツ		カク	カン	6
ガ ga	ガ	ガイ		ガチ	ガツ		ガク	ガン	6
サ sa	サ	サイ		サチ	サツ		サク	サン	6
ザ za	ザ	ザイ		ザチ	ザツ		ザク	ザン	6
タ ta	タ	タイ		タチ	タツ		タク	タン	6
ダ da	ダ	ダイ		ダチ	ダツ		ダク	ダン	6
ナ na	ナ	ナイ		ナチ	ナツ		ナク	ナン	6
ハ ha	ハ	ハイ		ハチ	ハツ		ハク	ハン	6
バ ba	バ	バイ		バチ	バツ		バク	バン	6
マ ma	マ	マイ		マチ	マツ		マク	マン	6
ラ ra	ラ	ライ		ラチ	ラツ		ラク	ラン	6
イ i	イ							イン	6
キ ki	キ							キン	5
ギ gi	ギ							ギン	5
シ si	シ			シチ	シツ	シキ		シン	5
ジ ji	ジ			ジチ	ジツ	ジキ	ジク	ジン	6
チ chi	チ				チツ	チキ	チク	チン	5
ニ ni	ニ			ニチ			ニク	ニン	4
ヒ hi	ヒ			ヒチ	ヒツ	ヒキ		ヒン	5
ビ bi	ビ			ビチ	ビツ	ビキ		ビン	5
ミ mi	ミ			ミチ	ミツ	ミキ	ミク	ミン	6
リ ri	リ			リチ	リツ	リキ	リク	リン	6

『大漢和辭典』(1984, 諸橋轍次)의 한자음 393개를
곽석량(2014)의 한자와 대조하여 정리함.

7. 한자음 음절 구조 제약

현대일본어 한자음

음절 내부 제약

- 초성 제약
- 중성 제약
- 종성 제약

음절 결합 제약

- 초성과 중성의 결합 제약
- 중성과 종성의 결합 제약

현대일본어 한자음

음절 내부 제약

- 초성 제약
- 중성 제약
- 특수 음소

음절 결합 제약

- 반모음 결합 제약
- 특수 음소 결합 제약

음절 간 결합 제약

- 제2음절 제약

8. 한자음 대응의 기본 원리

음절 구조	한국어 한자음	한자	일본어 한자음
개음절	기본 음절 구조	폐음절 (CVC)	개음절 (CV)
	한자 1자	1음절	1음절 또는 2음절
폐음절	음절 구조 대응	C_1V_1 $C_1V_1C_2$	C_1V_1 $C_1V_1-C_2V_2$ (단, V_2 는 u와 i)
	학 [hak] $C_1V_1C_2$...	學	$C_1V_1-C_2V_2$ [ga-ku] ガク

9. 한국어와 일본어의 한자음 음절 대응

구분	한국어 한자음	일본어 한자음	
중성	특수음소로 교체	-ㅁ [m]	N
		-ㄴ [n]	
	모음 삽입	-ㄱ [k]	ku/ki
		-ㄹ [l]	tsu/tsi
	모음으로 교체	-ㅂ [p]	u
-ㅇ [ŋ]	u/i		
중성	-ㄷ [d]계	어 [ə]	jo
		여 [jə]	jo/ei
		워 [wə]	e
	-ㅌ [t]계	애 [ɛ], 왜 [wɛ]	ai
	-ㄷ [d]계	에 [e], 예 [je]	ei
	-ㄴ [n]계	이 [i], 의 [i], 위 [y], 으 [i]	i
	예외	외 [ø]	ai
웨 [we]		i	

참고문헌

1. 기본 자료

郭錫良(2014), 『漢字古音手冊』, 商務印書館.

檀國大學校 東洋學研究所(2015), 『漢韓大辭典』, 檀國大學校出版部.

諸橋轍次(1984), 『大漢和辭典』, 大修館書店.

2. 단행본 및 논문

김무림·김유범·최미현(2017), 현대국어 한자음, 한국문화사.

나운연(2019), 한·중·일 현대 한자음의 대응 양상에 대한 연구, 인하대학교 대학원 박사학위논문.

반충무(2016), 韓·中·日 『千字文』의 漢字音 比較 研究, 단국대학교 석사학위논문.

이경철(2013), 일본어 한자음의 이해, 책사랑.

조광호(2014), 일본인 학습자를 위한 한국어 한자 어휘 교육 전략, 한국외국어대학교 석사학위논문

沼本克明, 김정빈 역(2008), 한국인을 위한 일본어 한자음의 역사, 한국학술원. (원서출판 1986)

샤우쥬나스·까찐스까스·마르띠나스(2011), 한·중·일 한자의 발음 비교 연구, 경희대학교 대학원 박사학위 논문.

Plausible Cognate Nouns in Mongolian and Korean (몽골어와 한국어 사이의 공통어휘로 보이는 명사들)

Jae-Ick Park

(Kosin University)

I. Background

(1) Various ethnological and anthropological researches support that Korea and Mongolia once had close relationship and had similar languages. History books in China give a clue that the two nations once used the same (or similar) language (Choi 2005).

(2) Halha county (Халха Сум) and Buriat (Бурят) folktales contain an anecdote on Koreans' leaving for the south. One of the Korean tribes, called Solongos, started leaving the Mongolian territory at least 5000 years ago.

(3) Historical and classificatory linguistic studies have proposed that Korean and Mongolian are seemingly related but supporting linguistic data are not sufficient enough (Ramstedt 1949, 1954; Poppe 1965; Lee 1972; Starostin 2003; Jung 2008, 2012; Kang 2012; Lee 2018). Some of the available data have not been universally accepted by linguistics. Manchu data, which could serve as the bridge to such affinity, have been buried or almost disappeared along with the fall of the nation.

(4) There have been methodological problems in genetic study on Korean and its related languages. Lexicostatistics estimates the length of time since two or more languages diverged from a common proto-language. Glottochronology has the assumption of a constant rate of change for basic elements in the language.

(5) Constant rate of change has been rejected by most linguists, and individual words or elements have their own rate of change and their own history (Starostin 2002).

(6) Bergsland & Vogt (1962), the authors make an impressive demonstration, on the basis of actual language data verifiable by extra-linguistic sources, that the "rate of change" for Icelandic constituted around 4% per millennium, whereas for

closely connected Riksmål (Literary Norwegian) it would amount to as much as 20%. (Morris Swadesh' s original proposal on "constant rate" was around 14% per millennium).

(7) Moreover, individual items on the 100 wordlist have different stability rates; for instance, 'mother' generally has a much lower chance of being replaced than 'black'.

(8) Over 70% of Mongolian vocabulary is native, while less than 30% of Korean vocabulary in general use is native. Is it fair to apply the same divergence rate per millennium?

(9) Mongolian, like German, makes full use of native resources (elements) in producing new words, but Korean, like English, easily borrows words from other languages (such as Chinese) or combine Chinese elements for its needs.

(10) Due to the absolutely long period of time of separation and the differences in sociocultural and political events integrated in the two nations, the rate of change (or retention) of words in Korean and Mongolian can be very different. We can apply any surviving native Korean elements to the Mongolian wordstock, but not vice versa to attain any cognates showing statistically meaningful relationship.

2. Aims of the Study

(11) While ignoring glottochronology/lexicostatistics and results from previous research, this study aims to provide plausible cognate nouns between the two languages in order to engage in the study on language classification in North–East Asian languages.

(12) The reason for ignoring the previous data is to search for plausible cognates without being influenced by other data, which could be correct or incorrect.

(13) The current data has mostly been introduced on the basis of current sound and meaning correspondence extracted from general Mongolian learning textbooks, Mongolian–Korean/Korean–Mongolian dictionaries, Mongolian explanatory dictionary, personal dialogues with Mongolians, and the Bible. A few items are provided with some explanatory remarks in reference to the literature on old forms or forms in other languages (홍문각HongMunGag 1790; 박재연Park 2016; Tadmor et al. 2010; Tsevel 2013; Suld–Erdene et al. 2018).

3. Plausible Cognates

3.1. Words for Body Parts

No.	English Meaning	Mongol Cyrillic	Mongolian Sound ¹⁾	Korean Sound	Korean Alphabet	Remarks
1	eye	нүд	nud	nun/*nidü	눈	고어 *니뒤
2	nose	хамар	hamar	koh	코	고ㅎ(고어)
3	ear	чих	chi	kwi	귀	
4	leg, foot	хөл	həl	həri, hıl	허티, 흙뒤	허티, 허뒤(고어'다리'), 흙뒤(고어 '아킬레스건'), '흙뒤'로 발음. 사잇소리 기는 소리 안남.
5	tongue	хэл	hel	hyə	혀	
6	rib	хавирга	havirag	kalbi	갈비	
7	hand	гар	gar	son, son karak	(손)가락	손, 가락, 팔
8	finger	хуруу	horoo	horong	호롱	손가락,(낙지)호롱
9	head	толгой	təlgəi	məri, təgari, təgol	대가리	머리
10	tent (yurt) top cover	дээвэр	deever→ degeburi	degubari	대구바리	꼭대기, 머리, 대구박? 머릿대골(훈몽자회), 몽:yurt 꼭대기 덮는 천
11	hit with the head	мөргөх	mərgəh	məri	머리	머리로 부딪히다
12	mouth	ам	am	agami	아가미	아금(강원도), 아가리, 물다, 아물다, 다물다
13	outfit, appearance	царай	tsarai	charim	차림	용모, 옷차림
14	face	нүүр	nuur	nuri	누리	누리, 얼굴, 세상
15	face, mask	тал	tal	tal	탈	체면, 누르탈, 박, 탈바가지

1) The phonemic description does not reflect the correct pronunciation neither in Mongolian nor in Korean. There is vowel fronting in Mongolian when followed by a front vowel or glide. Korean has no distinction between voiced and voiceless obstruents unless they are tense or aspirated.

3.2. Words for Human Beings

No.	English Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	enemy	дайн	daing	toi	되	원수, 적, 되蠻(훈몽자회), 되夷, 되胡
2	father	аав	aab	abai	아바이	아바이, 아버지
3	grandpa	өвөө	эвээ	эбэi	어버이	한-아버지, 할배, 할바이
4	grandma	эмээ	emeе	эmei	어매	한-어미, 할미, 할매, 할마이
5	child	ахай	ahai	ahai	아이, 아해	Ахай, миний хүү тэгээрэй! 애야, 그렇게 하려므나!
6	adult	эрчүүд, эрүүхчд	eru	эrim	어른	어른, 어른, 어로신내, 어루신, 어른사람
7	hunting	ан	ang	sanyang	사냥	san-hæng 山行, <san-yang? 'mount hunting?' Анд авсан. 사냥 갔다.
8	hunter	анчин	angching	san-ong-b achi	산-웅-바 치	사냥바치, 사냥꾼
9	secretary recorder	бичээч	bicheech	turubich	두루빛	총무=두루두루 글 쓰는 사람
10	small animal, kid	ишиг	ishig	isæk'i	이새끼, 삿기	새끼염소, 어린아이, 꼬마
11	man	хүн	hun	k'un	꾼	일꾼, 짐꾼, 농사꾼, 장사꾼

3.3. Words for Time and Direction

No.	English Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	side, wide area	тал	tal	tal	달, 들	양달, 들, 쪽, 편
2	left	зүүн	zuun	oin	왼	왼쪽
3	right	өрнө	erən	orin	오른	오른쪽
4	west	баруун	baroon	barin	바른	바른쪽
5	side, direction	зүг	zuk	c'ok	쪽	어느 쪽으로 'heden zukeer'
6	outside	гадна, гадаад	gadən	kət	겉	겉, 갓(경상), 밖, 고어'겉'
7	distance, gap	зай	zai	sai	사이	사이, 거리, 간격, 자리
8	this time, recently	сая	saya	sai	사이, 새	요새, 금새, 사이, 방금 Бисаяирлээ. 'I've just arrived.'
9	the time when	цаг	tsag	cak	적, 작	적, 고어'작'(가례9:43a)
10	on, about, time	дээр	deer	t'ai	때	대로, 위, 때
11	day, sun	нар	nar	nal	날	해, 하루
12	month, new year day	сар	sar	sari	사리	사리 '그믐달이나 보름달'
13	late time, evening	орой	oroi	oræ	오래	오래되다 '밤늦어지다', 어스름(고어: 첫어웁)
14	then, that time, short span	төд	təd	təd	덧	덧없다, 곧장, 즉각
15	back, north	хойт	hoit	huiduri	휘두리	휘두리, 뒤
16	front	өвөр	ävər	ap, alp	앞	앞(시간상)
17	front, south	урд	ord	ap, alp	앞, 앞	고어 '앞에'
18	corner	муж	moj	mo	모	모서리

3.4. Words for Animals

No.	English Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	horse	морь	mər	mal	말, 몰	몰(제주, 진해 등)
2	pony	жороо морь	ʃɔɾɔɔ məri	jorangmal	조랑말	차용어?
3	spotted horse	айг морь	alag məri	əlugmal	얼룩말	차용어?
4	cow	үнээ	unee	ungchi	웅치	웅치(함경도말, 심마니말)
5	ox	үхэр	une	(n)yuch	웃	쇼, 슈, 중, 뇨, 뇨(여러 방언)
6	sheep	хурга	horga	kəl	걸	도개걸웃모
7	goat	ямаа	yamaa	yəm/yam	염(소)	염생이, 염생이
8	goat	*Khitan small letter	im.a	yəm/yam	염(소)	Khitan Letters (Liao Dynasty 916-1125)염, 압(의성어?)
9	antelope	буга	bog	pok	복작이	цаабуга[tsaabog] '순록', 몽어유해2:32b 사슴 '보구'
10	piglet	тоорой	tɔɔri	to, tos	도, 돛	도개걸웃모, 돛, 도야지
11	cf. wolf in Evenki		irgichi	iri	이리?	어윙커에 11가지 늑대이름 중 하나
12	jackal	цөөвөр чоно	tsəəvɔɾ chɔnɔ	singyangi	승냥이?	승냥이~늑대?
13	dog	нохой	nɔhɔi	kahi	가히	개, 가히
14	tiger	бар	bar	pari	바리	납닥바리, 두루바리, 도루바리, 범파리
15	hare	туулай	toolai	thok'i	토끼	한자가 나중?
16	cuckoo	хөөхөө	həəhəə	p'ək'ugi	삐꾸새	삐꾸, 의성어?
17	crow	хэрээ	heree	kamagoi	가마귀	까마귀
18	caw	*Khitan small letter	ga.g	k'ak	각	각각(의성어)
19	hen	тахиа	tahia	talk	닭	달구벌, 달구통
20	hen	*Khitan small letter	t(e).q(o).a	talk	닭	달구새끼
21	cf. hen cackles, coos	гооглох, гоог гоог	gɔɔglɔh, gɔɔg gɔɔg	kugu, goggog	구구, 곡곡	암탉이 알 낳으려는 소리
22	owl	ууль	ool	olp'æmi	올빼미	올빼미
23	tail	сүүл	suul	sool	솔	꼬리, 털북, 솔
24	swan	хун	hon	koni	고니	몽어유해2:28a 곤이=훈, Cygnus (학명) 차용어?
25	fat	хос	hɔs	kob	곱	곱창

3.5. Words for Nature

No.	English Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	rain	бороо	bərəə	pora	(비)보라	차용어?
2	plain, mount	тал	tal	tal, til	달, 들	아사달, 들
3	refuse	хөрмөг	hərməg	kərim	거름	나뭇잎 쌓인 것
4	earthworm	хорхой	hərhoi	k'ək'ængi	꺼쟁이	지렁이, 꺼꾸랭이, 거취이, 꺼싱이, 거시랑, 끄시랑이
5	fence	хүрээ	huree	(k)uul	울	울타리, 굴, 구레(전라지명)
6	river	гол	gəl	kol	골	골, 가라골(강변동네 이름), 개울, 거랑, 가람
7	water	амурь амуть амү	amu	mul	물	강, 호수, 물
8	bamboo	хул	gol	kol	왕골	대나무
9	soil, earth	хөрс	hərs	hir	흙	고어hir(g),kir
10	reed	зэгс	zags	sasg	삿, 갈대	삿광주리, 삿갓, 삿갓, 삿자리, 산
11	reed	гал	gal	kal	갈(대)	불, 갈색
12	valley	гуу жалга	goo jalak	kol ch'gi	골짜기	
13	slag, waste	шаар	shaar	sireki	쓰레기?	재, 찌끼
14	straw, waste	сүрэл	surel	siregi	쓰레기?	티끌, 초개, 쓰레기?
15	whirlwind	хуй салин	hoi	hoi	회	회오리바람, 휘두르다
16	acacia	хуайс	hoais	kasi	가시	까시, 가시나무
17	wood	*goli 'log, block'	Starostin, 통나무, 그루?	kiru, k'il	그루, 끌	끌 (심마니말)
18	oak	царс	tsars	cham	참	참나무, 참새, 찰흙, 참쌀, 차나락(찰벼)
19	sticky rice	цавуулаг	tsavoolag	chaps'al	찰쌀	차쌀, 차나락(찰벼), 찰지다, 찰흙 tsavoo
20	wild grape	вил	mæl?	məru	머루	머르, 멀위
21	cloud	үүл *үгүл	ugul	kurIm	구름	Khitan small script [eu.ul]
22	mugwort	쉬하, 시랄지	shiha	(p)suk	쑥	부속(18C), 시랄지 (몽어유해2:3a) шарилж /sharilj/
23	mallow	아부가	abuga	aok	아욱	아욱, 아욱; 아복:아부가 (몽어유해2:3a)
24	monos fruit	мойл	moil	məru	머루	머뤼:뫼(몽어유해2:4b)

3.6. Suffixes for Nouns

No.	English Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	province	аймаг	aimag	mak	막	내리막, 오르막, 마지막
2	the firstborn	ахмаг	ahmag	mak	막	늘그막, 느즈막, 부뚜막, 두루막
3	cheating	хуурмаг	hoormag	mak	막	토막, 짤막, 올막줄막, 가막가막 ('깜빡깜빡' 제주)
4	piece, bit	үйрмэг	uirmeg	mak	막	조각, 부스러기
5	doctor	эмч	emchi	chi	치, 지	갈치, 장사치, 누룽지, 콩지
6	bow maker	нүмавач	numavachi	bachi	바치, 지	활바치, 갓바치, 강아지
7	herder	малчин	malching	jangi	쟁이, 장이	말쟁이, 소금쟁이, 뚜쟁이, 점쟁이
8	compose →composite	нийлэх→ нийлэг	niileh →niileg	V-k	악, 억, 옥, 옥	까막, 주먹, 부엌(부적), 비느럭 ('비늘'방언), 가죽
9	behappy →picnic	зулгаалах → зулгаалга	zolgaalah→ zolgaalar	V-k	악, 억, 옥, 옥	까막, 주먹, 부엌(부적), 비느럭 ('비늘'방언), 가죽
10	compete →competition	тэмцээх→тэ эмцээн	temtseeh→t emtseeng	V-ng	앙, 엉, 옹, 옹	마당, 도랑, 거랑, 고랑, 벼랑,
11	aged→elder	настай→нас тан	nastai→nast ang	V-ng	앙, 엉, 옹, 옹	이엉, 구멍, 지붕, 미중

3.7. Nouns for Instruments and Things

No.	English Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	castle, city	хот	hot	hol, hot	홀	홀, 잣, 성, 시
2	umbrella	шүхэр	shuher	shurub	슈룹	차용어?
3	red	хүлан, улаан	holang, olaang	ora	오라	오랏줄 '죄수 묶는 붉은 줄'
4	flour	гурил	goril	karu	가루	고어: 가라[kara](아래아)

5	bridge	гүүл	guul	kuul	굴	굴다리, 다리
6	animal feed, food?	хоол	hool	k'ol	꿀	먹이
7	border	хил	hil	gil	경계, 길?	경계, 국경, 지역경계가 끝 길?
8	wood stick	시담	s(h)idam	sirang	(쇠)스랑	serang 'long stick' Evenki, 시담 (몽어유해 1:37a 몽둥이, 목봉(木棒))
9	trident, harrow?	서러 сэрээ	seree	sirang	써레	서러(몽어유해 1:37a 삼지창), 써레?
10	table, shelf	ширээ	shiree	sirəng	시렁?	시렁(뭘 올려놓는 곳)
11	sharpening/metal file	хуурай/ зүлгүүр → зүлгэх	n. hoorai/ zulguur v. zulgeh	juul, s'ilh	줄 쏘다	줄로 쏘다, 실겘다(경상)
12	knitting hook	дэгээ	degee	t'ige	뜨개질?	코바늘, 고리, 갈고리, *gòlí 'hook' Starostin(2003)
13	spear	жаг	jag	cakdu	작두?	창날과 작두날이 비슷?
14	prison	шорнд	shərnd	corong	조롱	새조롱, 감옥, 철장
15	clothes	хувцас	hovtsas	kojang, gos	고장, 곳	속곳, 옷, 꼬장주
16	city wall, cliff	зан авир	zang avir	pyər	벼랑	비알, 비탈, 비르, 성벽, 잣, 비루박
17	wine	дарс	dars	t'ar	따르다?	술/물을 따르다
18	sack	уут	oot	ot	옷	옷, 밀가루옷, 봉투, 베갯잇
19	wrapping cloth	бось	bos	po	보	붓짐, 옷감, 천
20	bronze	хүрил	huril	kuri	구리	구리, 차용어?
21	bronze, brass	Hebrew?		nos	놋	cf. Nehushtan 느후스단 (열왕기하18:4)
22	carpet, cloth	хивс	hivs	kip	깁	비단옷감, 깁다?(silk?)
23	flint	цайур	tsaior	cha	차	차돌, 부싷돌
24	flint	хэт	het	hyə	혀다?	부싷돌, 켜다
25	lock	цоох	tsəoh	cha	차다	차다, 채우다, 자물쇠, 시계 차다
26	ring	гархи	garhi	kori	고리	고리, 골회, 가락지, *gòlí'hook' Starostin(2003)

3.8. Suffixes for Abstract Nouns

No.	Meaning	Mongol Cyrillic	Mongolian Sound	Korean Sound	Korean Alphabet	Remarks
1	form, style	хэв	hev	kəb	껍	껍, 껍데기, 겹, 거품
2	accusation, spirit?	ял	yal	əl	얼	얼, 죄얼(레위기4:4)
3	sin	нүгэл	nugel	nu	누	누를 끼치다
4	times, order	дахь, дэх	dəh, deh	tiwi	디위	디위, 첫째, 둘째
5	allotment, times	хувь	hov	kob	곶	곶, 호부, 곶, 고봉, гурванхувьхий '세곶(배) 만들어라'
6	revenge, fire	хал	hal	kal	갈	원한, 칼품다
7	power	аар	aag	ak	악	힘, 악쓰다
8	if, in case	гаруй	garoi	karyung	가령	가령, 명사, 접속사?
9	circle, perimeter	тойрог	tōirəg	tule	둘레	둘레, 두르다, 주위
10	help, assistance	дэм	dem	təm	덤	덤? 도움
11	work, job	үйл	uil	zil	질	일, 고어 '질'반치음
12	work, job	ажйл	ajil	zil	질	일, 고어 '질'반치음
13	when	хэжээ	hejee	ənje	언제	
14	what	хэдэн	heden	ət'ən	어떤	어떤, 무슨, 몇
15	who, the other	нөгөө	nəgəə	nuguu	누구	누구, 다른
16	name	нэр	ner	nirim	니름	니르다
17	name	*Khitan small letter	i.ri	irim	이름	이름, Khitan (Liao Dynasty 916-1125)

4. Conclusion

(14) There should be chances to find out more cognate nouns between Korean and Mongolian if one scrutinizes every element of nouns of past and present time.

(15) The study may extend its scope in a later stage to various aspects of the two languages. The findings in this study provides the foundation for further study in verbs and other grammatical categories. It may contribute to add more cognate to Sergei Starostin et al.'s data (2003), which is an extensive record of the language family.

(16) Further study of minute aspects of the two languages will contribute to the genetic classification of Korean as a language closely related to Mongol–Tungus languages.

(17) Students and researchers from both languages must cooperate for a better result from every element of real life.

[참고문헌]

- 강길운(Kang, K. U.). 2012. 「한국어계통론 I-II」 (Korean Genealogy). 한국문화사 (Seoul: Hankuk Munhwasa).
- 강신 외(Kang, S. et al.). 2009. 현대몽골어와 한국어의 문법 비교연구」 (Contrastive Study of Modern Mongolian and Korean Grammar). 서울: 한국문화사 (Seoul: Hankuk Munhwasa).
- 김기선(Kim, K. S.) 2009. 몽골어학: 몽골어와 한국어의 의성어, 의태어 대조 연구. (A Comparative study on Onomatopoeia between Korean and Mongolian Languages). 「몽골학」 (Mongol Studies) 26, 37-63.
- 동북아역사재단 북방사연구소 편(Northeastern Asian History Center, ed.). 2018. 「고조선의 언어계통 연구」 (The Study of Old-Chosun Language Classification). 서울: 동북아역사재단 (Seoul: Northeastern Asian History Center).
- 박재연 외(Park, J. H. et al.). 2016. 「고어대사전 (1-21)」 (Comprehensive Old Korean Dictionary 1-21). 서울: 선문대학교출판부 (Seoul: Sunmoon U Press)
- 성백인(Sung, B. I.). 1999. 「만주어와 알타이어학연구」 (The Study of Manchu and Altaic Languages). 서울: 태학사 (Seoul: Taehaksa)
- 엄순천(Eom, S. C.). 2018. 양웅의 방언에 수록된 고조선 지역 단어들과 어원키어 비교 (A Comparative Study of the Old Chosun and Evenki Words in Yangxiong's *Fangyan*). 「고조선의 언어계통 연구」 (The Study of Old-Chosun Language Classification). 서울: 동북아역사재단 (Seoul: Northeastern Asian History Center). 275-302.
- 유원수(Yu, W. S.). 2013. 현대 할하몽골어의 현대 한국어 차용 소고 (Some Remarks on Modern Korean Loan Words in Modern Khalkha Mongolian). 「알타이학보」 (Altaic Hakpo) 23, 111-126.
- 이기문(Lee, K. M.). 1972. 「개정 국어사개설」 (Revised Korean Language History). 서울: 민중서관 (Seoul: Minjungseokwan).

- 이성규(Lee, S. G.). 2018. 고조선어와 몽골어 비교연구 (A Comparative Study of the Old Chosun and Mongolian Words in Yangxiong`s *Fangyan*), 「고조선의 언어계통 연구」(The Study of Old-Chosun Language Classification). 서울: 동북아역사재단 (Seoul: Northeastern Asian History Center), 205-236.
- 정광 외 편(Jung, G. et al.). 2008. 「한국어와 알타이어 비교어휘 I」(A Comparative Vocabulary for Korean and Altaic Languages I). 서울: 제이앤씨 (Seoul: JNC).
- 정광 외 편(Jung, G. et al.). 2012. 「한국어와 알타이어 비교어휘 II」(A Comparative Vocabulary for Korean and Altaic Languages II). 서울: 지식과교양 (Seoul: Jisik kwa Gyoyang).
- 제주특별자치도(Jeju Province Office). 2009. 「개정증보 제주어사전」(Revised and Expanded Jeju Dictionary). 제주: 제주특별자치도 (Jeju: Jeju Province Office).
- 최기호(Choi, K. H.). 2005. 「한국어와 몽골어의 계통적 위치」(Genetic Linguistic Status of Korean and Mongolian). 「청람어문교육」(Chungnam Language and Literature Education) 32, 265-295.
- 홍문각(Hong Mun Kak). 1790. 「몽어유해」(MongEoYuHae: An Explanatory Dictionary of Mongolian Words). 서울: 서울대학교 규장각도서 3751호. (Seoul: SNU Kyujanggak Library Book Number 3751).
- Bergsland, K. and H. Vogt. 1962. On the Validity of Glottochronology. *Current Anthropology* 3, 115-153.
- Chuluunpurev, S. et al. 2013. *Old Mongolian Dictionary*. Mongolia: Munsudar Press.
- Poppe, N. 1965. *Introduction to Altaic Linguistics*. Wiesbaden: O. Harrassowitz; Ural-altaische Bibliothek.
- Poppe, N. 1987. *Introduction to Mongolian Comparative Studies*. Helsinki: Suomalais-Ugrilainen Seuran Toimituksia.
- Ramstedt, G. and P. Aalto. 1949. *Studies in Korean Etymology*. Helsinki: Suomalais-Ugrilainen Seuran Toimituksia.
- Ramstedt, G. and P. Aalto. 1954. *Additional Korean Etymologies*. Helsinki: Suomalais-Ugrilainen Seuran Toimituksia.
- Starostin, S. 2002. Methodology of Long-Range Comparison. Manuscript.
- Starostin, S., A. Dybo and O. Mudrak. 2003. *Etymological Dictionary of the Altaic Languages*. Leiden: Brill.
- Suld-Erdene, G et al. 2018. *Native Mongolian Dictionary*. Mongolia: Munsudar Press.
- Swadesh, M. 1971. *The Origin and Diversification of Language*. Chicago: Aldine.
- Tadmor, U., M. Haspelmath and B. Taylor. 2010. Borrowability and the Notion of Basic Vocabulary. In S. Wichmann and A. Grant (eds.) *Quantitative Approaches to Linguistic Diversity*. John Benjamins Publishing Company. 35-55.
- Tsevel, Y. 2013. *Mongolian Explanatory Dictionary*. Ulaanbaatar, Mongolia: International Ulaanbaatar University.

Park, Jae-Ick
 Department of English
 Kosin University
 Busan, Korea 49104
 Wisdom Hall 2101
 051-990-2298, 010-9422-2693
 jipark@kosin.ac.kr

A New Approach to Keyword Analysis: focusing on Old and New Testament

2021. 2. 18
하 명 호 (신라대학교)

Contents

1. Introduction
 2. Literature review
 3. Methodology
 4. Analysis results and Discussion
 5. Conclusion
- References

1. Introduction

- Scott (1997: 236) proposed a simple yet important definition for keywords: words that occur 'with unusual frequency' in a target corpus when compared with a reference corpus.
- The keywords that occur when two corpora are compared together should reveal the most significant lexical differences between them, in terms of aboutness and style (Baker, 2004: 347).

1. Introduction

- According to Egbert & Biber (2019), keyword analysis has become an indispensable tool for discourse analysts, being applied to identify the words that are especially characteristic of the texts in a target discourse domain.
- Keyword analysis is one of the most widely used methods in corpus linguistics and corpus assisted discourse studies (CADS).

1. Introduction

- This study has the following goals.
 - Compare word lists of NEWT and OLDT
 - Compare frequency-based keywords of NEWT and OLDT
 - Compare text dispersion keywords of NEWT and OLDT
 - Compare frequency-based and Text dispersion keywords of NLT (Positive/Negative)
 - Compare GSL/AWL and text dispersion keywords of NLT

2. Literature review

- The traditional, frequency-based method disregards text dispersion.
- Keywords are usually ranked according to the keyness statistic, with the assumption that words with higher keyness values are more strongly associated with the target corpus.
- Egbert & Biber (2019: 78) refer to this approach as 'corpus frequency keyness / keywords'.

2. Literature review

- Corpus frequency keywords can be, and often are, frequent in a corpus, but are not widely dispersed across the texts of that corpus. As a result, such words are not truly typical of the discourse domain represented by the corpus.
- Some scholars have identified serious limitations with the corpus frequency approach (Johnson and Ensslin, 2006; and Baker, 2004, 2010).

2. Literature review

- Egbert & Biber (2019) propose that the limitations of corpus frequency keyness, namely the lack of content-distinctiveness and content-generalisability, can be overcome by accounting for text dispersion.
- Text dispersion keyness is much more effective at capturing words that are more dispersed in the target corpus relative to the reference corpus.

2. Literature review

- A few potential limitations of text dispersion keyness.
 - Text dispersion keyness does not account for token frequency in any way.
 - It is necessary for corpora to be collected and organized in the form of texts.
 - Text dispersion keyness is heavily biased in favor of including content words and excluding grammatical (i.e., function) words in keyword lists.

3. Methodology

- This study uses Scott's (2020) WordSmith 8.0 to analyze keywords of Old and New Testament.

Table 1. Specifications of NEWT and OLD T

Corpus		Tokens (running words) in text	Types (distinct words)	Type/Token ratio (TTR)	Standardized TTR
Target	NEWT	191,292	6,524	3.41	34.95
	OLD T	569,109	12,230	2.15	34.15
Reference	BNC Baby-Aca	1,014,027	33,532	3.37	40.21

4. Analysis results and Discussion

Table 2. Word lists of NEWT and OLDT

N	NEWT			OLDT		
	Word	Texts	Freq.	Word	Texts	Freq.
1	ME	27	11,310	THE	39	43,627
2	AND	27	7,094	AND	39	21,125
3	TO	27	5,783	OF	39	19,522
4	OF	27	4,621	TO	39	14,813
5	YOU	27	4,537	YOU	39	10,542
6	IN	27	2,688	WILL	39	9,086
7	HE	27	2,589	IN	39	7,687
8	I	27	2,413	LORD	36	7,287
9	A	27	2,410	I	39	7,160
10	FOR	27	2,319	A	39	7,028
11	GOD	27	2,084	FOR	39	5,979
12	THEY	26	2,038	HE	37	5,890
13	WILL	27	1,979	YOUR	39	5,460
14	IS	27	1,954	THEY	39	5,247
15	THAT	27	1,901	WITH	39	4,441
16	JESUS	26	1,692	HIS	38	4,357
17	WHO	26	1,675	IS	39	4,184
18	WAS	25	1,528	FROM	39	4,177
19	BUT	27	1,524	ALL	39	3,989
20	HIM	26	1,517	THEM	39	3,772

4. Analysis results and Discussion

Table 3. Frequency-based keywords of NEWT and OLDT

N	NEWT			OLDT		
	Keywords	Texts	BIC	Keywords	Texts	BIC
1	YOU	27	10,138.05	LORD	36	13,783.04
2	GOD	27	7,353.52	YOU	39	13,435.68
3	JESUS	26	6,214.92	WILL	39	9,219.38
4	I	27	3,706.71	YOUR	39	7,865.57
5	HIM	26	3,690.26	I	39	7,193.06
6	HE	27	3,658.11	MY	39	5,835.94
7	ME	22	2,441.15	GOD	36	5,682.70
8	WHO	26	2,343.50	ME	37	5,302.39
9	WILL	27	2,178.39	ISRAEL	35	4,685.69
10	OUR	27	2,121.49	HE	37	4,512.54
11	CHRIST	23	2,029.42	KING	35	4,401.60
12	LORD	24	1,847.87	THEM	39	3,264.06
13	THEY	26	1,727.63	PEOPLE	38	3,124.20
14	DON	25	1,720.68	HIM	36	3,014.61
15	SAID	18	1,637.90	SON	34	2,815.73
16	MY	25	1,481.62	HIS	38	2,765.30
17	THEM	26	1,415.76	THEY	39	2,584.44
18	HIS	25	1,373.50	LAND	38	2,203.35
19	SON	23	1,340.10	ALL	39	2,178.57
20	FATHER	26	1,193.83	DAVID	19	1,920.42

4. Analysis results and Discussion

Table 4. Text dispersion keywords of NEWT and OLDT

N	NEWT			OLDT		
	Keywords	Texts	BIC	Keywords	Texts	BIC
1	JESUS	26	34.81	JUDAH	33	33.42
2	DON	25	26.12	SIN	28	27.72
3	PRAY	20	25.85	DON	34	27.15
4	SIN	20	25.85	ARMIES	27	26.58
5	GREETINGS	20	25.85	GODS	27	26.58
6	HOLY	23	23.30	WILDERNESS	26	25.43
7	BELIEVERS	18	22.86	SINNED	25	24.29
8	JOY	22	21.89	COVENANT	25	24.29
9	RIGHTEOUS	17	21.36	HOLY	31	23.91
10	GRACE	21	20.49	ISRAEL	35	23.47
11	RIGHTEOUSNESS	15	18.37	DIDN	24	23.15
12	GLORY	22	17.63	HEAVENS	24	23.15
13	DOESN	14	16.88	GOATS	24	23.15
14	DIDN	14	16.88	FLOCKS	24	23.15
15	AMEN	14	16.88	SWORDS	24	23.15
16	HEARTS	21	16.32	MOAB	23	22.01
17	SINS	21	16.32	POUR	23	22.01
18	HONOR	18	16.31	HONOR	29	21.75
19	CHRIST	23	15.58	VINEYARDS	22	20.87
20	SINFUL	13	15.38	EDOM	22	20.87

4. Analysis results and Discussion

Table 5. Frequency-based and Text dispersion keywords of NLT (Positive)

N	Frequency-based Keywords			Text dispersion Keywords		
	Keywords	Texts	BIC	Keywords	Texts	BIC
1	YOU	66	16,696.63	DON	59	31.80
2	LORD	60	12,375.75	SIN	48	31.41
3	WILL	66	9,091.14	HOLY	54	28.23
4	I	66	8,224.95	DIDN	38	23.91
5	GOD	63	8,157.89	JOY	47	23.26
6	YOUR	66	7,764.89	HONOR	47	23.26
7	HE	64	5,934.05	JUDAH	37	23.16
8	MY	64	5,682.33	PRAY	37	23.16
9	ME	59	5,677.28	RIGHTEOUS	35	21.66
10	HIM	62	4,436.49	COVENANT	35	21.66
11	ISRAEL	49	4,042.34	HEAVENS	34	20.91
12	KING	47	3,710.07	WILDERNESS	34	20.91
13	THEM	65	3,621.74	SINNED	34	20.91
14	THEY	65	3,274.84	SINS	50	20.60
15	PEOPLE	65	3,213.04	PROPHET	33	20.17
16	HIS	63	3,210.44	BLESSED	33	20.17
17	SON	57	3,007.83	ISRAEL	49	19.93
18	JESUS	26	2,853.17	ARMIES	32	19.42
19	WHO	64	2,812.84	DOESN	32	19.42
20	SAID	53	2,279.55	GODS	31	18.67

4. Analysis results and Discussion

Table 6. Frequency-based and Text dispersion keywords of NLT (Negative)

N	Frequency-based Keywords			Text dispersion Keywords		
	Keywords	Texts	BIC	Keywords	Texts	BIC
1	HOWEVER	27	745.13	DESCRIBED	4	37.40
2	HISTORY	14	758.60	DEVELOPED	4	37.40
3	FORM	16	822.89	PROBABLY	4	37.40
4	BETWEEN	41	852.70	USEFUL	5	38.25
5	<i>PARTICULAR</i>	3	855.34	BASIC	3	38.88
6	DIFFERENT	24	859.88	INDEPENDENT	3	38.88
7	EXAMPLE	31	877.33	COMPLEX	3	38.88
8	<i>GENERAL</i>	5	887.68	FAMILIAR	3	38.88
9	MORE	60	907.68	BASIS	3	38.88
10	OR	62	1,058.81	WIDELY	3	38.88
11	SUCH	54	1,142.62	REASONS	3	38.88
12	THAT	66	1,227.75	STRUCTURE	4	39.46
13	BY	64	1,331.90	SIGNIFICANCE	4	39.46
14	AN	62	1,407.04	<i>GENERAL</i>	5	40.26
15	OF	66	1,583.10	INTRODUCED	3	40.99
16	A	66	2,622.51	DIFFICULTY	3	40.99
17	IN	66	2,818.06	ATTEMPT	3	43.11
18	IS	66	2,831.48	SIMILARLY	3	43.11
19	WHICH	55	4,050.48	SYSTEM	3	45.23
20	#	34	12,128.19	<i>PARTICULAR</i>	3	45.23

4. Analysis results and Discussion

- 3,442 frequency-based keywords of NLT (Positive: 1,617, Negative: 1,825)
- 2,056 text dispersion keywords of NLT (Positive: 524, Negative: 1,532)

4. Analysis results and Discussion

- Currently, there are two word lists widely used in academic research; the General Service List (GSL) created by West (1953) and the Academic Word List (AWL) developed by Coxhead (2000).
- Both of them have been a cornerstone in the field of corpus linguistics to explore the lexical frequency and distribution of word lists in various registers (Chanasattru & Tangkiengsirisin, 2017).
- However, these two word lists have a limitation. GSL developed over 50 years ago is out of date and shows the imbalance between the 1st GSL and the 2nd GSL.

4. Analysis results and Discussion

- AWL contains too general or common words that do not meet the specialization in specific academic fields.
- Gilmore & Millar (2018: 2) point out that “items on the AWL often vary across disciplines in terms of range, frequency, collocation and meaning” (Hyland & Tse, 2007).
- This raised criticism against AWL, and many scholars have developed field-specific academic vocabulary in diverse academic fields (Hyland & Tse, 2007; Martínez, Beck & Panza, 2009; Valipouri & Nassaji, 2013; Yang, 2015).

4. Analysis results and Discussion

- Table 7 and 8 show the results analyzing OTC (Old Testament Corpus) and NTC (New Testament Corpus) through AntWordProfiler 1.4.0 developed by Anthony (2014).

Table 7. Word lists of OTC

Word lists	Tokens/%	Types/%	Word Families/%
1st GSL	466,541(82.18%)	2,726(22.29%)	810(1000)/81%
2nd GSL	30,582(5.39%)	1,916(15.67%)	800(1000)/80%
AWL	5,745(1.01%)	694(5.68%)	317(570)/55.61%
기타	64,846(11.42%)	6,893(56.37%)	
Total	567,714	12,229	1,927/74.98%

4. Analysis results and Discussion

- As seen in Table 7 and 8, the rate of GSL presents greatly high, 87.57% and 90.67% in both OTC and NTC, while that of AWL indicates remarkably low, 1.01% and 0.96% compared to 10% suggested by Coxhead (2000).

Table 8. Word lists of NTC

Word lists	Tokens/%	Types/%	Word Families/%
1st GSL	164,196(85.7%)	2,296(35.2%)	774(1000)/77.4%
2nd GSL	9,518(4.97%)	1,258(19.29%)	646(1000)/64.6%
AWL	1,838(0.96%)	443(6.79%)	243(570)/42.63%
기타	16,044(8.37%)	2,526(38.72%)	
Total	191,596	6,523	1,663/64.71%

5. Conclusion

- The above results confirm that text dispersion keyness is a better option than other methods.

References

- Anthony, L. (2014). AntWordProfiler (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net>.
- Baker, P. 2004. Querying keywords: questions of difference, frequency and sense in keywords analysis. *Journal of English linguistics*, 2(4), 346-359.
- Baker, P. 2010. Corpus methods in linguistics in L. Litosseliti (ed.) *Research Methods in Linguistics*, PP. 95-113. New York: Continuum.
- Chanasattru, S. & S. Tangkiengsirisin. (2017). The Word List Distribution in Social Science Research Articles. *Arab World English Journal* 8(4), 412-429.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* 34(2), 213-238.
- Egbert, J and D. Biber. 2019. Incorporating text dispersion into keyword analyses. *Corpora*, 14(1), 77-104.
- Gilmore, A. & N. Millar. (2018). The language of civil engineering research articles A corpus-based approach. *English for Specific Purposes* 51, 1-17.
- Hyland, K. & P. Tse. (2007). Is there an "academic vocabulary?" *TESOL Quarterly* 41(2), 235-253.
- Martínez, I. A., S. C. Beck & C. B. Panza. (2009). Academic vocabulary in agricultural research articles: A corpus-based study. *English for Specific Purposes* 28(3), 183-198.
- Scott, M. 1997. PC analysis of key words and key key words. *System*, 25(2). 233-245.
- Scott, M. 2020. *WordSmith Tools Version 8.0*. Liverpool: Lexical Analysis Software.
- Valipouri, L. & H. Nassaji. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes* 12(4), 248-263.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Yang, M. (2015). A nursing academic word list. *English for Specific Purposes* 37, 27-38.

Zoom 2

통사/화용

<https://pusan.zoom.us/j/85949609966?pwd=d01XVFNrYjI3aHd0Q1pRYnV3VUVYdz09>

회의 ID: 859 4960 9966

암호: 123456

Notes on Freezing Effect and Anti-Freezing Effect

Kiyang Kwon

(Youngsan University)

1. Introduction

1. Purpose:

- To consider how the MERGE-based system in Chomsky (2017, 2019a, b, c) and Chomsky et al. (2019) can account for the Freezing Effect and Anti-Freezing Effect.
- To suggest a new analysis of the Freezing Effect and Anti-Freezing Effect under the MERGE-based system

2. Consequences

- We can explore that Determinacy Principle can open up a new opportunity to account for the Freezing Effect and Anti-Freezing Effect.

2. Freezing Effect & Anti-Freezing Effect

1. **Generalized Freezing Principle (Wexler and Culicover 1980)** : If a DP moves out of its base position to a derived position, subextraction out of that DP should not be possible.

(1) a. *Who_j do you think that [[pictures of t_j]_i [John would like t_i]]?

b. *What_j did you [[sell t_i to Fred] [a beautiful and expensive painting of t_j]_i]?

(Culicover & Wexler 1977:21)

2. **Subject Condition (Huang 1982)**: Subextraction of DP from a derived subject is not possible.

(2) *Who_j did [pictures of t_j]_i [_{VP} t_i please you]? (Ishii 2003:22)

3. **DP-extraction from an embedded subject position in the finite clause is not allowed.**

(3) *John_i seems [that t_i [_{VP} t_i reads a book]]. (Goto & Ishii 2019:44)

4. Subextraction of *wh*-element out of an embedded derived subject position is not allowed on the presence of the complementizer *that*.

- (4) *[Which city]_j did they say [that [a picture of t_j]_i [_{VP} t_i caused a scandal]]?
(Luis 2020:3)

5. Extraction of *wh*-element out of a derived subject is not possible on the presence of the complementizer *that*. (Lasnik and Saito 1992, Rizzi 1990, and Ishii 2003)

- (5) *[What]_i did they say [that t_i [_{VP} t_i caused a scandal]]? (Luis 2020:3)

6. Anti-Freezing Effect: A'-movement out of a shifted position can be allowed. (Luis 2020)

- (6) a. [Them]_i, I believe t_i **with all my heart** t_i to have won the election.
b. (?) [Who]_j do you believe [a friend of t_j]_i **with all your heart** t_i to have won the election?
- (7) a. [Them]_i, Kima called t_i up t_i?
b. (?) [Which person]_j did Mary call [friends of t_j]_i up t_i? (Luis 2020:10)

3. Previous Analysis of Freezing Effect

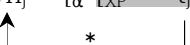
1. Freezing Effect and Labeling Algorithm

(8) Generalization of Extraction under Labeling (Goto 2016)

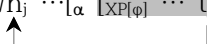
An unlabeled syntactic object is opaque for extraction but becomes transparent if it is labeled:

- a. * XP_i ... [α ... t_i ...] (α is not labeled, which is shaded with gray)
b. XP_i ... [α ... t_i ...] (α is labeled)

- (9) a. *Who_j do you think that [[pictures of t_j]_i [John would like t_i]]?
b. *What_j did you [[sell t_i to Fred] [a beautiful and expensive painting of t_j]_i]?

- (10) * Wh_j ... [α [XP ... t_j ...]_i [YP ... t_i ...]] (α = {XP, YP} is not labeled)
- 

2. Problems of Goto's (2016) Analysis

- (11) a. *Who_j did [pictures of t_j]_i [_{VP} t_i please you]?
b. Wh_j ... [α [XP_[φ] ... t_j ...]_i [YP_[φ] ... t_i ...]] (α = {XP_[φ], YP_[φ]} is labeled <φ>)
- 

- (12) a. Who_i is there [a picture of t_i] on the wall? (Stepanov 2007)
 b. *Wh_j ... [α [DP ... t_j ...] [_v*P ...]] (α = {DP, _v*P} is not labeled)
-

- (13) a. *Who_i was [α a picture of t_i] taken by Bill? (Matsubara 2008:472)
 b. Wh_j ... [α [XP_[φ] ... t_j ...]_i [YP_[φ] ... t_i ...]] (α = {XP_[φ], YP_[φ]} is labeled <φ>)
-

3. Freezing Effect and Anti-Locality Constraint

- (14) **Spec-to-Spec Anti-Locality Constraint** (Erlewine 2016, 2019)
 A-movement of a phrase from the Specifier of XP must cross maximal projection other than XP.

- (15) a. *[What]_i did they say [that t_i [_vP t_i caused a scandal]]? (Luis 2020:3)
 b. *[CP(embedded) *wh*-subject_i that [TP ... t_i ... [_vP ... t_i ...]]]
-

4. Problems of Anti-Locality Analysis

- (16) a. *[Which city]_j did they say [that [a picture of t_j]_i [_vP t_i caused a scandal]]?
 b. [CP Which city_j that [TP [DP a picture of t_j]_i ... [_vP t_i ...]]]
-

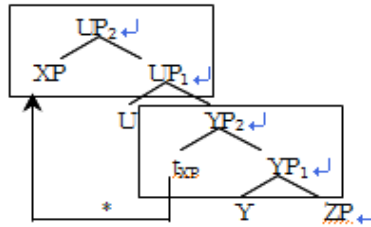
- (17) Revised Anti-locality (Luis 2020)
 Movement of an XP must cross a Projection Line.

- (18) Projection Line
 A PL of α = {x | x = a projection of α}

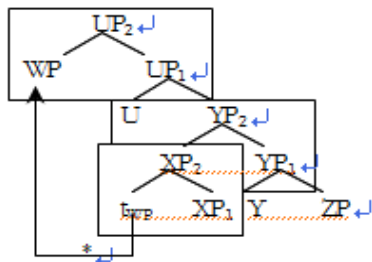
- (19) Crossing a Projection Line
 An element crosses a Projection Line (PL) on its way to a position iff
 a. all members of that PL dominate α's launching site; and
 b. no member of that PL dominates β.

(20) Luis' Argument (2020)

a. Illicit pattern of Extraction ↴



b. Illicit pattern of Subextraction ↴



(21) a. I believe **with all my heart** [[a friend of McNulty] to have won the election].

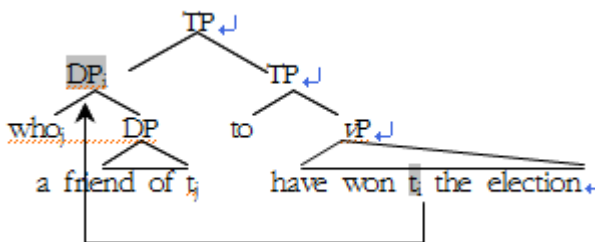
b. I believe [a friend of McNulty]_i **with all my heart** [t_i to have won the election].
(Luis 2020:9)

(22) **Anti-Freezing Effect in ECM** (Luis 2020:10)

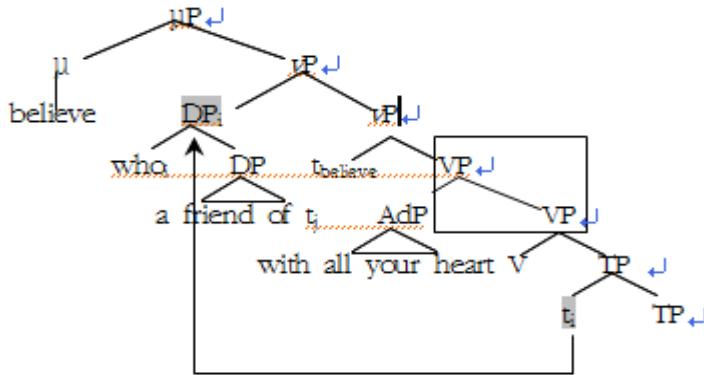
- a. [Them]_i, I believe t_i **with all my heart** t_i to have won the election.
- b. (?) [Who]_j do you believe **with all your heart** [a friend of t_j] to have won the election?
- c. (?) [Who]_j do you believe [a friend of t_j]_i **with all your heart** t_i to have won the election?

5. Anti-Freezing Effect and Revised Anti-locality (Luis 2020)

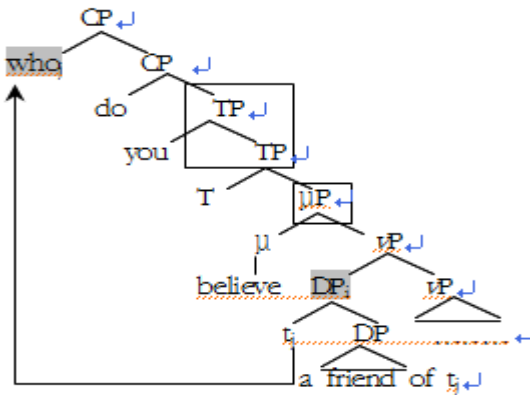
(23) Embedded TP



(24) Matrix Clause μ P



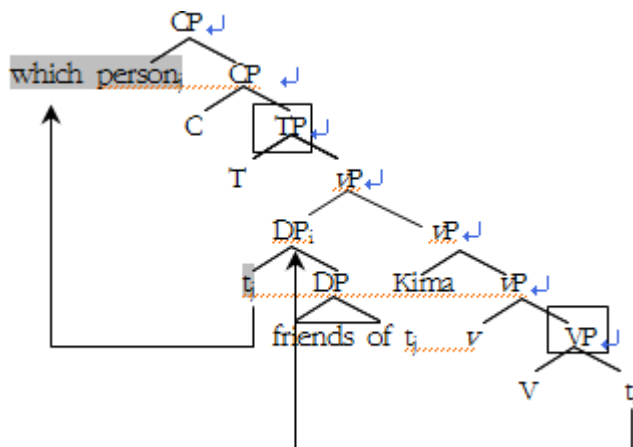
(25) Matrix Clause CP



(26) Anti-Freezing Effect in Object Shift (Luis 2020:10)

- a. Kima called up [friends of McNulty].
- b. Kima called [friends of McNulty]_i up t_i.
- c. [Them]_i, Kima called t_i up t_i?
- d. (?) [Which person]_i did Kima call up [friends of t_i]?
- e. (?) [Which person]_j did Kima call [friends of t_j]_i up t_i?

(27) Matrix Clause CP



6. Problems of Revised Anti-Locality Analysis

(28) a. Who saw James?

b. $[_{CP} \text{ who}_i [_{CP} C [_{TP} t_i [_{TP} T [_{VP} t_i [_{VP} \text{ watched James }]]]]]]$

(29) The Principle of Minimal Compliance (PMC, Luis 2020)

If a head H with probes $P_{[X],[Y]}$... Agrees with a Goal G in a feature X, G is accessible to subsequent syntactic operations regardless of locality conditions with respect to probes on H.

(30) $[_{TP} \text{ who}_i [_{CP} C [_{TP} t_i [_{TP} T [_{VP} t_i [_{VP} \text{ watched James }]]]]]]$

7. Problems of PMC + Anti-Locality Account (Kaqchikel의 Agent Focus 현상)

(31) Object wh-question does not require Agent Focus

Achike ✓x-Ø-u tëj / *x-Ø-tj-ö ri a Juan.
 what COM-B3sg-A3sg-eat/ COM-B3sg-eat-AF Juan
 'What did Juan eat?' (Erlewine 2015:2)

(32) Subject wh-question requires Agent Focus

Achike *x-Ø-u tëj / ✓x-Ø-tj-ö re wäy
 who COM-B3sg-A3sg-eat /COM-B3sg-eat-AF the tortilla
 'Who ate the tortilla?' (Erlewine 2015:1)

4. Freezing Effect and MERGE-Based System

1. Merge vs MERGE

- (33) a. Merge $(\alpha, \beta) \rightarrow \{\alpha, \beta\}$
 b. MERGE maps $WS = [X, Y]$ to $WS' = [\{X, Y\}]$

2. Workspace

- (34) a. $WS = [P, Q] \rightarrow \text{MERGE}(P, Q)$
 b. $WS' = [\{P, Q\}, P, Q]$

3. The Principle of Determinacy

- (35) If the structural conditions for a rule holds for some workspace, then the structural change must be unique. (Chomsky 2019a: 275)

4. Determinacy Principle + PIC Account

- (36) a. *Who_j do you think that [[pictures of t_j]_i [John would like t_i]]?
 b. [_{CP1} **who** [C1-that [_{CP2} [pictures of who] [C2 [_{TP} John [T-would [_{v*P} John [_{v*-R(LIKE)} [_{RP} [pictures of who] [R(LIKE) [pictures of who]]]]]]]]]]]]]
- (37) a. (?)[Who]_j do you believe [a friend of t_j]_i **with all your heart** t_i to have won the election?
 b. [_{CP} **who** C(do) [_{TP} you T [_{v*P} [a friend of who] v*-R(believe) [_{v*P} [_{ADV}P [with all your heart] [R(believe) [a friend of who] [_{TP} to have won the election]]]]]]]]]]
- (38) a. (?)[Which person]_j did Mary call [friends of t_j]_i up t_i?
 b. [_{CP1} **which person** C-did [_{TP} Mary T [_{v*P} [a friend of which person] [_{v*-R(call)} [_{v*P} [R(call up) [a friend of which person]]]]]]]]]

5. Conclusion

We have considered that the MERGE-based System in Chomsky (2017, 2019a, b, c) and Chomsky, Gallego, and Ott (2019) can explain the Freezing Effects in English. In addition, we have proved that the MERGE-based System can also successfully account for the Anti-Freezing Effects in ECM and object shift in English

6. References

- Boeckx, Cedric, and Norbert Hornstein. 2005. A Gap in the {ECM} Paradigm. *Linguistic Inquiry* 36 (3): 437-441.
- Brillman, Ruth, and Aron Hirsch. 2016. An anti-locality account of English subject/non-subject asymmetries. In *Proceedings of CLS 50*.
- Chomsky, Noam. 2013. Problems of projection. *Lingua* 130:33-49.
- Chomsky, Noam. 2015. Problems of projection: Extensions. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*, 1-16. Amsterdam: John Benjamins.
- Chomsky, Noam. 2019a. Some puzzling foundational issues: The Reading program. *Catalan Journal of Linguistics Special Issue*: 263-285.
- Chomsky, Noam. 2019b. Fundamental issues in linguistics, MIT lectures, April 10th and 12th.
- Chomsky, Noam. 2019c. UCLA lectures, April 29th, 30th, May 1st, 2nd.
- Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. 2019. Generative grammar and the

- faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics*: 229-261.
- Corver, Norbert. 2017. Freezing Effects. In *The Wiley Blackwell Companion to Syntax*, Second edi, edited by Martin Everaert and Henk van Riemsdijk. John Wiley & Sons.
- Goto, Nobu. 2016. Labelability = Extractability: Its theoretical implications for the free-Merge hypothesis. In *Proceedings of the North Eastern Linguistic Society* 46, 335-348. Amherst, Mass.: GLSA Publications.
- Goto, Nobu and Toru Ishii. 2019. Some Consequences of MERGE and Determinacy. ms. Toyo University and Meiji University.
- Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language and Linguistic Theory* 34 (2): 429-479.
- Erlewine, Michael Yoshitaka. 2019. Anti-locality and subject extraction. Ms. National University of Singapore.
- Huang, James 1982. *Logical relations in Chinese and the theory of grammar*. Doctoral dissertation, MIT.
- Ishii, Toru. 2003. Adjunct Condition and Empty Operator Movement. *Bungei Kenkyu* 91:21-52.
- Kwon, Kiyang. 2016. *that*-trace effect and labeling algorithm, *Studies in British and American Language and Literature* 120.127-146.
- Kwon, Kiyang. 2018. Ant-Locality Effect and Labeling Algorithm, *New Korean Journal of English Language and Literature* 60.2.175-193.
- Kwon, Kiyang. and Wonbin. Lee 2019. Notes on the MERGE-Based System, *New Korean Journal of English Language and Literature* 61.3.131-153.
- Lasnik, Howard and Mamoru Saito. 1992. *Move α : Conditions on its application and output*, Cambridge, MA: MIT Press.
- Luis Miguel Toquero-Pérez 2020. Revisiting extraction and subextraction patterns from arguments. MS. University of Southern California
- Matsubara, Fuminori. 2008. Remarks on Chomsky's (2008) Analysis of Extraction from subject. *English Linguistics* 25(2), 464-474
- Müller, Gereon. 2011. *Constraints on Displacement: A phase-based approach*. John Benjamins Pub Co.
- Rizzi, Luigi. 1990. *Relativized minimality*, Cambridge, Mass: MIT Press.
- Wexler, Ken, and Peter Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

Labelability of Elliptical Properties : Evidence from Fragmentary Questions in Korean

Haewon Jeon

(Pukyong National University)

Abstract: In this paper, I have presented some properties of FQs in Korean, the implication pointed by connectivity effects would seem to be : (i) there exist syntactic structure inside ellipsis sites, but it is unpronounced. (ii) Remnants in FQs in Korean are considered as a Contrastive Topic. In accordance with Takita et al. (2016), den Dikken (2008), and Ko (2011), I propose there exist a functional head H mediating the remnant (Contrastive Topic) and elided material in FQs in Korean that can salvage remnants from labeling failure at narrow syntax (NS). I also scrutinized to explore the nature of the interaction between narrow syntax (NS) and two interfaces with the goal of achieving an explanation of how elided FQ is realized. In Chomsky (2013, 2015), every syntactic object needs to be labeled for interpretation at the interfaces since those labels play certain roles in semantic and phonological computations. This paper aims to present the possibility that ellipsis should be licensed in narrow syntax. Furthermore, labels play a crucial role in ellipsis phenomena, specially in FQ in Korean.

Keywords: Ellipsis, Fragmentary Questions, Labels, Labeling Algorithm, Interfaces, Sensori-motor system, SM, CI, Transfer, Minimalist Program

1. Introduction

Ellipsis phenomena cover a range of constructions, such as predicate ellipsis, ellipsis in the nominal phrase and clausal ellipses found in sluicing, gapping, fragment answers. In this paper, I primarily focus on a fragment. Specifically, considerable attention is paid to how to derive a fragmentary expression that nonetheless yields a full-fledged clausal interpretation and how to get labels of elided parts. In Chomsky (2013), syntactic objects need to be labeled for interpretation at the interfaces since those labels, by assumption, play certain roles in semantic and phonological computations.

Concerning the derivation of ellipsis phenomena, Merchant's (2004) analysis, where the remnant undergoes focus movement to the left periphery followed by PF deletion of the rest of the clause, has been influential in the literature.

Richards (2016, 2020) claims that some parts of phonological features are constructed in the narrow syntax. I propose that labels of elliptical properties play a crucial role in licensing of ellipsis. Section 2 introduces the properties of FQ in Korean. Section 3 discusses labelability of FQs in the perspective of Labeling Algorithm of Chomsky (2013, 2015). Section 4 presents remaining issues to figure out the mismatch between PHON and SEM.

2. Properties of Fragment Question (FQ) in Korean

(1) a. A: What did she eat?

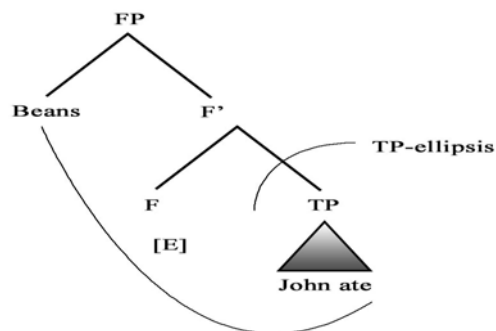
B: ~~She ate~~ beans. (fragment Answer)

b. A: What did she eat?

B: Someone ate beans, but I don't know who ~~ate~~ beans. (sluicing)

(1a) is the example of fragment answer (FA), which means that the answer to a question often makes use of fragment rather than a complete sentence (Morgan 1973, 1989, Merchant 2004). In analogy with FA in (1a) to sluicing in (1b), Merchant (2004:675) argues that FA in fact involves a full-fledged structure, in which the fragment moves from within TP to a clause-peripheral position, Spec FP, headed by F containing [E] (elide) feature, prior to the constituent TP ellipsis, as illustrated in (2).

(2)



(Merchant 2004:67)

In his syntactic implementation, he argues that the left-peripheral head which attracts the fragment to its Spec is endowed with a feature [E], which licenses the non-pronunciation of its complement. Put differently, the pronounced fragment DP *Beans* moves to a clausal peripheral specifier position of a functional projection, here called FP. The crucial [E] feature triggers non-pronunciation of

the clause TP *John ate*, more neutrally put, unpronounced material. This paper shows that the notion of fragment also applies to questions in Korean, so-called Fragment Question (FQ). The derivation of FQ in Korean can be assimilated to that of FA in English, involving focus movement and TP ellipsis.

- (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).

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

- (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?

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'). Furthermore, we raise the question how parallel or independent are the transferences of information to CI and SM? This paper also aims to present the possibility that labels 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.

3. Labelability of Elliptical Properties

Chomsky(2013) assumes that a label is required for interpretation at the interfaces.

...there is a fixed labeling algorithm LA that licenses syntactic objects so that they can be interpreted at the interfaces, operating at the phase level along with other operations.

(Chomsky 2013: 43)

In this system, labels play no role in executing the syntactic derivation, but rather are required for interpretation at the interfaces. Labeling is necessary for SOs to be interpreted at both of the interfaces. In other words, labels are required for both of the semantic and the phonological sides of grammar. Let us see how labels are determined under the Labeling Algorithm.

- (8) a. [α H YP]: $\alpha = H$
 b. [α XP YP] [XP...[α <XP> YP]]: $\alpha = YP$
 c. [α XP[F] YP[F]]: $\alpha = \langle F, F \rangle$

(by the most prominent shared feature F.)

Let us revisit the ellipsis phenomena. Jung (2017) assumes that only functional categories are eligible to be the licensors of ellipses. Those with “[E]-feature” (Merchant 2001) induces its complement to be deleted at PF. Lobeck (1995) elaborates a theory of ellipsis that elliptical categories in IP (VP Ellipsis), DP (N’ Ellipsis), and CP (Sluicing) are empty and subject to the same licensing conditions. She proposes that both types of empty pro nominals must be licensed under head-government to satisfy the Empty Category Principle, and identified through strong agreement.

- (9) 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)

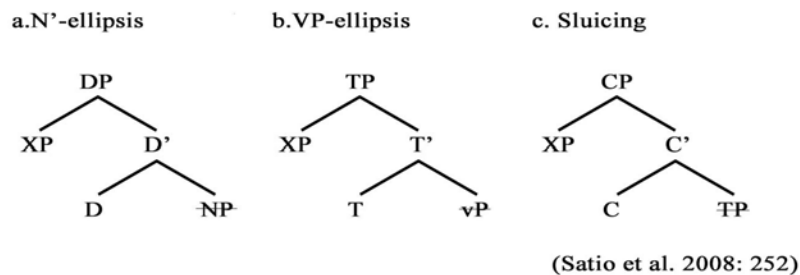
Given Lobeck (1990,1995)'s proposal that there exist common properties among N'-deletion, VP-deletion, and sluicing, they all involve the deletion of the functional head-complement. She further claims that these major cases of ellipsis bear the same structural requirement. In other words, these all involve 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-complement relation. The examples are presented below in (10).

- (10) 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)

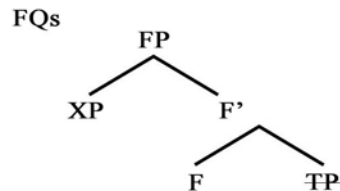
Each case of ellipsis in (10) are illustrated in (11).

(11)



FQs in Korean are depicted as below.

(12)



As sketched in (11–12), 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 (Contrastive 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 (14).

- (13) a. [vP V° [E] { $VP \dots$ }] (VP-deletion)
 b. [TP T° [E] { vP —{ $VP \dots$ }}] (vP-deletion)
 c. [CP C° [E] [TP { vP —{ $VP \dots$ }}]] (TP-deletion)

Lobeck (1991, 1995)

- (14) [FP F° [E] —{ TP —{ vP —{ $VP \dots$ }}}] (TP-deletion) : FQs in Korean

One important remaining issue is that this work does not show how labels are utilized at the SM interface under Chomsky(2013), which requires further scrutiny in the future.

4. Remaining Issues

One important remaining issue is that this work does not show how labels are utilized at the SM interface under Chomsky(2013), which requires further scrutiny in the future.

(15) a. A: Phathi-eyse motwu mwues-ul masyess ni?
 party-at-top everyone what-acc drank q

‘What did everyone drink at the party?’

B: Yumi-nun maykcwu-lul masyesse yo.

Yumi-top beer-acc drank part

‘Yumi drank beer.’

A: Minswu-nun? (FQ)

Minswu-top

‘Minswu? (Intended: What did Minswu drink?)’

b. [a Minswu-nun HP mwues-ul i [TP <Minswu-nun> [vP <mwues-ul> i [VP t i masita]]]

Phonetic Realization (o)

Phonetic Realization (x)

Semantic Interpretation (o)

Semantic Interpretation (o)

Shaded part represents ellipsis site: it does not realize phonetically; it is necessarily visible to LF computation for remnant *Minswu-nun* in order to be interpreted. In this section, we examined that, in the case of reconstruction, Transfer to CI occurs before the Transfer to SM. For expository convenience, let us revisit (4), here as in (16). FQs in Korean are defined as the locus of interpretation differs from the locus of pronunciation as in (18).

Elided Part

(16) [FP XP F° [TP ...]]

Chomsky presumes that the target domain for each application of Transfer is not the entire phase but only the completion of a phase head. Thus, it follows from labeling that deletion targets the head-complement. (18) illustrates the situation where Transfer applies to the phrase head completion, TP. At this point, Transfer sends to the SM interface separately from F° and XP. As a consequence, TP is not expected to have any phonological interaction with F° and XP. However, the only part that is linearized is XP, remnant of FQ. How can XP is visible at SM after Transfer? Though many other questions remain unsolved, it seems we could take some further steps towards approaching the Strong Minimalist Thesis.

Selected References

- 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.
- Maeda, Masako. 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*.
- Richards, Norvin. 2016. *Contiguity theory*. Cambridge, MA: MIT Press, 2016
- Richards, Norvin. 2020. Immobile wh-phrases in Tagalog – lingbuzz/005294

Assessing the linguistic ability of BERT on co-indexation of English reflexives

김현수 · 송상현

고려대학교

1. Introduction

최근 몇 년 동안 언어 연구에서 딥 러닝의 방법론은 다양한 관점으로 언어 현상에 관한 연구를 발전시켰다. 그중에서도 특히 언어 모델은 다양한 언어적 특징과 신경망 시스템의 성능을 학습하는 언어 처리를 통해 특정 통사 현상에 대한 지식을 학습할 수 있다는 사실이 여러 연구를 통해 밝혀졌다. 또한, 이를 기반으로 언어학과 딥 러닝이 서로에 어떻게 기여할 수 있는지에 대한 연구에서 입증되었듯이 (Lizén, 2019), 언어 모델의 성능을 측정하기 위해 서로 다른 언어 현상을 평가해왔다. 그렇다면 어떤 언어 모델들이 있는지 서술하는 것도 중요하겠지만 본 연구에서는 단 하나의 모델을 통해 연구를 진행했다.

Vaswani et al. (2017)은 어텐션 메커니즘을 기반으로 하는 새로운 단순하고 뛰어난 네트워크 아키텍처인 transformer 언어 모델을 발표했다. 이 연구는 언어 모델을 크고 제한된 훈련 데이터로 모두 파싱하는 영어 구성에 성공적으로 적용함으로써 다른 작업에 잘 일반화할 수 있도록 한다. 1년 후에 SQuAD 리더보드에 트랜스포머 메커니즘을 통해 BERT (Devlin et al, 2018)가 등장했다. BERT는 사전 교육 후 특정 목적에 맞게 미세 조정 및 사용할 수 있다는 장점이 있어 상대적으로 적은 리소스로 다양한 작업에 사용할 수 있다.

BERT에는 인코더 블록이 있는데, Base 모델은 12개의 인코더로 구성되고 Large 모델은 24개가 있다. 이는 전체 입력 시퀀스의 의미를 반복적으로 구축함을 의미한다. 당연히 인코더 블록의 수가 많을수록 단어 사이의 더 복잡한 관계를 더 잘 포착할 수 있을 것이다. 인코더 블록은 RNN과 유사한 특성을 가지며, 이전 출력을 현재 입력으로 사용하기 때문에 이 부분은 12개의 Base 모델과 24개의 Large 모델로 구성되며, 모두 반복적으로 반복되며, 블록 내의 각 입력 및 처리 결과는 다이어그램과 같이 별도의 연결로 처리되어 비선 활동을 거치지 않고 네트워크를 통해 구동된다. MLM(Masked Language Model)은 마스킹된 토큰과 일치하도록 학습된 결과의 직접 벡터를 가지고 있으므로 더 직관적으로 볼 수 있다. 마스킹은 모든 토큰을 마스킹하지 않을 정도로 전체 단어의 약 15%만 진행되지만, 이중 80%만 [MASK]로 처리되어 10%는 랜덤으로, 나머지 10%는 정상으로 처리된다.

위와 같은 언어 모델의 경우와 반대로 실험 언어학의 관점에서는 문장의 옳고 그름을 알아보는 것이 전통적인 연구 방법이다. 문법성과 수용성을 통해 실험이 진행되고 연구 결과를 얻는다. 수용성 판단은 생성 언어학의 주류방법론이었기 때문에, 언어학자들은 대부분 수용성 판단에 따라 이진법적인 방법론으로 해당 언어 표현이 비문인지 정문인지를 구분하는 방법을 통해 관련 분야에서 실험을 진행한다. 마찬가지로, 딥러닝 기반 언어 연구도 해

당 종류의 방법론과 관련하여 비슷하게 진행되고 있다 (Warstadt 외, 2018).

Warstadt et al. (2018)의 연구에 따라 데이터 중심 중립 네트워크가 인간 학습자의 네트워크와 유사한 제약 조건 내에서 수용성 분류를 실제로 수행할 수 있는지 테스트할 필요가 있다. 딥 러닝 언어 모델 연구의 관점에서, 그것은 비슷한 문제이자 해결해야 하는 문제이다 (Tan & Frank, 2019; McCoy et al. 2020). 그러나 Schütze(1996)가 주장하듯이 언어학자들은 종종 그 데이터를 명확하지 않은 방식으로 만들고 사용해왔다. 일단 특정 언어 현상을 통해 딥러닝 언어 모델이 언어 제약을 구분하고 배울 수 있는지를 골라야 할 것이다. 많은 선행연구에서 주어 동사 일치, 의존성 공백 문장, 부정극어 및 재귀사 같은 다양한 언어 모델의 학습에 대한 발견이 이루어졌다. 다양한 선행연구의 특징 중에서, 우리는 특히 재귀사에 대해 집중한다. 언어학에서 대용어에 관한 연구는 NLP 분야에서 흥미로운 실험 중 하나였다. 주로 Chomsky (1981)의 생성 언어학에서 연구되어온 현상에서 성분통어 및 공지시에 대한 이론으로 관련 논문을 연구하기 위한 여러 가지 방향을 제시하는 데 도움이 된다. 그는 영어의 모든 대용어는 동일한 조항 내에서 이전 NP에서 공지시되는 것이 필수라고 제안했다.

그에 따라 성분통어 및 공지시에 대한 딥러닝 언어 모델에 평가 방식은 재귀사와 관련하여 모델을 테스트하는 것이다. Marvin & Linzen(2018)은 RNN 및 LSTM과 같은 언어 모델의 예측에 대한 문법성을 평가하기 위한 데이터를 디자인했고 그에 따른 언어 모델간의 결과를 비교 대조했다. Goldberg (2019)에서는 BERT가 Marvin & Linzen의 이전 연구에서 동일한 데이터에 통사 지식 가졌는지 여부를 연구한 후 Tan & Linzen의 각 모델 결과와 비교하였다. 다른 양적 지표인 'Confusion score' '로 연구한 Tan & Frank (2019)의 연구에서는 성 및 수 일치를 통해 BERT의 통사적 지식 여부를 연구했다. 마지막으로, BERT의 원시 데이터에 대한 사전 훈련을 통한 계층적 구조에 대한 귀납적 편견을 가지는지 여부를 평가하는 Warstadt & Bowman(2020)의 연구가 있다. 위의 이전 연구를 기반으로, 우리는 현재 가장 많이 사용되는 언어 모델인 BERT가 어떻게 학습된 것을 기반으로 학습하고 사용할 수 있는지에 대한 실험을 수행할 것이며, 이는 인간이 데이터 세트와 예제 문장을 통해 언어를 배우고 인식하는 것과 동일한 방법이다. 데이터 구성 과정은 이전 선행연구의 데이터 설정을 참조하여 gradient를 이룰 수 있는 요소를 추가했다. 연구는 기본 구조의 문장, by 전치사구, for 전치사구, 주격과 목적격 관계대명사의 총 5 종류의 데이터 세트로 이루어졌다.

특히 BERT와 같은 모델은 특정 작업을 수행할 때 특정 작업을 충족하기 위해 훈련이 필요한 기존 방법과 달리 많은 양의 데이터를 기반으로 하는 일반 모델로 상세 작업에 따라 데이터를 학습하고 사용할 수 있다. BERT를 통해 재귀 대명사에 대한 언어 모델의 학습은 [MASK]라는 특수 토큰으로 모든 토큰을 커버한 다음 이를 예측하는 방식으로 훈련을 수행하는 Masked Language Model을 사용하여 조사되었다.

2. Method

2.1 Task

기존의 통사적인 연구에서의 수용성 판단은 이론적인 측면으로 구성되지만, 언어 모델은 샘플 효율적인 통계 방법의 매체를 통해 대량의 텍스트 데이터에서 자연어 이해를 추출한

다. 데이터 세트의 구성은 이전 연구를 기반으로 수정되고 인칭과 유생성이 추가된 예문으로 실험되었으며 언어 모델의 특성으로 인해 사람들의 수용성 판단을 결정하는 데 차이가 있었다. 대상 언어 모델은 위키백과에서 10만 개 이상의 문장으로 사전 훈련된 BERT이므로 [UNK] 토큰을 피하여 MLM 실험을 수행할 때 어휘를 고려해야 한다. 방법론적인 부분에서 우리는 언어 모델의 공동 인덱싱 능력에 대한 계산 구현의 정량적 지표인 surprisal을 사용했다.

각 실험에서 사람, 숫자, 성별, 유생성의 네 가지 언어적 현상으로 데이터를 설계했다. 각 제약조건은 여러 요인(즉, 데이터 유형, 제약 유형, 선행사, 재귀대명사)과 관련하여 문장의 수용성을 결정한다. 기본적으로 네 가지 요인을 기반으로 각 요소가 잘 형성된 기준을 충족하지 못할 때 기울기를 분석하여 제약 조건을 나눈다. 예를 들어, 표 1을 통해 각각의 언어적 현상에 대한 제약을 살펴볼 수 있다. 문장 (a)은 성별이 완벽하게 일치하며 수일치 그리고 인칭도 완벽하다. 하지만 예문 (b)는 인칭에 오류가 있으며, (c)에서 인칭은 3인칭으로 일치하지만 수일치에서 오류가 있고 성별은 알 수 없다. 예문 (d)는 성별이 일치하지 않으며 예문 (e)는 유생성에 문제가 있다.

	예문	제약
a	He hated himself.	None
b	*He hated myself.	Person
c	*He hated themselves	Number
d	.*He hated herself.	Gender
e	*He hated itself.	Animacy

표 1. 각 제약에 따른 예문

언어 실험의 제약조건이 수용성 판단의 기초가 되는 것은 사실이다. (Schütze, 1996) 그러나 이진법적인 분류는 많은 변수와 편향을 무시한다는 가정이 있을 수 있다. 위의 예에서 어떤 제약조건은 별 의미가 없을 수 있지만, 어떤 제약조건은 큰 역할을 할 수 있다. 따라서 제약 조건 사이에 차이가 있다는 가정하에 제약 조건에 따른 언어 모델의 반응을 연구하면 수용성 판단 테스트를 결정하는 데 더 유용할 수 있다 (Sorace & Keller, 2005).

2.1 Data

네 가지의 제약 뿐 아니라 영어 재귀사의 언어 모델 처리를 방해요소(attractor)으로 구성된 5가지 종류의 데이터 세트를 설계했다. 첫째, 단순한 반사형 문장, 방해요소와 수식어가 없으며 재귀대명사가 목적어에 위치한다. 이 데이터 세트는 단순 문장으로 나뉜 유일한 자료이다. 두 번째와 세 번째는 목적어에 방해요소가 나타나는 두 가지 유형의 전치사구로 구성된 데이터이다. 마지막 데이터는 관계대명사로 주격 및 목적격 관계절의 두 가지 구별되는 유형으로 구성된다.

선행 연구에서는 목적격 관계절만 사용했지만, 본 연구에서는 더 많은 유형이 추가되어 더 다양한 제약 조건을 측정하였다. 하나의 데이터 세트에서 제약의 점진성을 나타내면서 BERT의 구문 능력에서 어떤 제약 조건이 강하고 어떤 제약 조건이 약한지 결정하는 두 개

의 제약 조건 순위를 측정할 수 있다. 다음 장에서 순서대로 설명해야 할 총 4개의 제약조건 외에도, 데이터 세트는 대명사와 NP를 구별하는 유형, 재귀대명사와 선행사 사이의 공지시 관계 등 총 6개의 기준으로 구성되었다. 한 데이터셋 당 1088개의 예문으로 구성되었으며 총 5440개의 실험 문장을 통해 진행했다. 하나의 문장 단위로 전환될 때 8가지 종류의 재귀대명사가 있으므로, 8개의 반사 대명사의 [MASK] 처리하였고, 각 항목은 또한 8개의 동사로 구성되어 있으며, 모두 COCA에서 가장 빈도가 높은 AGT-PAT 동사를 고려하여 만들었다.

방해요소	문장	적형성
Simple	The boy liked ourselves.	nil(Person-Number)
for-PP	The boy liked the author for ourselves.	nil(Person-Number)
by-PP	The boy liked the author by ourselves.	nil(Person-Number)
Subj-relative	the accountant that believed the boy liked ourselves.	nil(Person-Number)
Obj-relative	the accountant that believed the girl liked yourselves.	nil(Person-Number)

표 2. 방해요소에 따른 예문

첫째, 다른 현상과 달리 언어의 인칭은 모두 동일하기 때문에 이에 대해 크게 걱정할 필요는 없었다. 하지만 3인칭 과 달리 1인칭과 2인칭에서만 대명사만 사용할 수 있다. 따라서 1인칭에는 대명사만 분류되고, 3인칭에는 대명사와 일반명사가 모두 사용된다.

- (1) a. I/you/he hit myself/yourself/himself.
- b. The boy/boys hit himself/themselves.

예문 (1)과 같이 인칭별로 세 부분으로 나누었다. (1a)와 같은 대명사 선행사의 경우 다른 설명에서는 공통 명사를 사용할 방법이 없다. (1a)와 대조적으로, (1b)는 각각 사용할 수 있다. 따라서, 첫 번째 사람과 두 번째 사람은 실험적인 문장을 만들기 위해 서로 다른 부분으로 나뉘었다. 본 연구에서는 언어 현상으로 구분하여 표시하였으며, 인칭을 구분하도록 설계하였다. 1인칭과 2인칭의 경우 대명사 외에 선행사를 선택할 수 있는 방법이 없어 대명사만으로 주어틀 구성했다. 영어에는 총 7가지 종류의 개인 대명사가 있으며, you를 제외한 모든 대명사는 수일치로 구분할 수 있다.

- (2) a. The boy hated himself.
- b. The boys hated themselves.

수일치는 Marvin & Linzen과 같은 선행연구에서도 실험되었다. 주어와 동사 수일치 외에도, 재귀사의 공지시에서 매우 중요한 문법적 판단 요인이다. 예문(2)과 같이 단수와 복수 두 부분을 구분하여 구성했다. 재귀대명사는 (2a)의 단수와 (2b)의 복수 선행사와 공지시 일치해야 한다.

영어에서 재귀대명사는 사람의 성별에 따라 달라진다. 따라서, 예문을 만들 때 이것을 확실히 할 선행사를 선택하는 것이 중요하다. 인간 실험의 경우 직관에 따라 일부 모호한 명

사를 선택할 수 있지만, 본 연구는 언어 모델을 겨냥한 실험이기 때문에 [MASK] 자리와 선행사와의 공지시 여부를 고려해야 한다. (Devline et al. 2019) 선례에 대한 다음 예에서와 같이, 소년과 소녀는 성별에 따라 구별될 수 있는 일반적인 명사 중 선택되었다.

- (3) a. The boy kicked himself.
b. The girl kicked herself.

(3a)의 경우 일반명사 the boy는 명사로서 himself의 선행사가 되고, (1b)의 the girl도 마찬가지로 herself의 선행사가 된다. 이 중 본 논문에서 사용된 BERT 모델은 예를 만들기 위해 사전 훈련된 어휘로 선택되었다. (2b)에서 보듯이, 다수의 재귀대명사는 성별에 영향을 받지 않고 그 자체로 모두 사용된다는 것을 알 수 있다. 예를 들어, 소년들과 소녀들 모두 재귀대명사 themselves와 공지시되고 이는 성별을 구분할 수 없는 재귀대명사이다. 따라서 이러한 성별 구분을 함께 검사할 수 없다. 숫자와 성별의 관계와 대조적으로, 각각의 제약에서 성별은 3인칭에서만 구분할 수 있다. 따라서 BERT는 제약 조건 순위를 통해 데이터의 그라데이션 분석 성별을 통해 알아 볼 수 없다. 따라서 Animacy와 관련된 공지시 일치의 경우는 다음 예와 같이 구성할 수 있다.

Dahl & Fraurud. (1996)는 애니매시스가 생물과 무생물이 구별되는 관계라고 주장하였다. 실험 설계와 관련하여 우리는 NA, 사람, 사람이 아닌것 또는 알수 없는 값으로 각 선행사를 구별했다. 영어에서 동물과 인간이 아닌 사물은 다음과 같은 대명사의 선행사로 사용된다.

- (4) a. The dog beat the boys by itself.
b. The table moved itself.
c. They are broken by themselves

예문 (4a)의 선행은 동물이고 (4b)는 물체이며, 둘 다 비인간으로 처리되었다. 그러나 (4c)의 경우 위의 정보만으로 볼 때 선행이 인간인지 비인간인지 알 수 없기 때문에 NA로 표시된다. 이것은 다른 값을 가진 숫자와 성별에 따라 BERT의 능력을 분석하는 다른 요소로 사용될 것이다.

각 요인 결과의 세부 정보를 조사하기 전에 각 데이터 세트의 surprisal 분석을 비교하는 것이 눈에 띈다. 표 1과 그림 1은 각각 다른 데이터 세트에서 각 제약 조건의 놀라운 평가를 보여준다. Simple 데이터 세트에서 surprisal 점수는 24.105223으로 점수가 매겨지며, 이는 BERT가 주목할 만한 주의를 기울였다는 것을 의미한다. BERT는 PP와 관계절과 같은 방해 요소 때문에 혼란을 느낀 것으로 보인다. 그것은 18.859413으로 높게 측정되었지만 다른 종류의 조건들에 비해 상대적으로 낮게 측정되었다. 용어와 용어의 전치사가 관용어로 사용된 것으로 추측되었다. 두 개의 상대적인 조항 모두 비슷한 수치를 보였고, 비록 방해요소가 있다 하더라도 두 조항 모두 공지시가 되지 않아 높은 수치를 보였다. 이러한 결과는 BERT의 구문 능력이 이전 연구에서 언어적 특징에 존재함을 나타냈다. 따라서, 언어 구조의 문법성을 위해 일부 제약 조건이 다른 제약 조건보다 더 중요하다는 것이 눈에 띈다. 한 요소만 일치하면 surprisal 분석 값이 약간 낮거나 거의 비슷하고 두 요인 및 모든 요인 사이에 차이점이 존재한다.

surprisal 평가의 요인 결과는 모든 문장 간의 surprisal 평가 점수에 차이를 보인다. 우

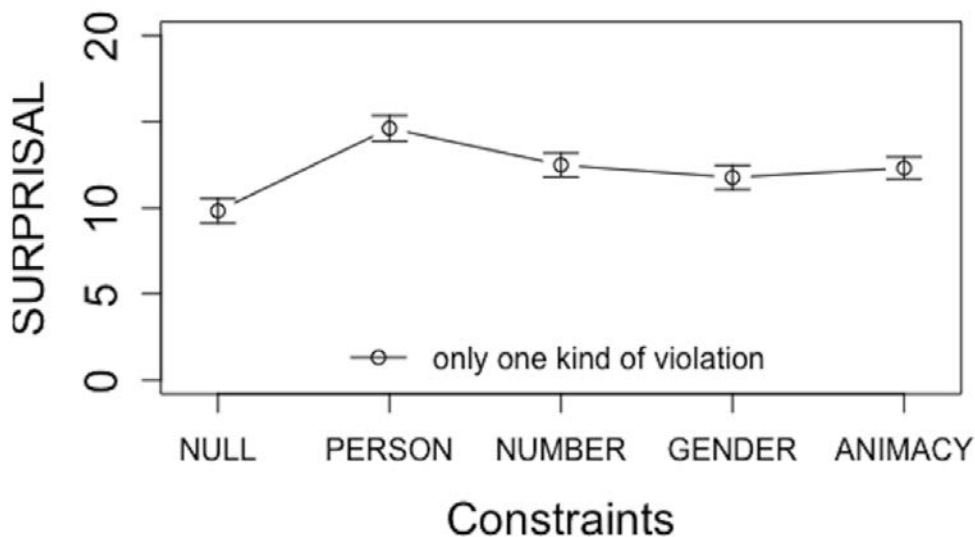
선 1, 2인칭 사례에서 대명사가 아닌 NP 선행사를 취하는 것은 불가능하다. 따라서 표 2에서는 NP 선행사 유형에 대한 결과를 분석했다. 모든 요소가 일치하는 각 문장은 수용성 판단을 테스트하기 위해 평가되며, 하나 또는 두 개의 요인이 일치하지 않도록 설계되었다.

3. Result

첫 번째 문장의 경우 재귀대명사 자체가 대명사를 공지시하는 문장이다. 해당 문장에는 아무런 제약도 없다. 이후의 문장들은 각각 다른 요소를 가지고 있다. 잘못된 각 요인에 따라 surprisal 평가 점수가 점차 커지는 것을 알 수 있는데, 이는 다음 그림 1과 같이 각 제약조건에 차이가 있다는 사실이다. 예를 들어, 성별의 경우 표 1과 같이 10.689967로, 제약이 없는 문장의 6.5980725보다 크다. 마찬가지로, 수일치의 surprisal 값은 11.97738로 성별보다 약간 크다. 12.207875의 인칭이 개별 제약 중 가장 큰 값을 가지고 있는 것을 알 수 있다. 더불어 제약이 두 가지 이상 겹칠 때는 13.3294735로 한 가지 제약만 있을 때 보다 비교적 값이 상승했다. 그에 따라 각 한 가지 제약만이 존재할 때의 surprisal 값을 그래프 1을 통해 비교할 수 있다. 표에서와 마찬가지로 전체 데이터의 값 또한 인칭이 가장 높고, 수일치와 유생성이 거의 비슷하며 성별이 가장 낮은 값으로 나타나는 것을 볼 수 있다.

Type	Sentence	wellformedness	surprisal
Pronoun	He hated himself.	main	4.969067
Pronoun	He hated herself.	nil(Gender)	11.172779
Pronoun	He hated myself.	nil(Person)	8.858944
Pronoun	He hated themselves.	nil(Number)	12.528681
Pronoun	He hated ourselves.	nil(Number, Person)	20.954342

표 3. 각 제약에 따른 surprisal 수치

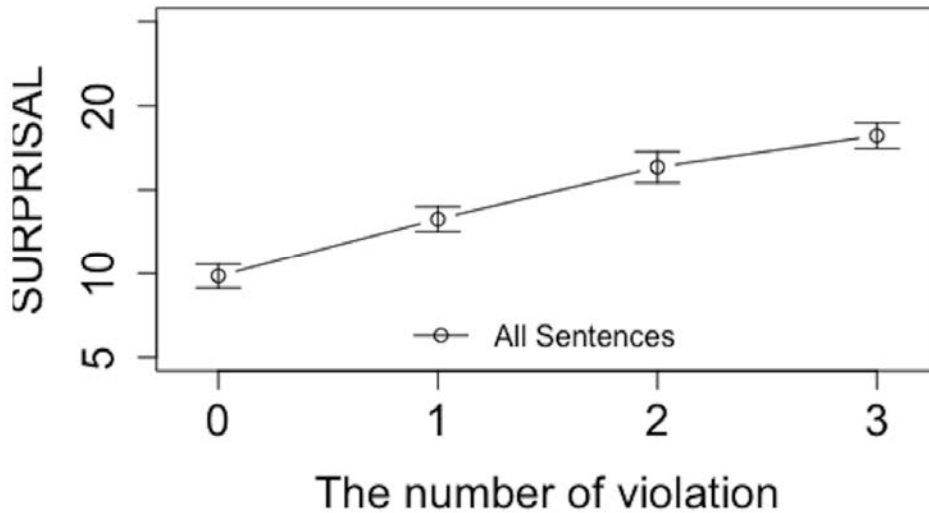


그래프 1. 각 제약에 따른 상호 도표

그렇다면 제약에 수에 따라 surprisal 값을 비교한 그래프를 살펴보자. 그래프 2에서 나타나듯, 모든 문장에서 제약의 개수가 늘어날수록 꾸준히 surprisal 값이 증가하는 것을 알 수 있다. 또한, 방해요소도 중요하게 살펴볼 것 중 하나이다. 3인칭 NP 구조의 선행사를 취하는 문장, 인칭, 수일치, 성별도 표시된다. 또한, NP 구조에서 알려지지 않은 세 가지 요인을 모두 비교할 수 있었으며, 마찬가지로 각 제약 위반의 개수와 방해요소에 따라 수치가 증가했다.

Type	Sentence	wellformedness	surprisal
Simple	The boy liked ourselves.	nil(Person-Number)	24.105223
for-PP	The boy liked the author for ourselves.	nil(Person-Number)	18.859413
by-PP	The boy liked the author by ourselves.	nil(Person-Number)	18.689775
Subj-relative	the accountant that believed the boy liked ourselves.	nil(Person-Number)	24.250294
Obj-relative	the accountant that believed the girl liked yourselves.	nil(Person-Number)	25.390387

표 4. 각 방해 요소에 따른 surprisal 수치



그래프 2. 제약의 수에 따른 상호 도표

마지막으로 각 제약 요소 수에 surprisal 값에 대한 gradient가 존재하는지를 살펴보았다. 모든 제약들에 대한 정보를 VT(Violation type)라는 열로 만들었고, 재귀대명사와 선행사의 공지시 정보를 담은 VI(Violation information)이라는 열과 비교하여 제약에 대한 차이가 존재하는지에 대한 surprisal 수치를 ANOVA 분석을 통해 측정했다. 이에 대한 결과로 $F=450.616$, $p < 2e-16$ 이라는 값이 나왔으며, TurkeyHSD 사후 검정 결과는 다음과 같이 나타났다.

	diff	lwr
GENDER-ANIMACY	-0.58567314	-1.45021898
NULL-ANIMACY	-2.52461984	-3.19548206
NUMBER-ANIMACY	0.16897004	-0.39506547
NUMBER-ANIMACY-ANIMACY	0.34155281	-0.28462643
NUMBER-GENDER-ANIMACY	0.11512445	-1.13926962
PERSON-ANIMACY	2.25808880	1.73632671
PERSON-ANIMACY-ANIMACY	3.10104842	2.43255569
PERSON-NUMBER-ANIMACY	5.68553502	5.19098594
PERSON-NUMBER-ANIMACY-ANIMACY	5.86144436	5.27323036
NULL-GENDER	-1.93894670	-2.86582726
NUMBER-GENDER	0.75464318	-0.09810972
NUMBER-ANIMACY-GENDER	0.92722595	0.03215506

검정을 통해 제약에 따라 층위가 나뉘지만, 각각의 요소들의 세부적인 부분을 살펴볼 수는 없다. 따라서 공지시에 있어 각각의 세부적인 요소들을 살펴볼 수 있는 분산분석을 시행했다. 예를 들어, 인칭에서의 요소는 1인칭, 2인칭 그리고 3인칭으로 나뉜다. 각각의 인칭에서 제약이 생길 때 다른 제약과의 surprisal 값을 비교한다면 어떤 인칭에서 제약이 큰지 알아볼 수 있을 것이다. 이 경우 $F=178.32$, $p < 2e-16$ 의 결과가 나타났다. 마찬가지로 수일치에서는 $F=433.3$, $p < 2e-16$, 성별에서는 $F=2.097$, $p < 0.148932$ 로 통계적으로 유의미하지 않은 수치가 나왔다. 이는 성별의 경우 3인칭과 단수에서만 나타나는 이유로 제약이 더 심하기 때문으로 간주한다. 유생성의 경우는 $F=362.85$, $p < 2e-16$ 의 결과를 보였다. 마지막으로 PP, 관계절과 같은 방해요소들의 경우 어떤 영향을 미치는지 검정을 통해 알아볼 수 있다. 각각의 제약의 정보와 그에 따른 방해요소의 존재 여부를 분산분석을 통해 검정하였다. $F=128.425$, $p < 2e-16$ 의 결과가 나타났다.

4. Conclusion

통사적 실험 및 이론 연구 방법론에 있어 여러 방법이 존재하고 사용되고 있다. 그 한 가지 방법인 gradience를 통해 신경망 모델의 데이터 실험을 진행하는 것이 본 연구의 주된 목적이다. 통사적 현상에서의 수용성 판단이 사실은 ‘옳다-그르다’의 판단으로 이루어지는 것이 아니다. 이런 가정에 따라 gradient 데이터가 언어학자들이 전통적으로 사용되는 이진법적인 방법보다 더 많은 정보를 제공하기 때문에 언어 이론에 기여할 수 있다는 주장을 기반으로 딥러닝 언어 모델도 마찬가지로 이런 데이터를 구분할 수 있는지를 연구하고 이를 통해 언어 연구에 이바지할 수 있다(Sorach & Keller, 2005).

본 연구에서는 재귀대명사의 공지시 여부에 대한 통사적 현상을 인칭, 성별, 수일치, 유생성과 같은 각각의 제약에 따라 순위가 매겨지며, 제약이 겹칠수록 더 많은 정보를 학습할 수 있다는 가정에 따라 실험이 디자인되었다. 그 결과로 BERT large 언어 모델에 의한 surprisal 점수를 통해 확인할 수 있었다. 첫 번째로 각 제약의 순위가 매겨졌다. 인칭의 경우 성별과 유생성과 같은 요소보다 강한 제약을 가지는 것으로 나타났다. 두 번째로 제약의 수가 늘어나면 늘어날수록 언어 모델의 surprisal 점수는 더 커지는 것을 알 수 있었다. 마지

막으로 각 테스트 아이템에 방해요소의 유무에 따라 점수가 달라지는 것을 알 수 있었다. 이는 언어 모델이 공지시하는 것을 방해함으로써 점수가 낮아질 수도 있었고, 높아질 수도 있다는 것을 입증했다.

언어 모델을 통한 통사 연구는 각 장단점을 가지고 있다. 우선 언어 모델은 전이학습에 따라 수식 연산을 통해 통계적인 방법으로 언어 현상을 해석한다. 그에 따라 통사적 현상의 빈도가 높다거나, 어휘의 빈도 유무에 따라 결과값이 달라질 수도 있다. 그러나 큰 수의 법칙에 따라 학습 데이터의 크기가 방대하다면 결과값을 신뢰할 수 있다는 설명이 된다. 또한, 사람을 대상으로 하는 실험은 많은 품이 든다는 것이 사실이다. 현재 가장 활발하게 연구되는 신경망 네트워크 모델을 통해 실험 방법을 먼저 적용하는 것은 언어 연구에 있어 인간의 대상으로 하는 연구만큼 유의미한 결과를 낼 수 있을 것이다.

Reference

- Chomsky, N. (1981). *Lectures on government and binding*. Language, Dordrecht: Foris.
- Dahl, O., & Fraurud, K. (1996). Animacy in grammar and discourse. *Pragmatics and Beyond New Series*, 47–64.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Goldberg, Y. 2019. Assessing BERT's syntactic abilities. arXiv preprint arXiv:1901.05287.
- Lin, Y., Tan, Y. C., & Frank, R. 2019. Open Sesame: Getting Inside BERT's Linguistic Knowledge. arXiv preprint arXiv:1906.01698.
- Linzen, T. 2019. What can linguistics and deep learning contribute to each other? Response to Pater. *Language*, 95(1), e99–e108.
- Marvin, R., & Linzen, T. 2018. Targeted syntactic evaluation of language models. arXiv preprint arXiv:1808.09031.
- McCoy, R. T., Frank, R., & Linzen, T. 2020. Does syntax need to grow on trees? Sources of hierarchical inductive bias in sequence-to-sequence networks. *Transactions of the Association for Computational Linguistics*, 8, 125–140.
- Schütze, C. T. 1996. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago, IL: Univer.
- Sorace, A., & Keller, F. 2005. Gradience in linguistic data. *Lingua*, 115(11), 1497–1524.
- Warstadt, A., & Bowman, S. R. 2020. Can neural networks acquire a structural bias from raw linguistic data?. arXiv preprint arXiv:2007.06761.
- Warstadt, A., Singh, A., & Bowman, S. R. 2019. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7, 625–641. doi: 10.17002/sil.35.201504.69.

CNPC Effects of Why-Stripping in Korean

이두원

(한국교통대학교)

1. 서론

Yoshida et al.(2015)에 따르면, (1B)의 why-스트리핑 구문(why-stripping)은 초점 잔여(focused remnant)인 NATTO가 초점 투사인 Spec-CP로 이동하고 남은 절 TP는 생략(elided)된다. 이때 why는 Spec-CP에서 기저 생성(base-generated)된다. 이와 같이 (2B)의 한국어 ‘왜’-스트리핑 구문도 이처럼 초점 잔여인 ‘사과를’이 초점 투사인 Spec-FP로 이동하고 남은 절 TP는 생략(elided)되며 ‘왜’는 의문사 위치인 Spec-IntP에서 생성된다.

(1) A: John was eating natto.

B: [_{CP1} Why [_{CP2} NATTO [_{TP} ~~he was eating t_{natto}~~]] (and not another food)?

(2) A: 철수가 사과를 먹었어.

B: [_{IntP} 왜 [_{FP} 사과를 [_{CP} [_{TP} ~~철수가 t_{사과를} 먹었나~~]]] (다른 음식이 아니라)?

본고에서 보이고자 하는 것은 이중절의 삽입절 요소가 조각구문(fragment)을 형성할 때, 초점 잔여가 섬을 넘어서 이동하면, 섬 치료 효과(island-repair effect)가 나타나는 반면에, 한국어 ‘왜’-스트리핑 구문은 이러한 섬 치료 효과가 나타나지 않는다는 것이다. 섬 효과를 보이는 한국어 스트리핑 구문들은 (4)-(6)에 예시된다. (4)-(6)의 스트리핑 구문에서 보듯, 선행하는 의문부사인 ‘왜’와 후행하는 초점 잔여는 섬 효과(island effect)를 보인다.¹⁾ 영어의 결속현상은 조각구문에서 섬 효과를 보인다. 이렇듯, PF상에서 생략에도 섬 치료 효과가 나타나지 않을 수 있는데, 본고는 Potter(2017)의 영어 조응어 조각구문의 장거리 결속에 관한 고찰과 Yoshida et al.(2015)의 영어 스트리핑 구문의 연구를 바탕으로 한국어의 스트리핑 구문의 섬 효과 현상을 분석하겠다.

앞으로 3장 3절에서 논의될 것처럼, 한국어 삽입절의 스트리핑 구문은 삽입절의 의문부사 ‘왜’와 초점 요소가 각각 상위절의 Spec-IntP와 Spec-FP에 위치할 때, 남은 두 개의 절이 생략된다.²⁾ (3)의 답문인 스트리핑 구문은 이렇게 형성된다. 그러나 (4)-(6)의 답문에서 의문

1) 의문부사 ‘언제’, ‘어디서’, ‘어떻게’ 또는 부사 ‘아마도’ 등과 후행하는 초점 잔여의 두 요소도 섬을 넘어서 스트리핑 구문을 형성할 때, 섬 효과가 나타난다. 이것은 한국어 스트리핑 구문이 섬 치료 효과가 나타나지 않는다는 또 다른 증거가 된다. 본고는 스트리핑 구문의 섬 효과는 의문부사(즉, 여기서는 ‘왜’)에 의해 야기된다고 주장할 것이다. 이것은 저자가 진행 중인 또 다른 논문의 내용이기도 하다.

2) 앞으로 논의하겠지만, 먼저 이해를 돕기 위해 (4)-(6)의 스트리핑 구문들 중 (6B)의 스트리핑 구문의 도출을 예로 들어 보자. (6B)의 삽입절의 스트리핑 구문의 경우에 삽입절을 지배하는 복합명사구 NP가 섬이다. 이것은 (73c)와 같이 도식화 될 것이다. (6B)의 삽입절 스트리핑 구문의 경우에 삽입절 Spec-CP는 의문사 위치인 Spec-IntP와 초점 위치인 Spec-FP로 분화될 수 없다. 삽입절

부사와 삽입절의 초점 요소가 각각 Spec-IntP와 Spec-FP에 위치할 때, 섬 효과가 나타나서 비문이 된다. 이것은 스트리핑 구문에 나타나는 섬 효과가 조응어에 국한되지 않음을 보인다.³⁾

(3) A: 철수는 영희가 (어제) 자기_{i/j}의 아버지/인호-를 비판했다고 들었다.

B: (왜/)언제 자기_{i/j}의 아버지/인호-를? (삽입절 해석)

B': (왜/)언제 영희가? (삽입절 해석)

(4) A: 철수는 영희가 (어제) 자기_{i/j}의 아버지/인호-를 비판했다는 사실/소문을 들었다.

B: *왜/??언제 자기_{i/j}의 아버지/인호-를? (삽입절 해석)

B': *왜/??언제 영희가? (삽입절 해석)

(5) A: 철수는 [NP [CP 장학회(의) 이사장님이 t_i 준] 장학금_i을] 받았어.

B: *왜/??언제 장학회(의) 이사장님이?

(6) A: 철수는 [NP [CP t_i (수십년 전에) 원주에서 토지를 쓴] 작가_i를] (가장) 만나고 싶어했어.

B: *왜/??언제 토지를?

위에서 보인 한국어 스트리핑 구문들이 PF상에서 섬 치료 효과를 보이지 않는 것은 다음 (7)과 (8)에서 질문 A의 의문사와 조각 답문 B의 초점 잔여가 섬 효과를 보이지 않는 것과 비교할 때, 주목할 만하다. 본고는 한국어 스트리핑 구문들이 섬의 생략에도 왜 여전히 섬 효과를 보이는지를 그 도출 과정을 통해 규명하겠다. 즉 (9B)에서처럼, 영어의 조응어가 PF상에서 섬 효과를 보이듯, 한국어 스트리핑 구문도 PF상에서 섬 효과가 나타난다는 것을 보일 것이다.⁴⁾

(7) A: 너는 [NP [CP 누가 t_i 준] 장학금_i을] 받았니?

B: 장학회(의) 이사장님이.

(8) A: 너는 [NP [CP t_i 어떤 소설을 쓴] 작가_i를] (가장) 좋아하니?⁵⁾

B: 토지를

이 섬인 명사구 NP에 의해 지배되기 때문이다. 이것이 (6B)와 같은 유형의 스트리핑 구문의 비문법성을 야기하고 섬 치료 효과는 나타나지 않는다. 중요한 점은 (6B)와 같은 유형의 스트리핑 구문의 초점 잔여와 (7) 및 (8)과 같은 조각구문의 초점 잔여가 삽입절로부터 상위절 Spec-FP로 이동한다는 것이다. 이 두 구문의 문법성의 차이는 스트리핑 구문을 형성하는 의문부사인 '왜'의 섬 효과로부터 온다. 이것이 본고가 추구하는 목적과 맞닿아 있다.

3) (3A)와 스트리핑 구문인 답문 (3B)에서 주절 주어인 '철수'보다 삽입절 주어인 '영희'가 삽입절 목적어 위치의 '자기'를 더 강하게 결속하는데 이것은 조응어(anaphor) 결속 현상의 일반적 현상이다.

4) (9B)의 조응어는 PF상에서 결속 조건 A를 위반한다. (4)-(6)의 스트리핑 구문들의 의문부사가 PF상에서 섬 효과를 보인다. 즉 초점 잔여들이 조각구문과 스트리핑 구문을 형성할 때, 생략에 의한 섬 치료 효과를 보이지만, 영어의 조응어와 한국어의 의문부사는 PF상에서도 섬 효과를 보인다는 의미이다. 결국, 이것이 한국어 스트리핑 구문의 비문법성을 야기한다. 이것이 본고가 밝히려는 바이다.

5) 부사 '가장'은 선행하는 '어떤 소설을'을 반드시 의문사로 기능하게 한다.

(9) A: James knows the girl who likes someone.

B: *Yeah, [_{CP} [_{himself}_i] [_{TP} James_i knows ~~the girl who_{island} likes e~~]].

2. 조각구문, 수문구문, 그리고 스트리핑 구문

본고는 스트리핑 구문의 섬 효과를 분석한다. 먼저 유사한 생략 구문들을 살펴보자. Merchant(2004)의 조각구문 생략 분석에서는 (10B)에서처럼, 완전한 문장(full-fledged sentence)에서 초점이 주어진 XP가 이동과 생략(movement-and-deletion) 과정을 거쳐서 도출된다.

(10) A: Who killed John?

B: The janitor [_{TP} ~~t_i killed John~~].

수문구문은 스트리핑 구문과는 달리 (11B) 및 (12a)와 같이 의문사만 남고 나머지 문장은 생략된다. (12b)는 초점 잔여인 'an apple'이 'not'과 함께 쓰이는 스트리핑 구문이다.

(11) A: He bought a camcorder.

B: Why [~~did he buy a camcorder~~]?

(12) a. John ate something, but I don't know what [~~John ate~~]. (수문구문)

b. John ate something, but not an apple [~~John ate~~]. (스트리핑 구문)

Yoshida et al.(2015)에 따르면, (1B)의 why-스트리핑 구문은 초점 잔여인 NATTO가 초점 투사인 Spec-CP로 이동하고 남은 절 TP는 생략(elided)된다. 이때 why는 Spec-CP에서 기저 생성된다.⁶⁾

(13=1) A: John was eating natto.

B: [_{CP1} Why [_{CP2} NATTO [_{TP} ~~he was eating t_{natto}~~]]] (and not another food)?

(14B), (15B), 그리고 (16B)를 Merchant(2004)는 조각구문으로 분류한다.⁷⁾ Potter(2017)는 이것들을 스트리핑 구문으로 분류한다. 그러나 여기서는 이것들을 조각구문으로 분류하겠

6) Yoshida et al.(2015)에 따르면, how come-스트리핑 구문은 why-스트리핑 구문보다 덜 수용적(less acceptable)이다.

7) Merchant(2004)는 (i)의 조각구문은 섬 치료 효과가 나타나지 않는다고 보고하는데 그 이유를 Spec-CP에 있는 *t가 생략되지 않아서라고 주장한다.

(i) A: Does Abby speak the same Balkan language that BEN speaks?

B: No, [_{FP} CHARLE_i [_{CP} *t_i [_{TP} ~~she speak the same Balkan language {that t_i speaks}]]]].~~

그러나 본고는 스트리핑 구문, 즉 조각구문이 섬 치료 효과를 보인다는 Potter(2017)의 입장을 따를 것이다. 앞으로 논의하겠지만, 한국어에서 조각구문은 일반적으로 섬 치료 효과를 보이지만, 스트리핑 구문은 영어의 조응어 조각구문 (16B)처럼 PF상에서 섬 효과가 나타난다.

다. 조각구문들은 (14B)에서처럼, 섬 치료 효과가 나타난다. 그러나 (15B)와는 달리, (16B)에서는 조응어의 초점 잔여가 PF상에서 연속순환이동(successive cyclic movement)을 할 수 없어서 섬 효과를 보인다. 이것이 (16B)의 비문법성을 야기한다. 즉 이중절의 삽입절로부터 Fox and Lasnik (2003)의 한번의 이동(one-fell-swoop)이 PF상의 조응어 이동에 섬 효과를 야기한다(Potter, 2017).

(14) A: Tim met [_{island} the researcher who likes *Glenlivet*].

B: No, Ardbeg Tim met [~~_{island} the researcher who likes e~~].

(15) A: James said that Mary likes someone.

B: Yeah, [_{CP} [_{himself}_i] [~~_{TP} James_i said [_{himself}_i] that Mary likes e~~]].

(16) A: James knows the girl who likes someone.

B: *Yeah, [_{CP} [_{himself}_i] [~~_{TP} James_i knows the girl who_{island} likes e~~]].

(17B)에서 조각은 상위절의 초점 위치로 이동하고 나머지는 생략되어 섬 치료 효과가 나타난다. 이것이 바로 영어의 조각구문 (14B)가 보여주는 바이다.

(17) A: [_{CP} John은 [_{NP} 자신의 동생에게 무엇을 준 사람을_{island}] 만났니?]

B: [사과를]_i [~~_{CP} John은 [_{NP} 자신의 동생에게 t_i 준 사람을_{island}] 만났어~~]. (조각구문)

(6B)의 두 스트리핑 구문의 비문법성 정도의 차이는 (18a)와 (18b)의 비문법성(의 정도의 차이)에서 비롯된다. (6B)에서 보듯, 삽입절 의문부사에 대해 생략에 의한 섬 치료 효과는 없다. 언급했듯이, 한국어 스트리핑 구문은 영어의 조응어 조각구문 (16B)처럼 PF상에서 섬 효과가 나타난다. 이것은 스트리핑 구문을 형성할 때, 초점 잔여와는 달리, 의문부사가 PF상에서 섬 효과를 보이기 때문이다.

(18) a. *왜 철수는 [_{NP} [_{CP} t_i 원주에서 토지를 쓴] 작가₁를] 만나고 싶어 했니?

(‘왜’의 삽입절 해석일 경우)

b. ??언제 철수는 [_{NP} [_{CP} t_i 원주에서 토지를 쓴] 작가₁를] 만나고 싶어 했니?

(‘언제’의 삽입절 해석일 경우)

3. 선행 연구들

1. Yoshida et al.(2015)의 Why-스트리핑 구문

언급했듯이, why-스트리핑 구문은 (18B)와 같이 Spec-CP₁에서 why가 생성되고 초점 잔여가 Spec-CP₂로 이동한 후 TP는 생략되어 형성된다(Yoshida et al., 2015). why-스트리핑 구문의 특징은 본래 문장의 특성이 반영되는 연결 효과(connectivity effects)를 보이는데, 그 중 하나가 다음 (19B)의 결속 현상이다. 즉 (19B)와 (19B')는 (20)의 완전한 문장의 문법성 반영이다.

- (19) A: He_i is selling all of these pictures.
 B: Why (even) PICTURES OF HIMSELF_i?
 B': *Why (even) PICTURES OF him_i?
- (20) a. Why is he_i selling pictures of himself_i? (=19B)
 b. *Why is he_i selling pictures of him_i? (=19B')

영어의 재귀대명사, 즉 조응어는 같은 절 또는 명사구 내에서 성분통어(c-command)하는 요소에 의해 결속되어야 하는데(Chomsky, 1981), 이것이 결속조건 A(Principle A)이다. (20a)는 조응어가 주어에 의해 결속되어서 결속조건 A를 충족하여 문법적이고 (20b)는 같은 절 내의 대명사는 자유로워야 한다는 결속조건 B(Principle B)를 위반하여 비문이다. 위에서 보인 why-스트리핑 구문 외에도 다음과 같은 스트리핑 구문들이 있다(Bromberger, 1992; Stepanov and Tsai, 2008).

- (21) A: Jane is learning tango.
 B: Not BALLETT?
- (22) A: What is Jane learning?
 B: Probably TANGO.
- (23) A: What is Jane learning?
 B: Why JANE? (=Why did JANE learn tango?)
- (24) A: What is Jane learning?
 B: Why LEARN? (=Why did Jane LEARN tango?)

선행하는 문장에 내포된 표현(implicit correlate)이 답문 (25B)의 why-스트리핑 구문에 나타날 수 있다. (26)과 같이 why-스트리핑 구문이 삽입절에도 쓰일 수 있다(Yoshida et al., 2015).

- (25) A: John served dinner.
 B: Why/How come (even) TO BILL?
- (26) John served dinner, but I don't understand why/how come (even) TO BILL.

다음에서처럼, PP, CP, AP, VP 등이 why-스트리핑 구문을 형성할 수 있다(Yoshida et al., 2015). (30)은 Collins(1991: 34)에서 인용하였다.

- (27) A: John danced with Mary.
 B: Why/How come WITH MARY?

(28) A: John believes many strange things. One day he said that ghosts exist. Another day he said that trolls exist.

B: Why/How come THAT TROLLS EXIST?

(29) A: John made too weak an espresso.

B: Why/How come TOO WEAK?

(30) Why GO TO THE STORE, when there is orange juice at home?

(31B)에서처럼, 민전치사(bare preposition)도 why-스트리핑 구문을 형성할 수 있다 (Yoshida et al., 2015).

(31) A: Veterans are honored after death, but not before.

B: Why/How come after?

2. Potter(2017)의 PF상에서의 결속과 섬 효과

결속조건 A는 다음 (32)에서 보듯, 장거리 조응어 결속(long distance reflexive binding)을 허용하지 않는다.

(32) *John_i thinks (that) Sue is unlikely to admire [pictures of himself_i].

(Potter, 2017: 267)

그러나, 다음의 예들은 재귀대명사를 포함하는 요소의 이동은 속해 있는 절 밖에서도 재귀대명사가 결속되는 것을 허용한다(Lebaux, 1990; Heycock, 1995; Fox, 2003).

(33) a. [Which pictures of himself_i]_k did John_i think Sue admired t_k?

(Heycock, 1995: 548)

b. John and Bill know [[which picture of themselves]_k Mary bought t_k].

(Fox, 2003: 22)

(34) a. [Pictures of himself_i]_k John_i thinks Mary is unlikely to admire t_k.

(Heycock, 1995: 548)

b. John and Bill know [[the pictures of themselves]_k Mary bought t_k].

(Modeled on (Fox, 2003: 22))

이동의 복사이론(copy theory) 하에서 직전 결속 효과는 직전 이동 착륙 지점의 복사에 의해 설명된다(Fox, 2003). 이때 가장 높은 단계에서 직전 재귀대명사의 복사가 (35)에서 보듯, 최상위절 주어에 의해 결속된다.

(35) [Which pictures of himself_i]_k did John_i [which pictures of himself_i]_k think [which pictures of himself_i]_k that Sue [which pictures of himself_i]_k admired [which pictures of himself_i]_k? (Potter, 2017: 267)

이와 같은 현상이 다음 c의 예들에서도 나타난다. 이것은 LF에서의 결속 기제, 즉 직전 결속 효과가 PF-이동에서도 나타난다는 것을 말해 준다. b의 예들은 같은 절 내에서 국부 결속되는데 반하여, c의 예들은 상위절 주어에 의해 결속된다. (36c)는 (37)과 같이 도출된다. (38c)와 (39c)도 (37)과 같이 연속순환이동에 의해 도출된다. (F는 focus를 의미한다.)

(36) a. Did John_i say that the girl_j admired pictures of Mary_F?
 b. No, pictures of herself_j
 c. No, pictures of himself_i

(37) No, [_{CP} [pictures of himself_i]_k [_{TP} John_i say [~~pictures of himself_i]_k that the girl_j admired [~~pictures of himself_i]_k]]]~~~~

(38) a. John_i thinks that the girl_j admired pictures of someone.
 b. Yeah, pictures of herself_j
 c. Yeah, pictures of himself_i

(39) a. John_i thinks that the girl_j admired pictures of Mary_F.
 b. No, (just) pictures of herself_j.
 c. No, (just) pictures of himself_i. (Potter, 2017: 268-69)

(33), (34) 및 (35)에서 보인 LF상의 연속순환이동이 (37)에서처럼 PF상에서도 적용되어야 한다. 다음 예들에서는 초점 잔여 이동이 섬 효과를 보인다. 이것이 다음 c의 각 예들의 비문법성을 야기한다. 즉 (40c')에서 보인 것처럼, 중간의 섬으로 인한 장거리 재귀사 결속이 비문을 야기한다.⁸⁾ (41c)와 (42c)도 (40c')와 같이 도출되는데 섬 효과를 보여 비문이 된다.

(40) a. Did John_i meet the girl_j who admired pictures of Mary_F?
 b. No, pictures of herself_j
 c. *No, pictures of himself_i
 c'. *No, [_{CP} [pictures of himself_i]_k [_{TP} John_i meet the girl who_{island} admired [~~pictures of himself_i]_k]]]~~

(41) a. John_i met the girl_j who admired pictures of someone.
 b. Yeah, pictures of herself_j.
 c. *Yeah, pictures of himself_i.

8) 여기서 Fox and Lasnik (2003)의 한 번의 이동은 재귀대명사의 결속을 허용하지 않아서 섬 효과를 야기하여 비문을 초래한다.

- (42) a. John_i met the girl_j who admired pictures of Mary_F.
 b. No, (just) pictures of herself_j.
 c. *No, (just) pictures of himself_i. (Potter, 2017: 269-70)

다음의 예들은 조응어 초점 잔여가 섬 효과를 보이는 것이 picture 명사구(picture of NP)의 유형에 국한되지 않는다는 것을 보여준다. (43b)에서 조응어는 같은 절 내에서 국부 결속되는데 반하여, (44b)에서 조응어는 연속순환이동시 상위절 주어에 의해 결속된다. (45b)에서 조응어는 PF상에서 Fox and Lasnik (2003)의 한 번의 이동으로 섬 효과를 야기한다.

- (43) a. James likes someone.
 b. Yeah, himself.
- (44) a. James said that Mary likes someone.
 b. Yeah, himself.
- (45) a. James knows the girl that likes someone.
 b. *Yeah, himself. (Potter, 2017: 270)

다음은 고유명사(proper name)와 같은 지시적 표현(R-expression)은 동지시된 요소들에 의해 성분통어될 수 없다는 Chomsky(1981)의 결속조건 C(Principle C)가 스트리핑 구문에 적용되는 예를 보자.⁹⁾ (46b)의 고유명사인 초점 잔여는 원 위치(base position)에서 고유명사 James가 대명사 he에 의해 성분통어되기 때문에 문장은 비문법적이지만, (47b)의 초점 잔여는 원 위치에서 성분통어되지 않아서 정문이다. 즉 PF에서 결속조건이 적용된다는 의미이다.

- (46) a. He_i admired the professor who supervises Mary.
 b. *No, James_i, he_F admired the professor who supervises e.
- (47) a. One of his_i colleges admired the professor who supervises Mary.
 b. No, James_i, one of his_F colleges admired the professor who supervises e.

이상으로 PF상에서 결속조건 A, B, C가 적용되는 것을 보았다.

9) Chomsky(1981)의 결속조건 C의 영어 정의는 (i)과 같다.

(i) R-expressions, such as proper names, cannot be C-commanded by co-indexed elements.

4. 복합명사구제약과 스트리핑 구문

1. 이중절에서 ‘왜’의 중의성

복합명사구제약과 ‘왜’-스트리핑 구문을 살펴보기 전에 본고의 논의의 목적을 위해 우선 이중절에서 ‘왜’의 중의성을 살펴보겠다. 한국어의 이중 절에서 (48)의 ‘왜’-수문구문은 상위절과 삽입절 사이에서 중의적인데 반하여, (49)와 (50)의 ‘왜’-스트리핑 구문의 생략에 의한 도출과정에 작용하는 기제는 잔여 요소들이 같은 절 내에 있어야 한다는 것이다(Frazier, 1987; Gorrell, 1995; Pritchett, 1992).

(48) A: 철수는 영희가 인호를 비판했다고 말했어.

B: 왜? (중의적)

(49) A: 철수는 영희가 인호를 비판했다고 말했어.

B: 왜 인호를? (삽입절 해석)

(50) A: 철수는 영희가 인호를 비판했다고 말했어.

B: 왜 철수가 (상위절 해석)

(48B)의 ‘왜’-수문구문의 중의성과 같이 (52)의 ‘왜’도 중의적이다. 따라서 (48B)는 (55)와 같이 두 가지로 도출된다. (52)의 이중절에서 ‘왜’는 그 도출 과정이 (54)의 영어 why의 도출 과정과 같이 (55)처럼 도출된다. Yoshida et al.(2015)에 따르면, (51B)에서처럼 영어의 why는 초점 요소와 함께 쓰일 때만 그 절의 Spec-CP에서 기저 생성된다. (52)의 ‘왜’는 (55a, b)에서처럼, 상위절 또는 하위절에서 생성되어 의문사 위치인 상위절 Spec-IntP로 이동한다. 이때 중의적 해석이 나온다.

(51) A: Everyone hates John.

B: Why JOHN <everyone-hates>?

(52) 왜 철수는 [영희가 인호를 비판했다고] 말했니? (중의적)

Shlonsky and Soare(2012)에 의하면, why는 (54a, b)에서처럼, 이중절에서 상위절 또는 하위절에서 생성되어 Spec-CP로 이동한다. 이때 (53)에서 Mary의 말한 이유와 John의 떠난 이유의 두 가지 해석이 나온다.

(53) Why did Mary say John left? (중의적)

(54) a. [_{CP} Why_i t_i did Mary say [_{CP} John left]]? (상위절의 Mary의 말한 이유)

b. [_{CP} Why_i did Mary say [_{CP} t_i John left]]? (하위절의 John의 떠난 이유)

(48B)의 '왜'-수문구문에서 상위절과 삽입절 사이의 중의적 해석을 다시 보자. (55a)와 (55b)에서 상위절 CP이하가 생략된다. 이때 상위절은 Rizzi(1997, 2001)의 분화된 CP-구조(split CP)를 가질 수 있는데, '왜'는 의문사의 지정어(Spec-IntP)에 위치한다(Ko (2005) 참고).¹⁰⁾

- (55) a. [_{IntP} 왜 [_{CP} ~~철수는~~ [_{CP} ~~영희가~~ 인호를 비판했다고] 말했니]]?
 (상위절의 철수가 말한 이유)
- b. [_{IntP} 왜_i [_{CP} ~~철수는~~ [_{IntP/CP} ~~t_i~~ 영희가 인호를 비판했다고] 말했니]]?
 (삽입절의 영희가 비판한 이유)

2. 복합명사구제약

Ross(1967)는 (56a)의 비문법성을 설명하기 위해 명사구(NP)에 의해 관찰되는 문장에 포함된 어떤 요소도 그 NP 밖으로 이동할 수 없다는 복합명사구제약(complex NP constraint, CNPC)을 제안한다.¹¹⁾ 관계절이 삽입된 (56b)의 비문법성도 복합명사구제약으로 설명된다.

- (56) a. *What_i do you know [_{NP} the fact [_{CP} John wrote t_i]]?
 b. *Who_i did you get scholarship_j that t_i gave t_j?

다음은 한국어에 복합명사구제약이 어떻게 적용되는지를 보자. 복합명사구제약은 의문사가 삽입절에 쓰이는 다음 한국어 예문들의 문법성의 차이를 밝히지 못한다(Lee, 2011). Lee(2011)는 한국어에서 복합명사구가 삽입절로부터 이동한 문장에서는 논항의 의문사가 그 삽입절에 쓰일 수 있다고 주장한다. 즉, (57A)와 (57A)에서 삽입절의 주어나 목적어로 쓰이는 의문사의 복합명사구제약 효과는 사라진다. 이것은 영어에는 적용되지 않는다. 이러한 차이는 곧 논의할 언어 간의 섬 위치의 역 방향에 기인한다. 따라서 (56b)는 여전히 복합명사구제약에 의해 비문이 된다.¹²⁾ 그러나 (57)에서 명사구인 '사실을'은 삽입절로부터 이동하지 않아서 의문사에 대한 복합명사구제약 효과가 나타난다.¹³⁾

- (57) a. *너는 [_{NP} [_{CP} 누가 자고 있다는] 사실을] 알고 있니? (누구=의문사)
 b. *너는 [_{NP} [_{CP} 소설가 박경리가 어떤 소설을 쓴] 사실을] 알고 있니?

10) Ko(2005)는 한국어 의문부사 '왜'는 Spec-CP에 병합된다고 주장한다. 그 위치가 Spec-IntP이다. 그녀는 또한 '왜'가 이동할 수 없다고 주장한다. 그러나 본고는 '왜'가 영어의 why처럼, (55b)와 같이 삽입절에서 이동할 수 있다는 입장을 취한다. 이것의 경험적 증거는 (48B)와 (52)이다.

11) Ross(1967)의 복합명사구제약(complex NP constraint, CNPC)의 영어 표현의 정의는 다음과 같다.

(i) No element contained in a sentence dominated by an NP may be extracted from that NP.

12) 더 이상의 복합명사구제약 효과 여부 논의는 Yoon(1998)과 Lee(2011)를 참고하기 바란다. 여기서는 한국어의 복합명사구제약 효과가 일괄적으로 나타나지 않는다는 것을 보이는 것이어서 영어와의 비교는 더 이상 다루지 않겠다.

13) 앞으로 논의하겠지만, (57)-(59)의 문법성 차이와 관계없이 이들 구문에 대한 스트리핑 구문은 모두 비문이다. 이것을 밝히고자 하는 것이 이 글을 쓰는 이유이다.

(어떤 소설=의문사)

(58) A: 너는 [NP [CP 누가 t_i 준] 장학금_i] 받았니? (누구=의문사)

B: 장학회(의) 이사장님이.

(59) A: 너는 [NP [CP t_i 어떤 소설을 쓴] 작가_i] (가장) 좋아하니?

B: 토지를

위의 (58B)와 (59B)에서 보였듯, 한국어의 조각구문은 섬 효과를 보이지 않는다(Park, 2005; Kim, 2013). (61B)와 (62A')에도 섬 치료 효과가 나타난다. 핵-선두 영어와 핵-후미 한국어에서 보이는 섬 효과는 그 강도에서 차이를 보인다. 영어에서 섬은 왼쪽에, 한국어에서 섬은 오른쪽에 위치한다. 즉 한국어에서는 (60)에서 보듯, 섬, 즉 NP인 '사람을'은 이동하는 요소의 역방향에 있다. 물론 이때 통사상 영어나 한국어의 삽입절 Spec-CP는 들러서 이동할 수 없는 섬으로 기능하지만 서로의 역방향이 문법성에 영향을 주는 듯하다.¹⁴⁾ (64A)에서 보듯, 조응어 결속에도 이것이 작용한다. 조각구문의 기존 연구인 Park(2005)과 Kim(2013)은 (60)의 비문법성 표기를 ?*으로 판단하지만 본고는 이보다 약화된 ?정도로 그 비문법성을 판단한다. 앞으로 다시 언급하겠지만, (60)에서 이동한 요소가 미약하게나마 섬 효과를 보이지만, 조각구문에서는 (61B)에서처럼 섬의 생략으로 섬 효과는 사라진다. (57)과 같이 복합명사구제약을 보이는 예문에서도 그 조각 답문은 (62A')에서 보듯, 생략에 의한 섬 치료 효과가 나타난다. 영어의 조응어와는 달리 한국어의 조응어는 장거리 결속이 허용된다. 즉 하위절 Spec-CP를 통한 연속순환이동 없이 조응어 결속이 이루어진다.¹⁵⁾

(60) ?[사과를]_i [CP John은 [NP 자신의 동생에게 (어제) t_i 준 사람을_{island}] 만났어].(61) A: [CP John은 [NP 자신의 동생에게 (어제) 무엇을 준 사람을_{island}] 만났니?]B: [사과를]_i [~~CP John은 [NP 자신의 동생에게 (어제) t_i 준 사람을_{island}] 만났어].~~

(조각구문)

(62) A: 철수는 영희가 (어제) 자기_{i/j}를 비판했다는 사실/소문을 들었다.B: 철수는 영희가 (어제) 누구를 비판했다는 사실/소문을 들었다고?¹⁶⁾A': 자기_{i/j}를 (조각구문)

3. 스트리핑 구문의 복합명사구 제약

Reinhart(1991)는 다음 스트리핑 구문 (63)의 비문법성을 보고한다(또한, Nakao, Yoshida, and Ortega-Santos(2012)를 참고할 것). Yoshida et al.(2015)도 영어의 스트리핑 구문은 복합명사구제약에 효과를 보인다고 보고한다.

14) 역방향성에서 기인하는 문법성의 차이는 본고의 주된 논의가 아니라서 더 이상 다루지 않겠다.

15) 여기서 중요한 점은 한국어 조응어 조각구문은 위에서 살펴본 영어 조응어 조각구문과는 달리 섬 효과를 보이지 않는다는 것이다.

16) 언급했듯이, 반향 의문문은 섬 제약을 보이지 않는다.

(63) *John likes [girls who speak *French*], but *not Italian*.¹⁷⁾

조각구문과는 달리, (64)와 (65)의 문법성 대조에서 보듯, 스트리핑 구문은 섬 치료 효과가 나타나지 않는다.

(64) A: 철수는 영희가 (어제) 자기_{i/j}의 아버지/인호-를 비판했다고 들었다.

B: 왜/언제 자기_{i/j}의 아버지/인호-를? (삽입절 해석)

B': 왜/언제 영희가? (삽입절 해석)

(65) A: 철수는 영희가 (어제) 자기_{i/j}의 아버지/인호-를 비판했다는 사실/소문을 들었다.

B: ?*왜/??언제 자기_{i/j}의 아버지/인호-를? (삽입절 해석)

B': ?*왜/??언제 영희가? (삽입절 해석)

(61B)와 (62A')에서처럼, 복합명사구 제약 효과를 보이지 않는 조각구문과는 달리, (65B, B')의 스트리핑 구문은 섬 효과를 보인다. 즉 조각구문은 생략에 의한 섬 치료 효과를 보이는 반면에, 스트리핑 구문은 생략에 의한 섬 치료 효과를 보이지 않는다. 삽입절의 주어와 목적어 뿐 만 아니라, (66B)와 (67B)에서 보듯, 그 외의 삽입절 성분, 즉 부가어 성분의 스트리핑 구문도 섬 치료 효과를 보이지 않는다.¹⁸⁾ (66B')와 (67B')는 주절 주어의 스트리핑 구문이라서 섬 제약을 받지 않는다.

(66) A: 철수는 [NP [CP t_i 인호에게 저녁을 (차려) 준] 여자들_i] 만났어.

B: *왜 인호에게

B': 왜 철수가

(67) A: 영희는 [NP [CP 철수가 숙희와 춤을 추었다는] 사실을] 알았다.

B: *왜 숙희와

B': 왜 영희가

(68)-(70)의 스트리핑 구문들에서 '왜'는 삽입절과의 연결 효과를 보여서 삽입절에서 생성되어야 하나 삽입절이 복합명사구문이라서 분화된 CP를 가질 수 없다. 이들 스트리핑 구문에서 의문사 위치인 Spec-IntP와 초점 위치인 Spec-FP가 설정될 수 없다. 따라서 '왜'는 삽

17) Culicover and Jackendoff(2005: 244-245)는 부사가 쓰이는 영어의 스트리핑 구문이 복합명사구 제약을 포함하여 섬 제약에 대해 다음과 같이 치료 효과를 보인다고 보고한다.

(i) A: John met a guy who speaks a very unusual language.

B: Yes, probably Albanian.

(ii) A: John met a woman who speaks French.

B: Not Spanish?

이것은 Reinhart(1991)의 (63)에 대한 문법적 판단과 상충된다. 본고는 한국어 스트리핑 구문이 섬 효과를 보인다고 판단한다. 이것은 스트리핑 구문이 섬 효과를 보인다는 면에서 영어의 스트리핑 구문 (63)의 문법적 판단과 그 맥을 같이 한다. 본고의 논지와는 달리 영어의 스트리핑 구문이 섬 치료 효과를 보인다는 입장에서의 논리 전개를 보기 위해서는 Yoo(2014)를 참고할 것.

18) 논의의 목적을 위해 여기서는 여격 요소인 '인호에게'도 부가어 성분으로 표현하겠다.

입절에서 생성될 수 없다. 따라서 초점 잔여가 삽입절로부터 상위절 초점 위치인 Spec-FP로 이동한 후 문장의 나머지 성분이 생략되더라도 스트리핑 구문은 비문이 된다. 본고에서는 삽입절이 복합명사구를 형성할 때, 의문부사 ‘왜’가 삽입절에서 생성될 수 없어서 비문이 되는 스트리핑 구문도 섬 효과를 보인다고 표현한다.

(68) A: 철수는 [NP [CP 장학회(의) 이사장님이 (어제) t_i 준] 장학금_i] 받았어.

B: *왜/??언제 장학회(의) 이사장님이?

(69) A: 철수는 [NP [CP t_i 수십년 전에 원주에서 토지를 쓴] 작가_i] (가장) 만나고 싶어했다.

B: *왜/??언제 토지를?¹⁹⁾

(70) A: John은 [NP [CP 자신의 동생에게 (어제) 사과를 준] 사람_i] 만났어.

B: *왜/??언제 사과를?

위에서 예시한 조각구문 (61B)와 스트리핑 구문 (69B)의 도출과정을 비교해 보자. (61B)에서 조각은 상위절의 초점 위치로 이동하고 나머지는 생략되어 섬 치료 효과가 나타난다.²⁰⁾ 이것은 생략에 의해 나타나는 섬 치료 효과와 무관한 (60)의 비문법성과 대비된다. (69B)의 스트리핑 구문은 (73c)와 같이 도출되는데, 이때 생략에 의한 섬 치료 효과는 없다. (73a)에서 보듯, 명사구에 의해 지배되는 삽입절의 CP는 그 위로의 층위(layer)를 형성할 수 없어서 삽입절 내에서 ‘왜’가 생성될 수 있는 Spec-IntP가 없다. 따라서 (69B), 즉 (73c)는 복합명사구의 섬 제약 때문에 도출될 수 없다. 삽입절 VP에 부가되는 의문부사 ‘언제’ 이것을 증명한다. (73c)의 스트리핑 구문이 보이는 비문법성의 차이는 (73a)에서 보듯, ‘왜’와 ‘언제’의 복합명사구 제약으로 인한 비문법성의 차이의 반영이다. (72B)에서 보듯, 초점 조각은 섬 치료 효과를 보인다는 맥락에서 (73c)의 스트리핑 구문의 초점 잔여도 생략에 의한 섬 치료 효과가 나타난다고 볼 수 있다. 그렇다면, (73c)의 스트리핑 구문의 섬 효과는 PF상에서 의문부사로 인해 발생한다고 볼 수 있다.²¹⁾

19) 스트리핑 구문으로의 답변 (i)은 문법적이다. 언급했듯이, 의문사는 상위절에서 생성되고 이동하는 요소는 섬을 넘어서 이동하지 않기 때문이다.

(i) 왜 철수가

20) (iB)의 스트리핑 구문이 문법적으로 도출될 수도 있다는 의문이 생길 수 있다. 즉 (iB)에서 ‘왜’가 상위절에서 생성된다고 가정하면, 조각 ‘사과를’은 (63B)처럼 도출 과정에서 섬 치료 효과가 나타날 수 있다고 가정할 수 있다.

(i) A: [CP John은 [NP 자신의 동생에게 사과를 준 사람_iisland] 만났어].

B: *왜 사과를.

그러나 의문사 ‘왜’가 상위절 Spec-IntP에서 생성되고 조각이 하위절로부터 초점 이동을 겪는다는 논리는 위에서 언급했듯이, 잘 못된 것이다. 생략에 의한 잔여 요소들의 도출과정에 작용하는 기제는 잔여 요소들(즉, (iB)의 ‘왜’와 ‘사과를’)이 같은 절 내에 있어야 한다(Frazier, 1987; Gorrell, 1995; Pritchett, 1992). (73)에서 보였듯이, 스트리핑 구문이 조각구문의 도출과는 달리 섬 효과를 보이는 것은 의문부사 ‘왜’의 섬 효과가 생략 후에도 나타나기 때문이다. 즉 스트리핑 구문은 의문부사 ‘왜’ 때문에 PF상의 섬 효과를 보인다.

21) (73b)의 초점 잔여인 ‘토지를’의 섬 효과, 즉 비문법성은 (72B)에서처럼 (73c)의 생략에 의해 치료된다고 본다.

(71=60) ?[사과를]_i [_{CP} John은 [_{NP} 자신의 동생에게 (어제) _{t_i} 준 사람을 island] 만났어].

(72=61) A: [_{CP} John은 [_{NP} 자신의 동생에게 무엇을 준 사람을 island] 만났니?]

B: [사과를]_i [_{CP} John은 [_{NP} 자신의 동생에게 _{t_i} 준 사람을 island] 만 ~~났어~~]. (조각구문)

(73) a. [*왜/??언제 [_{CP} 철수는 [_{NP} [_{CP} _{t_i} 원주에서 토지를 쓴] 작가_i를] (가장) 만나고 싶어 했니]]? (의문부사의 삽입절 해석)

b. [_{IntP} *왜/??언제 [_{FP} (??)토지를 [_{CP} 철수는 [_{NP} [_{CP} _{t_i} 원주에서 _{t_{토지를}} 쓴] 작가_i를] (가장) 만나고 싶어했니]]]? (의문부사와 초점 잔여의 삽입절 해석)

c. *왜/??언제 토지를 [_{CP} 철수는 [_{NP} [_{CP} _{t_i} 원주에서 _{t_{토지를}} 쓴] 작가_i를] (가장) 만나고 싶어했니]? (스트리핑 구문의 삽입절 해석)

5. 결론

한국어의 '왜'-스트리핑 구문에서 의문부사 '왜'는 삽입절에서 생성되어 '왜'를 후행하는 초점 잔여가 상위절 Spec-FP로 초점이동을 한다. 그리고 남은 주절과 삽입절은 생략된다. 스트리핑 구문에서 이처럼 의문부사와 초점 잔여가 상위절 Spec-IntP와 Spec-FP로 각각 이동하는데, 이때 섬을 포함한 나머지 문장 성분들이 생략되어도 섬 치료 효과는 나타나지 않는다. 왜냐하면, 복합명사구 NP에 의해 지배되는 삽입절에 의문부사 '왜'가 쓰일 수 없기 때문이다. 이것이 복합명사구 제약 효과가 나타나는 문장이 삽입절로부터 '왜'-스트리핑 구문을 형성할 수 없는 이유이다. 섬을 포함한 두 개의 절이 생략되더라도 섬 치료 효과는 나타나지 않는다. 결국, 본고는 한국어에서 삽입절 요소가 조각구문을 형성할 때, 초점 잔여가 섬을 넘어서 이동하면, 섬 치료 효과를 보이는 반면에, 한국어 스트리핑 구문에서는 이러한 섬 치료 효과를 보이지 않는다는 점을 보였다. PF상에서 영어의 조음어 조각이 섬 제약으로 연속순환이동을 하지 못하면 섬 효과를 야기한다. 영어의 이중절에서 조음어 잔여가 PF상에서 삽입절로부터 연속순환이동을 할 수 없다면, 그 조각구문은 섬 효과를 보인다. 즉 Fox and Lasnik(2003)의 한 번의 이동이 PF상의 영어 조음어 이동에 섬 효과를 야기한다. 이러한 측면에서 한국어 스트리핑 구문에서는 섬 치료 효과를 보이지 않는다는 점을 보였다.

참고문헌

- Chomsky, N. 1981. *Lectures on Government and Binding*. Foris: Dordrecht.
- Collins, C. 1991. Why and How Come. In L. S. Cheng & H. Demirdache (Eds.), *More Papers on Wh-Movement: MIT Working Papers in Linguistics 15*, 31-45. Cambridge: MIT Press.
- Fox, D. 2003. On Logical Form. In R. Hendrick (Ed.), *Minimalist Syntax*. UK: Blackwell. 82-123.
- Culicover, P. W. and R. Jackendoff. 2005. *Simpler Syntax*. New York: Oxford University Press.
- Fox, D. 2003. On Logical Form. In R. Hendrick (Ed.), *Minimalist Syntax*. Oxford: Blackwell. 82-123.
- Fox, D. and H. Lasnik. 2003. Successive-Cyclic Movement and Island Repair: The Difference between Sluicing and vp-Ellipsis. *Linguistic Inquiry* 34(1), 143 - 154.

- Frazier, L. 1978. *On Comprehending Sentences: Syntactic Parsing Strategies*. Unpublished Ph. D. Dissertation. University of Connecticut, Storrs.
- Gorrell, P. 1995. *Syntax and Parsing* (Vol. 76). Cambridge: Cambridge University Press.
- Heycock, C. 1995. Asymmetries in Reconstruction. *Linguistic Inquiry* 26(4), 547-570.
- Kim, J.-B. 2017. Why-Stripping in English and Korean: A Direct Interpretation Approach. *Studies in Generative Grammar* 27(4), 733-751.
- Kim, R. 2013. On the Negativity of Negative Fragment Answers and Ellipsis Resolution. *Studies in Generative Grammar* 23(3), 447-467.
- Ko, H. 2005. Syntax of Why-in-situ: Merge into [Spec, CP] in the Overt syntax. *Natural Language and Linguistic Theory* 23(4), 867-916.
- Lebaux, D. 1990. Relative Clauses, Licensing, and the Nature of the Derivation. In J. Carter (Ed.), *Proceedings of NELS 20*, 318 - 332. Amherst: GLSA.
- Lee, D.-W. 2011. Presence and Absence of the Complex NP Island Effects. *Studies in Linguistics* 21, 255-276.
- Merchant, J. 2001. *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford: Oxford University Press.
- Merchant, J. 2004. Fragments and Ellipsis. *Linguistics and Philosophy* 27, 661-738.
- Nakao, C., M. Yoshida, and I. Ortega-Santos. 2012. On the Syntax of *Why*-Stripping. In N. Arnett and R. Bennet (Eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics*, 270-280. Somerville, MA: Cascadilla Proceedings Project.
- Park, B.-S. 2005. Island-Insensitive Fragment Answers in Korean. In J. Alderete et al. (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics*, 317-325. Somerville, MA: Cascadilla Proceedings Project.
- Potter, D. 2017. *The Island (In)sensitivity of Stripping*. Ph. D. Dissertation. Northwestern University, Evanston, Illinois.
- Pritchett, L. 1992. *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.
- Reinhart, T. 1991. Elliptic Conjunctions - Non-Quantificational LF. In A. Kasher (Ed.), *The Chomskyan Turn*. Oxford: Blackwell. 360-384.
- Rizzi, L. 1997. The Fine Structure of the Left Periphery. In L. Haegeman (Ed.), *Elements of Grammar: A Handbook of Generative Syntax* Dordrecht: KluIr. 281-337.
- Rizzi, L. 2001. On the Position "Int(errogative)" in the Left Periphery of the Clause. In G. Cinque and G. Salvi (Eds.), *Current Studies in Italian Syntax: Essays Offered to Lorenzo Renzi*. Amsterdam: Elsevier. 267-296.
- Ross, J. R. 1967. *Constraints on Variables in Syntax*. Ph. D. Dissertation. MIT.
- Shlonsky, U. and G. Soare. 2011. Remarks and Replies: Where's 'Why'? *Linguistic Inquiry* 42(4), 651-669.
- Stepanov, A. and W.-T. D. Tsai. 2008. Cartography and Licensing of Wh-Adjuncts: A Cross-linguistic Perspective. *Natural Language and Linguistic Theory* 26, 589-638.
- Yoo, E.-J. 2014. A Base-Generated Approach to Why-Stripping in English. *Korean Journal of Linguistics* 39(4), 921-943.
- Yoon, J.-M. 1999. Decomposing WH-Questions: WH-marking and WH-scoping Movement. *Studies in Generative Grammar* 9(1), 153-198.
- Yoshida, M., Nakao, C., and I. Ortega-Santos. 2015. The Syntax of Why-Stripping. *Natural Language and Linguistic Theory* 33(1), 323-370.

The Split VP Hypothesis Re-Revisited

Kyungchul Chang

(Pusan National University)

In this talk, we review Chang's (2017) article on the split VP hypothesis that the VP headed by a transitive verb such as *read an ebook* splits into at least two layers: VP core and v*P shell. In this article, it is argued that the verb-final account for the VP core such as [read [an ebook <read>]] should be equally valid, compared with the verb-initial standard such as [read [<read> an ebook]], as long as we pursue the fundamental minimalist assumption that merge is an unordered set operation (Chomsky 1995), that is, if there is no ordering of merged elements in a set such as $\{\alpha, \beta\}$. It is also shown that the former approach may outperform the latter, at the very least within the v*P domain, particularly with regard to the syntactic alternation of English phrasal verbs and their gapping in coordinate clauses. In this regard, we touch on the literature which supports the V-final VP core hypothesis, including Chang (2016), Edelstein (2020), van Gelderen (2013) and Sportiche (2005). In this talk, we also look at another idea of Chang's (2017) that the V-final VP core is derived from what is called a 'synthetic noun-verb (N-V) compound verb' (e.g. [ebook read] for *ebook reading* and *ebook reader*). Then, we discuss what kinds of data could support this hypothesis.

References

- Chang, K. 2016. On the Syntactic Alternation of English Transitive Phrasal Verbs. *Studies in Modern Grammar* 88, 1-18.
- Chang, K. 2017. The Split VP Hypothesis Revisited. *The New Korean Journal of English Language and Literature* 59, 261-279.
- Chang, K. 2018. English Phrasal Verbs Are Not Primitives. *Studies in Modern Grammar* 100, 41-63.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Edelstein, E. 2020. *A Minimalist Account of Structure and Variation*. Edinburgh: Edinburgh University Press.
- van Gelderen, E. 2013. *Clause Structure*. Cambridge: Cambridge University Press.
- Sportiche, D. 2005. Division of Labor between Merge and Move: Strictly Locality of Selection and Apparent Reconstruction Paradoxes, ms. University of California Los Angeles.

Two types of node-sprouting

한국언어과학회 겨울학술대회

2021년 2월 18일

발표자: 대구대학교 최재훈

The goals of this paper

- To investigate how Distributed Morphology(DM) deals with agreement and concord.
- Through the lens of Korean discourse particle -yo.

•연구목적

- 본 연구는 agreement와 concord 현상을 분산형태론의 틀 안에서 더 정확하게 이해하고 문법적 기제를 마련하는 것을 목표.
- 이러한 목표 달성을 위해 본 연구에서는 agreement와 concord 현상의 논의에서 주목받지 못하는 한국어를 경험적 근거로 삼음.

Main proposals

- Both agreement and concord are post-syntactic phenomena.
- Both involve node-sprouting(Choi & Harley 2019)
- Agreement
 - Node-sprouting targets morphological structure by phase.
- Concord
 - Node-sprouting targets prosodic structure across phase.

Puzzle

- The discourse particle –yo in Korean exhibits interesting distribution.

- 1) 인호가 어제 서울에서 친구를 만났어.
- 2) 인호가 어제 서울에서 친구를 만났어요.
- 3) 인호가(요) 어제(요) 서울에서(요) 친구를(요) 만났어*(요).

Puzzle

- The discourse particle **-yo** in Korean exhibits interesting distribution.

- 1) a. *(가방이 바로 여기**요** 있어요.)
b. (가방이 있어요), (바로 여기**요**.)
- 2) a. *(이것을 당장**요** 처리하세요.)
b. (당장**요**.)
- 3) a. *(그건 그 사람 잘못**이요** 아니에요.)
b. (그건 그 사람이 잘못**이요**) (절대로 아니에요.)

A prosodic analysis of **-yo**

- Yim & Dobashi (2016):
 - **-yo** is attached to prosodic boundary.
 - No constraint in the narrow syntax.
 - **-yo** can be overgenerated in syntax, but filtered out at PF.
- Problem
 - 1) *인호가(요) 어제(요) 서울에서(요) 친구를(요) 만났어.

DM analysis of -yo

- The sentence-final -yo sprouts on the closest phase head when c-commanded by an addressee argument.
- The sentence-medial -yo sprouts on the prosodically independent unit (i.e., an internally-complex opaque head) across phases when c-commanded by an addressee argument.

Zoom 3

코퍼스/담화/텍스트

<https://pusan.zoom.us/j/9709769748?pwd=TU5ucGxKcFBVkpWYkpSOXk5Sm5HUT09>

회의 ID: 970 976 9748

암호: 123456



키워드 네트워크 분석을 통한 글로벌 해양저널의 주요 주제와 경향 분석
(Prominent Themes and Trends in Global Maritime Journals
Using Keyword Network Analysis)

2021. 02. 18

2021 한국언어과학회 온라인 겨울학술대회
(주제: 빅데이터와 언어학: 현재와 미래)

장세은 (한국해양대)

본 발표는 아래 논문의 일부분을 발췌한 것임

Jhang, Se-Eun · Ho-Min Park · Yaochen Deng · Sunghwa Lee (2020) Themes and Trends in Global Maritime Journals Using Keyword Network Analysis, *KMI International Journal of Maritime Affairs and Fisheries*, 12(2): 37-81.



Contents

- Abstract & Keywords
(Purpose and New Findings of Research)
- Previous Studies
- Data and Methodology
- Results and Discussion
- Conclusion and Suggestions
- References

Abstract & Keywords



- This study **identifies** research themes and trends of international journal data in global maritime affairs, fisheries, marine and transport policy, and logistics over the last 20 years from 2000 to 2020 using keyword network analysis through degree centrality.
- This study pays special attention to **six different types of patterns** through the **Delta-C algorithm**. First, we discuss highly remarkable research themes that are shared throughout all the three periods defined as Type A. Second, we focus on interest-increased, interest-decreased, and newly emerging research themes shown in the most recent period (the third period) from Type B to Type F.
- This study shows **two new findings**. First, in **Type A** representing **consistently shared themes**, the main research themes change from *growth* and *fishery management* in fisheries and *sustainability* and *governance* in maritime sectors in the 2000s; to *growth* and *aquaculture* in fisheries and *accessibility*, *China* and *sustainability* in maritime sectors in the early 2010s; and to *aquaculture* and *growth* in fisheries and *accessibility*, *climate change*, and *China* in maritime sectors in the late 2010s. Second, in **Type F** as **new trends**, the top 10 keywords in newly issues illustrate that issues surrounding *climate change* and *Green House Gas emission* attract more attention in the literature, the subjects of *machine learning* and *artificial Intelligence (AI)* become popular in accordance with the development of internet of things (IoT) in the late 2010s, and *Belt-Road initiative* demonstrates the enlargement of China's economic potential in the 2010s.
- **Keywords:** international maritime journals, trend analysis, keyword network analysis, degree centrality, Delta-C algorithm, **data**

Previous Studies on *Network Analysis and Trend Analysis*



- Lee and Kang (2011): network analysis used in the field of **Sociology**
 - two-way mode data, such as articles and keywords, journals and keywords, and journals and authors
 - articles from 20 academic journals in the Korean Journal Citation Index from 2004 to 2020
 - by analyzing degree centrality and betweenness centrality, they revealed **what journals** retained high centrality.
- Choi and Kang (2014): research trend analysis in the **Korean Educational Technology** using keyword network analysis
 - by examining 645 articles in the Journal of Educational Technology for three time periods: 1985-1994, 1995-2004, and 2005-2013.
 - Keywords associated with 'structuralism' have steadily increased and social media-related keywords have emerged in recent years.
- Jhang et al. (2015): research themes and trends of the **offshore industry** by analyzing author keywords
 - A total of 800 articles – 200 articles per period – were selected for the analysis.
- Jhang and Lee (2016): themes and trends in **maritime economics and logistics** by analyzing keywords from 303 articles in international journals from 2000 to 2014
 - invention of the **Delta-C algorithm** to identify the differences in centrality by five-year periods
 - Degree centrality revealed research themes whereas betweenness centrality disclosed newly emerging themes in each period.

More Recent Previous Studies

- Khitous et al. (2020): themes and emerging research trends in circular economy (CE)
 - by utilizing citation network analysis (based on references), keywords co-occurrence network, global citations score, and burst detection of keywords, based on references and keywords systematic literature network analysis.
 - Eight main trends of CE research that have been dominated by environment and engineering scholars.
- Bai et al. (2020): themes and trends in e-learning research
 - by analyzing keywords collected from 7214 articles published in 10 journals over two decades by dividing into two time periods (1999-2008 and 2009-2018)
 - Knowledge constructions were visualized by employing a clustering method, a social network analysis, and a strategic diagram; popular topics, core topics and bridge topics were presented.

Delta-C algorithm

- Based on Jhang and Lee (2016), this study further develops the Delta-C application to be able to identify themes and trends as distinguishable types according to increasing and/or decreasing interest as well as newly emerging themes in a specific period and shared themes in all the periods.
- This Delta-C extended application allows us a new way to closely investigate various trends of research.

Data and Methodology

- Data
 - selection of **15 international maritime journals**
 - extraction of 26,359 articles of the journals listed in the **Web of Science** by entering key words such as *maritime, fishery, marine environment, port, shipping, shipbuilding*, etc. into the search bar in the periods of **2000 to 2020**
 - These 15 journals were confirmed by maritime and marine specialists who are working at Korea Maritime Institute (KMI)

A list of keywords was made in an individual journal per period in order to obtain information for synchronic comparison with other journals and diachronic comparison of other periods for remarkable themes and trends

Figure 2. A screen shot of a list of keywords per journal in the third period (2015–2020)

Journal Title/Keywords	Freq	Per thousand	Freq	Per thousand	Freq	Per thousand	Freq	Per thousand	Freq	Per thousand
Maine Policy										
Fishery management	105	14.50476585	309	14.107656	21	9.067357513	63	11.08374	15	23.21981424
Fishery	91	12.57079707	292	12.874949	14	6.044965009	47	8.268825	9	13.93188854
maine protected area	56	12.43265844	200	8.1311692	12	5.18134715	43	7.565095	9	13.93188854
maine spatial planning	54	8.841000138	135	6.1633592	11	4.78568221	36	6.33956	7	10.8591331
small-scale fishery	52	8.012156375	114	5.2847865	9	3.88810363	33	5.89771	6	9.287925697
governance	50	6.90701356	104	4.740268	8	3.454231434	33	5.89771	6	9.287925697
aquaculture	47	6.492609476	98	4.4742729	8	3.454231434	31	5.43960	6	7.39930808
climate change	44	6.078167595	92	4.2003379	8	3.454231434	29	5.02041	4	6.191950464
sustainability	38	4.834921951	88	4.0177145	7	3.022452504	26	4.926108	4	6.191950464
conservation	28	4.006078188	87	3.9720586	7	3.022452504	26	4.574243	4	6.191950464
management	28	3.66793756	81	3.6981235	7	3.022452504	26	4.574243	4	6.191950464
co-management	25	3.591656306	77	3.5155002	7	3.022452504	24	4.22379	3	4.64362848
ecosystem service	24	3.315375052	68	3.1045975	7	3.022452504	23	4.04644	3	4.64362848
china	23	3.177234425	68	3.1045975	6	2.590673575	22	3.870514	3	4.64362848
ecosystem-based management	23	3.177234425	65	2.96763	6	2.590673575	21	3.694581	3	4.64362848
stakeholder	21	2.90095317	64	2.9219742	6	2.590673575	21	3.694581	3	4.64362848
bycatch	21	2.90095317	62	2.8763183	6	2.590673575	20	3.518649	3	4.64362848
sustainable development	20	2.762812543	62	2.8066625	6	2.590673575	17	2.990852	3	4.64362848
stakeholder engagement	20	2.762812543	58	2.6480891	6	2.590673575	16	2.814919	3	4.64362848
policy	19	2.624671916	58	2.6480891	6	2.590673575	15	2.638987	3	4.64362848
blue growth	19	2.48633289	56	2.5887274	6	2.590673575	15	2.638987	3	4.64362848
indica	18	2.48633289	54	2.4654157	6	2.590673575	15	2.638987	3	4.64362848
common fishery policy	17	2.348390662	48	2.1914806	6	2.590673575	14	2.463064	2	3.095975232
monitoring	15	2.21055035	47	2.194406	6	2.590673575	14	2.463064	2	3.095975232
european union	15	2.21055035	47	2.1458248	5	2.156894646	13	2.287122	2	3.095975232
maine governance	15	2.21055035	47	2.1458248	5	2.156894646	13	2.287122	2	3.095975232
artisanal fishery	15	2.072109407	46	2.1001689	5	2.156894646	13	2.287122	2	3.095975232
food security	15	2.072109407	46	2.1001689	5	2.156894646	13	2.287122	2	3.095975232

Note that the number of keywords per thousand is needed to be compared with the frequency of keywords of journals due to different numbers of articles in even the same period.

The order of ranking based on the normalized number of keywords per thousand differs as MRE is jumped to first place from fourth place, as follows: MRE (13.93), Aquaculture (12.87), MP (6.49), OCM (3.69), and AJAE (1.29).

Table 2. Periodic information

	P1 (2000–2009)	P2 (2010–2014)	P3 (2015–2020)	Total
Number of articles	7,445	6,376	12,538	26,359
Number of author keywords as type	37,554	32,442	65,535	135,531
Number of author keywords as lemma	15,878	16,439	29,826	62,143 ⁶
Type and Lemma Ratio	0.4233	0.5067	0.4551	0.4617

6 Lemma number of author keywords given in Table 1 was 60,569. However, it is natural that this number differs as the number of lemmas of each period may overlap with that of each journal.

Methodology

Degree Centrality

Degree centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has) according to Freeman (1978/1979).

The higher the degree, the more central the node is.

Degree is a simple centrality measure that counts how many neighbors a node has.

Freeman's general formula for centralization

$$C_D = \frac{\sum_{i=1}^n [C_D(n) - C_D(i)]}{[(N-1)(N-2)]}$$

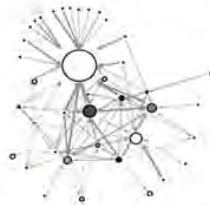
maximum value in the network

where n = number of points

$C_x(P_i)$ = one of the point centralities defined above

$C_x(P^*)$ = largest value of $C_x(P_i)$ for any point in the network

(1a) high centralization: one node trading with many others



(1b) low centralization: trades are more evenly distributed



In **Netminer 4.4**, degree centrality is calculated through the main menu of Analyze >> Neighbor >> Degree.

For visualization of networks, setting up a threshold of minimum frequency of **10-times** occurrences so that we have **a total of 1,149 keywords**

71

Delta-C

- Although degree centrality is a good measure of the total connections a node has, it does not show difference between quantity and quality. That is, degree centrality does not necessarily indicate the importance of a node in connecting others or how central it is to the main group.
- So this study utilizes Delta-C in order to examine research trends by identifying the differences of degree centrality in specific periods.
- Delta-C is an algorithm that was initially proposed by Jhang and Lee (2016), in which C is the abbreviation of Centrality.

The formula of the Delta-C

$$\Delta D(C) = \frac{E(c) - O(c)}{N}$$

$E(c)$ refers to the period of the most recent year or years.

$O(c)$ refers to the period of the past year or years, and

N denotes the sum of the total centrality.

A **plus value** represents higher degree centrality of the recent years, whereas a **minus value** represents higher degree centrality of the past.

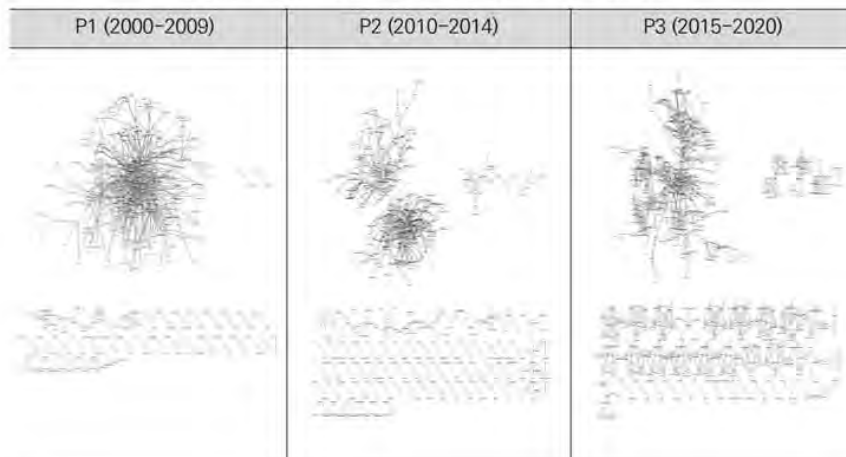
Table 3. Types of keywords categorized by Delta-C value of periods

TYPE	P1 (2000–2009)	P2 (2010–2014)	P3 (2015–2020)	Delta-C
A	Presence of DC	Presence of DC	Presence of DC	*9
B	No Occurrence of Keywords	Presence of DC	Higher DC than P2	$P3 - P2 = \text{Plus value}$
C	No Occurrence of Keywords	Presence of DC	Lower DC than P2	$P3 - P2 = \text{Minus value}$
D	Presence of DC	No Occurrence of Keywords	Higher DC than P1	$P3 - P1 = \text{Plus value}$
E	Presence of DC	No Occurrence of Keywords	Lower DC than P1	$P3 - P1 = \text{Minus value}$
F	No Occurrence of Keywords	No Occurrence of Keywords	Newly introduced keywords	N/A

Six types of trends are selected by focusing on the presence of keywords occurring in P3.

Results and Discussion

Figure 3. One-mode keyword networks of three periods



- Two-mode networks of keywords and paper titles are converted into one-mode keyword networks.
- Each period has one giant clump (giant component = 대집단 클럼프/컴포넌트).
- The second period features two giant clumps, where the node *fishery* connects with the node of *aquaculture* in the other giant clump, and the most stings (선형).
- The third period has the most cliques (파당).
- Let us take **one giant clump** composed of interesting nodes in the following discussion.

Table 4. Keyword information for drawing networks of degree centrality in three periods

P1 (2000–2009)	In-Degree Centrality	Out-Degree Centrality
1. growth	0.336283186	0.336283186
2. atlantic salmon	0.075221239	0.075221239
...
226. sea cucumber	0.004424779	0.004424779
227. phytase	0.004424779	0.004424779
P2 (2010–2014)		
1. growth	0.154929577	0.154929577
2. marine protected area	0.049295775	0.049295775
...
284. coral reef	0.003521127	0.003521127
285. arctic	0.003521127	0.003521127
P3 (2015–2020)		
1. management	0.074324324	0.074324324
2. shipping	0.067567568	0.067567568
...
148. sea level rise	0.006756757	0.006756757
149. complex system	0.006756757	0.006756757

For undirected networks, this study does not need two measures of degree (in-degree and out-degree) because there is **no direction between a node of keywords and a node of articles**

Type A as consistently shared themes

Table 5. Top 30 keywords sorted by high degree centrality in each period

P1 (2000–2009)	P2 (2010–2014)	P3 (2015–2020)
growth*	accessibility	aquaculture*
fishery management	growth*	growth*
aquaculture*	transport*	management
co management	China	accessibility
fishery	shipping	climate change
atlantic salmon*	travel behavior	fishery
temperature	aquaculture*	China
shrimp	sustainability*	growth performance
rainbow trout	GI	sustainability
fatty acid	commuting	governance*
survival	mobility	marine protected area*
larva	cycling	conservation
marine protected area*	port	fishery management
reproduction	liner shipping	shipping
fish	governance*	public transport
lipid	mode choice	port
sustainability*	container shipping	fish
digestibility	climate change	built environment
nutrition	travel behaviour	policy
governance*	walking	atlantic salmon*
protein	transport logistics	liner shipping
salmon	container terminal	gene expression
participation	high speed rail	impact
salmo salar	airport	land use
transport*	atlantic salmon*	mobility
tilapia	public transport	container terminal
stress	logistics	transport*
feeding	supply chain	rainbow trout
microsatellites	time geography	competition
penaeus monodon	marine protected area*	Europe

Seven keywords with an asterisk* are shared throughout the three periods within the top 30 keywords sorted by high degree centrality of each period: *aquaculture, Atlantic salmon, governance, growth, marine protected area, sustainability, and transport.*

The main research themes change from *growth and fishery management in fisheries and sustainability and governance in maritime sectors in the 2000s*; to *growth and aquaculture in fisheries and accessibility, China and sustainability in maritime sectors in the early 2010s*; and to *aquaculture and growth in fisheries and accessibility, climate change, and China in maritime sectors in the late 2010s.*

Table 6. Occurrence frequency of shared keywords in each period

Shared keywords	P1 (2000–2009)		P2 (2010–2014)		P3 (2015–2020)	
	Freq.	Normalized Freq.	Freq.	Normalized Freq.	Freq.	Normalized Freq.
aquaculture	206	5.485	156	4.808	410	6.256
atlantic salmon	178	4.739	93	2.866	140	2.136
governance	20	0.532	50	1.541	150	2.288
growth	528	14.059	227	6.997	319	4.867
marine protected area	27	0.718	90	2.774	166	2.532
sustainability	36	0.958	73	2.250	169	2.578
transport	33	0.878	47	1.448	45	0.686

➤ **Frequency-based analysis** indicates that blue-colored keywords show the most occurrences in each period in alphabetic order.

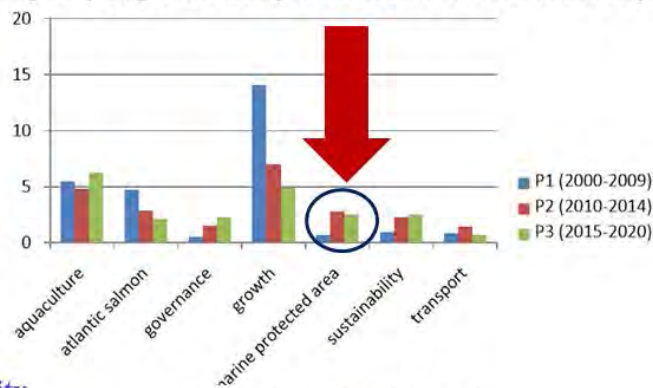
➤ However, what about the gray-colored keyword, *marine protected area*?

Answer: A theme of *marine protected area* is more prominent in P2 than P3.

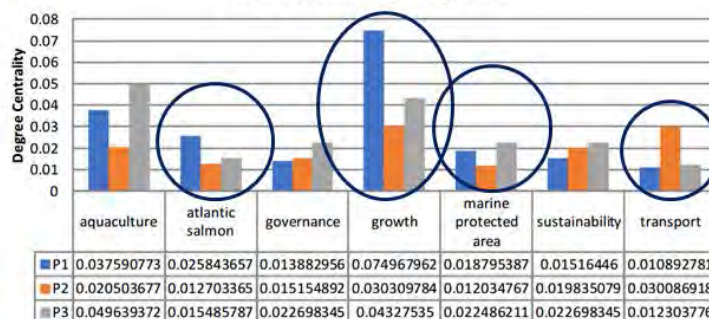
The most straightforward approach is to make a comparison of **normalized frequency**

Figures 4-5. Frequency/Degree centrality based trends of seven shared keywords over time

Normalized Frequency based trends



Degree centrality based trends



Type B as highly interest-increased themes

Table 7. Top 10 keywords belonging to Type B

Keywords (2015-2020)	Degree Centrality	Freq.	Normalized Freq. ¹⁰	Delta-C against P2 (2010-2014)	Delta C against P1 (2000-2009)
1. ecosystem service	0.009546033	80	1.220	0.001180780	no occurrence
2. big data	0.006151888	27	0.411	0.001007571	no occurrence
3. public transit	0.008061095	32	0.488	0.000826014	no occurrence
4. mixed integer linear programming	0.004242681	26	0.396	0.000697701	no occurrence
5. inland port	0.004879084	11	0.167	0.000691785	no occurrence
6. maritime	0.004242681	21	0.320	0.000656749	no occurrence
7. buyer supplier relationship	0.004242681	20	0.305	0.000656749	no occurrence
8. maritime transport	0.009970301	33	0.503	0.000644456	no occurrence
9. customer satisfaction	0.005939754	16	0.244	0.000640972	no occurrence
10. transit	0.004879084	15	0.228	0.000568928	no occurrence

Degree centrality value of a theme, *maritime transport* (0.00997001) is a little higher than that of another theme, *ecosystem service* (0.009546033) in P3. Nonetheless, the latter is ranked the first because the Delta-C of the latter (*ecosystem service* (0.00118078)) has much bigger plus difference between degree centrality value of P1 and P2 than that of the former (*maritime transport* (0.000644456)).

Type C as extremely interest-decreased themes

Table 8. Top 10 keywords belonging to Type C

Keywords (2015-2020)	Degree Centrality	Freq.	Normalized Freq.	Delta-C against P2(2010-2014)	Delta C against P1 (2000-2009)
1. high speed rail	0.004879084	48	0.732	-0.00151964	no occurrence
2. Finland	0.00106067	2	0.030	-0.001197477	no occurrence
3. Hong Kong	0.000424268	7	0.106	-0.001150609	no occurrence
4. supply chain security	0.002121341	4	0.061	-0.000838767	no occurrence
5. hub	0.000212134	4	0.061	-0.000821018	no occurrence
6. fuzzy logic	0.000212134	4	0.061	-0.000780066	no occurrence
7. corridor	0.000424268	5	0.076	-0.000741086	no occurrence
8. black sea	0.000212134	5	0.076	-0.000739114	no occurrence
9. punctuality	0.000212134	2	0.030	-0.000739114	no occurrence
10. port security	0.001484938	3	0.045	-0.000709994	no occurrence

The degree centrality value of a theme, *high speed rail* (0.004879084) is the highest in a group of Type C in P3. Nonetheless, this theme is ranked the first in the group of **interest-decreased themes** because the Delta-C of this keyword (*high speed rail* (-0.00151964)) has the bigger difference between degree centrality value of P2 and P3 than other keywords.

Type D as retro trends (복고풍) with interest-increased themes

Table 9. Top 10 keywords belonging to Type D

Keywords (2015-2020)	Degree Centrality	Freq.	Normalized Freq.	Delta-C against P1 (2000-2009)	Delta C against P2 (2010-2014)
1. impact	0.013364446	68	1.037	0.002416511	no occurrence
2. maritime logistics	0.005939754	20	0.305	0.000934461	no occurrence
3. shipping market	0.005091218	7	0.106	0.00077854	no occurrence
4. Malaysia	0.003818413	15	0.228	0.000583905	no occurrence
5. marine	0.003182011	29	0.442	0.000545457	no occurrence
6. stochastic frontier analysis	0.003394145	12	0.183	0.000505944	no occurrence
7. bulk shipping	0.003394145	9	0.137	0.000466698	no occurrence
8. fuzzy set theory	0.003394145	4	0.061	0.000466698	no occurrence
9. dynamic model	0.003182011	5	0.076	0.000427717	no occurrence
10. Kenya	0.003394145	16	0.244	0.000427451	no occurrence

The difference between Type B and Type D lies in a **discontinuous series** of some themes as **retro trends**. For example, a theme, *impact*, has comparatively high degree centrality value in P1 but it does **not occur** in P2 but re-appears in P3 with a degree centrality value (0.013364446) **higher** than other themes shown in a group of Type D in P3.

Type E as retro trends with interest-decreased themes

Table 10. Top 10 keywords belonging to Type E

Keywords (2015-2020)	Degree Centrality	Freq.	Normalized Freq.	Delta-C against P1 (2000-2009)	Delta-C against P2 (2010-2014)
1. pollution	0.001060670	16	0.244	-0.001100240	no occurrence
2. coastal management	0.003818413	61	0.930	-0.001025210	no occurrence
3. law sea	0.001060670	10	0.152	-0.001021746	no occurrence
4. legitimacy	0.000848536	13	0.198	-0.001021480	no occurrence
5. oxytetracycline	0.000848536	8	0.122	-0.000825246	no occurrence
6. globalisation	0.000212134	3	0.045	-0.000785200	no occurrence
7. fishing effort	0.000424268	13	0.198	-0.000746220	no occurrence
8. decommissioning	0.000212134	4	0.061	-0.000706707	no occurrence
9. exclusive economic zone	0.000848536	8	0.122	-0.000629013	no occurrence
10. texture	0.000636402	8	0.122	-0.000628746	no occurrence

The degree centrality value of a theme, *pollution* (0.001060670) is not the highest in P3 and this theme is ranked the first in a group of **interest-decreased** themes because the Delta-C of this keyword (*pollution* (-0.001100240)) has the bigger difference between degree centrality value of P1 and P3 than other keywords.

Type F as new trends



Table 11. Top 10 keywords belonging to Type F

Keywords (2015-2020)	Degree Centrality	Freq.	Normalized Freq.	Delta-C against P2 (2010-2014)	Delta-C against P1 (2000-2009)
1. literature review	0.009970301	45	0.686	no occurrence	no occurrence
2. sea level rise	0.006151888	54	0.823	no occurrence	no occurrence
3. lipid metabolism	0.005727620	45	0.686	no occurrence	no occurrence
4. machine learning	0.005727620	20	0.305	no occurrence	no occurrence
5. GHG emission	0.005303352	13	0.198	no occurrence	no occurrence
6. Belt Road Initiative	0.005091218	28	0.427	no occurrence	no occurrence
7. demand elasticity	0.005091218	4	0.061	no occurrence	no occurrence
8. comanagement	0.004242681	23	0.350	no occurrence	no occurrence
9. AI	0.004030547	11	0.167	no occurrence	no occurrence
10. random forest	0.004030547	8	0.122	no occurrence	no occurrence

The frequency of keywords is not directly related to how much some keywords are important, since the **frequency** of *sea level rise* is the **greatest** in this group of Type E, but its **degree centrality value** is **not the highest**. Furthermore, the frequency of *machine learning* (20) and *GHG emission* (13) is **less than** *Belt Road Initiative* (28) and *co-management* (23) but the former's degree centrality values (0.005727620 and 0.005303352) are **greater** values than the latter's (0.005091218 and 0.004242681) respectively.

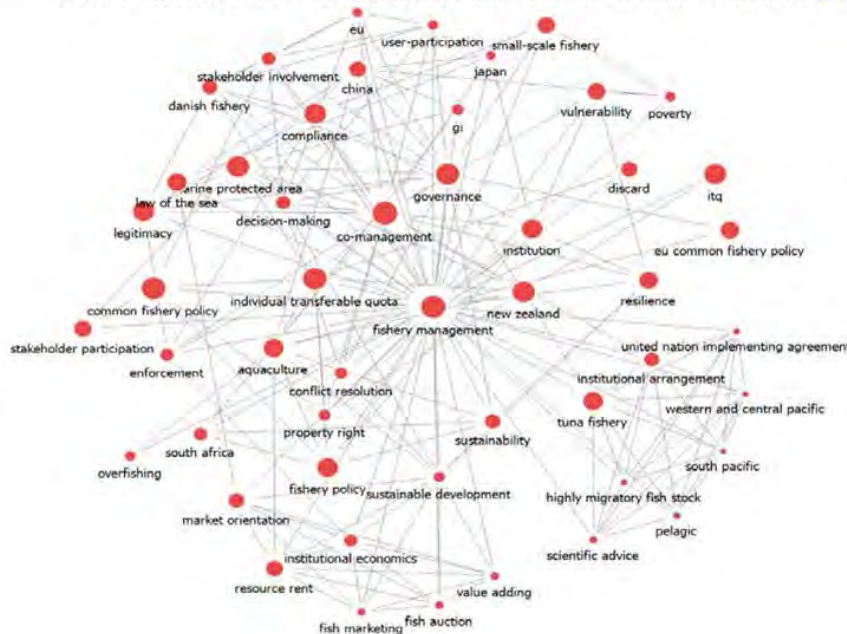
It can be interpreted that these top 10 keywords having emerged as **new issues** of concern illustrate that the issues of *sea level rise* and *Green House Gas emission* attract **more attention** in the literature, the subjects of *machine learning* and *artificial Intelligence (AI)* become popular in accordance with the development of internet of things (IoT) in the late 2010s, and *Belt Road Initiative* demonstrates the enlargement of China's economic potential in the 2010s.

Networks of prominent themes in each period and in type patterns



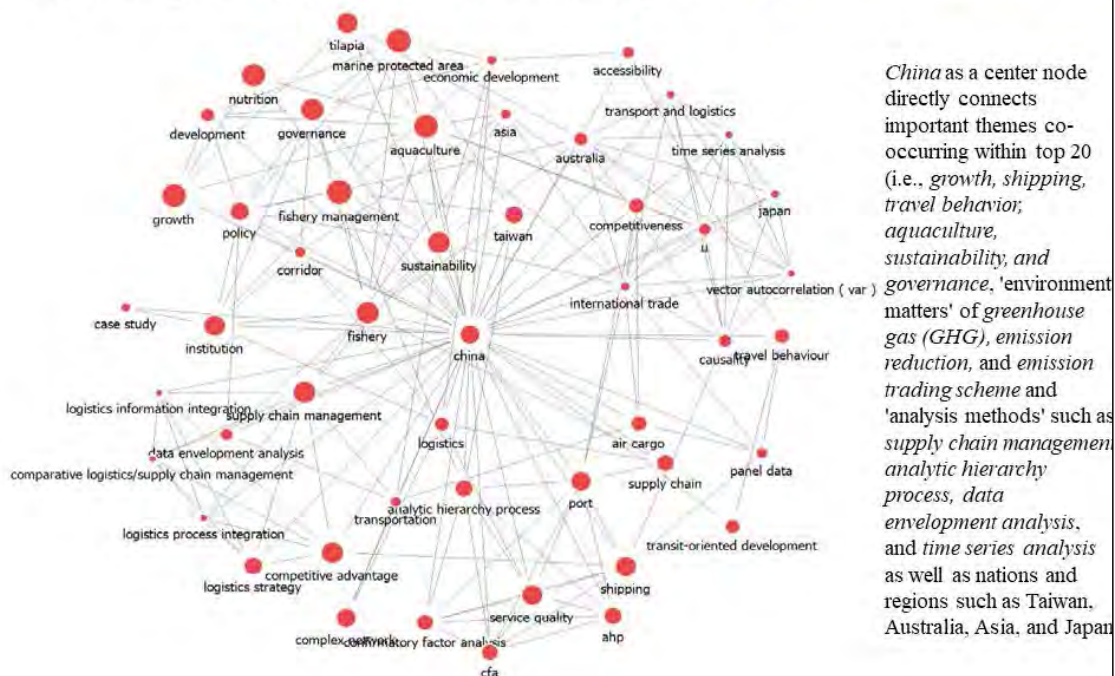
Prominent themes in each period & three interesting themes (fishery management, China, and climate change) belonging to Type A

Figure 6. Degree centrality visualization of fishery management ranking second in P1



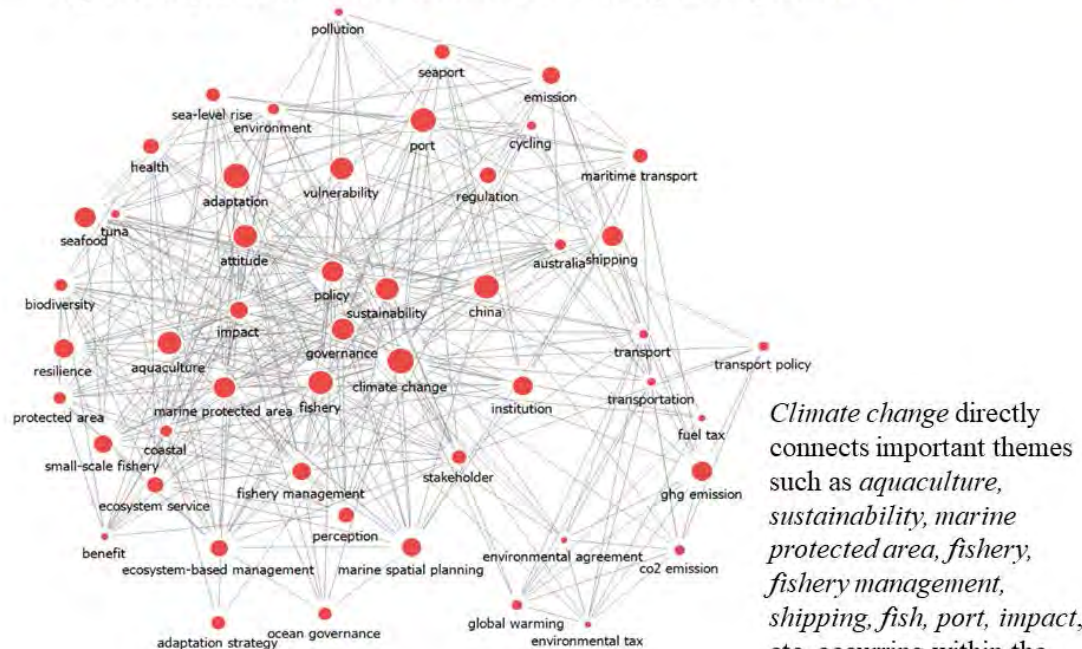
A center node, *fishery management* directly connects many important nodes such as *governance*, *aquaculture*, *sustainability* and *co-management* co-occurring within the top 20 as well as nations such as China, New Zealand, South Africa, and EU.

Figure 7. Degree centrality visualization of *China* ranking fourth in P2



China as a center node directly connects important themes co-occurring within top 20 (i.e., growth, shipping, travel behavior, aquaculture, sustainability, and governance, 'environment matters' of greenhouse gas (GHG), emission reduction, and emission trading scheme and 'analysis methods' such as supply chain management, analytic hierarchy process, data envelopment analysis, and time series analysis as well as nations and regions such as Taiwan, Australia, Asia, and Japan

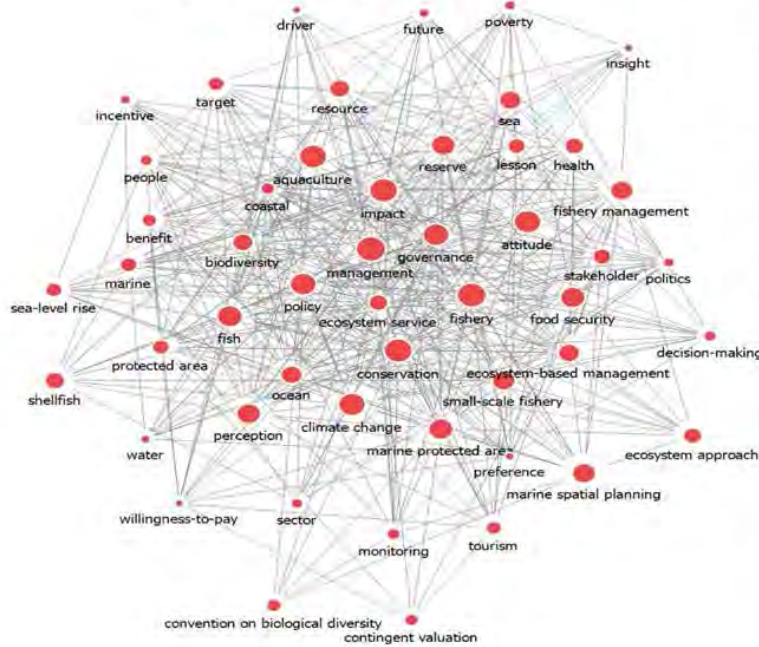
Figure 8. Degree centrality visualization of *climate change* ranking fifth in P3



Climate change directly connects important themes such as aquaculture, sustainability, marine protected area, fishery, fishery management, shipping, fish, port, impact, etc. occurring within the Top 30 in P3.

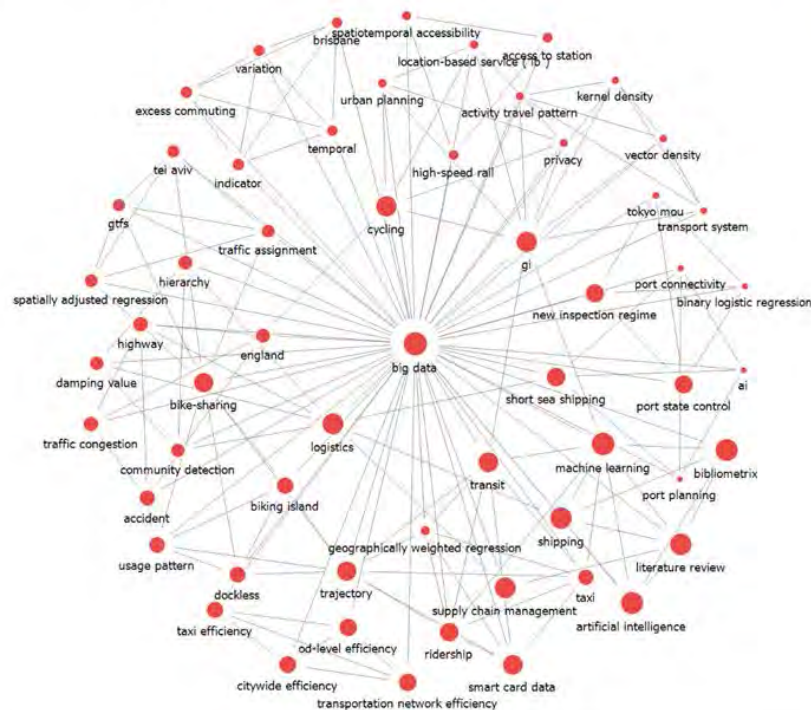
Prominent themes in type patterns : four interesting themes belonging to Type B (*ecosystem service* and *big data*) and Type F (*sea level rise* and *AI*) in P3

Figure 9. *ecosystem service*



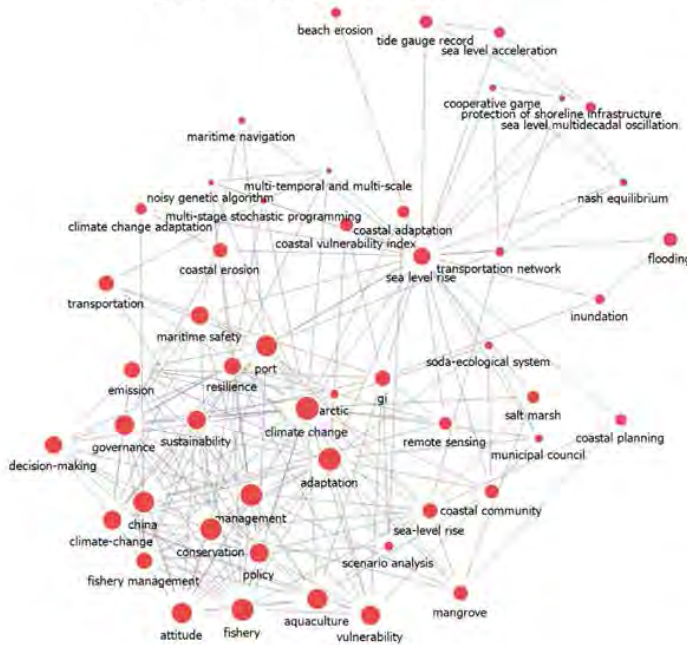
Ecosystem service ranking first has more nodes than *big data* ranking second. The former connects a nation such as *Kenya* as well as regional areas such as *Gulf Mexico*, *Baltic Sea*, *marine protected area*, and *wetland*, whereas the latter connects a nation such as *England* and new trends of Type F such as *machine learning*, *artificial intelligence*, *general transit feed specification (gtfs)*, *decision making*, etc.

Figure 10. *big data*



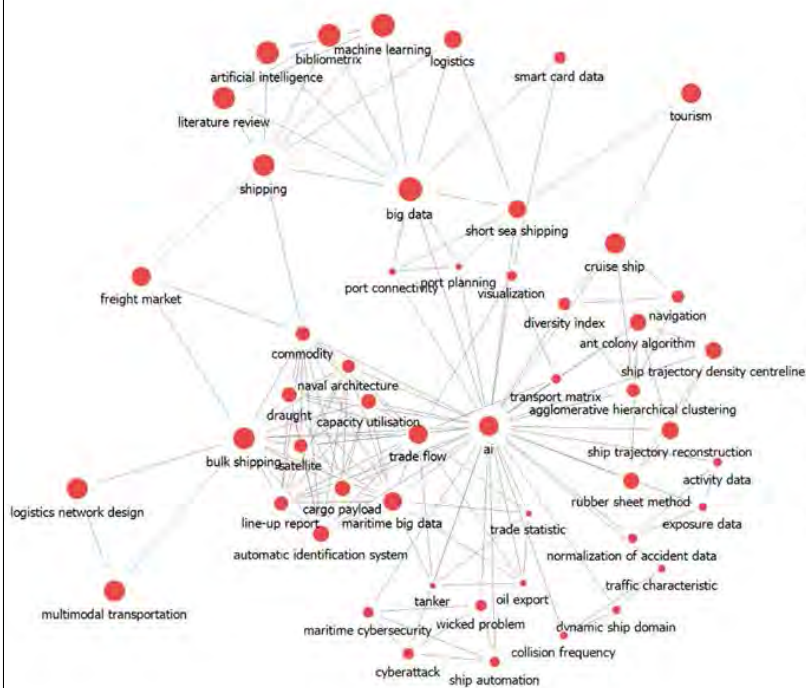
Two themes newly introduced to the most recent period (P3):
sea level rise ranking second and *AI* ranking ninth.

Figure 11. *sea level rise*



Sea level rise is associated with 'climate' such as *climate change* and *climate change adaption*, 'erosion' such as *beach erosion* and *coastal erosion*, 'maritime services' such as *maritime safety*, *port*, *Global Initiative (GI)*, and *maritime navigation*, and 'coastal problems' with *coastal adaption*, *coastal flooding*, *coastal planning*, *coastal management*, *coastal community*, *coastal wetland*, *coastal vulnerability index*, etc.

Figure 12. *AI*



AI is mainly associated with 'maritime navigation affairs' such as *navigation*, *automatic identification system*, *draught*, *bulk shipping*, *maritime big data*, etc. and 'shipbuilding' such as *cruise ship*, *ship trajectory reconstruction*, *ship trajectory density centerline*, *rubber sheet method*, etc.

Conclusion

- We have paid special attention to **six different types of patterns** through the **Delta-C algorithm**
- First, we discussed highly remarkable shared research themes (i.e., **seven shared keywords** such as *aquaculture*, *Atlantic salmon*, *governance*, *growth*, *marine protected area*, *sustainability*, and *transport*) throughout the three periods as **Type A**.
- Second, we focused on interest-increased, interest-decreased, and newly emerging research themes shown in the third period (P3) from **Type B to Type E**. As for Type B to Type E, we discussed the top 10 themes of each type.
- This study showed **two new findings**.
- First, in Type A, representing consistently shared themes, the main research themes changed from *growth* and *fishery management* in fisheries and *sustainability* and *governance* in maritime sectors in the 2000s; to *growth* and *aquaculture* in fisheries and *accessibility*, *China* and *sustainability* in maritime sectors in the early 2010s; and to *aquaculture* and *growth* in fisheries and *accessibility*, *climate change*, and *China* in maritime sectors in the late 2010s.
- Second, in Type F as new trends, the top 10 keywords illustrated that the issues of *sea level rise* and *Green House Gas emission* attract more attention in the literature. It can also be interpreted that the subjects of *machine learning* and *artificial Intelligence (AI)* become popular in accordance with the development of internet of things (IoT) in the late 2010s and *Belt Road Initiative* demonstrates the enlargement of China's economic potential in the 2010s.

31

Suggestions

- The findings of this study lead us to **some suggestions** for future research.
- First of all, the **Delta-C algorithm** used in this study is very helpful to identify **new themes** emerging in a designated period.
- Second, this algorithm is also able to show **the trend of how much a theme of interest has been increased or decreased** in a designated period.
- However, this study has also a **data limitation**. We realize that 15 international maritime related journals are **not enough** to inform us of all the trends and important themes in the maritime fields during the last 20 years. In a future study, we need far more maritime related journals included in order to offer more accurate and significant trends and themes in global maritime journals.

32

References

- Bai, Y., Li, H., and Y. Liu (2020) Visualizing Research Trends and Research Theme Evolution in E-learning Field: 1999-2018. *Scientometrics*. <https://doi.org/10.1007/s11192-020-03760-7>
- Choi, J. W. and Kang, W. S. (2014) Themes and Trends in Korean Educational Technology Research: A Social Network Analysis of Keywords. *Procedia: Social and Behavioral Sciences* 131:171-176
- Freeman, L. C. (1978/1979) Centrality in social networks: conceptual clarification. *Social Networks* 1: 215-239
- Jhang, S. E., Lee, S. M., Lee, S. S., and Kim, J. H. (2015) Themes and Trends in Offshore Industry Research through Incorporation of Corpus and Language Network Analysis: A Social Network Analysis of Author Keywords in English Academic Articles. *Journal of Language Sciences* 22(3):171-198.
- Jhang, S. E. and Lee, S. H. (2016) A Study of Themes and Trends in Research of Global Maritime Economics through Keyword Network Analysis. *Journal of Korea Port Economic Association* 32(1):79-95.
- Khitous, F., Strozzi, F., Urbinati, A., and Alberti, F. (2020) A Systematic Literature Network Analysis of Existing Themes and Merging Research Trends in Circular Economy. *Sustainability* 12(4):1633. <https://doi.org/10.3390/su12041633>
- Lee, J. M. and Kang, J. H. (2011) Mapping Knowledge Domains: The Structure of Knowledge Production and the Status of Theory in Korean Sociology 2004-2010. *Society and Theory* 19:89-144.

33



34

한국어 영화대사와 영어 자막 번역에 나타난 감정 분석 비교: 건축학개론을 중심으로

이성화, 박호민
한국해양대학교
한국언어과학회 동계 학술대회
2021. 02. 18

2

Outline

- 연구 목적 및 연구 질문
- 감정 분석의 정의 및 선행연구
- 데이터 및 연구 방법
- 결과 및 논의
- 요약 및 결론
- 참고 문헌

연구 목적 및 연구 질문

- 연구 목적
 - 영상 번역에서 원천 텍스트(Source Text)인 한국어 대사와 목표 텍스트(Target Text)인 영어 자막에 대한 감정 분석(Sentiment Analysis)을 비교한다.
- 연구 질문
 - 감정 분석이 영화 등장인물의 갈등 구조와 감정 변화를 보여주는가?
 - ST에서 나타나는 감정의 누적값과 등장인물의 감정 평균값이 TT의 그것과 일치 혹은 유사한가?
 - ST와 TT 간 감정값의 차이가 있다면 원인은 무엇인가?
 - 감정 분석이 번역 평가의 지표로 사용될 수 있을까?
- 연구 가설
 - 문화적 차이가 존재하나 인간의 감정은 보편적(universal)이라고 할 수 있다. 문화적 차이를 뛰어 넘어 보편적 감정값을 잘 반영한 번역이 목표 관객에게 호소력이 높을 것이다.

감정 분석(Sentiment Analysis) 정의 및 선행연구

- 감정 분석이란 ‘오피니언 마이닝(opinion mining)’으로도 불리며, 어떤 사안이나 인물, 이슈, 이벤트에 대한 사람들의 의견이나 평가, 태도, 감정 등을 분석하는 것을 말한다(Liu, Hu, and Cheng, 2005)
- Piryani, Madhavi, and Singh(2017)은 2000-2015년 사이에 감정 분석에 관한 연구 동향 연구에 따르면, 데이터 종류로 리뷰가 39.30%로 가장 높고 트위터가 11.06%, 뉴스기사가 7.29% 기사.
- 영화 리뷰:
 - 상훈, 최정, 김종우(2016): 영역 맞춤형 감정 사전을 개발하여 장르별 영화 리뷰를 분석
 - 김유영, 송민(2016): 기계학습 기반의 감정 분류기를 사용하여 ‘네이버 영화’에서 수집한 리뷰 데이터로 감정 점수를 도출하여 관객 리뷰와 흥행 간의 상관성을 논함
 - 남길임, 강현아(2019) : 단어 분석에서 나아가 구 단위별 감성표현 분석 및 주제 특정적(영화, 의류) 감성 분석의 필요성을 논함
- 셰익스피어 희곡에 관한 국내외 연구
 - Nalisnick, 2013; Nalisnick and Baird. 2013; 이경은, 2018; 장세은, 이경은, 박호민, 송원문, 정해룡, 이수상, 김재훈, 2019; 서혜진, 이종현, 신정아, 2019; 장세은, 박호민, 송원문, 이수상, 김재훈, 2020 등

감정 분석(Sentiment Analysis) 정의 및 선행연구

- 영화의 감정 분석:
 - 유은순(2013)은 영화 대사의 감정표현과 플롯 간의 연관성 연구를 감정사전이나 딥러닝이 아닌 수의적으로 정량 분석을 수행하였다.
- 소설 작품에 대한 감정 분석연구
 - Reagan(2017): 구텐베르크 프로젝트에 있는 소설 1,327 작품을 선정하여 기본 감정 곡선 여섯 가지를 도출하여 인기 있는 작품은 어떤 모드에 집중해 있는지 분석
 - 강우규, 김바로(2019): 고전소설인 <구운몽>을 디지털 감정 분석을 시도한 연구로, 국문학과 학생들이 어휘, 문장에 감정을 직접 입력하여 감정 데이터를 구축
 - 문석형, 강주영(2019): 아서 코난 도일과 애거서 크리스티의 추리 소설을 각각 2권씩 사용하여 감성 분석과 토픽모델링을 통해 소설 전개를 관찰하고 네트워크 분석을 통해 등장인물 간의 관계 변화 관찰하여 문법 체계 분석을 통해 두 작가 간의 문체적 특성을 비교
- 원문과 번역에 대한 감정 분석
 - Mohammad, Salameh, and Kiritchenko (2016): 영어의 감정 사전을 아랍어로 번역하여 아랍어 소셜미디어 텍스트를 분석한 결과와 아랍어 텍스트를 영어로 번역해서 영어 감정 사전을 사용한 결과를 비교 분석
 - 한국어와 영어 간 번역 텍스트를 감정 분석에 사용한 사례는 찾아 볼 수 없다.

Data: 영화 <건축학개론>

- 한국어 대사와 그에 상응하는 영어 자막 번역
- 한국어 대사는 대본에 있는 것이 아니라 실제로 발화된 대사를 전사한 것
- 영어 자막은 번역학 대학원생인 비전문가의 번역



Data: 건축학개론 PLOT

- 30대 중반의 주인공 서연(여) 과 승민(남)이 건축의뢰인과 건축가로 만나서 집을 짓는 과정에서, 15년 전 대학 1학년 때를 회상한다.
- 현재와 과거로 교차하며 전개 된다.
- 본 연구를 위해 총 11개의 scene으로 분류하였다.
 - 현재: 1, 3, 5, 7, 9, 11
 - 과거: 2, 4, 6, 8, 10



엑셀 파일 Raw Data

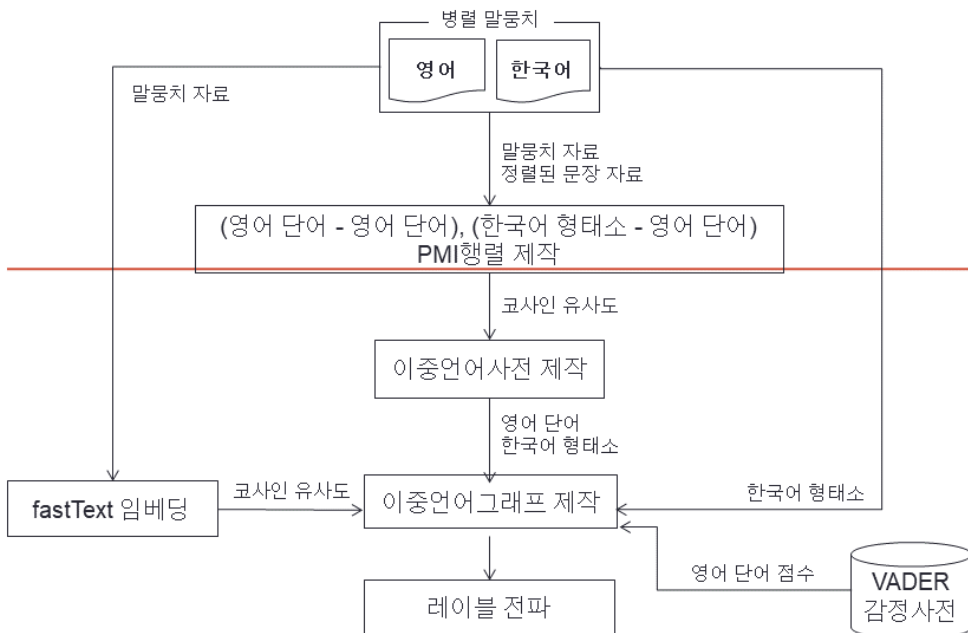
시점	모든 등장인물	남녀주인공	한글 원본 건축학개론	영어 자막번역 Architecture 101	SCENE
1 현재	사장		어, 피곤하지?	Tired, huh?	1
2 현재	사장		짜식	Hey, you.	1
3 현재	사장		밤 샌 거 꼭 이렇게 극적으로 티를 내요	You always exaggerate of hardship.	1
4 현재	사장		하여튼 보면 참 영리해	Anyway, you're smart.	1
5 현재	사장		아	Hey	1
6 현재	사장		너 이거 ?? 바꿨지?	Did you change this part?	1
7 현재	사장		아 이거 너무 날라가는 거 아니야?	Don't you think it's too much?	1
8 현재	승민	승민, 현재	아 디자인을 예쁘게 채워도 뭐라 그래	How come you complain about this nice design?	1
9 현재	사장		예쁘기만 하면 땀이야?	Does appearance cover everything?	1
10 현재	승민	승민, 현재	아이 건축주를 이런 거 좋아하잖아요	Clients like this type.	1
11 현재	승민	승민, 현재	일단 뭐 저기 디자인 좀 날라가줘야 뭐 있나보다 할 거 아니야	A nice design would make a good impression.	1
12 현재	사장		아 찜질방이나 좀 갔다와	Go to sauna.	1
13 현재	사장		2시 미팅이니까	We have a meeting at 2 p.m.	1
14 현재	승민	승민, 현재	근데 내가 건축주를 만나서 뭐해요 내가	Why should I meet the clients?	1
15 현재	동료		일마 디자인한 놀이 같이 가야지	Because the architect should go.	1
16 현재	승민	승민, 현재	나도 퇴근이라는 것 좀 합시다	Let me get off work!	1
17 현재	승민	승민, 현재	이틀 동안 집어를 못 갔네	I stayed on for two days here.	1
18 현재	은재		아 진짜!	Come on.	1
19 현재	은재		사무실에서 피지 말라니까	I told you not to smoke in the office.	1
20 현재	은재		간접흡연이 얼마나 안 좋은데	You know how bad is the second-hand smoking.	1
21 현재	승민	승민, 현재	아니 그럼 직접흡연하는 나는 얼마나 안 좋겠니?	How worse wouldn't it be for me who is inhaling it?	1
22 현재	승민	승민, 현재	안 그래?	Right?	1
23 현재	직원		손님 왔어요	You have a visitor.	1
24 현재	승민	승민, 현재	손님? 누구?	A visitor? Who?	1
25 현재	직원		똥장님 친구라는데요	She said she's your friend.	1
26 현재	승민	승민, 현재	내 친구?	My friend?	1
27 현재	승민	승민, 현재	친구 누구?	What friend?	1
28 현재	직원		그걸 내가 어떻게 알아요	How do I know that?	1
29 현재	서연	서연, 현재	오랜만이네	It's been a long time.	1
30 현재	서연	서연, 현재	동문주소록 되게 정확하다-	The college address book is so exact.	1

Methodology 1: VADER 감정 사전

Hutto and Gilbert(2014) : VADER 감정사전, 긍정 감정어 3,345개, 부정 감정어 4,172개 총 7,517개
(cf. AFINN 감정사전 2,477개)

- 평가자 10명에게 감정어의 점수를 -4 ~ +4 사이 정수로 평가하게 하여 평균으로 점수를 매긴 감정사전 시스템
 - ✓ 해당 표준편차가 2.5 미만일 때만
 - ✓ 평가자 10명은 품질 관리를 위한 특별한 교육 이수
- 초기 감정어 선정은 기존에 검증된 감정사전의 감정어를 사용.
이후 전문가 두 명이 검증한 텍스트에서 새롭게 나타나는 감정어로 사전 확장
- 언어적인 규칙도 활용
 - ✓ 1. 대문자 2. 문장부호 3. 부정어(negation) 파악 4. 이모티콘, 축약어 5. '매우'와 같은 증폭/감쇠어)
- 트위터 데이터 한정으로 사람보다 더 높은 정확도를 보임
 - ✓ 다른 분야에서는 사람보단 낮지만 비교한 11가지 기계학습적 방법보다 높은 성능을 보임)

Methodology 2: KMOU 한국어 감정 사전 제작시스템



KMOU 한국어 감정사전 결과

감정어	점수	감정어	점수	항목	제작된 감정사전	VADER 감정사전
우울하/VA	-1.6	depress	-2.2	감정어 수	50,068	10,258
울/VV	-1.8	cry	-2.1	긍정적 감정어 수	20,759	4,237
화나/VV	-2.0	angry	-2.3	부정적 감정어 수	29,759	6,021
대립하/VV	-1.7	confront	-0.7			
간악하/VV	-3.4	wicked	-2.4			
기쁘/VA	2.2	please	1.3			
스마트하/VA	1.7	smart	1.7			
명예롭/VA	2.3	honorable	2.5			
대단하/VA	1.9	great	3.1			
행복/NNG	2.1	happiness	2.6			

KMOU 한국어 감정사전
예시 단어 10개

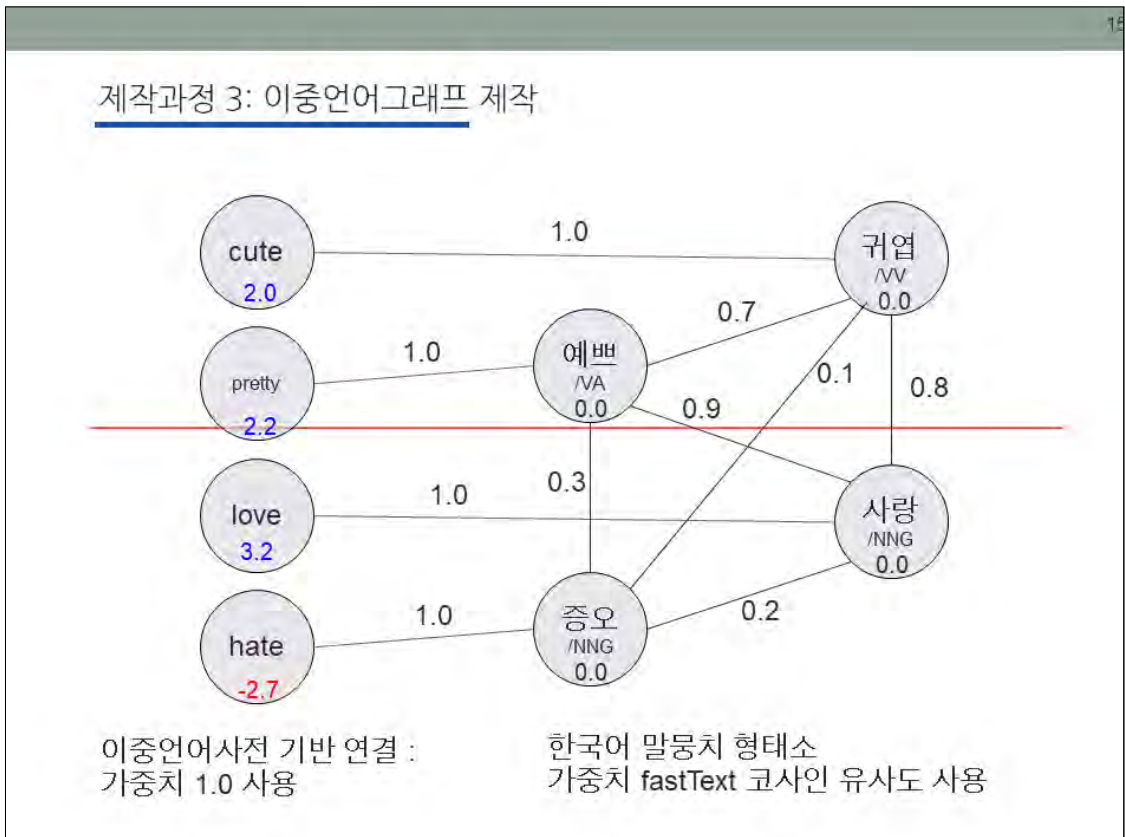
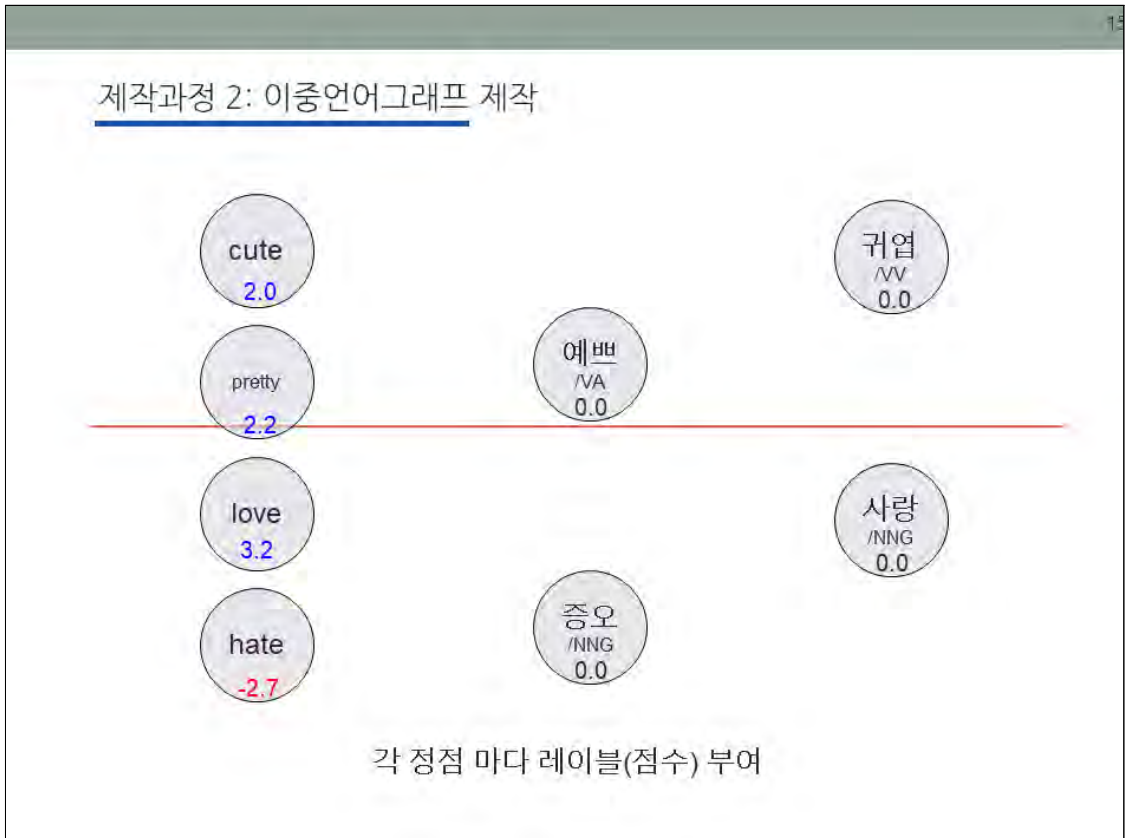
VADER 감정사전
예시 단어 10개

제작과정 1: 이중언어그래프 제작

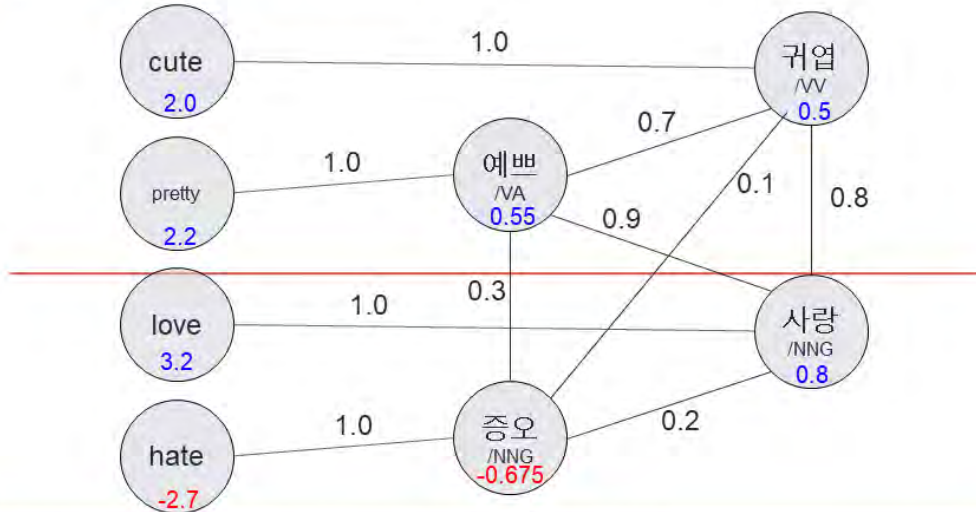


씨앗 사전의 감정어
이중언어사전의 영어 단어

추후 제작될 감정사전의 감정어
이중언어사전의 한국어 형태소

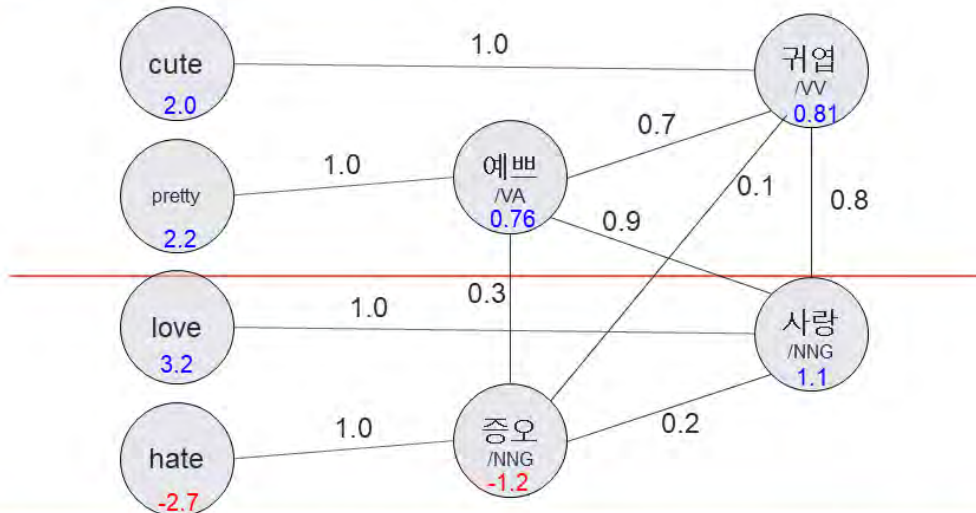


제작과정 4: 레이블 전파 - 1회 후



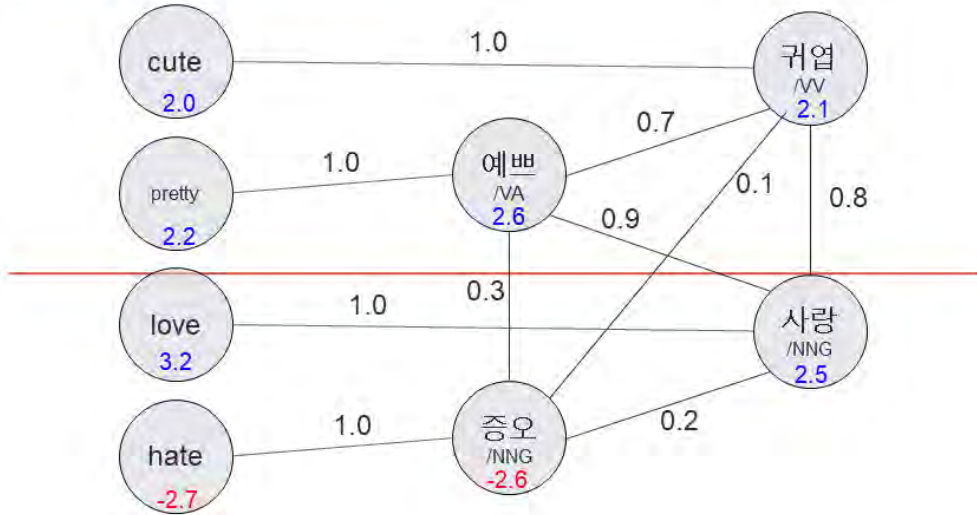
예쁘	pretty	귀엽/VV	사랑/NNG	증오/NNG	$2.2 / 4 = 0.55$
/VA					$2.2 \times 1.0 + 0.0 \times 0.7 + 0.0 \times 0.9 + 0.0 \times 0.3$

제작과정 5: 레이블 전파 - 2회 후



예쁘	pretty	귀엽/VV	사랑/NNG	증오/NNG	$3.06 / 4 = 0.76$
/VA					$2.2 \times 1.0 + 0.5 \times 0.7 + 0.8 \times 0.9 + -0.675 \times 0.3$

제작과정 6: 레이블 전파 - 최종 결과



Methodology 3: 감정분석을 통한 감정 추이 시각화작업과정

- 전체 감정누적 추이선 그래프 작업과정 프로그램 소스 코드

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sys
import os

class ACIO(object): # ArchitectureIO
    def __init__(self, sheetdata, name):
        self.name = name
        self.number = sheetdata['번호']
        self.hero = sheetdata['인물']
        self.korean = sheetdata['한국어 대사']
        self.korean_vader = sheetdata['한국어 vader']
        self.english = sheetdata['영어 대사']
        self.english_vader = sheetdata['영어 vader']
        self.act = sheetdata['ACT']
        print(name + " initialization complete.")

# 엑셀파일 경로지정
TOTAL_ALL = pd.ExcelFile("1. 전체.xlsx")
TOTAL_HERO = pd.ExcelFile("2. 전체 인물.xlsx")
PRESENT_HERO = pd.ExcelFile("3. 현재 인물.xlsx")
PAST_HERO = pd.ExcelFile("4. 과거 인물.xlsx")

# 인스턴스 생성하기
architectureIO_total_all = ACIO(TOTAL_ALL.parse('Sheet1'), "전체")
architectureIO_total_hero = ACIO(TOTAL_HERO.parse('Sheet1'), "전체 인물")
architectureIO_present_hero = ACIO(PRESENT_HERO.parse('Sheet1'), "현재 인물")
architectureIO_past_hero = ACIO(PAST_HERO.parse('Sheet1'), "과거 인물")
ACIO_list = [architectureIO_total_all, architectureIO_total_hero, architectureIO_present_hero, architectureIO_past_hero]

fig = plt.figure()
ax = fig.add_subplot(1, 1, 1)

# 진행 상황
act_list = []
flag = 1
for i in range(len(ACIO_list[0].act)):
    if str(ACIO_list[0].act[i]) != flag:
        act_list.append(i)
        flag = str(ACIO_list[0].act[i])
act_list.append(len(ACIO_list[0].act))

temp_series = pd.Series(ACIO_list[0].korean_vader)
temp_series = temp_series.cumsum()
lines = ax.plot([i for i in range(1, act_list[-1] + 1)], temp_series, label="KOREAN")
ax.set_xticks(act_list)
ax.set_xticklabels(['ACT1', 'ACT2', 'ACT3', 'ACT4', 'ACT5', 'ACT6', 'ACT7', 'ACT8', 'ACT9', 'ACT10', 'ACT11'])
ax.legend(fontsize=10)

temp_series = pd.Series(ACIO_list[0].english_vader)
temp_series = temp_series.cumsum()
lines = ax.plot([i for i in range(1, act_list[-1] + 1)], temp_series, label="ENGLISH")
ax.set_xticks(act_list)
ax.set_xticklabels(['ACT1', 'ACT2', 'ACT3', 'ACT4', 'ACT5', 'ACT6', 'ACT7', 'ACT8', 'ACT9', 'ACT10', 'ACT11'])
ax.legend(fontsize=10)

plt.show()
    
```

Methodology 3: 감정분석을 통한 감정 추이 시각화작업과정

• 주인공 간의 감정 추이 막대 그래프 시각화작업과정

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

fig = plt.figure()
ax = fig.add_subplot(1, 1, 1)

# 감정별 평균 감정인 바 그래프
color_list = ['ACT1', 'ACT2', 'ACT3', 'ACT4', 'ACT5', 'ACT6', 'ACT7', 'ACT8', 'ACT9', 'ACT10']
temp_df = pd.DataFrame(index = ['ACT1', 'ACT2', 'ACT3', 'ACT4', 'ACT5', 'ACT6', 'ACT7', 'ACT8', 'ACT9', 'ACT10'])

temp_df['SUNG MIN - (K)'] = [7.5001 / 25.0, 3.1025 / 12.0, 2.6165 / 11.0, 2.4020 / 7.0, 7.512 / 23.0, 1.8641 / 5.0, 0.6907 / 8.0, 0.1027 / 1.0, 1.4722 / 5.0, 0.0256 / 1.0]
temp_df['SUNG MIN - (S)'] = [5.4206 / 18.0, 4.0464 / 12.0, 4.2511 / 8.0, 3.7158 / 9.0, 4.1316 / 14.0, 4.0133 / 3.0, 2.2527 / 7.0, 0.3612 / 1.0, 0.9391 / 3.0, 0.0]
temp_df['SUNG MIN - (D)'] = [-3.5562 / 16.0, -4.8495 / 13.0, -3.4904 / 9.0, -3.8534 / 8.0, -12.0772 / 36.0, -1.9909 / 7.0, -2.0026 / 10.0, -0.0516 / 1.0, -3.0494 / 8.0, -0.9406 / 3.0]
temp_df['SUNG MIN - (I)'] = [-0.2795 / 1.0, -0.0001 / 5.0, -0.5289 / 3.0, -0.2755 / 1.0, -3.5709 / 10.0, 0.0, -2.2571 / 7.0, 0.0, -0.4336 / 2.0, -0.3182 / 1.0]
temp_df['SUN YOUNG - (K)'] = [19.3622 / 31.0, 5.078 / 15.0, 0.8718 / 5.0, 4.0263 / 12.0, 7.9621 / 25.0, 7.3345 / 23.0, 2.6919 / 9.0, 0.8779 / 3.0, 2.7614 / 7.0, 0.1544 / 2.0]
temp_df['SUN YOUNG - (S)'] = [5.5633 / 15.0, 6.6549 / 17.0, 2.7292 / 8.0, 2.7255 / 8.0, 2.9123 / 10.0, 8.205 / 17.0, 3.9341 / 13.0, 0.2253 / 1.0, 4.2499 / 12.0, 0.0]
temp_df['SUN YOUNG - (D)'] = [-7.9318 / 23.0, -7.4516 / 25.0, -2.5116 / 13.0, -5.6459 / 13.0, -11.5329 / 36.0, -7.9910 / 31.0, -4.3634 / 15.0, -0.236 / 1.0, -6.7076 / 15.0, -0.236 / 1.0]
temp_df['SUN YOUNG - (I)'] = [-3.9314 / 12.0, -2.181 / 5.0, -0.1834 / 2.0, -0.8685 / 5.0, -9.0051 / 25.0, -3.1295 / 9.0, -0.4215 / 1.0, 0.0, -0.2891 / 2.0, 0.0]

temp_df.plot.bar(color = color_list)
plt.grid()
plt.legend(fontsize=xx+large)
plt.show()
```

• 주인공 간의 감정 평균값의 선 그래프 시각화작업과정

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

fig = plt.figure()
ax = fig.add_subplot(1, 1, 1)

# 감정별 평균 감정인 선 그래프
temp_df = pd.DataFrame(index = ['ACT1', 'ACT2', 'ACT3', 'ACT4', 'ACT5', 'ACT6', 'ACT7', 'ACT8', 'ACT9', 'ACT10'])

temp_df['SUNG MIN - (K)'] = [7.5001 / 25.0, 3.1025 / 12.0, 2.6165 / 11.0, 2.4020 / 7.0, 7.512 / 23.0, 1.8641 / 5.0, 0.6907 / 8.0, 0.1027 / 1.0, 1.4722 / 5.0, 0.0256 / 1.0]
temp_df['SUNG MIN - (S)'] = [5.4206 / 18.0, 4.0464 / 12.0, 4.2511 / 8.0, 3.7158 / 9.0, 4.1316 / 14.0, 4.0133 / 3.0, 2.2527 / 7.0, 0.3612 / 1.0, 0.9391 / 3.0, 0.0]
temp_df['SUNG MIN - (D)'] = [-3.5562 / 16.0, -4.8495 / 13.0, -3.4904 / 9.0, -3.8534 / 8.0, -12.0772 / 36.0, -1.9909 / 7.0, -2.0026 / 10.0, -0.0516 / 1.0, -3.0494 / 8.0, -0.9406 / 3.0]
temp_df['SUNG MIN - (I)'] = [-0.2795 / 1.0, -0.0001 / 5.0, -0.5289 / 3.0, -0.2755 / 1.0, -3.5709 / 10.0, 0.0, -2.2571 / 7.0, 0.0, -0.4336 / 2.0, -0.3182 / 1.0]
temp_df['SUN YOUNG - (K)'] = [19.3622 / 31.0, 5.078 / 15.0, 0.8718 / 5.0, 4.0263 / 12.0, 7.9621 / 25.0, 7.3345 / 23.0, 2.6919 / 9.0, 0.8779 / 3.0, 2.7614 / 7.0, 0.1544 / 2.0]
temp_df['SUN YOUNG - (S)'] = [5.5633 / 15.0, 6.6549 / 17.0, 2.7292 / 8.0, 2.7255 / 8.0, 2.9123 / 10.0, 8.205 / 17.0, 3.9341 / 13.0, 0.2253 / 1.0, 4.2499 / 12.0, 0.0]
temp_df['SUN YOUNG - (D)'] = [-7.9318 / 23.0, -7.4516 / 25.0, -2.5116 / 13.0, -5.6459 / 13.0, -11.5329 / 36.0, -7.9910 / 31.0, -4.3634 / 15.0, -0.236 / 1.0, -6.7076 / 15.0, -0.236 / 1.0]
temp_df['SUN YOUNG - (I)'] = [-3.9314 / 12.0, -2.181 / 5.0, -0.1834 / 2.0, -0.8685 / 5.0, -9.0051 / 25.0, -3.1295 / 9.0, -0.4215 / 1.0, 0.0, -0.2891 / 2.0, 0.0]

temp_df.plot.line()
plt.grid()
plt.legend(fontsize=xx+large)
plt.show()
```

엑셀에 정리한 감정값

A	B	C	D	E	F	G
번호	남녀주인공	한글 원본	한글 vader	영어 자막번역	영어 vader	ACTs
1		어, 피곤하지?	-0.25	Tired, huh?	-0.4404	1
2		확실	0	Hey, you.	0	1
3		방생 거 꼭 이렇게 극적으로 티를 내요	-0.6486	You always exaggerate of hardship.	-0.4404	1
4		하어튼 보면 참 영리해	0.2263	Anyway, you're smart.	0.4019	1
5		아	0	Hey	0	1
6		너 이거 ?? 바꿨지?	0	Did you change this part?	0	1
7		야 이게 너무 날라가는 거 아니야?	-0.5267	Don't you think it's too much?	0	1
8	송민_현재	아 디자인을 예쁘게 해줘도 뭐라 그래	0.25	How come you complain about this nice design?	0.0772	1
9		예쁘기만 하면 뭐야?	0.25	Does appearance cover everything?	0	1
10	송민_현재	아이 건축주들 이런 거 좋아하잖아요	-0.3182	Clients like this type.	0.3612	1
11	송민_현재	일단 뭐 저기 디자인 좀 날라가줘야 뭐 있나보다 할 거 아니야	-0.1531	A nice design would make a good impression.	0.765	1
12		야 찜질방이나 좀 갔다와	0.1779	Go to sauna.	0	1
13		2시 미팅이니까	0	We have a meeting at 2 p.m.	0	1
14	송민_현재	근데 내가 건축주들 만나서 뭐해요 내가	0.3612	Why should I meet the clients?	0	1
15		일마 디자인한 놀이 같이 가야지	-0.3182	Because the architect should go.	0	1
16	송민_현재	나도 퇴근이라는 것 좀 합시다	0.0258	Let me get off work!	0	1
17	송민_현재	이틀 동안 집을 못 갔네	-0.2263	I stayed on for two days here.	0	1
18		아 진짜!	0.2942	Come on.	0	1
19		사무실에서 피지 말라니까	0	I told you not to smoke in the office.	0	1
20		간접흡연이 얼마나 안 좋은데	-0.0772	You know how bad is the second-hand smoking. How worse wouldn't it be for me who is inhaling	-0.5423	1
21	송민_현재	아니 그럼 직접흡연하는 나는 얼마나 안 좋겠니?	-0.0772	It?	-0.4767	1
22	송민_현재	안 그래?	-0.34	Right?	0	1
23		손님 왔어요	0.4404	You have a visitor.	0	1
24	송민_현재	손님? 누구?	0.5204	A visitor? Who?	0	1
25		말장님 친구라는데요	0	She said she's your friend.	0.4939	1
26	송민_현재	내 친구?	0	My friend?	0.4939	1
27	송민_현재	친구 누구?	0	What friend?	0.4939	1
28		그걸 내가 어떻게 알아요	-0.25	How do I know that?	0	1
29	서연_현재	오랜만이네	0	It's been a long time.	0	1
30	서연_현재	동문주소록 되게 정확하다-	0.34	The college address book is so exact.	0	1

결과 및 논의 1: 전체 감정 누적값 추이 선 그래프



- 원천 언어는 하향선을 보여주는 것에 비해 목표 언어는 상향선을 나타냄

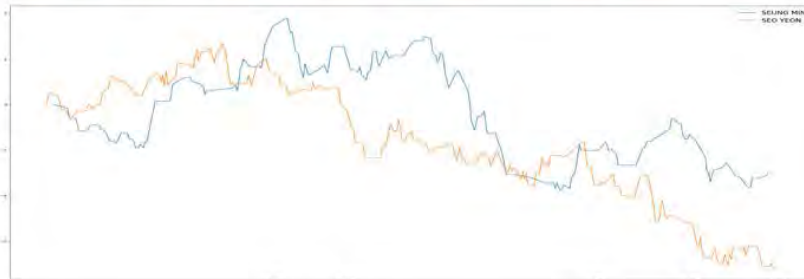
결과 및 논의 2: 남녀주인공 현재+과거 감정 누적값 추이 그래프



원천어
남녀 주인공이
갈등이 과거(2, 4,
6) 현재(5, 7)에
걸쳐 잘 나타남

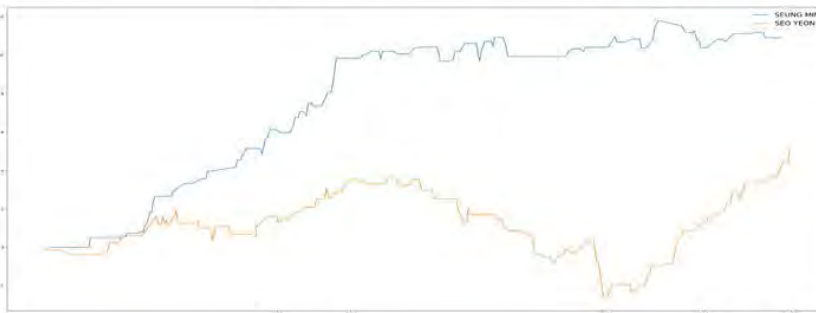
목표어
원천어에서 보였
던 갈등이 나타나
지 않음

결과및 논의 3:
남녀주인공 현재 감정 누적값 추이 그래프



원천어

서연: 극 초반 부분의 긍정적 감정이 극 후반에서 부정으로 바뀐다
승민: Scene 3-5 사이에 걸쳐 긍정적 인 감정이 최고조가 된다

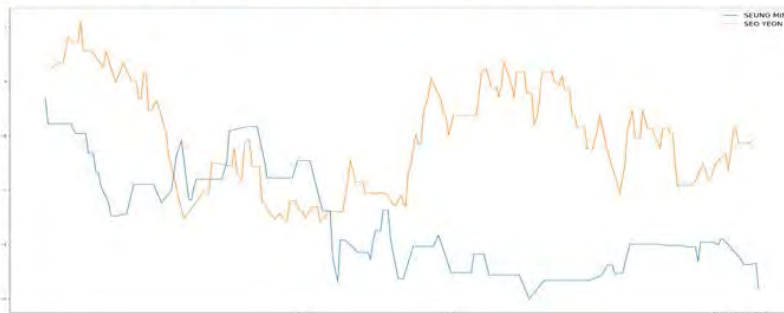


목표어

서연: 감정 선이 후반(7-9)에서 긍정으로 반등한다.

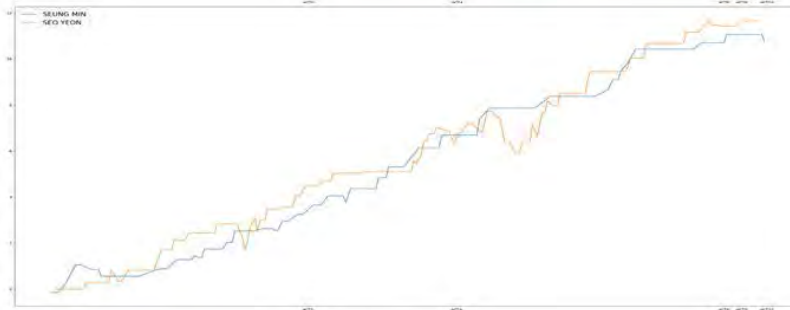
승민: Scene 3 이후 큰 감정 변화 없이 긍정에 머무름

결과및 논의 4:
남녀주인공 과거 감정 누적값 추이 그래프



원천어

서연은 전반적으로 긍정값을 승민은 대체로 부정값을 보여준다.



목표어

원천어에서 보여줬던 두 주인공 간의 감정값 차는 찾아볼 수 없다.

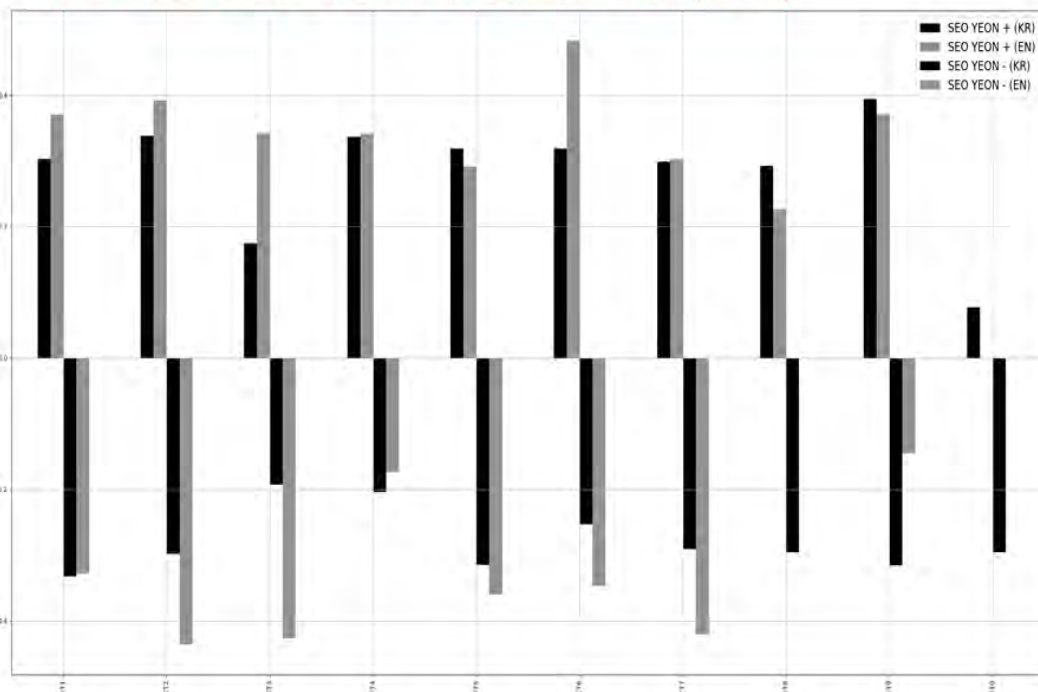
결과 및 논의 5 & 6:

남녀 주인공의 감정 평균값 산출 및 막대그래프

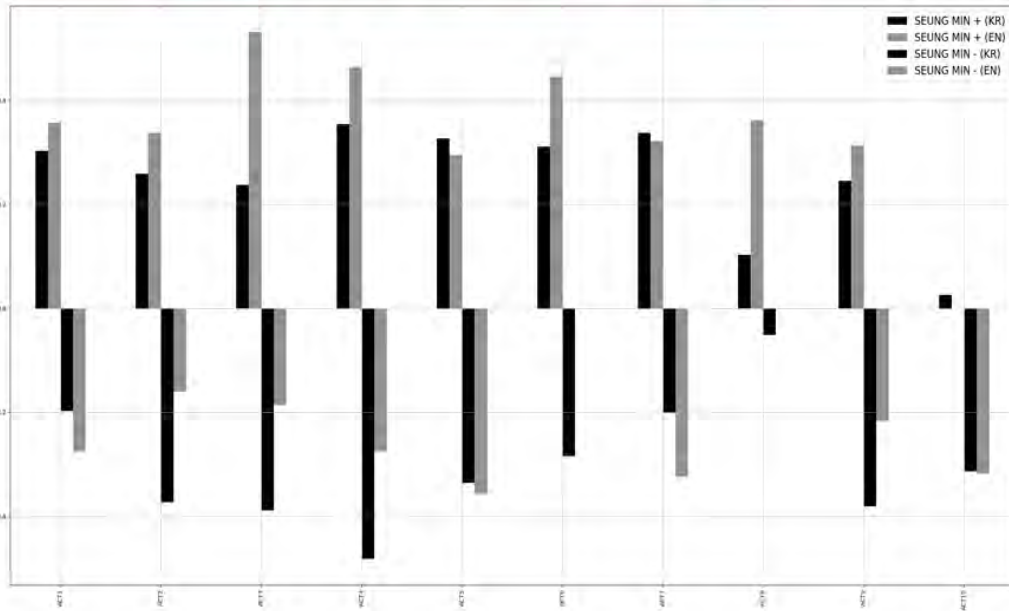
- 막대그래프의 각 막대는 해당 SCENE, 인물의 감정값이 있는 대사만 계산에 사용함
- 각 SCENE의 ST와 TT의 감정값을 비교해 볼 수 있다.
- 결과 및 논의 7에서 가장 차이가 나는 남자 주인공 SCENE 6을 살펴본다

결과 및 논의 5:

여주인공 Scene 별 긍정·부정 평균 감정값 막대그래프 (ST/TT)



결과 및 논의 6:
남주인공 Scene 별 긍정·부정 평균 감정값 막대 그래프 (ST/TT)



결과 및 논의 7:
남주인공 SCENE 6 부정값 고찰

884	승민_과거	아 뭐야 왜 이렇게 많이 때려	-0.6908	0	Why do you beat me so much?	6
887	승민_과거	오늘?	0	0	Today?	6
891	승민_과거	일일일일	0	0	one, one, one, one	6
892	승민_과거	아유 오늘 생일이면	0	0	Then you should...	6
893	승민_과거	친구들하고 파티하고 그래야되는 거 아니야?	0.34	0.7003	...have a birthday party with your friends.	6
897	승민_과거	짤	0	0.4767	Cheers	6
899	승민_과거	난 건축하고 넌 피아노 치고 있었지	-0.3818	0	I will be an architect, and you will be a pianist.	6
906	승민_과거	그게 뭔데?	0	0	What does it mean?	6
911	승민_과거	그럼 뭐할건데?	0	0	Then what will you do?	6
915	승민_과거	아나운서가 돈 많이 벌어?	-0.4404	0	Is an anchor a money-making job?	6
921	승민_과거	공짜로?	0.34	0.5106	For free?	6
939	승민_과거	화장실...	0	0	A restroom...	6
991	승민_과거	어, 들었어	0.1027	0.296	Yeah, I did.	6
993	승민_과거	너 목소리 되게 좋더라	0.1779	0.4215	I liked your voice.	6
1089	승민_과거	아니야	0	0	No.	6
1090	승민_과거	나 무거운거 드는거 좋아해	-0.1531	0.3612	I like carrying heavy things.	6
1093	승민_과거	어	0	0.4019	Yes.	6
1096	승민_과거	이사하니까 좋아?	0.5423	0.5719	Are you happy?	6
1107	승민_과거	이게?	0	0	This one?	6
1110	승민_과거	나 혼자 갈수 있는데	-0.0258	0	I can go by myself.	6
1117	승민_과거	자기	0	0	Hey.	6
1118	승민_과거	첫눈오는날	-0.0258	0	On the first day of snow...	6
1119	승민_과거	뭐해?	0	0	What are you going to do?	6
1122	승민_과거	그러니까 글썸	0	0.2732	Well...	6
1123	승민_과거	보통 겨울에 때되면 오던데	-0.2732	0	It usually comes when it's time.	6
1124	승민_과거	우리 그날 만날까?	0.3612	0	Should we meet then?	6
1129	승민_과거	뭐 그러던가	0	0	Ok.	6

결과 및 논의 8:
ST와 TT 감정값 차이가 큰 25개 대사 I (어휘의 차이)

번	남녀주인	ST	한글 vad	차이값	영어 vad	TT	SCER
1		대학 졸업하고 아나운서 시험 준비하다가 계속 떨어지고	0.6124	1.308	-0.6956	After college, I studied to become an anchor but to fail.	1
2	117 서연_현재	아 이형 졸렸따.	0.7269	1.2364	-0.5095	I'm so jealous of you.	4
3	520	별로 안 좋아하니까 그러지	0.7783	1.0538	-0.2755	I just don't like it.	4
4	589 승민_과거	아우 쌍 년 진짜	0.2263	0.886	-0.6597	God damn it bitch.	8
5	1280	고양에서 더 새 물 받았지?	0.5719	0.813	-0.2411	A new start in Jeju. Isn't it cool?	7
6	1184 승민_현재	진짜 쌍년이지?	0.2263	0.8122	-0.5859	I'm a real bitch.	5
7	834 서연_현재	뭔 소리야 그게?	0.0772	0.758	-0.6808	What the hell are you talking about?	4
8	656	그건 지금도 그래요- 그래요?	0.5574	0.72	-0.1626	~ He still isn't popular. ~ Isn't he?	1
9	177	아 니가 분명히 첫눈오는날 만나자고 그랬었다고	0.7184	0.7184	0	You asked to see me on the first day of snowing.	5
10	750 서연_현재	천천히 해 천천히	0.7184	0.7184	0	Take your time.	7
11	1217 승민_현재	대인생졌나?	0.7096	0.7096	0	Do you have a girlfriend?	2
12	261	그럼 니가 원했던 것처럼 놓아지니까 시뮬해보이고	0.7096	0.7096	0	This is what you wanted originally.	5
13	739 승민_현재	집 지어줄 사람이 그렇게 없었어?	0.4019	0.6979	-0.296	Was there no one to build a house?	9
14	1337 승민_현재	만나면 반갑다고 뽀뽀	0.6908	0.6908	0	It's a pecking, ok?	6
15	965	살기 좋은	0.34	0.6812	-0.3412	It's not as good as you think it is.	2
16	377 서연_과거	내가 쪽팔려서 그래 쪽팔려서	0.3182	0.6794	-0.3612	It's because I'm embarrassed.	5
17	816 서연_현재	내가 끝날게 그렇게 하게 해줘	0.6705	0.6705	0	Just let me finish this.	7
18	1202 승민_현재	-니가 그랬어 -안그랬다니까	0.6597	0.6597	0	-You did. -No.	5
19	747 승민_현재	뭐야, 별로야?	0.5106	0.6484	-0.1378	Why? You don't like it?	9
20	1313 승민_현재	니가 처음으로 우리집에 온 손님이야	0.6369	0.6369	0	You are the first guest in my place.	6
21	1095 서연_과거	아 씨발 좆같애!	0	0.6352	-0.6352	What the fucking shit!	5
22	876 서연_현재	축같은 년	0	0.624	-0.624	Fucking bitch.	8
23	1281	알이 들어가면 알탕이고 칼비가 들어가면 갈비탕인 데	0.6124	0.6124	0	Dishes are usually named after their ingredients.	5
24	842 서연_현재	다음주까지 숙제가 있습니다 리포트	0.5994	0.5994	0	You have an assignment due next week.	2
25	213	년 엄마한테 그제 뭐냐?	0	0.5949	-0.5949	Why are you so rude to your mom?	5

결과 및 논의 9:
ST와 TT 감정값 차이가 큰 25개 대사 II (문장구조적 차이)

번	남녀주인	ST	한글 vad	차이값	영어 vad	TT	SCER
1		구내 오락 학교 다닐 때 첫사랑 있었거든데	-0.6908	-1.3277	0.6369	I heard that he met his first love in the University.	3
2	479	엄마가 죽긴 왜 죽어?	-0.7783	-1.263	0.4847	You're not going to die.	11
3	1412 승민_현재	철길 건너가 먼저 떨어지는 사람이 지는 거	-0.6369	-1.2228	0.5859	The winner is who keep balances on the railwa	6
4	972 승민_과거	오락 시간 될 때 아무 때냐	-0.6908	-1.2014	0.5106	Whenever you're free.	6
5	1023 서연_과거	우리는 여기서 밥 먹고 술먹고 학교 가고 당구장 가 고	-0.8402	-1.1802	0.34	We eat, drink, go to school play billiards, etc.	2
6	186	집 짓길 잘한 거 같아	-0.7184	-1.1588	0.4404	I think I did a good job.	9
7	1308 승민_현재	그날자 명장았다 이렇게 예쁜 온채씨를	-0.6808	-1.1562	0.4754	He is a very lucky guy to marry you.	3
8	465 서연_현재	니가 신경 좀 써드려	-0.4215	-1.1484	0.7269	Take a good care of your mom.	5
9	794 서연_현재	돈도 많이 벌고 엄청 유명해질 거야	-0.4019	-1.1449	0.743	I will make a fortune, and gain fame.	6
10	914 서연_과거	적어도 돈 많은 남자랑 결혼할걸?	-0.5859	-1.1433	0.5574	I can get married to a rich man, at least.	6
11	917 서연_과거	우리엄마가 너때때 돌아가신 것도 아닌데	-0.6486	-1.1333	0.4847	My mom didn't die because of you.	2
12	388 서연_과거	아니 손목 때리기는 보통 사이에선 하지 않나?	-0.5859	-1.0626	0.4767	It is a special penalty between lovers.	6
13	977 승민_과거	힘내 새끼야	-0.5423	-1.0529	0.5106	Cheer up, bro.	8
14	1298	손목 때리기, 손목 때리기	-0.6597	-0.9997	0.34	Let's play hitting on the wrist.	8
15	1295	아 중추를 이렇게 하면 어떻게 생각은 하는데	-0.6369	-0.9981	0.3612	How about extending the house like this?	3
16	421 승민_현재	서울 집 다 정리하고	-0.6124	-0.9942	0.3818	As soon as I clear my things in Seoul	9
17	1303 서연_현재	엄마는 이 집이 언제까지 원아?	0.5859	-0.9997	0.4232	Keep your side of this house?	11
18	1414 승민_현재	엄마가 너 미워해?	-0.5106	-0.9691	0.4585	Doesn't your mom hate you?	2
19	248	이렇게 예쁜 여자친구를 두고	-0.5859	-0.9677	0.3818	You actually have a stunning girlfriend.	3
20	475 서연_현재	작은 살림을 가지고 봄을 기다려보는 것도 나쁘지 않을 것 같아서요	-0.5106	-0.951	0.4404	It may be good to wait and see what flowers bloom in spring.	6
21	987	응 아니 아직 안먹었어	-0.5423	-0.9442	0.4019	Yes, I mean no.	4
22	582 승민_과거	내 걱정하지 말고 새아가한테나 잘해줘	-0.296	-0.9337	0.6377	Don't worry about me and be nice to your wife	11
23	1406	도면에서 이미 너도 나 봤잖아 오케이 했고	-0.5267	-0.9286	0.4019	You knew the design and said yes.	5
24	742 승민_현재	자연스럽게 이렇게 뭐 이렇게 되겠지	-0.5574	-0.9186	0.3612	They do like this.	6
25	953	말단 뭐 저기 디자인 좀 날라가줘야 뭐 하나보다 할 거 아니야	-0.1531	-0.9181	0.765	A nice design would make a good impression.	1

결과 및 논의 10:
ST와 TT 동일 혹은 유사한 감정값 대사 예시 25개

1	번호	남녀주인공	ST	한글 vader	차이값	영어 vader	TT	SCENE
2	170	송민_현재	대체 어떤 게 세련된 넥타인데?	0.4404	0	0.4404	What kind of tie is good?	1
3	569		어, 곧장 집에 갈꺼지?	0.2263	0	0.2263	Are you going straight home?	4
4	865	서연_현재	아 알았어	0.2263	0	0.2263	Okay.	5
5	910	서연_과거	우리 아빠한테만 자랑이지	0.4767	0	0.4767	Only my father is proud of me.	6
6	920	서연_과거	공짜로 알았지?	0.5106	0	0.5106	For free.	6
7	1154	서연_현재	요즘 굶어?	-0.4215	0	-0.4215	Are you starving?	7
8	1277		아 놔 이 시발놈이	-0.34	0	-0.34	Hey you crazy guy.	8
9	1428		난 이게 좋아	0.3612	0	0.3612	I like outside.	11
10	813	송민_현재	아 그게 아니고	0	0	0	Oh, I mean...	5
11	815	송민_현재	야야 그게 아니고	0	0	0	Hey, I didn't mean that.	5
12	828	송민_현재	버티긴 왜 버티	0	0	0	Why did you do that?	5
13	830	서연_현재	그리고 변호사가 그러더라	0	0	0	And the lawyer said.	5
14	840	서연_현재	매운탕	0	0	0	Spicy soup.	5
15	843	서연_현재	이거는 그냥 매운탕	0	0	0	This is just spicy soup.	5
16	849	서연_현재	그냥	0	0	0	Just...	5
17	852	서연_현재	그냥 맵기만 하네	0	0	0	It's just hot and spicy.	5
18	853	송민_현재	어 사무실에 별일있지?	0	0	0	How is it going there?	5
19	858	송민_현재	아니야	0	0	0	No, it's not.	5
20	273		미팅나갈때 조금씩 아껴써	-0.3818	0.0201	-0.4019	Use it when you have a blind date.	2
21	761	송민_현재	오늘따라 궁금한것도 참 많고	0.34	0.0218	0.3182	You have many things you're curious about.	5
22	381	서연_과거	앱구정동도 진짜 멋있더라	0.6486	0.0246	0.624	Appgujeong is also really amazing.	2
23	774	서연_현재	참이상하네	-0.1531	0.0248	-0.1779	It's weird.	5
24	395	서연_과거	들올래?	0.1027	0.0255	0.0772	Do you want to listen?	2
25	1169	서연_현재	나 받고 싶은거 있는데	0.1027	0.0255	0.0772	I have something I want.	7
26	847	서연_현재	맘에 안들어	-0.25	0.0255	-0.2755	I don't like it.	5

결론

- ST의 감정 분석에서 극중 남녀 주인공의 갈등 구조 및 각 인물의 감정 변화가 잘 반영되었다.
- 그에 비해 TT의 감정 분석 값은 비교적 단조롭게 나타나 극중 인물의 갈등 구조를 설명하지 못하였다.
- ST와 TT의 감정 분석값의 차이는 어휘의 선택에서 기인하는 경우와 문장의 구문(not, no)의 차이에서 오는 것으로 보인다.
- 문화적 차이가 감정 분석값에 미치는 영향은 확인되지 않았다.
- 감정 분석이 번역 평가의 지표가 될 수 있는 가능성에 대한 추가 연구가 필요하다.

참고문헌

- 강유규, 김바로(Kang, W. K. and B. R. Kim). 2019. 고전소설에 대한 디지털 감정 분석방법론 탐색 - <구운몽>을 대상으로(A Study on the Methodology of Digital Emotion Analysis for Classical Novels - For The Cloud Dream of the Nine). 『동아시아고대학』(The East Asian Ancient Studies), 56, 349-377.
- 문석형, 강주영(Moon, S. H. and J. Y. Kang). 2019. 텍스트 마이닝 기법을 활용한 고전 추리 소설 작가 간 문체적 차이와 문체 구조에 대한 연구(A study on detective story authors' style differentiation and style structure based on Text Mining). 『지능정보연구』(Journal of Intelligence and Information Systems), 25.3, 89-115.
- 김유영, 송민(Kim, Y. Y and M. Song). 2016. 영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축(A Study on Analyzing Sentiments on Movie Reviews by Multi-Level Sentiment Classifier). 『지능정보연구』(Journal of Intelligence and Information Systems), 22.3, 71-89.
- 남길임, 강현아(Nam, K. I. and H. A. Kang). 2019. 말뭉치언어학적 관점에서 본 감성표현 추출의 쟁점 - 사용자 리뷰 말뭉치를 중심으로(Corpus Linguistic Perspectives on Issues Related to the Extraction of Sentiments Expressions from User Review Corpora). 『어문논총』(Eo-Mun-Lon-Chong), 82, 207-236.
- 서혜진, 이종현, 신정아(Seo, H. J., J. H. Lee and J. A. Shin). 2019. 딥러닝을 이용한 셰익스피어 작품의 감정분석 (Sentiment Analysis of Shakespeare's Plays Using a Deep Learning Technique). 『영어학』(Korean Journal of English Language and Linguistics), 19.4, 817-836.
- 유은순(You, E. S.). 2013. 영화 대사의 정량적 분석을 통한 등장인물의 감정과 서사간의 상관성 연구 (Analysis of the Correlation between Narrative and Emotions Displayed by Movie Characters through a Quantitative Analysis of Dialogues in a Movie). 『한국콘텐츠학회 논문지』(The Journal of the Korea Contents Association), 13.6, 95-107.
- 이경은(Lee, K. E.). 2018. *대화 코퍼스를 활용한 셰익스피어 사랑 비극의 감정 분석과 네트워크 분석 (Sentiment Analysis and Network Analysis of Dialogue Corpora in Shakespeare's Love Tragedies)*. 한국해양대학교 석사학위 논문(M.A. Thesis, Korea Maritime and Ocean University).

참고문헌

- 이상훈, 최정, 김종우(Lee, S. H., J. Cui and J. W. Kim). 2016. 영역별 맞춤형 감성사전 구축을 통한 영화리뷰 감성분석(Sentiment analysis on movie review through building modified sentiment dictionary by movie genre). 『지능정보연구』(Journal of Intelligence and Information Systems), 22.2, 97-113.
- 장세은, 박호민, 송원문, 이수상, 김재훈(Jhang, S. E., H. M. Park, W. M. Song, S. S. Lee and J. H. Kim). 2020. 감정분석의 시각화를 통한 셰익스피어 4대 비극작품 주요 등장인물의 막간 감정 추이와 감정 관계 연구(A Study of Emotional Changes across Acts and Emotion Relationship between Major Characters in Shakespeare's Four Great Tragedies through Visualization of Sentiment Analysis). 『언어과학』(Journal of Language Sciences), 27.2, 25-60.
- 장세은, 이경은, 박호민, 송원문, 정해룡, 이수상, 김재훈(Jhang, S. E., K. E. Lee, H. M. Park, W. M. Song, H. R. Jung, S. S. Lee and J. H. Kim). 2019. 대화 코퍼스를 통한 셰익스피어 비극 작품과 주요 남녀 등장인물 간의 감정분석 (Sentiment Analysis of Shakespeare's Tragedy Plays and Their Major Male and Female Characters through Dialogue Corpora). 『언어과학』(Journal of Language Sciences) 26.1, 115-147.
- Mohammad, S., S. M. Mohammad, and S. Kiritchenko. 2015. Sentiment after Translation: A Case study on Arabic Social Media Posts. *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, 767-777.
- Nalisnick, E. T. 2013. *Automatic Methods for Tracking Sentiment Dynamics in Plays*. Master of Science Thesis. Lehigh University.
- Nalisnick, E. T. and H. S. Baird. 2013. Character-to-Character Sentiment Analysis in Shakespeare's Plays. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* 2, 479-483.
- Piryani, R., D. Madhavi and V. Singh. 2017. Analytical Mapping of Opinion Mining and Sentiment Analysis Research During 2000-2015. *Information Processing & Management*. 53.1. DOI: 10.1016/j.ipm.2016.07.001
- Reagan, A. 2017. *Towards a Science of Human Stories: Using Sentiment Analysis and Emotional Arcs to Understand the Building Blocks of Complex Social Systems*. Ph. D. Dissertation. The University



Exploring lexical bundles in rhetorical move of research article abstracts in Applied Linguistics

Xiaolei LIANG (PhD student)
(Korea Maritime and Ocean University)

2021 Online Winter Conference
The Korean Association of Language Sciences
18th of February, 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 study attempts to explore lexical bundle features (bundle type and functions) in rhetorical moves in research article(RA) abstracts published in international journals by professional writers in order to establish possible connections between lexical bundles and research article(RA) rhetorical moves in applied linguistic abstracts.
- This paper also examines the frequency and structural characteristics of 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 study aims to explore relationship between rhetorical moves in RA abstracts frequent multiword units (lexical bundles) in each rhetorical move in RA abstracts.

So this study will address the following research questions:

- Q1: What are the **frequent multiword units** in each of the RA rhetorical moves of applied linguistic RA abstracts?
- Q2: What bundle **structure and function** can be identified in association to each move in applied linguistic RA abstracts?



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)(e.g. Biber et al. 1999) to 20 or 40 occurrences pmw (e.g. 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 be 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)
- However, literature shows that there is need to explore the connection between lexical bundles and research article(RA) rhetorical moves in Applied linguistic abstracts, which is the main reason for this study



3. Corpus and Methodology

3.1 Corpus

- A RA abstract corpus was built out of RA abstracts by L1 professionals.
- L1 corpus contains 604 RA abstracts randomly selected from privileged international journals
- The RA were published from 2005 to 2015,
- All in total were 105,238 in word tokens and 8,359 in word type.



3.2 Analysis framework and data process procedure

- Swales (1990) and Santos (1996) five-move scheme will be used as move analysis framework in this 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'.
- Biber, *et al* (1999) classification of lexical bundles category and function will be employed for multiword unit classification and function analysis in this study.
- Biber *et al* (1999, 2004) classified lexical bundles into four general classes: *verb phrase* fragments (VP-based), *dependent clause* fragment, *noun phrase* (NP-based) and *prepositional phrase* fragments (PP-based).



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

- 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; Wordsimth 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.

11

Bundle structure classification scheme *

Bundles	Types	Examples
NP-based	N1 noun phrase with post-modifier fragment	<i>the results of this study</i>
	N2 noun phrase without post-modifier fragment	<i>a retrospective cohort study</i>
PP-based	P1 prepositional phrase with embedded of phrase	<i>at the time of</i>
	P2 other prepositional phrase	<i>between the two groups</i>
VP-based	V1 (noun phrase+) copular be (+noun/ adjective phrase)	<i>were not significantly different</i>
	V2 (noun phrase+) passive verb (+prepositional/ gerund phrase fragment)	<i>were included in the</i>
	V3 noun phrase+ verb+ that-clause fragment	<i>these findings suggest that</i>
	V4 (noun phrase+) verb/adj+ to-clause fragment	<i>were more likely to</i>
	V5 (pronoun+) active verb +(noun / preposition/adverb phrase fragment)	<i>did not differ between</i>
	V6 (noun phrase+) passive verb+ to-clause fragment	<i>were randomised to receive</i>
Others		<i>significantly higher in the</i>

12

TAGs For Bundle functions*

Pan, Rappen & Biber (2016, 2020)

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

13

4. Results and Discussion

The most frequently occurred 4-word bundles were: *in this paper we, this paper reports on, the results of the, the results suggest that*, etc.

- 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*

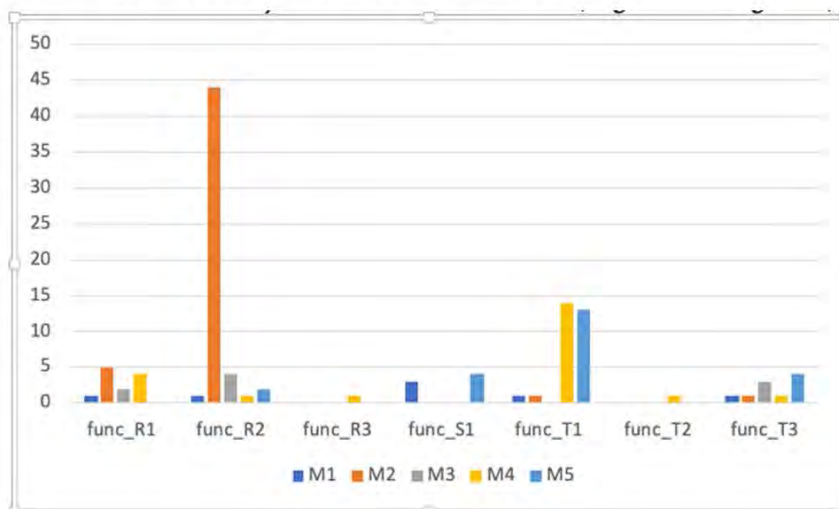
14

Bundle structure information in Moves

Category	Structure Category			
	NP	VP	PP	t_freq(ty/to)
Moves				
M1	3	3	1	7/34
M2	17	27**	7	51/289
M3	5	2	2	9/36
M4	4	13**	5	22/111
M5	5	16**	2	23/112
Total	34**	61**	17	112/582

* = 0.05, ** =0.01, ***=0.001

Bundle functional distribution in Moves



Bundle functions in Moves

Category	Functional Classifications						
	func_R1 *	func_R2	func_R3	func_S1 *	func_T1 *	func_T2	func_T3
Moves							
M1	1	1	NA	3	1	NA	1
M2	5	44**	NA	NA	1	NA	1
M3	2	4	NA	NA	NA	NA	3
M4	4	1	1	NA	14	1	1
M5	NA	2	NA	4	13	NA	4
Total	12	52**	1	7	19**	1	10

- Research-oriented: R1,R2,R3; Stance-oriented: S1; Text-oriented: T1,T2, T3
- * = 0.05, ** =0.01, ***=0.001

17

5. Conclusion

Q1: Frequently used lexical bundles 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*

18

Bundle structural preferences in Moves

Q2: bundle structural and functional distribution in Moves

- In general, abstract writers used VP-based and NP-based bundles when they wrote their RA abstract.
- There are significant differences between NP-based and VP-based bundles in M2, M4 and M5;
- For functional differences, M1 saw more stance-oriented bundles, while M2 favored Research-oriented (R2) bundles.
- In M3, writers preferred research-oriented and text-oriented bundles more than others
- In M4 and M5, writers used text-oriented bundles more frequently used.

19

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 [J]. *Applied Linguistics*, 25(3), 371-405.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing [J]. *Language Learning & Technology*, 14(2), 30-49.
- Hyland, K. (2008). 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 [J]. *Journal of English for Academic Purposes* 21:60-71
- Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics [J]. *Text-Interdisciplinary Journal for the Study of Discourse*, 16(4), 481-500.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings* [M]. Cambridge:Cambridge University Press.

20

영어 참고서에 나타난 번역 양상 연구 중학교 학년별 중심으로

신순란

부산대학교 대학원 영어 영문학과

목차

1. 서론
 2. 이론적 배경
 - 2.1. 해석이론
 - 2.1.1. 해석이론의 번역과정
 - 2.2. 문법-번역식 번역과 번역투
 - 2.2.1. 문법-번역식 번역과 독해식 접근법
 - 2.2.2. 번역투
 - 2.2.3. 번역투 유형
 3. 연구 방법
 - 3.1. 연구 대상
 - 3.2. 연구 도구
 - 3.3. 연구 절차 및 분석
 4. 결과 분석
 - 4.1. 중학교 학년별 영어 참고서 내 번역투 유형과 빈도 비율
 - 4.1.1. 번역투 유형에 따른 학년별 번역 양상
 - 4.1.2. 중학교 학년 변동에 따른 번역투 변동 양상
 - 4.2. 번역투 적용과 대안 번역
 - 4.2.1. 영어 참고서내 번역투 유형과 해결책
 - 4.2.2. 영어 참고서내 번역투와 대안 번역
 5. 결론
- 참고 문헌

1. 서론

2015년 제 7차 교육개정 과정의 읽기 비율

초등 20%, 중등 26%, 고등 28.5% 로 점점 증가

→ 제 2 외국어 학습에서 읽기 교육이 차지하는 비율과 중요성은 점차 증가하고 있는 추세

Crystal(2003) : 인터넷에 실린 정보의 80%정도가 영어로 작성, 읽기 교육의 중요성은 점차 커지고 있다.

김학양(2002) 김현아(2007) 이정원(2015) : 제 2외국어 학습에서 유창한 읽기 능력은 그 어떤 분야보다 중요.

한영자(1997) 류일영(2001) 박노철(2004) 박재욱(2009) : 영어 교육에서 **번역 교육의 필요성**을 제기

→ 문법-번역식(Grammar-Translation Method)의 문제점 지적

→ **해석적 번역 교육(Interpretive Theory of Translation)의 필요성을 제기.**

류일영(2001) 박재욱(2009) : 영어 참고서내 읽기 부분과 관련, 한국어와 영어의 차이를 고려하지 않은채 원문의 단어 그대로를 번역하는 번역투의 문제점을 지적하며 장기간에 걸친 참고서의 **번역투(Translationese)번역의 노출은 학생들의 의사소통 능력과 영어해석 능력에 부정적인 영향을 끼칠 수 있다.**

영어 참고서

: 해석적 번역의 적용이 시급한 분야.
 학습자의 영어 능력 수준과 상관없이 한글 번역본이 있어 학습자들이 영어 참고서에 의존.

신순란(2020) : 중학교 1학년 영어 참고서중 3종을 선별 중학교 1학년 영어 참고서내 문법-번역식 번역투에 대해 전반적으로 살펴보고 문제점을 지적, 해석적 번역에 근거한 대안 번역 제시.

이은숙(2008), 박재욱(2009) : 특정품사나 어휘 등으로만 이루어진 번역투 연구 범
 위와는 차별됨

한계점 : 중학교 1학년이라는 연구범위의 제한으로 영어 단어의 길이, 난이도, 내용의 수준 등에서 번역투 연구 결과를 일반화 하기에는 부족.

연구 범위의 확대 필요성 : 중학교 학년 전체의 문법-번역식 번역투의 양상과 중학교 학년별 특징을 대조, 분석.

본 논문의 연구 목적

: 중학교 1학년 영어 참고서내 번역투에 대한 연구와 대안 번역을 제시한 신순란(2020)의 연구를 바탕으로 **중학교 영어 참고서에 나타난 번역투 문장과 번역투의 양상을 학년별로 확장하여 비교, 분석하고 특징을 살핀 후 의사소통 능력향상을 위한 해석적 번역을 바탕으로 대안 번역을 제시하고자 한다.**

본 논문의 의의:

중학교 전 학년에 걸친 문법-번역식 번역투의 일반적인 양상을 체계적으로 살펴볼 수 있고 본격적인 영어 학습이 시작되는 중학교 학습과정에서 일차적인 단어 대 단어의 대응에 의한 어색한 번역투가 아닌 원천언어인 영어와 모국어의 차이에 기반한 **여러 층위의 번역이 있음을 학습자들이 인지하고** 상황맥락과 모국어 사용 용례에 맞는 번역 교육의 올바른 방향을 제시하는데도 도움이 될 것으로 사료된다.

2. 이론적 배경

해석적 번역과 문법-번역식 번역

2.1. 해석이론

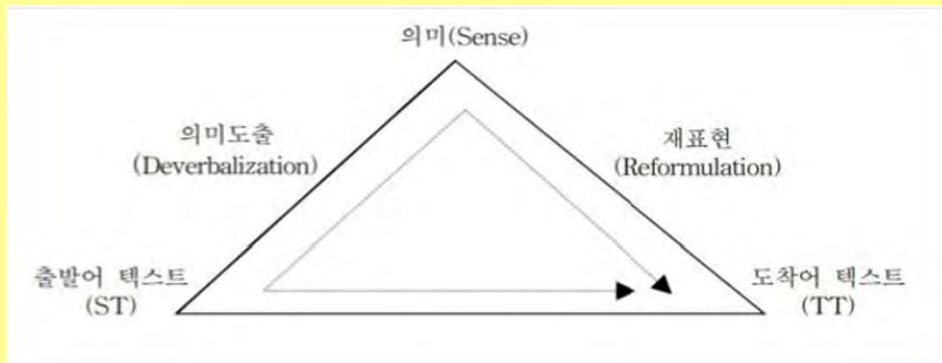
1970년대 파리 통번역대학원인 에지트(ESIT:lecolesuperiur tsettraducteursdeParis)에서 실시했던 번역교육이론.

해석이론의 핵심 : 번역이란 단어나 문장의 일대일 대응에 의한 코드 전환이 아니라 발화자 메시지의 의미(sense)를 파악한 후, 번역가가 자신의 경험과 지식을 활용하여 '해석'하고 상황과 맥락에 맞게 '재표현' 한다는 것.

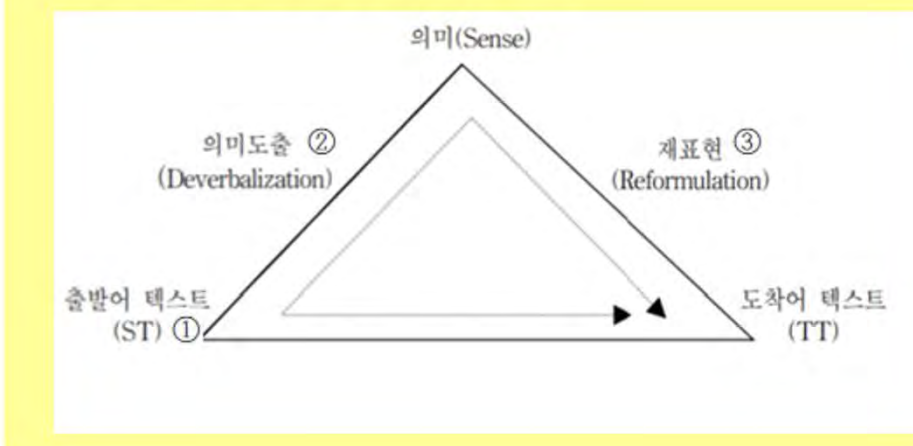
셀레스코비치(Seleskovitch), 레데르(Lederer) : 전문통역사이자 통번역 이론가. 해석이론의 창시자
듀리외(Durieux) : 기술번역분야 뿐 아니라 문학번역을 포괄하는 체계적인 번역 교육 방법으로 확장. 수립에 기여

2.1.1. 해석이론의 번역 과정

<그림 1> 번역의 삼각형 모델 (최정화, 1998, p22)



<그림 2> 번역의 삼각형 모델을 이용한 해석이론의 번역과정



1) 의미 도출 단계

(1)현동화 (Actualization) : 텍스트의 내용 즉 메시지를 이해할 수 있는 것은 특정한 개별 단어가 아니라 그 단어를 통해 우리가 이미 인지하고 있는 개념 혹은 이미지나 경험을 떠올리는게 하는 것.

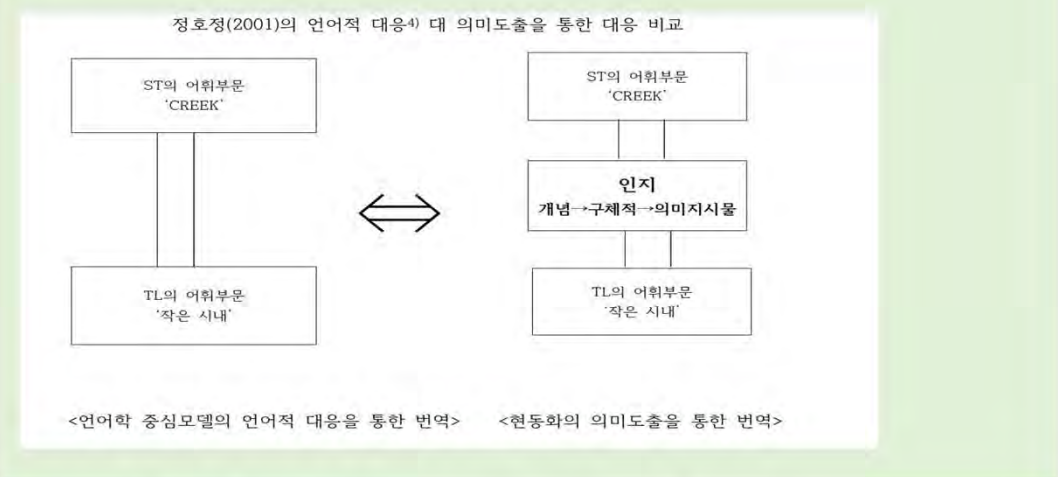
정호정(2011) : **현동화의 개념을 작은 시냇물을 뜻하는 Creek 을 통해 설명.**

Creek : 보통 강보다 작은 물의 흐름 이라는 개념 + 어릴적 기억에 있던 시원한 물소리와 시각적 이미지 환기

→ 더 이상 개념으로만 존재하는 단어가 아니라 경험과 연결된 구체적 의미를 갖는 지시물.

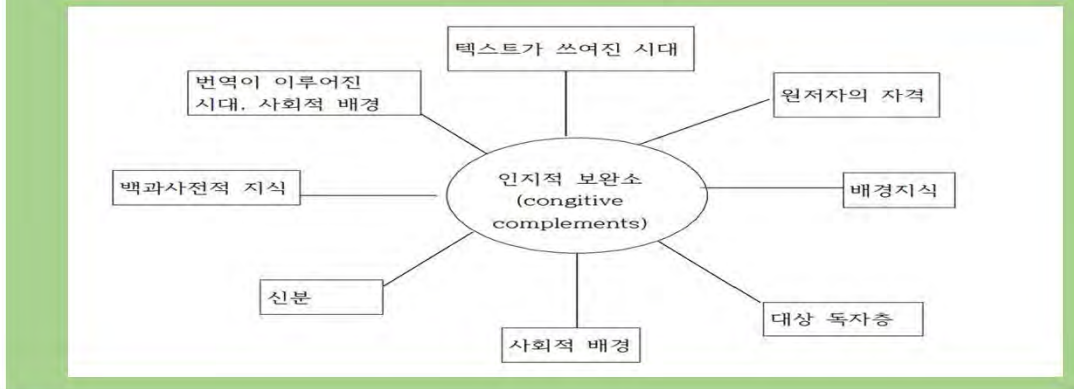
<그림 3> 해석이론의 의미도출 과정

정호정(2001)의 언어적 대응 대 의미도출을 통한 대응 비교



(2) 인지적 보완소 (Cognitive Complements) : 발신자의 메시지를 정확히 이해하기 위해서는 언어뿐만 아니라 언어외적인 요소들 즉 백과사전적 지식, 배경지식, 텍스트가 쓰여진 시대, 사회적 배경, 원 저자의 자격, 신분, 번역이 이루어지는 시대나 사회적 배경, 대상 독자층 등에 대한 이해가 필요(Lederer, 2001).

<그림 4> 인지적 보완소 (Cognitive Complements)의 구성요소 (Lederer, 2001)



2) 재표현 (Reformulation) :

해석적 번역 과정의 마지막 단계인 재표현 과정에서 중요한 것은 **두 언어 사이의 등가 창출**.

- 번역행위란 서로 다른 언어로 된 텍스트를 이해하고 등가(equivalence)를 재현하는 것.
- Lederer(2001): 원천 텍스트와 목표 텍스트의 등가가 알맞게 성립되기 위해서는 낱말이나 문법 구조의 일대일 '대응'이 아니라 원천 텍스트와 번역 텍스트에 대한 **총괄적 등가관계**가 이루어져야 한다.
- Nigel(2005) : Vinay와 Darbelent의 7가지 번역방법을 소개하여 등가를 구현하는 방법을 소개. → 번역의 과정을 차용어, 모방어, 축역, 전이, 변조, 등가, 변환으로 나누고 특히 언어구조나 언어 계통이 상이한 두 언어에서는 품사전환이나 낱말의 순서, 낱말 전환등
- 이은숙(2007) : 번역에서 달성해야 하는 것은 단순한 언어차원의 '대응'이 아니라 '전체 텍스트에서의 포괄적 등가' 따라서 원천 텍스트와 목표 텍스트의 포괄적 등가를 이루기 위해서는 어순이나 품사의 과감한 전환 혹은 인지적 맥락에 의한 생략 등 언어적 등가의 달성뿐 아니라 공손전략(politeness)등의 사회언어학적 차원의 등가가 포함되어야 함.

<표 1> W. 콜러의 5가지 등가 분류

등가의 종류	등가의 내용
1. 외연적 등가 (Denotative equivalence)	하나의 텍스트의 초언어적 내용의 등가와 관련 되어 있다. 풀러는 이것을 '내용적 불변'이라 부른다.
2. 내포적 등가 (Connotative equivalence)	언어 사용역, 사회언어, 표현의 사용지역 등의 문체를 존중하는 것을 말하며 '문체적 등가'라고도 한다.
3. 텍스트-규범적 등가 (Text-normative equivalence)	번역된 텍스트의 유형에 적합해야 함을 의미한다. 예를 들어 요리법을 법학개론처럼 쓰지 않는다.
4. 화용적 등가 (Pragmatic equivalence)	'소통적 등가'라고도 하며 텍스트 또는 메시지 수신자 중심이다. 번역이 이해되려면 독자의 지식에 맞추어져야 한다.
5. 형식적 등가 (Formal equivalence)	'표현적 등가'라고도 하며 텍스트의 형식과 미학적인 부분과 관련되어 있으며 원천텍스트의 단어과 개개의 문체적 특징을 포함한다.

2.2. 문법-번역식 번역과 번역투

2.2.1. 문법-번역식 번역과 독해식 접근법

선형성(1990) : 지금까지 **우리나라의 영어 교육에 도입된교수법은 첫째 문법-번역식 (Grammar-Translation Method, GTM)교수법인 영문해석법 또는 영문 독해법**과 또하나는 이것에 대한 끊임없는 대응의 신교수법

→문법-번역식 번역은 최근까지도 우리나라 제 2외국어 교육 현장에서 사용되는 있는 만큼 긴 역사와 함께 많은 영향을 끼치고 있다.

김태영(2016) : 중세 시대의 고전과 성서가 대부분 그리스어와 라틴어로 작성되었기에 지식과 교양을 쌓기 위한 하나의 방법으로 19세기 말 유럽과 미국에 도입되었으며 전통 교수법(traditional method)라고도 한다.

Larsen-Freeman, Anderson(2016) : 문법-번역식 번역의 특징

<표 2> 문법-번역식 번역의 특징

학습목표	목표 언어로 쓰인 읽기 자료를 정확히 이해하기 위해 모국어로 번역할 수 있는 능력을 기르는 것.
학습 초점	구어(spoken language)보다는 문어(written language)학습에 중점을 둔다.
학습 영역	단어와 문법이 강조되고 읽기 위주의 학습을 하고 발음이나 말하기, 듣기 영역에는 큰 관심을 두지 않는다.
학습 지도	교사는 권위자의 역할을 하며 학생들의 문법 오류를 정확하게 수정해주고 답을 제시하는 것을 중요하게 여긴다.

<그림 5> 듀리외(2003)의 독해식 접근법



- ① 원문 텍스트 ②대응어 찾기 ③텍스트의 이해 ④최종텍스트의 작성

듀리외(2003)의 독해식 접근법 : 원문 텍스트인 ① 이 전달하고자 하는 바를 ②번의 대응어 찾기를 통해 텍스트 이해 단계인 ③ 번 과정없이 바로 ④ 번인 최종 텍스트의 작성이 끝난다.

→ 독해식 접근법에서는 텍스트는 단지 단어의 나열일 뿐이고 전체 발화문의 의미는 개별 단어의 총합으로 여기기 때문에 각 텍스트의 언어적 코딩 과정이 끝난 이후에는 원문 텍스트는 더 이상 고려의 대상이 되지 않음.

→

문법-번역식 번역과 독해식 접근법은 동의어

2.2.2. 번역투

국립국어원 : 번역투란 번역한 글에서 나타나는 자연스럽지 못한 말투. 따라서 번역투는 그 자체로 부정적인 의미를 내포.

-**이영옥(2001)**: 명사중심의 영어와 동사중심의 한국어의 특징을 간과한 채 원문의 영어 문장 구조를 그대로 한국어로 번역시 한국어 체계와는 맞지 않는 어색한 문장이 형성.

-**김정우(2003)** : 초·중·고등학교 국어 교과서에 나타난 영어번역투를 연구하고 분류. 외국어 문장을 번역하는 과정에서 모국어 사용용례에 맞지 않게 외국어 문장을 그대로 직역할때 발생하는 부자연스런 국어 문장 표현으로 정의.

-**이용성(2004)**: 한국어와 영어가 지닌 구문의 특징과 번역투가 발생하는 두 언어 구조를 분석하여 자연스런 영한 통역과 번역을 위한 10가지 지침을 마련.

-**이근희(2005)**: 대학교 학부생을 대상으로 번역실습실험을 통해 번역투의 유형과 빈도를 조사하고 번역투의 개념, 유형, 발생원인 및 번역투의 기능을 종합적으로 연구. → 번역투 발생원인을 모국어에 대한 지식 부족, 중·고등학교 시절 학습된 번역투의 고착화등을 원인으로 제시.

윤주희(2011): 특수 교육분야의 번역서를 연구대상으로 하여 번역투 유형을 분류하고 번역투 원인을 연구.

Shuttleworth & Cowie(1999): 번역투란 두드러지게 원천언어의 특징에 의존한 탓에 도착언어에 스인 목표언어가 매우 부자연스럽고 이해하기 어려우며 우습게 여겨지기조차 할 때 사용하는 경멸조의 언어.

신순란(2020) 중학교 1학년 영어 참고서에 나타난 번역투와 관련하여 가장 최근의 연구 수행. 여러 학자의 번역투 유형을 재분류하여 번역투를 새로이 구축함. 학창시절 지속적인 번역투의 노출은 학습자의 의사소통 능력을 약화시킬 뿐 아니라 학습자의 번역 학습에도 향후 부정적인 영향을 끼칠 수 있다.

<표 3> 번역투 유형에 따른 분류

번역투 유형	김정우(2003)	이용성(2004)	이근희(2005)	윤주희(2011)	신순란(2020)
대명사		○	○	○	○
동사직역	○		○	○	○
복수 '들'	○	○	○	○	○
소유격 '의'	○	○	○	○	○
무생물 주어	○	○	○	○	○
파·사동	○	○	○	○	○
전치사	○		○		○
품사전환					○
어순재배치					○

3. 연구 방법

3.1. 연구 대상

중학교 학년별 영어 참고서 12종 중 출판사 세 곳의 참고서 (A, B, C) 를 선택하여 총 9권을 선택하고 이 중 홀수과만을 분석 대상으로 삼음

➔ 각 영어 참고서의 구성이 말하기 듣기, 문법, 읽기 등 구성이 동일하고 학년별 단어수가 일정하기 때문.

- 각 한년별 분석 대상 문장 수 : 1학년 329 2학년 413 3학년 398

- 2, 3학년의 문장 개수 차이는 15개정도로 큰 차이가 없고

- 1학년의 경우 주어 +동사 가 하나인 단문이 대부분

- 2학년 부터 관계대명사나 접속사를 사용한 복문이 두드러짐.

전체 분석 대상 문장: 1,140

<표 4> 중학교 1학년 영어 참고서 분석대상

출판사	저자	분석과	제목	문장 수
A	Kim 외 8명	Lesson 1	My friend on a world map	26
		Lesson 3	Let's go to the flea market	29
		Lesson 5	The Impossible Team	38
		Lesson 7	Who Brings the Gifts?	27
B	Yoon 외 8명	Lesson 1	My heart Map	16
		Lesson 3	Full House	35
		Lesson 5	The world of Hate	22
		Lesson 7	Kelly's Road trip to Namhae	29
C	Jeong 외 5명	Lesson 1	Welcome to My World	20
		Lesson 3	We are Special	35
		Lesson 5	We Are Not Alone!	21
		Lesson 7	A Tasty Painting	29
총 문장수				329

<표 5> 중학교 2학년 영어 참고서 분석대상

출판사	저자	분석과	제목	문장 수
A	Kim 외 8명	Lesson 1	All about selfies!	42
		Lesson 3	Traditional Korean symbols	44
		Lesson 5	The Spirit of Audrey	29
		Lesson 7	Jobs of the Future	38
		Lesson 1	My Favorite Time of the Day	28
		Lesson 3	Let's Talk about Upcycling	28
		Lesson 5	Be Smart with Your Smartphones!	33
B	Yoon 외 8명	Lesson 7	The best New Thing	45
		Lesson 1	The Tickets	32
C	Jeong 외 5명	Lesson 1	The Tickets	32
		Lesson 3	Fun Places for Selfies	34
		Lesson 5	Shapes Around Us	26
		Lesson 7	Write Your Feelings	34
		Lesson 1	Write Your Feelings	34
총 문장수				413

<표 6> 중학교 3학년 영어 참고서 분석대상

출판사	저자	분석과	제목	문장 수
A	Kim 외 8명	Lesson 1	Learn a New Language. Find a New World	41
		Lesson 3	Waking up to an Earthquake	38
		Lesson 5	Join Hands. Save the Earth	31
		Lesson 7	Life with Robots	35
B	Yoon 외 8명	Lesson 1	Find a Dream and Live It!	38
		Lesson 3	English Words of Foreign Origin	28
		Lesson 5	Hidden Peoples in Sports	32
		Lesson 7	Living with Big Data	36
C	Jeong 외 5명	Lesson 1	How Do You "Ha-Ha" in Your Texts?	39
		Lesson 3	Laugh First and Then Think	21
		Lesson 5	A Picture Letter from a Mother of Three	36
		Lesson 7	How to Be a Smart News Reader	21
총 문장 수				398
중학교 3학년 연구대상 총 문장수				1140

3.2. 연구 도구

중학교 학년별 영어 참고서에 나타난 번역투 양상을 종합적으로 살펴보기 위한 본 연구의 목적에 따라 **신순란(2020)**이 **구축한 분석틀**을 따랐으며, 구축된 분석틀은 다음과 같다. **신순란(2020)**은 **김정우(2003), 이근희(2005), 김지혜(2006)**가 분류한 번역투 유형 중 두 학자 이상이 공통으로 분류한 번역투를 분석틀로 삼았으며 또 기존 번역투에 포함되지 않았던 **'품사 전환', '어순 재배치'**의 번역투를 새롭게 번역투 분석틀에 포함시켰다.

<표 7> 중학교 영어 참고서에 나타난 번역투 분류와 예(신순란,2020)

	번역투 예문
대명사	그녀는 서둘러 수업을 받으러 갔다.
동사	더 나은 학교생활을 만들기 위해 저에게 동참하는 것은 어떠세요?
무생물 주어	그건 그들을 신나게 한다.
~들	셀피를 찍거나 게시하거나 보기 전에 스스로에게 이 질문들을 해 보라
전치사	3D 모형 제작자에 <u>관하여</u> 들어 본 적이 있으신가요?
피·사동문	그것은 사람들을 행복하게 만들었다.
~의	당신이 <u>자신의</u> 사진을 찍으며, 그것이 셀피입니다.
품사 전환	이 손님방은 전통적인 한국의 물건들로 가득 차 있어.
어순 재배치	방금 너를 봤어. 무척 너는 빠르더라!

3.3. 연구 절차 및 분석

본 연구는 다음과 같은 절차에 의해 수행되었다. 먼저 2020년 5월 첫째 주 부터 약 3주간에 걸쳐 중학교 영어 참고서를 선정하고 구매한 후 기존 신순란(2020)의 분석틀을 바탕으로 상향식(bottom up)방식으로 문법-번역식 문장을 분류하였다. 2020년 6월 부터는 다시 한번 연구의 객관성을 확보하고 코딩의 타당성을 검증하기 위해 현재 통·번역 전문가로 활동하시는 분과 해외에서 10년이상 거주 경험이 있는 번역학 전공자에게 의뢰하였다. 9월 셋째 주 부터는 해석적 번역을 바탕으로 한 대안 번역을 제시하였으며 다시 한번 대안번역의 적절성을 확인하기 위해 위에서 언급한 두분께 코딩을 의뢰하였다.

4. 결과 분석

4. 1. 중학교 학년별 영어 참고서에 나타난 번역투 유형과 빈도 및 비율 (신순란, 2020)

<표 8> 중학교 1학년 영어 참고서에 나타난 번역투 유형과 빈도 및 비율

번역투 유형	빈도수	비율 (%)
대명사	111	34
소유격 '의'	35	11
전치사	6	1.82
피·사동문	5	1.51
동사직역	5	1.51
무생물 주어	5	1.51
복수 '들'	5	1.51
품사 전환	4	1.22
어순 재배치	4	1.22

<표 9> 중학교 2학년 영어 참고서에 나타난 번역투 유형과 빈도 및 비율

번역투 유형	빈도수	비율 (%)
대명사	167	40.4
소유격 '의'	68	16.4
전치사	10	2.42
피·사동문	7	1.7
동사직역	9	2.2
무생물 주어	8	1.9
복수 '들'	63	15.3
품사전환	11	2.7
어순 재배치	12	2.9

<표 9> 중학교 2학년 영어 참고서에 나타난 번역투 유형과 빈도 및 비율

번역투 유형	빈도수	비율 (%)
대명사	167	40.4
소유격 '의'	68	16.4
전치사	10	2.42
피·사동문	7	1.7
동사직역	9	2.2
무생물 주어	8	1.9
복수 '들'	63	15.3
품사전환	11	2.7
어순 재배치	12	2.9

4.1.2. 번역투 유형에 따른 학년별 번역 양상

<표 11> 번역투 종류와 학년별 양상 (대명사)

번역투 유형	빈도수	비율 (%)	증감율 (%)
중학교 1	111	34	
중학교 2	167	40	50.4
중학교 3	182	46	9

<표 12> 학년별 번역투 종류와 양상(소유격의')

번역투 유형	빈도수	비율 (%)	증감율 (%)
중학교 1	35	11	
중학교 2	68	16.4	94.3
중학교 3	50	13	-26.5

<표 13> 학년별 번역투 종류와 양상(복수 '들')

번역투 유형	빈도수	비율 (%)	증감율 (%)
중학교 1	5	1.5	
중학교 2	63	15.3	1160
중학교 3	89	22.3	41.3

<표 14> 학년별 번역투 종류와 양상(전치사)

번역투 유형	빈도수	비율 (%)	증감율 (%)
중학교 1	6	1.82	
중학교 2	10	2.42	66.7
중학교 3	18	4.5	80

<표 15> 학년별 번역투 종류와 양상 (피·사동문)

번역투 유형	빈도수	비율 (%)	증감율 (%)
중학교 1	5	1.51	
중학교 2	7	1.7	40
중학교 3	25	6.3	257

<표 16> 학년별 번역투 종류와 양상(동사직역)

번역투 유형	빈도수	비율 (%)	증감율 (%)
중학교 1	5	1.51	
중학교 2	9	2.2	80
중학교 3	16	4.0	78

<표 17> 학년별 번역투 종류와 양상(무생물 주어)

번역투 유형	빈도수	비율(X)	증감율(X)
중학교 1	5	1.5	
중학교 2	8	1.9	60
중학교 3	19	4.8	138

<표 18> 학년별 번역투 종류와 양상 (품사전환)

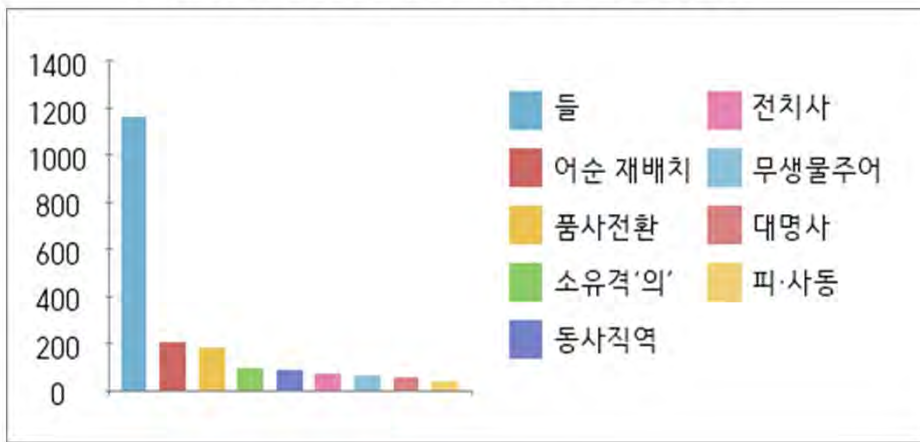
번역투 유형	빈도수	비율(X)	증감율(X)
중학교 1	4	1.2	
중학교 2	11	2.7	175
중학교 3	26	6.5	136

<표 19> 학년별 번역투 종류와 양상 (어순 재배치)

번역투 유형	빈도수	비율(X)	증감율(X)
중학교 1	4	1.2	
중학교 2	12	2.9	200
중학교 3	39	9.8	225

4.1.3. 중학교 학년 변동에 따른 번역투 변동 양상

<그림 6> 중학교 1학년부터 2학년으로 학년 이동에 따른 번역투 양상



<그림 7> 중학교 2학년부터 3학년으로 이동에 따른 번역투 양상



4.2. 번역틀 적용과 대안 번역

< 표 20 > 영어참고서내 번역투 유형과 해결책

번역투 유형	해결책
대명사	① 한국어의 경우 불명요소가 모두 필요한 것은 아니다. ② 상호 인지의 경우 생략하는 것이 더 자연스럽다.
소유격 '-의'	① 수량을 나타내거나 비유적 표현 등을 제외하고는 생략한다. ② 소유, 기원, 시간의 경우 소유격 '-의'를 사용하지 않는 것이 자연스럽다. ③ 한국어에서는 어떤 행위의 시작이나 대상은 -에게서 라는 표현이 자연스럽다.
접지사	① 접지사 for의 경우 -향하여, -대신에, -에 적합할 -에 대비한 등 총 24가지의 의미가 있으므로 문맥에 맞게 꼭꼭 사용해야 한다. ② 접지사 about의 경우 -에 대하여 라는 기본적인 의미외에도 -역기까지, -내내, -거의 등 8개의 다른 의미가 있으므로 문맥에 따라 사용해야 한다. ③ 접지사 with의 경우 -와 함께 라는 핵심적 개념외에도 동반, 수단, 대상, 비교 등 총 20개의 의미가 있으므로 문맥에 맞게 꼭꼭 사용해야 한다.
과-사동문	① 한국어에서는 가급적 수동태를 지피고 능동문을 사용한다. ② 주어와 동작주를 바꾸어 번역한다.
동사격	① 주어와의 호응 관계를 고려해서 적절히 번역해야 한다. ② have, make, provide 등의 동사를 사동격 의미 그대로만 적용하지 않도록 해야 한다.
부정형 주어	① 한국어에서는 부정형 주어용어의 사용하지 않는다. ② 한국어로 번역 시 무조건사 -에, -것을 붙이거나 적절한 문장 전환을 해야 한다.
부수 '-를'	① 문맥상 자연스럽게 부수임을 알 수 있는 경우에는 '를'을 붙이지 않는다. ② 이미 부수 개념이 포함되어 있을 때도 사용하지 않는다.
형사 전환	① 원문과 목표 텍스트의 언어구조가 다를 때는 원문의 동사에 의미가 없게 된다. ② 영어 형용사는 한국어의 부사로 바꾸는 것이 자연스럽다. ③ 영어 명사는 우리말 동사나 부사로 바꾸면 자연스럽다.
주어 생략	① 부사와 동사는 되도록 연결하여 사용한다. ② 문맥이 불다면 한국어 사용용례에 따라 의미를 띄고 맞게 한다.

4.2.2. 영어 참고서내 번역투와 대안 번역

1) 어순 재배치

① He made the hangers, the plants, and all the round things square.

-- 그는 옷걸이와 식물과 모든 둥근 물건들을 사각형으로 만들었다.

대안 번역) 그는 옷걸이와 식물과 둥근 물건을 모두 사각형으로 만들었다.

** <표 20> 의 어순 재배치 해결책 ② 에 의해 형용사 + 형용사+명사 한국어 용례에 따라 어순을 매끄럽게 정리하는 것이 자연스럽다.

2) 품사 전환

This guest room is full of traditional Korean things.

-- 이 손님방은 전통적인 한국의 물건들로 가득차 있어.

대안 번역) 이 손님방에는 전통적인 한국물건이 많아.

** <표 20> 품사 해결책 ③ 명사는 동사나 부사로 바꾸면 자연스럽다에 의해 부사로 전환

3) 전치사

Here are some tips for these problems. --- 이런 문제들을 위한 몇가지 조언이 여기 있다.

대안 번역) 이런 문제를 해결할 몇가지 조언이 여기 있다.

4) 피. 사동

① The left one was created by Chusa in the Period of Joseon Dynasty.

-- 왼쪽 작품은 조선 왕조 시대에 추사에 의해 창작되었다.

대안 번역) 왼쪽 작품은 조선 왕조 시대에 추사가 창작하였다.

5) 무생물 주어

① "what?" This will not hold my clothes. ---"뭐야? 이것은 내옷을 걸고 있지 못할 거야"

대안 번역) 뭐야, 여기에 옷을 걸지 못할 것 같은데.

6) 동사 직역

① Selfie museums **exist** in Korea, too. --- 셀피 박물관은 한국에도 **존재합니다**.

대안 번역) 셀피 박물관은 한국에도 **있습니다**.

5. 결론

첫째, 학년이 올라감에 따라 번역투 소유격"의 " 를 제외한 **모든 번역투 유형에서 문법-번역식 번역투가 증가하였음을** 알 수 있었다.

→ 학년 증가에 따른 번역투가 심화, 확대되어 중학교 학습자가 번역투를 더욱 고착화 시킬 위험이 있다.

둘째, **번역투 복수 '~들'(1160%), '피.사동'(257%), '어순 재배치'(225%), ' 품사전환'(175%)**은 학년이 올라감에 가장 많은 증가율을 보인 번역투 유형이다.

→ 특히, 피.사동의 경우 2학년에서 3학년으로 이동시 문장수의 감소에 비해 증가율이 257%로 급등을 보여 향후 번역투 개선이 시급한 유형임을 알 수 있었다.

셋째, 문법-번역식 번역의 한계점에도 불구하고 여전히 **영어 참고서내 읽기 부분에서는 이를 벗어나지 못하고 있음**을 보여주었다.

넷째, 번역투 '동사 직역', '전치사', '무생물 주어'는 중학교 전 학년에 걸친 빈도와 비율에서 모두 5% 미만의 낮은 분포를 보였다.

→ 향후 영어와 한국어의 차이를 바르게 인지하여 세심하게 번역한다면 중학교 학습자에게 좀 더 유익한 참고 자료로서 기여 할 수 있을것이라 사료된다.

연구의 한계점

중학교 영어 참고서 종류 12종 중 3종만의 연구대상으로 삼았기에 중학교 영어 참고서의 번역투 양상을 일반화 하는데는 한계가 있을 수 있다.

본 연구를 통해 학습자가 번역에는 다양한 층위가 있음을 인지하고 단어 대 단어의 일차적 사전적 대응이 아닌 외국어와 모국어의 언어적 차이에 기반한 포괄적 등가를 구축하여 자연스럽게 매끄러운 번역이 활성화되고 아울러 교실 영어에서도 올바른 번역에 대해 다시 한번 고민해 보는 계기가 되기를 기대해 본다.

참고문헌

- 김미향. 2006. 교과서 분석을 통한 의사소통 중심의 영어교육: 중학교 교육을 중심으로. 아주대학교 교육대학원 석사학위논문.
- 김지혜. 2006. 번역투의 개념제고: 국어 교과서, 국어학자와 번역학자의 글을 중심으로. 부산외국어대학교 통역번역 대학원 석사학위논문.
- 김정미. 1992. 외국어 교육에서 번역에 관한 연구. 이화여자대학교 석사학위논문.
- 김정우. 2003. 교과서 문장의 영어 번역투 사용실태-초·중·고등학교 국어교과서를 중심으로. 『인문논총』 1.16, 139-151.
- 김태영. 2016. 영어 교육을 일면 영어가 보인다. 서울: 글로벌 콘텐츠.
- 김학양. 2002. 고등학교 영어 읽기 지도의 실태에 관한 연구. 전남대학교 석사학위논문.
- 김현아. 2007. 영어 독해 능력 향상을 위한 효율적인 독해 전략 지도. 『교육연구논문집』 28.3, 185-199.
- 김효중. 2004. 해석학적 번역이론과 텍스트 기능. 『번역학연구』 5.1, 77-95.
- 박노철. 2004. 영어교육에서 번역교육의 필요성연구. 『영어교육연구』 3.28, 39-73.
- 류일명. 2001. 영어 교육에 있어 번역의 필요성. 세종대학교 석사학위논문.
- 박시현, 이향. 2003. 전문번역 어떻게 가르칠 것인가?. 서울: 고려대학교 출판부.
- 박재욱. 2009. 영어교육에 있어 번역의 필요성과 번역 전략 적용 가능성. 부산외국어대학교 석사학위논문.
- 방혜영. 2009. 고등학교 영어교과서를 통한 전치사 for의 용례 분석 및 교수방안. 경희대학교 석사학위논문.
- 서계인. 2004. 실전 영어번역의 기술. 서울: 북라인.
- 신순란. 2020. 중학교 1학년 영어 참고서에 나타난 번역의 문제와 대안 번역. 『영어교육』 32(1), 89-106.


- 이근희. 2005. 번역산책-번역투에서 번역의 전략까지. 서울: 한국문화사.
- 이근희. 2005b. *영한 번역에서의 '번역투' 연구*. 세종대학교 박사학위논문.
- 이근희. 2005c. 영한 번역에서의 효율적인 번역 교육을 위한 소고. 『번역학연구』, 6(1), 99-119.
- 이영옥. 2001. 한국어와 영어간 구조의 차이에 따른 번역의 문제-수동구문을 중심으로. 『번역학 연구』 1.2, 47-76.
- 이영옥. 2002. 한국어와 영어간 구조의 차이에 따른 번역의 문제-인용문의 번역을 중심으로. 『번역학연구』 3(1), 59-81.
- 이용성. 2004. 통번역을 위한 영한 대조분석. 『부산외국어대학교 논문집』 26(2), 9-38.
- 이은수. 2011. *전치사 about 과 around 의 의미 분석*. 충남대학교 석사학위논문.
- 이은숙. 2007. 해석이론과 등가에 관한 연구. 『번역학 연구』, 8.2, 245-261.
- 이은숙. 2008. 영어 수동문의 한국어 번역 연구: 현행 고등학교 영어 교과서 번역사례 중심으로. 『번역학 연구』, 9.2, 167-190.
- 이정원. 2015. 영어 교육에 게재된 읽기 관련 논문 분석. 『한국영어교육학회』 70, 53-80
- 이향. 2009. 해석이론의 특징과 한계. 『번역학 연구』 10.1, 121-139.
- 윤주희. 2011. *최신 특수교육 번역 분석: 번역투 유형과 가독성 개선 전략*. 계명대학교 석사학위논문.
- 전성기. 2001. 번역의 오늘. 서울: 고려대학교 출판부.
- 정호정. 2001. 번역학에 있어서의 통합모델-해석이론을 중심으로-. 『통역과 번역』 3, 243-266.
- 조인정. 2005. 영한 번역의 문제점 :수동태를 중심으로. 『한국번역학회』. 6.1. 121-142.
- 조창현. 2017. *고등학교 영어 교과서에 나타난 전치사 'with' 의 용례 및 의미분석*. 건국대학교 석사학위논문.

- 한영자. 1996. *외국어 교육에 있어 해석적 번역의 활용에 관한 연구*. 이화여자대학교 석사학위논문.
- 한정은. 2014. 학부 중한번역 수업에서의 번역이론 적용가능성-해석이론 중심의 고찰. 『통역과 번역』 16.1, 225-252.
- 최진영. 2017. *독해능력 향상을 위한 영어문법 지도 방안 연구*. 건국대학교 교육대학원 석사 논문.
- Baker, M. 2005. *In other words* (2nd ed.). London & New York: Routledge.
- Crystal, D. 2003. *English as a global language* (2nd ed.). New York: Cambridge University Press.
- Durieux, C. 1988. *Fondement didactique de la traduction Technique*. Paris: Didier Erudition.
- Halla, S. 2015. Venuti versus Nida. *Federation Internationale des Traducteurs*, 61(1), 78-92.
- Koller, W. 1979. *Translation Studies*.
- Larsen-Freeman, D & Anderson, M. 2016. *Techniques and Principles in language teaching*. Oxford: Oxford university press.
- Lederer, M. 1994. *La traduction aujourd'hui*. Paris: Hachette.
- Munday, Jeremy. 2001. *Introducing Translation Studies: Theories and Applications*. London & NY: routledge.
- Newmark, P. 1981. *Approaches to translation*. Oxford: Pergamon Press.
- Nigel, A. 2005. *Translation, linguistics, culture: A French-English handbook*. Bristol: Multilingual Matters Ltd.
- Rivers, Wilga M. 1981. *Teaching Foreign-language Skills*. Chicago & London: The Univeristy of Chicago Press.
- Shuttleworth, M. & Moria, C. 1999. *Dictionary of translation studies*. Manchester: St. Jerome Publishing.
- Snell-Hornby, Mary. 2006. *The turns of Translation Studies*. Amsterdam: John Benjamins publishing.

감사합니다. !! ~~~~



Contents



1. Introduction
 - Research Background
 - Research Purpose and Significance
 - Research Questions
 - Paper Layout
2. Literature Review
3. Data and Research Methodology
4. Results and Discussion
5. Conclusion

References

1. Introduction

1.1 Research Background

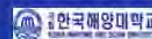
- The appropriate use of **collocation** is now widely recognized as a prerequisite for proficient language use (Cowie 1998; Sinclair 1991; Wray 2002). Nevertheless, achieving decent accomplishments in collocation use for second language learners can not be that easy (Granger 1998; Howarth 1998; Nesselhauf 2003, 2004). Several researchers have suggested that second language learners depend heavily on creativity and thus make over liberal assumptions about the collocational equivalence of semantically similar items.
- The term **semantic prosody** was firstly employed by Louw (1993) but inspired by Sinclair (1991) who elaborated the phrase in later work (Sinclair 2004). The technical term has also been used by Stubbs (1996), Tognini-Bonelli (2001) and Partington (2004), among others. It is a major part of lexical patterns that is crucial in collocation. It is the association of lexical items with a semantic field, interpretable on the basis of a connotation.

1. Introduction

1.2 Research Purpose and Significance

● Collocation has always been regarded as the most troublesome issue for second language learners to handle. Given the insurmountable difficulties which lie in second language learners' cognitive systems, improvements can only be made with respect to proper instruction, continuous efforts and positive exposure. So the collocational behavior of Chinese English learners and making contrastive analysis with native speakers will surely consolidate the language nature and support second language acquisition.

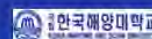
▫ This study puts its emphasis on the collocational pattern of GET-Adjective, probing into the semantic preference of the habitual combination, by comparing the usages of Chinese English majors with those of native speakers.



1. Introduction

1.3 Research Questions

1. What are the respective characteristics of the semantic preferences embodied in the use of GET-Adjective patterns by Chinese English majors and native speakers?
2. How do Chinese English majors use GET-Adjective patterns differently from the way native speakers employ them?



1. Introduction

1.4 Paper Layout

- Section 1 makes preparation for the entire research. Research significance and purpose are also presented in this part.
- Section 2 deals with the fundamental concepts and theories which will be referred to during the process of the study, including basic knowledge concerning language transfer and semantic preference.
- Section 3 is responsible for raising research questions in a logical way and lists some feasible methods. Corpora involved and tools applied are also mentioned in this section.
- Section 4 implements the measures and procedures mentioned before. Retrieved data and obtained results are here for discussion. Specifically, the semantic preferences of GET-Adjective manifested in both learners' and native speakers' language will be elaborated.
- Section 5 summarizes the findings and the discoveries that this study aims to dig and thus leaves behind some pedagogical implications for second language acquisition.

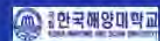
2. Literature Review



2.1 Semantic Prosody and Semantic Preference

Semantic preference, is defined by Stubbs (2001) as “the relation, not between individual words, but between a lemma or word-form and a set of semantically related words” (P. 65). He cites the item large which often co-occurs with words for quantities and sizes, such as numbers, scale, part, amounts, quantities. He also explains that an item shows semantic preference when it co-occurs with “a class of words which share some semantic feature (such as words to do with medicine or change)” (ibid: P. 88).

2. Literature Review



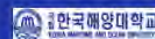
2.2 Language Transfer

Odlin (1989) offers the definition of transfer as a basis for his own thoughtful treatment of such phenomena:

Transfer is the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired (P. 27).

Transfer regarded in this way is far deviated from the original use of the term in behaviorist theories of language learning. This definition, although somewhat vague, provides an adequate basis for the material to be considered in this study.

3. Data and Research Methodology

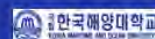


3.1 Corpora Employed in This Study

- **CEM**

Chinese English Majors' Corpus (CEM) is a learners' corpus containing 1.77 million words. The corpus consists of five major genres: routine translation exercises by English majors from 2004 to 2006, routine composition exercises, essays composed at Test for English Majors (Band 4) from 2004 to 2006, translations finished at Test for English Majors (Band 8) from 2004 to 2006 and essays composed at Test for English Majors (Band 8). CEM is the learner corpus for this study.

3. Data and Research Methodology



- **BROWN**

The Brown University Standard Corpus of Present-Day American English (or just Brown Corpus) was compiled in the 1960. It contains 500 samples of English-language text, totaling roughly one million words, compiled from works published in the United States in 1961. Each sample began at a random sentence-boundary in the article or other unit chosen, and continued up to the first sentence boundary after 2,000 words. In a very few cases miscounts led to samples being just under 2,000 words.

- The original data entry was done on upper-case only [keypunch](#) machines; capitals were indicated by a preceding asterisk, and various special items had special codes.

- BROWN plays the role of comparable corpus in this research.

3. Data and Research Methodology



3.2 AntConc

AntConc is a [UNICODE](#) compliant freeware [concordance](#) program for [Windows](#), [Mac OS X](#), and [Linux](#) systems developed by Laurence Anthony of [Waseda University](#), Japan. AntConc can generate [KWIC](#) concordance lines and concordance distribution plots. It also has tools to analyze word clusters (lexical bundles), [n-grams](#), collocates, word frequencies, and keywords. The functions of concordance lines and collocates are mainly involved in this study.

3. Data and Research Methodology



3.3 *I* Value

- *I* value is short for Mutual Information value, which depend on two calculations. One is the Observed, the number of instances of the co-occurring word found in the designated span of the node word. The other is the Expected, the number of the instances expected in that span, given the frequency of the co-occurring word in the corpus as a whole.
- *I* value is the Observed divided by the Expected, converted to a base-2 logarithm. *I* value measures the amount of non-randomness present when two words co-occur. It shows how strongly two words seem to associate in a corpus, based on the independent relative frequency of the two words. An *I* value of 3 or higher can be taken to be significant.

3. Data and Research Methodology



3.4 Procedures for Data Collection

- The first step: extract the lemma GET in both CEM and BROWN with the help of concordance line function of Antconc to grasp a universal feature of GET. Overall frequencies of GET were presented across the two corpora.
- By using the collocate function of Antconc, adjective collocates with highest *I* value among both corpora were retrieved. After the thorough analyses into the semantic preferences of these adjectives, general characteristics between the usage of GET-Adjectives by Chinese English majors and by native speakers could be summarized.

3. Data and Research Methodology



3.4 Procedures for Data Collection (Cont')

- Neutral GET-adjective patterns with high frequencies and their *I* value higher than 3 in CEM and BROWN were picked out for further research.
- Larger contexts of the chosen neutral patterns were drawn, and the author exerted a closer inspection into them. Then, the semantic preferences of the whole context, rather than the adjectives collocating with GET, could be perceived.

4. Results and Discussion



4.1 Overall Characteristics of GET-Adjective Patterns Distributed Across CEM and BROWN

In order to catch a glimpse of the overall features of GET-Adjective patterns, the holistic raw frequencies of them were computed and analyzed. The results are shown respectively in CEM and BROWN.

4.1.1 The Characteristics of GET-Adjective Patterns in CEM

Table 4.1 Raw and Normalized (per 100, 000 words) Frequencies of GET in CEM

Word	Raw	Normalized
GET	3587	2019

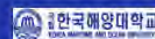
4. Results and Discussion



4.1 Overall Characteristics of GET-Adjective Patterns Distributed Across CEM and BROWN

GET	Frequency	I Value	Adjectives
1st	1st	9.28192	Infected
2nd	3rd	9.28192	Involved
3rd	1st	9.28192	Improved
4th	1st	9.28192	Assaulted
5th	46th	9.22052	Rid
6th	1st	8.28192	Unfavourable
7th	1st	8.28192	Slender
8th	1st	8.28192	Marrying
9th	1st	8.28192	Killed
10th	1st	8.28192	Disastrous
11th	1st	8.28192	Cancelled
12th	4th	8.28192	Acquainted
13th	98th	7.84135	Prepared
14th	1st	7.69696	Unwanted
15th	1st	7.69696	Contented
16th	1st	7.69696	Colder
17th	6th	7.54495	Marned
18th	1st	7.28192	Sleepless
19th	1st	6.95999	Stricken
20th	2nd	6.82249	Fired
21th	1st	6.69696	Dead
22th	2nd	6.69696	Accustomed
23th	2nd	6.58148	Unsatisfactory
24th	17th	6.51373	Ready
25th	1st	6.47456	Heavier
26th	1st	6.47456	Dreadful
27th	1st	6.28192	Milder
28th	3rd	6.16644	Electric
29th	11th	6.11199	Familiar
30th	1st	5.95999	Warmer
31th	1st	5.82249	Rejected
32th	1st	5.82249	Isolated
33th	2nd	5.82249	Exhausted
34th	1st	5.69696	Impatient
35th	1st	5.69696	Adapted
36th	2nd	5.58148	Enrolled
37th	1st	5.58148	Concrete
38th	3rd	5.54495	closer
39th	148th	5.54082	bad
40th	1st	5.47456	upper
41th	1st	5.47456	remote
42th	1st	5.37503	fruitful
43th	1st	5.37503	dropped
44th	1st	5.28192	closed
45th	1st	5.11199	unsatisfied
46th	8th	5.09209	angry
47th	4th	5.05310	ill
48th	1st	4.95999	cheated
49th	2nd	4.85565	endless
50th	3rd	4.84451	low
51th	1st	4.82249	splendid
52th	2nd	4.79007	hurt
53th	1st	4.75836	introduced
54th	2nd	4.72733	older
55th	3rd	4.69696	terrible
56th	4th	4.69696	improved
57th	39th	4.67099	high
58th	2nd	4.66721	confused
59th	1st	4.63806	hired
60th	2nd	4.58148	greater
61th	1st	4.47456	overall
62th	1st	4.42394	detailed
63th	5th	4.41403	nervous
64th	3rd	4.40745	involved
65th	1st	4.37503	puzzled
66th	1st	4.37503	immediate
67th	1st	4.28192	anxious
68th	2nd	4.19446	extra
69th	1st	4.19446	caught
70th	1st	4.19446	balanced
71th	1st	4.15264	charming
72th	1st	4.11199	inter-viewed
73th	8th	4.11199	higher
74th	1st	4.11199	countless
75th	1st	4.07247	frustrated
76th	5th	4.04160	further
77th	1st	3.92437	solid
78th	2nd	3.92437	excited
79th	1st	3.88960	upset
80th	5th	3.88960	lost
81th	15th	3.82687	used
82th	1st	3.79007	unlucky
83th	2nd	3.79007	tired

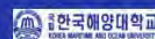
4. Results and Discussion



4.1 Overall Characteristics of GET-Adjective Patterns Distributed Across CEM and BROWN

- It can be seen from the above table that adjectives with neutral meanings make up the majority of the entire list. Among 83 adjective collocates in CEM, 53.125% of occurrences are neutral words, 29.6875% possessing explicit negative meanings, and only 17.1875% positive. In the light of statistics, Chinese English majors intend to employ neutral adjectives to collocate with GET and co-occurring with positive adjectives is a rare phenomenon.
- A more subtle and specific observation was made to carry on further investigations. Of all the 83 collocates listed above, *bad*, *more*, *prepared*, *rid* and *high* are five most frequently occurred adjectives. Only *bad* shows obvious negative meanings, while the other four collocates tend to be neutral.

4. Results and Discussion



4.2 Comparison of Top 25 High Frequency GET-Adjective Patterns in CEM and BROWN

Table 4.7 Top 25 Collocates in CEM and BROWN¹⁰

CEM ¹¹	Frequency ¹²	<i>t</i> Value ¹³	BROWN ¹⁴	Frequency ¹⁵	<i>t</i> Value ¹⁶
bad ¹⁷	148 ¹⁸	5.54082 ¹⁹	rid ²⁰	10 ²¹	7.01888 ²²
more ²³	108 ²⁴	3.18838 ²⁵	ready ²⁶	7 ²⁷	4.0987 ²⁸
prepared ²⁹	98 ³⁰	7.84135 ³¹	married ³²	6 ³³	7.54495 ³⁴
rid ³⁵	46 ³⁶	9.22052 ³⁷	tired ³⁸	5 ³⁹	5.11199 ⁴⁰
high ⁴¹	39 ⁴²	4.67099 ⁴³	drunk ⁴⁴	4 ⁴⁵	9.28192 ⁴⁶
ready ⁴⁷	17 ⁴⁸	6.51373 ⁴⁹	started ⁵⁰	4 ⁵¹	3.52703 ⁵²
used ⁵³	15 ⁵⁴	3.82687 ⁵⁵	caught ⁵⁶	4 ⁵⁷	5.77942 ⁵⁸
better ⁵⁹	12 ⁶⁰	3.07084 ⁶¹	involved ⁶²	4 ⁶³	4.40745 ⁶⁴
familiar ⁶⁵	11 ⁶⁶	6.11199 ⁶⁷	worse ⁶⁸	3 ⁶⁹	3.32772 ⁷⁰
angry ⁷¹	8 ⁷²	5.09209 ⁷³	wet ⁷⁴	2 ⁷⁵	6.95999 ⁷⁶
higher ⁷⁷	8 ⁷⁸	4.11199 ⁷⁹	fired ⁸⁰	2 ⁸¹	6.82249 ⁸²
married ⁸³	6 ⁸⁴	7.54495 ⁸⁵	hit ⁸⁶	2 ⁸⁷	5.32772 ⁸⁸
nervous ⁸⁹	5 ⁹⁰	4.41402 ⁹¹	bigger ⁹²	2 ⁹³	4.75836 ⁹⁴
further ⁹⁵	5 ⁹⁶	4.0416 ⁹⁷	busy ⁹⁸	2 ⁹⁹	3.31613 ¹⁰⁰
lost ¹⁰¹	5 ¹⁰²	3.8896 ¹⁰³	interested ¹⁰⁴	2 ¹⁰⁵	3.01513 ¹⁰⁶
acquainted ¹⁰⁷	4 ¹⁰⁸	8.28192 ¹⁰⁹	lost ¹¹⁰	2 ¹¹¹	3.56767 ¹¹²
ill ¹¹³	4 ¹¹⁴	5.0531 ¹¹⁵	overexcited ¹¹⁶	1 ¹¹⁷	9.28192 ¹¹⁸
improved ¹¹⁹	4 ¹²⁰	4.69696 ¹²¹	mad ¹²²	1 ¹²³	7.69696 ¹²⁴
poor ¹²⁵	4 ¹²⁶	3.1578 ¹²⁷	downtown ¹²⁸	1 ¹²⁹	7.28192 ¹³⁰
electric ¹³¹	3 ¹³²	6.16644 ¹³³	harsher ¹³⁴	1 ¹³⁵	6.95999 ¹³⁶
closer ¹³⁷	3 ¹³⁸	5.54495 ¹³⁹	stuck ¹⁴⁰	1 ¹⁴¹	6.11199 ¹⁴²
low ¹⁴³	3 ¹⁴⁴	4.84451 ¹⁴⁵	impatient ¹⁴⁶	1 ¹⁴⁷	5.69696 ¹⁴⁸
terrible ¹⁴⁹	3 ¹⁵⁰	4.69696 ¹⁵¹	shorter ¹⁵²	1 ¹⁵³	5.37503 ¹⁵⁴
involved ¹⁵⁵	3 ¹⁵⁶	4.40745 ¹⁵⁷	sick ¹⁵⁸	1 ¹⁵⁹	4.95999 ¹⁶⁰
practical ¹⁶¹	3 ¹⁶²	3.11199 ¹⁶³	angry ¹⁶⁴	1 ¹⁶⁵	3.09209 ¹⁶⁶

4. Results and Discussion



4.2 Comparison of Top 25 High Frequency GET-Adjective Patterns in CEM and BROWN

- There are six words appearing both in CEM and BROWN. Out of those words, two lexical terms *lost* and *angry* show negative meaning to certain extent while the other four *rid*, *married*, *involved* and *ready* are neutral.
- Two negative-inclined terms of the six overlapping words are out of the concern. Of the four adjectives with neutral sense, *rid* is always employed in the regular idiomatic pattern GET *rid of*. The study rules it out. Besides, GET *married* presents undisputed neutral sense and is required to be excluded out of this research, too.
- As a result, two neutral adjectives collocating with GET are left for investigation. They are *involved* and *ready*.

4. Results and Discussion



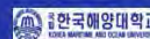
4.3 Semantic Preference of the Neutral Patterns GET-Involved and GET-Ready

4.3.1 Semantic Preferences of GET-Involved Manifested in CEM

- | | | |
|--|-----------------|--------------------------------------|
| 1 [↔]] from each class, then most of the students will | get | in the preparing work and create |
| | involved | [fm, -] a great a |
| 2 [↔] way. In the past, there were only a few students | got | in the arts festival. [su, -] |
| | involved | performing [fm, -] |
| 3 [↔] or me. While a lot of work should be done when we | get | in job-hunting, including setting up |
| | involved | a career goa |

By examining the above three instances in a broader context, connotational meanings are understood better. Prevailingly, GET-involved expresses neutral meaning. Line 1 and 2 are from the articles under the same topic PHONES WON'T KILL LETTER WRITING.

4. Results and Discussion

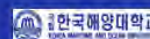


4.3 Semantic Preference of the Neutral Patterns GET-Involved and GET-Ready

4.3.2 Semantic Preferences of GET-Involved Manifested in BROWN

1 [↖] Rector gave him his final instructions. He was to ↖	get	in no arguments; he was to try to
	involved ↖	make no convert [↖]
2 [↖] how to be sponsored by <u>Qatmut</u> Grits. But if Joyce ↖	got	in murder or salacious scandal, the
	involved ↖	role would pr [↖]
3 [↖] hes, and abundance of names. "You are bound to ↖	get	with people when you have
	involved ↖	children", Fran had to [↖]
4 [↖] , "so it is good to know that those with whom you ↖	get	are not just dreary little housewives
	involved ↖	and dull hu [↖]

4. Results and Discussion



4.3 Semantic Preference of the Neutral Patterns GET-Involved and GET-Ready

4.3.2 Semantic Preferences of GET-Involved Manifested in BROWN

Line 3 and 4 are put together because they are from the same paragraph and share the identical context. These two sentences show unambiguously neutral meaning. However, it is noteworthy that GET-involved in Line 1 and 2 can be perceived in an obvious negative sense. This is different from the aforementioned three samples from CEM. By contrastively analyzing GET-involved in two corpora, it can be concluded that Chinese English majors seem to highly overuse this pattern in the neutral sense. As is shown from the above results in BROWN, native speakers employ GET-involved to express the negative meaning of "being trapped in an awkward situation" more frequently (67% vs. 33%).

4. Results and Discussion



4.3 Semantic Preference of the Neutral Patterns GET-Involved and GET-Ready

4.3.3 Semantic Preferences of GET-Ready Manifested in CEM

1 st umbrella [ʌn, -] for rainy day [ru, -] We should - ²	get ready - ³	for tomorrow and behave a good habit [ec3, 3:] A ²
2 nd only help you to earn money but also help you to - ³	get ready - ⁴	for your future career. Nowadays what society nee ²
3 rd ry opportunities [ɒp, -] just belong to those who - ³	get ready - ⁵	for themselves. [mɪ, -] The competition is so fr ²
4 th person they want. And of course all of us should - ³	get ready - ⁶	for the interview when we are hunting for a job] ²
5 th p, -] funny [ʃn, -] leader. Why don't you [ɒp, -] - ²	get ready - ⁷	to grasp this chance to show you [ɒp, -] off? Alt ²
6 th career in a [ag, -] ideal company. So, let's say, - ²	get ready - ⁸	for a head-on challenge! This world never lack;] ²
7 th proper activities and so on. All that is doing is - ²	getting ready - ⁹	for our detailed aim [cc, -], [m7, -] [m0, -] A ²
8 th hits we had sung in the contest. We had no sooner - ²	got ready - ¹⁰	when the secretary of our headmaster came in and - ²
9 th well. What's more, he is warm-hearted and always - ²	gets ready - ¹¹	to help others. It's his fine qualities that inf ²
10 th e] "Tomorrow is another day." What I should do is - ²	getting ready - ¹²	for tomorrow. An Unforgettable Day Almost every ²
11 th I got up one hour earlier than usual in order to - ²	get ready - ¹³	for my luggage. I did it all by myself, for my ps ²
12 th left, please check your answer [ap, -] again, and - ²	get ready - ¹⁴	to hand in your paper [ap, -] the short-laired - ²
13 th she one nervous and ashamed. So Be Prepared and - ²	get ready - ¹⁵	for the potential opportunities, and Be Honest to ²
14 th g my plan, my idea and my products to public, and - ²	get ready - ¹⁶	for the acid test. So there is my point: trying ²
15 th faced to the reality with a shining smile. He has - ²	got ready - ¹⁷	to start a new journey in the test of life and to ²
16 th roughly. It seems that no matter how much I do to - ²	get ready, - ¹⁸	I still feel a bit overwhelmed. Actually, I feel ²
17 th g to the Mid-Autumn Festival. Therefore, I had to - ²	get ready - ¹⁹	some riddles which were used in the shooting game ²

4. Results and Discussion

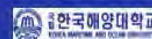


4.3 Semantic Preference of the Neutral Patterns GET-Involved and GET-Ready

4.3.3 Semantic Preferences of GET-Ready Manifested in CEM

A close observation of the above 17 samples demonstrates that GET-ready is mainly employed by Chinese English majors to express the meaning of “get sb./sth. prepared for upcoming positive matters”. Specifically speaking, 10 instances (i.e. lines 1, 2, 3, 4, 5, 6, 9, 10, 13, 15) out of 17 are perceived positive tendency. The positive preference is eminently shown in the following instances. The serial numbers before the following instances correspond to the numerical orders of the concordance lines above.

4. Results and Discussion



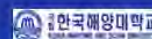
4.3 Semantic Preference of the Neutral Patterns GET-Involved and GET-Ready

4.3.4 Semantic Preferences of GET-Ready Manifested in BROWN

1 ^{c2} he ground, watched them sprout, bud, blossom, and ^{c2}	get ready ^{c2}	to bear. She was ready to kill the beef, dress it ^{c2}
2 ^{c2} the pool where they can amuse each other until we ^{c2}	get ready ^{c2}	to merge sides. All dressing (undressing to be ^{c2}
3 ^{c2} cross the Bay. The men fortified a blockhouse and ^{c2}	got ready ^{c2}	to fight, but meanwhile appealed to the King and ^{c2}
4 ^{c2} pleasantly. Then they returned to their hotel and ^{c2}	got ready ^{c2}	for bed. It was late, and Blackman was ready to g ^{c2}
5 ^{c2} vening Eileen had been as demure as a little girl ^{c2}	getting ready	for her first communion. It just about blew us h ^{c2}
6 ^{c2} dropped the paper and stood up. "I guess I better ^{c2}	get ready ^{c2}	to go". Winston watched her fumbling to untie h ^{c2}
7 ^{c2} I'll saddle the horses and bring them round. You ^{c2}	get ready ". ^{c2}	He burst from the hot confinement of the rog ^{c2}

After scrupulous inspection and examination, all of the seven concordance lines are set in neutral environment. Surprisingly, no positive or negative sense of semantic preferences can be discerned in GET-ready pattern of BROWN. In retrospect, it seems evident that Chinese English majors overuse this phrase in its positive preference.

5. Conclusion



The most apparent result this study has revealed is that great differences exist in the semantic preference of GET-Adjective utilized by Chinese English majors and native speakers. Specific points of this thesis's main findings can be concluded below.

Chinese English majors are inclined to overuse GET compared to native speakers.

23.43% of the holistic usage of GET in CEM is GET-Adjective pattern. However, the same pattern occurred in BROWN accounts for only 8.85% of all the GET-occurrences. These figures reveal that GET-Adjective patterns are also overused by Chinese English majors.

5. Conclusion

Chinese English majors intend to employ neutral adjectives to collocate with GET and co-occurring with positive adjectives is a rare phenomenon. Similarly, neutral adjectives are predominantly used by native speakers to collocate with GET. Still, adjectives with positive senses do not generally occur with GET.

Chinese English majors are prone to overuse the positive connotation (17.19 vs. 5.08) of those adjectives co-occurring with GET. There is no remarkable difference in the percentage of semantic preference with negative sense across two corpora (29.69 vs. 34.34). In terms of neutral sense, Chinese English majors slightly underuse the pattern vis-à-vis native speakers. Most of the adjectives concerning with GET pattern are neutral.

References

- Carter, R. & McCarthy, M. (1999). The English GET-passive in Spoken Discourse: Description and Implications for an Interpersonal Grammar. *English Language and Linguistics*, 3, 41-58.
- Corder, S. P. (1967). The Significance of Learners' Errors. *International Review of Applied Linguistics*, 5, 9-161.
- Corder, S. P. (1974). Error Analysis. In Allen & Corder (eds.), *The Edinburgh Course in Applied Linguistics*. London: Oxford University Press.
- Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (eds.), *Syntax and Semantics*. New York: Academic Press.
- Halliday, M. & Matthiessen, C. (2004). *An Introduction to Functional Grammar*. London: Arnold.
- Harley, T. A. (2001). *The Psychology of Language*. Hove, UK: Psychology Press.
- Skinner, B. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.

온라인 영화 댓글의 특성: 영화 <기생충>의 영어 댓글과 한국어 댓글 비교

김보라 · 박채빈 · 백다운

(부산대학교)

Bora Kim, Chaebin Bak, and Dayoon Baek (2021), Characteristics of Internet Movie Comments: Comparisons between English and Korean Comments on Movie, *Parasite*. As the digital world has been rapidly expanded by human interaction, digital discourses, which are language patterns occurring in digitally mediated interactions, can be useful for language learning and teaching. By learning the patterns of digital discourses, L2 learners engage with extended communities beyond traditional academic domains, explore critical cultural and political practices, and examine cultural and interpersonal phenomena (Skyles, 2019). Using data gathered from online movie review sites to identify the digital discourse patterns of movie comments, this paper explores the characteristics of English and Korean comments on the movie *Parasite*, comparing the two in terms of each characteristic. The results show significant disparities between English and Korean comments due to linguistic and cultural differences. This suggests that L2 learners engage with the target culture when building L2 competence and that exploring the digital discourse patterns of movie comments can be considered valuable in the language classroom. (Pusan National University)

Key Words: digital discourse, digital literacy, human interaction, movie review comments, language, culture

1. 서론

IT 기술의 발달과 인터넷 사용의 대중화로 인하여, 오늘날 현대인들은 개인 모바일 기기를 통하여 언제 어디서든 인터넷에 쉽게 접속할 수 있다. 따라서, 현대인들은 디지털 세계에 살고 있다고 해도 과언이 아니다. 디지털 세계에서는 디지털 매체를 이용하여 원하는 정보를 얻을 수 있는데, 특히 인터넷상에서 특정 정보에 대해 댓글을 작성함으로써 다른 사람들과 서로의 생각과 의견을 공유할 수 있다. 디지털 세계의 학습자들은 디지털 매체를 통해 전 세계사람들과 세계 공용어인 영어로 의사소통 할 수 있는 기회가 많고, 온라인상의 교류가 활발해짐에 따라 온라인상에서의 영어 사용이 일상화되었다(김양은, 2009; 김윤순, 임희정, 2018). 디지털 매체를 통한 의사소통의 증가는 글을 쓰는 공간이 종이에서 온라인으로 변화하였고, 문자 중심의 의사소통 능력이 기존의 음성 중심의 의사소통만큼 중요해졌음을 의미하고, 시대의 흐름에 맞는 영어 글쓰기 교육의 필요성을 시사한다.

하지만, 이미 여러 연구에서 지적된 바 있듯이, 현재 우리나라 영어 교육과정의 영어 글쓰기 교육과 쓰기 활동은 미흡한 실정이며, 이는 지금의 영어 글쓰기 교육이 시대의 흐름을 반영하고 있지 못하고 있음을 암시한다고 볼 수 있다. 영어학습자들에게 영어 글쓰기 기회는 흔치 않은 편이며, 영어학습자들은 제한적인 문법 지식과 어휘, 모국어의 간섭, 논리적 전개 능력의 부족으로 인하여 영어 글쓰기에 부담을 느낄 수 있는데(Fu-Lan, 2006; Hydayati, 2018), 이러한 점은 영어 글쓰기 활동이 저조한 이유가 될 수 있다. 이는 또한 시대의 흐름을 반영한 영어 글쓰기 교육이 미미함을 암시하기도 한다. 효과적인 글쓰기와

1) 디지털 매체는 디지털 코드를 기반으로 동작하는 전자 매체를 일컫으며, 디지털 매체로는 휴대전화, 콤팩트 디스크, 블루레이 디스크, 디지털 영상, 디지털 텔레비전, 전자책, 인터넷, 미니 디스크, 비디오 게임 등이 있다(위키백과, 2020).

관련하여, 김윤순, 임희정(2018)은 온라인을 통한 상호작용은 영어학습자의 흥미를 높이고, 의미 있는 글쓰기 활동을 기대할 수 있을 거라고 언급하며, Wong Fillmore and Snow(2000)은 실생활과 밀접한 관계가 있는 글쓰기 활동의 기회를 더 많이 제공하는 것을 권장한다.

디지털 매체를 통한 상호작용에서 나타나는 언어 패턴을 디지털 담론(digital discourse)이라 하는데(Sykes, 2019), 이는 모든 언어 학습자의 언어 레퍼토리에 필수적이다(Thorne et al., 2009, 2015). Sykes(2019)는 온라인상의 해시태그와 디지털 게임을 통해 형성되는 디지털 담론을 분석한다. 그의 연구는 해시태그와 디지털 게임을 통해 목표 문화를 배울 수 있고, 목표 문화를 배우는 것이 언어 능력 향상에 도움이 되므로, 외국어 학습자에게 디지털 담론이 유용하다는 점을 강조한다. 본 연구는 인터넷을 통해 전 세계사람들과 영어로 소통할 수 있는 환경이 갖춰진 디지털 세계에서 영어학습자가 디지털 담론을 통해 영어권 문화를 배울 수 있고, 나아가 영어학습자의 글쓰기 실력 향상에도 영향을 미칠 수 있으므로, 디지털 담론이 영어학습자의 영어 레퍼토리에 필수적이라고 본다. 특히, 이전 교실 안에서의 교육 틀에서 벗어나, 디지털 매체를 활용하여 실제 온라인 상황에서 영어 글쓰기를 시도하는 것이 접근성이 높아 다양한 측면에서의 문화를 많이 접할 수 있고, 실용적인 교육 효과를 기대할 수 있다고 본다. 본 연구는 영화 <기생충>에 대한 온라인 영화 댓글에서 나타나는 영어 댓글과 한국어 댓글의 특징을 분석한다. 언어와 문화는 분리할 수 없는 밀접한 관계이므로(Brooks, 1964: 85; Brown, 2000: 177), 같은 영화라 할지라도 언어적 특성과 문화에 따라 댓글의 특징적 요소들의 차이를 살펴보고자 한다. 그리고, 이러한 차이는 언어학습에 있어 문화 학습의 필요성, 빠르게 변화하는 시대에 맞는 실용적인 교육의 필요성을 시사할 것이다.

2. 이론적 배경

2.1. 디지털 담론

전 세계적으로 인터넷 보급률 증가와 IT 기술의 발전은 시간적, 공간적 개념을 넘어 기존의 면대면 형태의 의사소통이 아닌 새로운 형태의 상호작용이 생겨났다. 온라인상의 교류가 활발해지면서 온라인상의 의사소통을 위한 디지털 리터러시 능력이 부각되기 시작하였고, 이어서 디지털 담론이 등장하게 되었다(Sykes, 2019). 즉, 기술의 발달에 따라 의사소통의 형태에 변화가 생겼으므로, 원활한 의사소통을 위해 언어 교육에도 변화가 필요하다는 것을 의미한다. 실제로, 학습자의 연령과 관계없이 SNS(social network service)를 활용한 글쓰기 학습의 긍정적인 효과를 보여준 연구가 존재한다. 허영미(2012)의 연구는 SNS를 활용한 영어 쓰기 활동이 영어 쓰기 능력 향상에 효과적이라는 것을 보여줌으로써, 실제 교육 현장에서 SNS가 새로운 교육 도구로 사용될 수 있다고 언급한다. 특히, SNS를 통한 학습자 간의 상호작용은 자연스럽게 협동 학습을 하는 계기가 되며, PC와 웹, 스마트폰 등을 활용할 수 있으므로 학습의 접근성 또한 높일 수 있다는 장점이 있다. 김경옥, 박선호(2015)도 의사소통 환경의 변화에 따른 쓰기의 필요성이 점점 증가하므로, 초등교육에서도 효율적인 글쓰기 학습 방안에 대한 모색이 필요하다고 여기고, 초등학교 6학년생을 대상으로 연구한 바 있다. 이들의 연구는 SNS를 활용한 과정 중심 쓰기 학습이 효과적임을 보여주는데, 특히 온라인상에서 자신의 의견을 언어와 함께 소리, 사진, 동영상에 삽입해 표현하는 소통의 기회를 제공함으로써 글쓰기에 대한 흥미를 유발하는 방법이 수업의 효과를 극대화하는 하나의 방안이 될 수 있다고 한다.

글쓰기 과정이 적절한 어법과 어휘 사용에 대한 부담감이 있지만, 학습자의 흥미를 고려한 글쓰기 교육은 학습에 효과적이라는 점은 이전 연구에서도 많이 제시된 바 있다. Ismail(2011)은 영어 쓰기 향상을 위한 가장 중요한 두 요소로 많은 연습 기회와 학습자의 흥미를 언급한다. Gilakjani(2012)도 여기에 동의하며, 학습자가 학습 자료에 흥미를 느끼지 않으면 학습 동기가 매우 낮다고 한 바 있다. 즉, 흥미는 학습 동기를 높여주는 가장 중요한 요인으로, 학습자의 집중도와 의욕을 높이고 학습 성과에도 긍정적인 영향을 기대할 수 있게 한다. 따라서, 효과적인 영어 쓰기 능력을 기르기 위해 학습자의 학습 동기를 끌어올릴 수 있는 교수 매체를 선정하는 것이 중요한 요소가 된다. 안병섭(2015)의 연구는 디지털 시대의 문화

를 통해 본 매체 언어 교육의 방향에 대해 다룬다. 그에 따르면, 매체와 매체 언어는 현시대의 문화를 구성하는 매우 중요한 요소이므로, 현행 매체 언어 교육의 요소를 살펴보고, 향후 교육과정에서 이를 반영하고 나아가야 할 방향을 설정해야 한다고 언급한다. Sykes(2019)는 해시태그와 디지털 게임을 통한 디지털 담론을 분석하는데, 디지털 담론이 L2 언어의 능숙도를 향상시키므로, L2 문화와 연계되는 교육 도구로 활용하기에 매우 유용하다고 설명한다. 특히, 해시태그는 마케팅과 홍보, 대인 상호작용, 텍스트 구성이라는 3가지 담화 기능을 가지며, 각각의 담화 기능은 다양한 언어학습 상황에서의 언어 분석과 다양한 관점을 고려한 문화간 의사소통 능력을 가르치기에 효과적이라고 한다. 이는 언어와 문화가 매우 밀접하다는 점과 디지털 매체를 통해 문화를 포함한 언어 교육이 모두 가능함을 시사한다.

글쓰기 학습에 효과적인 디지털 매체의 상호작용 장치로는 댓글이 있다. 댓글은 인터넷과 함께 등장하였는데(권상희, 김익현, 2008), 작성자의 참여가 쉬우며 현재 가장 활성화되어 있으므로(김은미, 선유화, 2006), 연구 가치가 있다고 볼 수 있다. 김병선(2020)의 연구는 유튜브 댓글작성이 대학생들의 영어 쓰기 향상에 미치는 효과를 보여준다. 그의 연구는 대학생들이 영어로 제작되거나 영어 자막 처리된 유튜브 동영상에 시청하고, 시청 소감을 영어 댓글로 작성하게 한다. 유튜브는 디지털 네이티브 세대(Prensky, 2001)인 대학생들이 선호하는 매체이므로, 학습자 친화적 매체로 적극적인 학습에 도움이 되었다. 특히, 학습자 주도의 학습 상황에서 자연스러운 오류 수정 작업이 진행되었고, 타인의 댓글도 읽을 수 있는 점은 모방을 통한 영어 쓰기 능력에 도움을 주었다.

본 연구는 언어와 문화를 집약적으로 반영하며, 실질적인 글쓰기 연습이 가능하고, 디지털 세대의 흥미를 자극할 수 있는 댓글을 분석함으로써 댓글을 통한 디지털 담론의 특성을 알아보고자 한다. 특히, 문화적인 요소가 명백히 드러나는 영화라는 소재를 활용하여 영어 댓글과 한국어 댓글의 특성들을 비교·분석하여, 언어와 문화적 차이에 따라 나타나는 댓글들의 특성들을 다양한 관점에서 살펴보고자 한다.

2.2. 온라인 댓글의 특성

현대 사회에서는 인터넷 검색을 통해 원하는 정보를 일방적으로 얻을 수 있을 뿐만 아니라, 댓글을 작성함으로써 해당 정보에 대해 자신의 평가나 의견을 자유롭게 표현하거나, 이를 통해 추가적인 정보를 제공할 수도 있으며, 다른 사람과 해당 정보에 대해 상호작용 할 수 있다. 이처럼 댓글의 긍정적인 역할이 많은 만큼 댓글의 중요도 또한 높다고 볼 수 있다. 온라인 글에서 나타나는 특성 요소에는 길이, 태도, 공감, 관련성, 정보성, 신뢰성, 중요성, 흥미성, 객관성, 가독성, 유용성, 악성 댓글 등이 있다(Lang, Potter and Grabe, 2003; Wasko and Faraj, 2005; 김소담, 양성병, 2015; 김은미, 선유화, 2006; 김혜미, 이준웅, 2011; 박창호, 조수선, 2013; 안태형, 2013; 여은호, 박경우, 2011; 이형주, 이호성, 양성병, 2013; 진혜선, 신형덕, 2013). 댓글의 다양한 특성 요소 중에서, 김소담, 양성병(2015)은 온라인 뉴스 사이트에서의 일반댓글과 소셜댓글에 대하여 댓글의 길이, 댓글의 태도, 댓글의 공감도, 댓글의 품질(유용성과 가독성), 악성 댓글, 기사 관련성 등의 측면에서 비교 분석한다. 본 연구는 영화 <기생충>에 대한 영어 댓글과 한국어 댓글에서 나타나는 특성에 대해 댓글의 가독성과 유용성, 악성 댓글, 영화 관련성 등의 측면을 중점으로 비교 분석한다. 나머지 요소들은 본 연구의 분석 취지와 맞지 않거나, 중복되는 요소로 판단되었으므로 제외되었다. 특히, 댓글의 길이와 관련하여, 영어 댓글 분석을 위해 이용한 해외 사이트는 분석을 위해 선정된 영화에 대해 관심도가 높고 영화에 대한 지식이 있는 사람들이 주로 댓글을 작성하므로, 댓글의 길이가 길고 태도 또한 정성스러운 양상을 보였다. 댓글의 공감도와 관련하여, 해외 사이트는 댓글의 공감을 나타내는 기능이 없었으므로 본 연구의 분석대상에서 제외되었다.

2.2.1. 댓글의 유용성과 가독성

댓글의 유용성(helpfulness)과 가독성(readability)은 댓글의 내용이 정보성, 신뢰성, 이해 가능성, 가독성, 흥미성을 가지는 정도를 의미하므로, 댓글의 효과에 영향을 미치는 중요한 요소이다. 댓글은 내용적인 관점에서 유용성, 표현적 관점에서 가독성으로 구분된다(김소담, 양성병, 2015). Lang, Potter and Grabe(2003), Wasko and Faraj(2005), 김은미, 선유화(2006)에 의하면, 댓글의 유용성이 높다는 것은 분명

한 출처나 사실을 기반으로 작성되어 근거가 있는 내용을 포함하고, 댓글을 이해하기에 도움이 되는 정보와 흥미적인 요소가 있고, 정보 자체의 중요성이 높은 것을 의미한다. 댓글의 가독성이 높다는 것은 은어나 이모티콘, 줄임말, 신조어 등이 거의 없어서 읽기 쉽고, 맞춤법 오류와 오타가 거의 없어 내용의 이해가 쉬운 것을 의미한다. 댓글의 유용성과 가독성이 높다는 것은 댓글의 품질이 높다는 것을 의미한다. 이 경우의 댓글 수용자는 풍부한 정보를 얻을 뿐만 아니라, 댓글의 영향으로 자신의 태도나 판단을 변화시킬 수 있다(김은미, 선유화, 2006). 본 연구는 영화 관련 정보의 분명한 출처나 정보에 기반으로 근거가 있고, 영화 이해에 도움이 되고, 흥미적인 요소가 있으며, 중요한 정보를 담는지에 따라 유용성을 분석한다. 댓글 읽기에 불편함(은어, 이모티콘, 줄임말, 신조어)이 없는지, 내용 이해에 방해되는 요소가 없는지에 따라 가독성을 분석한다. 유용성이 낮은 댓글의 예로는 “진짜 꿀잼 스토리 탄탄 올해 들어서 개인적으로 1등 작품”, “재밌어서 두번 봤습니다.”, “남들이 명작이라니까 얼떨결에 뵈도모르고 박수치고 감.” 등이 있으며, 가독성이 낮은 댓글의 예로는 “진짜 꿀잼(은어) 스토리 탄탄 올해 들어서 개인적으로 1등 작품”, “냥(줄임말) 이견...말이 필요가 없음”, “영화 본지는 오래 되었는데 이제야 평점 남기내요!!(오타) 정말 새 장르가 탄생한 느낌...? 신선했습니다...!!!” 등이 있다.

2.2.2. 악성 댓글

악성 댓글은 인터넷상에서 상대방이 올린 글에 대한 비방이나 헐담을 하는 악의적인 댓글이며, 특정 대상에 대한 비방의 내용을 담고 있는 비방 댓글, 특정 대상을 저주하거나 폭력적인 언어로 상대를 위협하는 저주나 협박의 댓글, 노골적인 표현으로 성적 수치심을 불러일으키는 댓글, 광고, 유언비어와 같이 게시글과 무관한 내용의 댓글 등이 있다. 즉, 악성 댓글은 댓글이 타인에게 피해를 주는 정도를 의미하므로, 근거를 갖춘 부정적 평가와는 구별되어야 한다. 본 연구는 댓글이 누군가를 비방하거나, 저주하거나, 협박하거나, 또는 사회적 통념을 위배하는 내용을 의미하거나, 이러한 단어가 사용되었는지에 따라 악성 댓글을 분석한다(안태형, 2013). 또한, 특정 대상에 대한 비방의 내용을 담는 비방 댓글의 경우, 비방의 내용은 범위가 넓고, 비방의 정도가 다양하며, 비방성의 유무 판단이 개인에 따라 차이가 존재하기에 댓글을 읽는 이가 불편함을 느낄 수 있는 수준의 글을 비방글로 처리하였다. 악성 댓글의 예로는 “ocn에서 하는 거 다시 보는 중인데 정말 구역질나고 짜증남”, “나라망신. 진짜 기생충 같네...”, “반지하 냄새 행주 냄새 지하철 타면 나는 냄새...스멜스멜!!!” 등이 있다.

2.2.3. 영화 관련성

영화 관련성은 댓글에서 영화의 주요 키워드가 언급되고, 영화와 연관된 내용이 포함된 정도를 의미한다. 영화와 관련 없는 댓글은 댓글의 내용이 영화와 전혀 관련이 없거나 영화의 내용을 제대로 이해하지 못하여, 정상적인 의사소통이 이루어지지 않는 댓글에 해당한다. 이런 유형의 댓글은 자신이 알리고 싶은 내용을 영화의 내용과는 무관하게 무작위로 댓글을 작성하는 특징을 보인다. 본 연구는 영화 내용 또는 영화 관련 정보가 댓글과 하나의 통일적 텍스트 환경을 구성할 수 있다고 보고, 영화와 댓글에 나타난 이야기 구조가 일치하는지에 대해 분석한다. 영화 관련성이 높은 댓글은 영화의 주요 키워드가 다수 언급되고, 댓글의 내용 자체가 영화와 직접적으로 연관된 댓글에 해당하며, 영화 관련성이 낮은 댓글은 영화 관련 키워드가 전혀 언급되지 않고, 댓글의 내용 자체가 영화와 전혀 관련성이 없는 댓글에 해당한다. 영화 관련성이 낮은 댓글의 예로는 “대한민국에 기생해서 대한민국의 심장에 칼까지 꽂은 586 운동권 쓰레기들을 이렇게 멋지게 풍자한 영화가 있다니 감동하고 갑니다!!”, “인기 영화관계자들이 부자인가 그 외 사람들을 쓰레기 취급하는군”, “못사는 사람들은 선하고 그들이 저지른 살인을 미화하는 느낌이 들어 불편했다” 등이 있다.

본 연구는 댓글의 유용성과 가독성, 악성 댓글, 영화 관련성을 중점적으로 영화 <기생충>에 대한 한글 댓글과 영어 댓글을 분석하고자 한다. 언어적 차이와 문화적 차이에 따라 댓글에서는 어떠한 특징들이 보이는지 살펴보고자 하며, 두 댓글의 차이는 디지털 시대의 언어 교수법에도 변화가 필요함을 시사할 것 이라고 여겨진다.

3. 연구 방법

4. 연구 결과

5. 결론

참고문헌

- 권상희, 김익현(Kweon, S-H. and I-H. Kim). 2008. 온라인 댓글 인식과 댓글 활동의 관계에 관한 연구: 댓글의 신뢰도와 인터넷뉴스 수용자의 수용경향 중심으로(A Study of the Relationship between Perception and Activities in the News Replies: Focused on News Perception and Credibilities). 「한국언론정보학보」 (Korean Journal of Communication and Information), 13.1, 243-270.
- 김경옥, 박선호(Kim, K. O. and S. H. Park). 2015. 모바일 기반 SNS를 활용한 과정 중심 초등 영어 쓰기 학습의 효과(The Effects of Process-based Writing through the Use of Mobile SNS Software on Elementary English Education). 「멀티미디어 언어교육」 (Multimedia-Assisted Language Learning), 18.1, 11-32.
- 김병선(Kim, B. S.). 2020. 대학생들의 영어 쓰기 경험에 관한 연구: 유튜브 댓글작성을 중심으로(A Study on College Students' Experiences of Writing in English: Based on Leaving Comments on YouTube). 「미래영어영문학」 (The Mirae English Language and Literature Association), 25.3. 187-207.
- 김소담, 양성병(Kim, S-D. and S-B. Yang). 2015. 온라인 뉴스 사이트에서의 일반댓글과 소셜댓글의 비교분석 (A Comparative Analysis between General Comments and Social Comments on an Online News Site). 「한국콘텐츠학회」 (The Korea Contents Society), 15.4, 391-406.
- 김양은(Kim, Y.). 2009. *디지털 시대의 미디어 리터러시 (Media Literacy Skills in the Digital Era)*. 커뮤니케이션북스.
- 김윤순, 임희정(Kim, Y. S. and H-J. Ihm). 2018. 웹기반 만화 제작과 SNS 활동이 초등영어 과정중심 쓰기 학습에 미치는 영향(The Effects of Process-based Writing through Webtoon Making Activities on Elementary English Writing). 「멀티미디어 언어교육」 (Multimedia-Assisted Language Learning), 21.2, 77-103.
- 김은미, 선유화(Kim, E-M. and Y-H. Sun). 2006. 댓글에 대한 노출이 뉴스 수용에 미치는 효과(The Effect of Replies in Internet News on the Audience). 「한국언론학부」 (Korean Journal of Journalism & Communication Studies), 50.4, 33-64.
- 김혜미, 이준웅(Kim, H. M. and J. W. Rhee). 2011. 인터넷 뉴스와 댓글의 뉴스 프레임 융합 효과 연구 = 해석의 복잡성 및 태도의 극단성 분석을 중심으로(The Impacts of Frame Fusion of News Text and Comments on News Interpretation and Attitude). 「한국언론학부」 (Korean Journal of Journalism & Communication Studies), 55.2, 32-55.
- 안병섭(Ahn, B. S.). 2015. 디지털 시대의 문화 지체를 통해 본 매체 언어 교육의 방향(A Direction of Media Language Education Considering the Cultural Lag from Digital Media). 「우리어문연구」 (The Study

- of Korean Language and Literature), 53, 333-361.
- 허영미(Huh, Y. M.). 2012. SNS기반 환경에서의 영어 쓰기 활동이 초등학생의 영어 쓰기 능력 및 정서적 영역에 미치는 영향(The Effect of SNS(Social Network Service)-based English Writing Instruction on Korean Elementary School Students' English Writing Abilities and Affective Domains). 중앙대학교 교육대학원 석사학위논문(M.A. Thesis. Chung-Ang University).
- Brooks, N. 1964. *Language and Language Learning, Theory and Practice*. Harcourt, Brace & World.
- Brown, H. D. 2000. *Principles of language learning and teaching*, New York: Longman.
- Fu-Lan, L. 2006. Using Computer-mediated-communication to Overcome the Anxiety in ESL Writing. *US-China Foreign Language* 4(1), 97-101.
- Gilakjani, A. 2012. Visual, Auditory, Kinaesthetic Learning Styles and their Impacts on English Language Teaching. *Journal of Studies in Education* 2(1), 104-113.
- Hidayati, K. H. 2018. Teaching Writing to EFL Learners: An Investigation of Challenges Confronted by Indonesian Teachers. *Langkawi: Journal of The Association for Arabic and English* 4(1), 21-31.
- Ismail, S. 2011. Exploring Students' Perceptions of ESL Writing. *Language Teaching* 4(2), 73-83.
- Lang, A., D. Potter, and M. E. Grabe. 2003. Making News Memorable: Applying Theory to the Production of Local Television News. *Journal of Broadcasting & Electronic Media* 47(1), 113-123.
- Prensky, M. 2001. Digital Natives, Digital Immigrants Part 2: Do They really Think Differently?. *On the Horizon* 9(5), 1-6.
- Sykes, J. M. 2019. Emergent Digital Discourses: What Can We Learn From Hashtags and Digital Games to Expand Learners' Second Language Repertoire?. *Annual Review of Applied Linguistics* 39, 128-145.
- Thorne, S., R. Black, and J. Sykes. 2009. Second Language Use, Socialization, and Learning in Internet Interest Communities and Online Gaming. *The Modern Language Journal* 93, 802-821.
- Thorne, S., S. Sauro, and B. Smith. 2015. Technologies, Identities, and Expressive Activity. *Annual Review of Applied Linguistics* 35, 215.
- Wasko, M. M., and S. Faraj. 2005. Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS quarterly* 29(1), 35-57.
- Wong Fillmore, L. and C. Snow. 2000. *What Elementary Teachers Need to Know about Language*. Washington, DC: Center for Applied Linguistics.

Zoom 4

언어 교육

<https://pusan.zoom.us/j/82875282876?pwd=WkVSSWxaRTJOTUU2Z2tRVGhQTyt1Zz09>

회의 ID: 828 7528 2876

암호: 123456

The Underuse of English Progressive Aspect of Korean ESL Learners: A Collostructional Analysis

노강산 · 송상현

(고려대학교)

1. Introduction

The progressive aspect, which denotes ongoing actions or events in general, is actively used by English native speakers in their daily lives. In case of Korean ESL learners, however, English progressive seems to be underused or repelled to some extent. By taking the approach of collostructional analysis, the current study aims to shed light on Korean ESL learners' underuse of English progressive construction in an attempt to explain the underlying factors leading to Korean ESL learners' avoidance behavior.

2. Background

The underuse of certain expressions or structures has been studied with care since the pioneering study by Kleinmann (Kleinmann, H. H. 1977). In this study, Kleinmann is in support of the traditional method of contrastive analysis is still valid and furthermore argues that *an intersection of linguistic and psychological variables* determines the learner behavior, such as avoidance behavior. Even though via the use of contrastive analysis, the structural differences between L1 and L2 have been dealt with for a long time, the significance of the affective state of L2 learners resulting from cognitive difficulties has been marginalized. Following the study of Kleinmann, the current study delves into underuse behavior observed from Korean ESL learners, especially with regards to English progressive construction. While the affective side of ESL learners are to be dealt with by considering cognitive difficulties arising from L2 structure complexity, the traditional method of contrastive analysis will also be applied, considering the structural differences between Korean (L1) and English (L2).

3. Methods

3.1. Collostructional Analysis

Collostructional analysis (CA), which was proposed by Anatol Stefanowitsch and Stefan Th. Gries, is a set of methods that aims to measure and explain the degree of attraction and repulsion that words seem to exhibit to constructions. The current research adopts the point of view of their 2003 version (Gries, St. Th., & A. Stefanowitsch, 2003) in which they defined their method as follows.

[W]e propose a type of collocational analysis which is sensitive not only to various levels of linguistic structure, but to the specific constructions found at these levels. We will refer to this method as collostructional analysis. Collostructional analysis always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected);

As Gries and Stefanowitsch insist, CA is a method which aims to analyze the degree of attraction and repulsion that lexemes turn out to show to the particular construction which they are a part of. One of the advantages of CA is that it makes use of the Fisher exact test in order to maximize its precision. Since the Fisher exact test is quite difficult to use due to its computational complexity, it is applied for the sake of precision. By using collostructional analysis, the current study plans to excavate specific collexemes¹⁾ that are either attracted or repelled by English progressive.

3.2. Contrastive Analysis Hypothesis

Contrastive analysis hypothesis (CAH) is the method of analyzing a pair of languages by focusing on their structural differences. CAH has been widely used. Theoretically based on two strands of schools known as behaviorism and structuralism, CAH immensely affected both language acquisition and language pedagogy throughout the 20th century. Although it is now under the criticisms from many schools of thought, the current study adopts the point of view of CAH, hoping to find out structural differences between L1 and L2 which seem to result in learner behavior in the end.

3.3. Corpora

The learners' corpora containing the written response of Korean ESL learners were used as the basis of collostructional analysis. One is GLC (Gaecheon Learner Corpus) and the other is YELC (Yonsei English Learner Corpus).

1) Collexemes refer to lexemes that are attracted to or repelled by particular constructions.

4. Results & Discussions

Attracted			Repulsed		
RANK	LEXEME	p.value	RANK	LEXEME	p.value
1	thrill	9.67E-151	1	do	4.90E-149
2	excite	3.12E-76	2	think	1.69E-147
3	frighten	1.17E-72	3	try	1.71E-124
4	eat	2.14E-69	4	make	3.51E-110
5	interest	1.13E-64	5	go	1.02E-105
6	increase	7.58E-42	6	look	1.28E-77
7	amaze	3.84E-41	7	talk	1.63E-75
8	speed	1.40E-32	8	struggle	9.26E-56
9	bore	1.34E-29	9	watch	1.27E-42
10	drive	4.23E-28	10	use	2.69E-42
11	lack	6.70E-21	11	give	3.89E-38
12	fascinate	8.19E-21	12	have	1.42E-35
13	say	2.00E-20	13	see	7.80E-30
14	touch	5.50E-17	14	buy	6.10E-28
15	exist	6.98E-11	15	feel	3.34E-25
16	depend	8.41E-11	16	fight	6.84E-25
17	disgust	6.47E-10	17	happen	1.23E-22
18	ring	1.01E-09	18	hurt	1.28E-21
19	include	1.49E-09	19	learn	1.45E-21
20	annoy	4.56E-09	20	take	5.55E-21

Table 1. Collexemes most strongly attracted to or repelled by the progressive construction

By using collostructional analysis, the list of collexemes that are attracted to or repelled by the progressive construction was made as in table 1. It shows the collexemes that are either strongly attracted to or repelled by the progressive construction. The left column consists of the top 20 collexemes that are strongly attracted to the progressive while that of the right consists of the top 20 collexemes that are strongly repelled by the progressive.

It should be noted that from the top 20 attracted collexemes on the left, a series of 11 specific collexemes (e.g. *thrill*, *excite*, *frighten*, *interest*, *increase*, *amaze*, *bore*, *fascinate*, *touch*, *disgust*, *annoy*) are usually used as deverbal adjective (or participle adjective) as in (1), (2), and (3).

- (1) This movie is about an interesting adventure.
- (2) Last night, I saw a really exciting match between Korea and Germany.
- (3) What a boring story!

From (1) to (3), adjectival *ing* (e.g. *interesting*, *exciting*, *boring*) shows their adjectival characteristic by modifying nouns or noun phrases. After having a look at the corpus of Korean ESL learners, such examples were often found as in (4), (5), and (6).

- (4) She often tell us interesting story. (YELC)
- (5) But the most interesting thing is the clean sky. (YELC)
- (6) Her teaching is very interesting although math is boring subject. (YELC)

Note also that the other collexeme like *say* can be used as gerundive nominals as in (7).

- (7) Saying your opinion to others is very important.

From the corpus of Korean ESL learners, similar examples were to be found as in (8) and (9).

- (8) The right of saying their own ideas is really important nowadays.
- (9) The freedom of saying makes that possible.

In addition, verbs like *depend* and *include* are often used as the part of phrases as in (10) and (11).

- (10) Depending on your manner, I will decide what to do.
- (11) 4 students, including myself, were stuck in the library.

The corresponding examples can be found from Korean ESL learners corpus as in (12) and (13).

- (12) I think values are very variable depending on what situation people are in, age, sex, nation, etc.
- (13) (a) First of all, from 1970s, in order to develop Korean society, we redeemed western country especially USA as our model and has been pursued almost everything including their clothing, food, industry, technology, economical system, education, and so on. (YELC)
- (b) Should animals excluding human being be used in medical experiments? (YELC)

The examples cited above show that Korean ESL learners use the most of the top 20 attracted collexemes not as the progressive, but rather as deverbal adjective, gerundive nominals or phrases. Thus, it is likely that the top 20

collexemes were mistakenly regarded as the collexemes of the progressive just because of their *ing* form while in fact most of them were not. To sum up, even the list of the top 20 attracted collexemes seems to reveal the fact that Korean ESL learners have a tendency to avoid using English progressive construction in contrast to their relatively frequent use of *ing* form as deverbal adjective, gerundive nominals, or phrases.

Now, turning to the right column, the list of the top 20 repelled collexemes is to be seen. Before delving into analysis, it seems to be worth noting that the traditional approach towards the progressive has been arguing that stative verbs are generally repelled by the progressive construction and to what extent this restriction is valid has been fiercely debated, ending up giving this conundrum the name ‘the progressive puzzle’ (Lee, 2006). According to the traditional approach, stative verbs – verbs which denote static features, not dynamic actions – are not compatible with the progressive, which denotes ongoing actions or events. The incompatibility between them is provided in (14) and (15).

- (14) (a) I see you.
 (b) *I am seeing you.
- (15) (a) She has many guitars.
 (b) *She is having many guitars.

In the list of the top 20 repelled collexemes, some stative verbs are to be seen (e.g. *think, look, watch, have, see, feel*). While in case of English native speakers the amount of repelled stative verbs is greater (Gries, St. Th. & A. Stefanowitsch, 2003), it seems that Korean ESL learners also tend to avoid stative verbs when it comes to the progressive construction. Pedagogically, Korean ESL students are also in general taught that verbs denoting stativity cannot be used in progressive constructions. A long list of verbs that are known to be incompatible with the progressive (e.g. *love, like, hate, prefer, need, know, etc*) is often given to students in classroom setting, thereby making them avoid using stative verbs when they have to use the progressive. While it is true that in general stative verbs is incompatible with the progressive construction, there seems to be a series of exception as in (16).

- (16) (a) I was having a good time.
 (b) She was having breakfast when I came in.

As it can be seen from the examples above, certain stative verbs like *have* seem to be often compatible with the progressive construction when their meanings are extended. In case of (16) (a), *have* does not literally mean ownership, but rather denotes the experience of an event. In case of (16) (b) too, *have* extended its

meaning to denote the event of eating. Pedagogically considering, it can be said that Korean ESL learners are not given enough information that some stative verbs, although in a limited way, can coexist with the progressive construction, contributing to their avoidance behavior.

However, what should be noted is that Korean ESL learners also tend to avoid using dynamic verbs (e.g. *do, try, make, go, talk, struggle, use, give, buy, fight, happen, hurt, learn, take*) as well. This means that aktionsart (or the lexical aspect) alone cannot explain why Korean ESL learners have a tendency to avoid English progressive construction. In addition to the pedagogical factor, the cognitive factor might also explain why English progressive construction is being avoided by Korean ESL learners. English progressive construction, which consists of *be* verb and *ing* form, might lead Korean ESL learners to avoid using it due to its structural complexity. If Korean ESL learners are to put English progressive into use, they need an understanding of the combination between subjects, appropriate *be* verbs corresponding to subjects and tense, and *ing* form of the lexemes. This kind of cognitive pressure is highly likely to be a burden to the most ESL learners who are not familiar with English progressive construction, thereby resulting in the underuse of the construction. Instead of using the progressive, they can be tempted to choose the alternative option of simple present (or past) tense. From the learners' corpus, some examples lacking proper *ing* forms were found as in (16), (17), (18), (19), and (20).

- (16) Many student thinks they are lock the school and their teachers are angry to their act. (YELC)
- (17) Before I graduate highschool, I was usually spent my saturday playing soccer with my friends.
- (18) People think they are not happen accident.
- (19) Suddenly, a computer was not work! (YELC)
- (20) I disagree with the using of real name on the internet, because it can be block the right of freedom to express someone's opinion.

The examples cited above manifest the cognitive difficulties that Korean ESL learners seem to encounter when they try to use English progressive construction. Especially, in case of (17), the ESL learner is having a difficulty matching the past tense and the *ing* form of *spend*. It shows that ESL learners can be deterred from using the progressive due to its structural complexity. It is possible that the ESL learners decided to go for play-it-safe strategy in order to minimize their errors by using the simple present or past tense.

Finally, as the last cause of the underuse of English progressive construction, the structural differences between Korean progressive construction and that of English language is proposed. Consider (21) for example.

- (21) (a) I will move out next week.
 (b) I am moving out next week.

In (21)(a), the simple future tense involving the auxiliary *will* is given. In order to deliver the same message, the progressive construction is being used in (21)(b). In this way, in English language it is possible to denote the actions or events that will take place in the near future. However, things are a little different in case of Korean as in (22).

- (22) Korean: nanun taum cwuey isalul kanta
 English: I-SUB next on week moving-ACC go-decla
 'I will move out next week.'

Although in English the progressive construction can be used to express the actions or events taking place in the near future as in (21), the example from (22) indicates that for Korean speaker it is not so. If Korean speaker put the progressive into use, it will result in an ungrammatical sentence as in (23).

- (23) Korean: *nanun taum cwuey isalul kako issta
 English: I-SUB next on week moving-ACC go-comp ing-decla
 'I am moving out next week.'

What should be noted here is the fact that English progressive construction can be applied to express the actions or events that is about to happen in the near future while Korean progressive construction cannot. In other words, there seems to be a non-overlapping area between the scope of English progressive construction and that of Korean progressive construction. While to what extent the scope of English progressive and that of Korean progressive overlaps is in need of further research, it seems certain that this kind of structural disparity plays a part when it comes to Korean ESL learners' underuse of English progressive.


5. Conclusion

Through collostructional analysis, the list of attracted and repelled collexemes were found and analyzed, making it possible to hypothesize why Korean ESL learners tend to avoid using English progressive construction. With the assistance of the method, it was revealed that the attracted collexemes actually were not used as progressive, but as other constructions like deverbal adjective, gerund, and phrases. With respect to the repelled collexemes, the pedagogical factor regarding the lexical aspect, the cognitive difficulties resulting from structural

complexity, and the structural differences between the two languages' progressive construction were proposed as potential contributors to Korean ESL learners' underuse of English progressive construction. While more needs to be said about the underuse of other constructions, it is hoped that the current study can contribute to the understanding of underuse problems and the application of collocation analysis.

References

- Gries, St. Th., & Stefanowitsch. (2003). *Collocations: Investigating the interaction of words and constructions*. International Journal of Corpus Linguistics 8:2, 209–243
- Kleinmann H. H. (1977). *Avoidance behavior in adult second language acquisition*. Language Learning, 27, 93–107
- Lee, S.-A. (2005). *Ing forms and the progressive puzzle: a construction-based approach to English progressives*. J. Linguistics 43, 153–195.



A Corpus-based Study on Chinese EFL Learners' Use of Amplifier Collocations

Presenter: Shaojie Li (PhD student)
(Korea Maritime and Ocean University)

2021 Online Winter Conference
The Korean Association of Language Sciences
18th of February, 2021

CONTENT

01 Introduction	02 Methodology
03 Results and discussion	04 Conclusion

01

Introduction


Introduction



Intensification is a medium for expressing such interpersonal information as **“impressing, praising, persuading, insulting, and generally influencing the listener’s reception of the messages”** (Partington 1993: 178) in daily communication.

It plays an important part in both spoken and written interaction.

Introduction

 Intensifiers are divided into two subsets.



• Amplifiers

Amplifiers “scale upwards from an assumed norm” (Quirk *et al.* 1985: 589).
e.g. *a very interesting book* ~~is~~ *an interesting book*



• Downtoners

Downtoners “have a generally lowering effect, usually scaling downwards from an assumed norm” (ibid).
e.g. *You are probably right.* ~~is~~ *You are right.*

Introduction

Amplifiers

Maximizers

denote “an absolute degree of a scale” (ibid: 590), and cannot be modified or compared for degree.

Boosters

denote “a high degree on the scale” (ibid), but do not signify the extreme limit.

- _____
- The focus of the present research is on investigating the accurate use of amplifier collocations at two different proficiency levels.
- _____
- Intensified items involve adjectives, adverbs and verbs, with the amplifiers functioning as adverbs.

Introduction



Research
Questions



■ What are the general features of the amplifier collocations used by Chinese EFL learners?

■ What are the features of the amplifier collocations used by the learners at intermediate level (IL)?

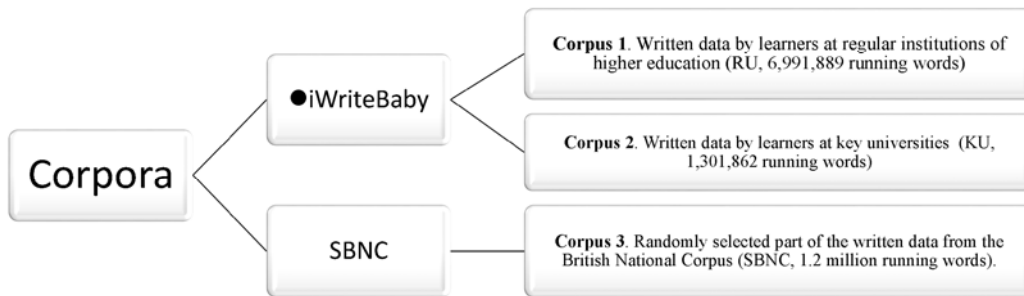
■ What are the features of the amplifier collocations used by the learners at advanced level (AL)?

02

Methodology

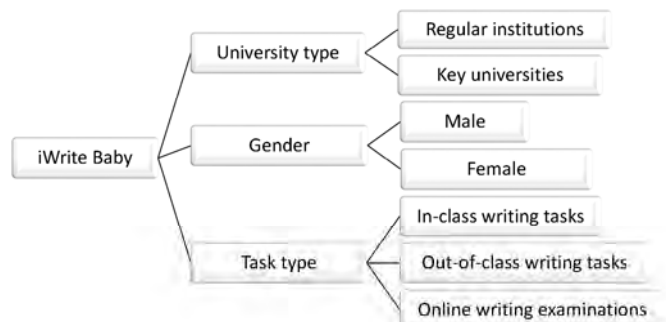
Methodology

Corpora used in the present study



Methodology

The compositions were written by Chinese EFL learners of 154 different disciplines from 69 universities. More than 1,000 topics are included in the corpus.



Methodology



The amplifiers involved in the present research were restricted to adverbs, mostly those ending with *-ly*.



By referring to Bolinger(1972), Quirk et al.(1985), Kennedy(2003), as well as teaching syllabuses of college English China, 31 amplifiers are selected as the research subjects.

Table 1. Amplifiers selected in the study

Maximizers	absolutely, altogether, completely, entirely, fully, perfectly, thoroughly, totally, utterly
Boosters	awfully, badly, bitterly, clearly, considerably, deeply, enormously, extremely, far, greatly, heavily, highly, incredibly, intensely, particularly, severely, so, strongly, terribly, very, very much, violently

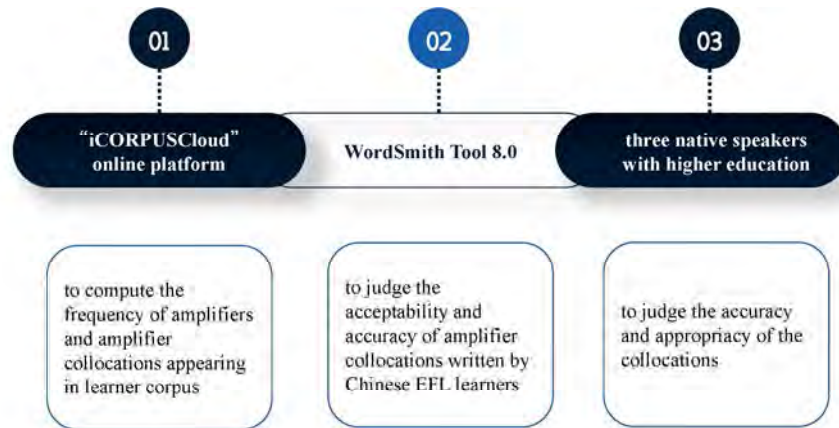
Methodology



- Irrelevant amplifier functions of these words were ruled out manually.

For instance, *so* was included as a booster (e.g. *so much, so beautiful*) but not as a conjunction (e.g. *So you can choose ...; So when you find the job...; so that we can control...*) or other meaning senses as an adverb (e.g. *I think so.*).

Methodology



Methodology

iCORPUSCloud

The screenshot displays the iCORPUSCloud search interface. On the left, search filters are set: 1. 语料: WriteBaby 1.0 bet; 2. 子库名称: All in One; 3. 子库: All in One. The main area shows a detailed word frequency table for the word 'to'.

词	词频	词频/总词频	词频/总词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频	词频			
to	225,501	19.225	0.628	0.628	812	841	827	415	29	2940	to	7285	72	1424	1532	1526	2053						
the	313,805	9.589	0.677	0.677	963	807	1482	391	11	3861	the	3839	1071	815	1132	1494	1347						
and	164,989	6.729	0.916	0.916	417	614	782	221	221	6682	and	2639	7	685	375	390	578						
we	106,687	6.132	1.066	1.066	267	225	194	230	11	905	we	7227	1698	1863	911	481	315						
that	61,944	3.563	1.231	1.231	186	156	211	131	47	733	that	6038	1623	1250	501	741	425						
is	149,752	7.279	0.905	0.905	441	447	383	391	190	3675	is	3695	12	1548	891	756	618						
I	83,171	4.514	1.074	1.074	208	325	190	190	38	1733	I	4791	2174	1006	496	281	194						
on	45,863	3.543	1.274	1.274	110	125	173	61	11	540	on	5055	4495	67	127	139	171						
can	93,192	4.889	0.827	0.827	264	107	108	151	9	719	can	4177	57	1337	1208	660	314						

03 Results

Results

Results - Amplifiers

Table 2. Raw and normalized frequency of amplifier tokens in IL and AL (per 100,000 words)

Maximizers	IL		AL		Boosters	IL		AL	
	Raw	Normalized	Raw	Normalized		Raw	Normalized	Raw	Normalized
completely	445	121	6.96	34.16	very	16299	2320	233.03	178.21
totally	348	141	4.98	26.73	so	9087	1634	129.96	125.51
absolutely	270	91	3.86	20.74	very much	937	146	13.33	11.21
fully	270	58	3.86	20.74	deeply	791	180	11.31	13.83
perfectly	156	52	2.23	11.98	greatly	682	129	9.75	9.91
amirately	128	23	1.83	9.83	clearly	608	176	6.7	13.52
thoroughly	39	14	0.56	3	highly	461	79	6.59	6.07
altogether	9	4	0.13	0.69	extremely	384	95	5.49	7.3
utterly	5	4	0.11	0.61	far	338	94	4.83	7.5
TOTAL	1673	508	23.93	39.62	strongly	268	74	3.83	5.68
					badly	264	52	3.78	3.99
					particularly	164	44	2.35	3.38
					heavily	78	13	1.12	1
					severely	49	17	0.7	1.31
					intensely	31	6	0.44	0.46
					enormously	25	4	0.36	0.31
					terribly	22	8	0.31	0.61
					considerably	11	8	0.16	0.61
					awfully	10	3	0.23	0.08
					incredibly	12	5	0.17	0.38
					violently	3	3	0.04	0.23
					literally	2	1	0.03	0.08
					TOTAL	30521	5060	316.52	390.98

There is a significant difference between the learners at IL and AL in using boosters ($p=0.004<0.05$). ↩

Table 3. The normalized frequency of boosters in the two corpora (after removing so, very, very much, per 100,000 words)

	IL	AL
Amplifiers	60.20	70.04

↪ The learners at AL use significantly more maximizers than the learners at IL ($p=0.008<0.05$). However, different from the maximizers, the learners at AL use fewer boosters than those at IL ($p=0.153>0.05$).

Results



Results-Booster collocations

Table 4. The normalized frequency of boosters in the two corpora (per 100,000 words)

Boosters	IL		AL	
	Raw	Normalized	Raw	Normalized
clearly	608	8.70	176	13.52
deeply	791	11.31	180	13.83
greatly	682	9.75	129	9.91
so	9087	129.96	1634	125.51
very	16293	233.03	2320	178.21
very much	932	13.33	146	11.21

In the two learner corpora, there are six boosters occurring more than 100 times, among which the top three boosters, *so*, *very*, *very much*, are picked out for further analyses.

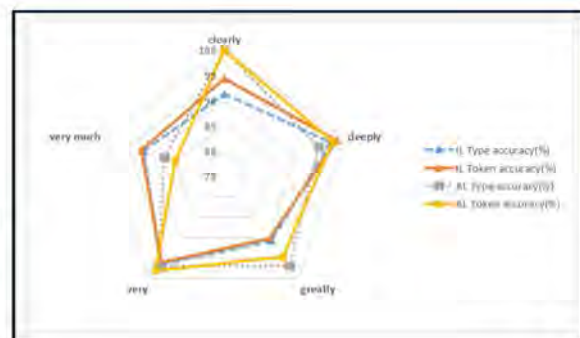
Results



Results-Booster collocations

very & very much

- 3 of the first 5 high-frequency words emphasized by *very much* in the two corpora are the same: *like*, *love* and *miss*. 5 high-frequency collocates, *enjoy*, *love*, *like*, *miss* and *thank you* are all the same.
- When Chinese EFL learners use *very*, 7 of the top 10 high-frequency collocates emphasized in the two corpora are the same, namely *important*, *good*, *much*, *happy*, *hard*, *like* and *convenient*, most of which are the same as Lorenz's (1999) research results on German English learners.



Results



Results-Booster
collocations

so

The use of *so* is mainly concentrated in 8 patterns:

- SO + ADJ;
- SO + ADJ + NOUN(S);
- SO + ADJ/ADV + THAT;
- SO + ADV;
- NOT + SO + + ADJ/ADV + + AS;
- SO + ADJ/ADV + AS + TO;
- SO + ADJ/ADV + TO + DO;
- SO MUCH SO THAT.

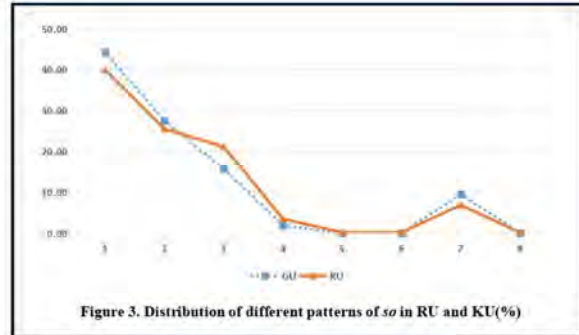


Figure 3. Distribution of different patterns of *so* in RU and KU (%)

04

Conclusion

Conclusion

1) Chinese EFL learners overuse “all-rounders” *so, very, and very much*.

In daily input from reading, listening and speaking, the three amplifiers appear more frequently. Therefore, learners regard them as core words or lexical ‘teddy bear’, and become part of productive vocabulary.

2) The learners at IL use more boosters and fewer varieties of collocates than those at AL.

From the perspective of syntax and semantics, the top three amplifiers are relatively simple.

The learners at IL are not familiar with the use of other amplifiers. In their writing, they use the universal amplifiers just to play safe.

3) The learners at AL demonstrate significantly higher accuracy in using amplifier collocations.

The learners at AL were enrolled by key universities, and their *Gaokao* scores were higher than those at regular institutions. Generally speaking, the learners in the key universities are more motivated, have more appropriate learning strategies and more learner autonomy, which contributes to their accurate use of amplifier collocations.

Conclusion



• Pedagogical implications

English teachers and textbook writers should pay attention to the use of amplifier collocations according to the specific characteristics of students at different proficiency levels in the process of English teaching and textbook compilation so that learners can remember them as chunks instead of as individual “bricks”, which will help learners to communicate in English more authentically and accurately.

References



- Allerton, D. J. (1987). English intensifiers and their idiosyncrasies. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honor of Michael Halliday* (pp. 15-31). Amsterdam: John Benjamins.
- Altenberg, B. (1991). Amplifier collocations in spoken English. In S. Johansson & A. Stenstrom (Eds.), *English computer corpora: Selected papers and research guide* (pp. 127-147). Berlin: Mouton de Gruyter.
- Bäcklund, U. (1973). *The collocation of adverbs of degree in English*. Uppsala: Acta.
- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47, 56-63.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Essex: Pearson Education Limited.
- Bolinger, D. (1972). Degree words. The Hague: Mouton.
- Bley-Vroman, R. (2002). Frequency in production, comprehension, and acquisition. *Studies in Second Language Acquisition*, 24, 209-213.

References



- Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal*, 35, 115-122.
- Ellis, R. *The Study of Second Language Acquisition[M]*. Oxford: Oxford University Press, 1994.
- Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143-188.
- Gass, S., & Mackey, A. (2002). Frequency effects and second language acquisition. *Studies in Second Language Acquisition*, 24, 249-260.
- Gitsaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco: International Scholars Publications.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.). (pp. 145-159).
- Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (Ed.). (pp. 199-209).
- Greenbaum, S. (1970). *Verb-intensifier collocations in English: An experimental approach*. The Hague: Mouton.

References



- Hulstijn, J. H. (2002). What does the impact of frequency tell us about the language acquisition device? *Studies in Second Language Acquisition*, 24, 269-273.
- Hunston, S. *Corpora in Applied Linguistics*[M]. Cambridge: Cambridge University Press, 2002.
- Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly*, 37, 467-487.
- Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition of second language learners. *Language Learning*, 26, 125-134.
- Lorenz, G. R. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.). (pp. 53-66).
- Lorenz, G. R. (1999). *Adjective intensification: Learners versus native speakers: A corpus study of argumentative writing*. Amsterdam: Rodopi.

References



- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223-242.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins, 2005.
- Partington, A. (1993). Corpus evidence of language change: The case of the intensifier. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.). (pp. 177-192).
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Stubbs, M. *Words and Phrases*[M]. Oxford: Blackwell, 2001.

References



- Tarone, E. (2002). Frequency effects, noticing, and creativity. *Studies in Second Language Acquisition*, 24, 287-296.
- Xiao Gang(2019). A Corpus-based study on amplifiers in academic writing by English majors. Qingdao University of Science & Technology.
- Xu Jiajin(2019). The construction of iWriteBaby Chinese learners' English corpus. *Corpus Linguistics*, 6, 105-117.
- Wei Y. & Lei L. (2011). The use of amplifiers in the doctoral dissertations of Chinese EFL learners. *Chinese Journal of Applied Linguistics*, 1, 47-61, 127.
- Zhao Bei(2017). A corpus-based comparative study on amplifiers in English academic writing. Zhejiang Gongshang University.

THANKS

디지털 미디어를 통한 한국인 영어학습자들의 교실 밖 영어학습 현상과 정의적 요인의 관계 연구

이향미
(부산대학교)

1. 서론

•정보통신기술과 디지털 미디어의 급속한 발전은 영화, TV, 음악, 컴퓨터게임 그리고 소셜 네트워크(SNS)와 같은 매체들을 통한 영어학습 기회를 증가시켰다.

•Richards(2009)는 스웨덴이나 핀란드와 같은 북유럽 국가들의 학생들이 남부 유럽이나 아시아 학생들 보다 영어학습 능력이 높다고 보고하면서 디지털 미디어 사용을 그 원인으로 지적하였다.

•Socket과 Tofoli(2012) 또한 음성 및 문자 소프트웨어, 블로그, 온라인 커뮤니티, 소셜 미디어, 멀티플레이어 온라인 게임 등 다양한 디지털 자원이 공교육과는 무관하게 학습자에게 영어학습을 지원한다고 언급하면서 영어학습에 있어 디지털 미디어 활용의 중요성을 제기하였다.

•Lai와 Gu(2011)는 영어 학습을 위한 디지털 미디어 기술은 교실 밖에서 가장 잘 실현되고 극대화될 수 있으며 이러한 기술의 가치에 대한 학습자들의 수용과 효과적인 사용이 영어학습에 중요하다고 언급하였다.

•Taher와 Tam(2012) 또한 정보통신기술과 디지털 미디어의 발전을 통해 다른 사람들과의 상호작용, 목표 언어인 영어에의 노출 등 영어 학습자가 교실 밖에서 비공식적으로 영어를 배울 수 있는 많은 방법이 있다고 언급하였다.

I. 서론

- AbuSa'aleek(2015)는 영어학습자를 대상으로 SNS를 이용한 영어학습의 가능성과 학습자들의 인식에 대해 조사하였고 그 결과 facebook과 같은 소셜 네트워크를 이용한 영어학습 플랫폼이 학습자들의 동기와 자신감과 같은 정의적 요소의 향상에 기여하였으며 학습자들이 긍정적인 태도를 배양하는데 도움을 주었다고 보고하였다.
- Lee(2019) 교실 밖 영어 디지털 학습에 대하여 인내, 동기부여, 자신감 그리고 불안감과 같은 정의적인 요인과 영어로 의사소통하려는 의지(WTC: Willingness To Communicate) 사이의 관계를 살펴 보았으며 그 결과 디지털 미디어를 통한 영어학습 활동과 정의적 요소는 WTC와 유의미한 상관관계가 있음을 언급하였다.
- 국내 영어교육과 관련하여 디지털 미디어를 이용한 전화영어, 화상영어, 앱 또는 채팅으로 이어지는 교실 밖 영어 산업은 크게 발전하고 있는 반면 이를 이용한 영어학습의 방법과 효과 그리고 한국인 영어학습자의 정의적 요소와의 관계에 대한 국내 연구는 많이 부족하다.
- 또한 최근 일 년 동안 Covid-19 대유행으로 인해 국내외 교육기관의 모든 교과목들은 줌(zoom)과 같은 온라인 화상회의 형식의 디지털 미디어 기술을 활용하고 있어, 이러한 학습 형태의 변화가 디지털 미디어를 활용한 교실 밖에서의 영어학습 활동과 그에 대한 인식변화에도 영향을 주었을 것이다.
- 본 연구는 한국인 영어학습자들의 디지털 미디어를 이용한 영어학습 활동에 대해 살펴보고자 한다. 또한 한국인 영어학습자들의 정의적 요소와 디지털 미디어를 이용한 교실 밖 영어학습 활동과의 관련성을 분석해 볼 것이다. 한국인 영어학습자들의 디지털 미디어에서의 수용적 활동과 생산적 활동을 살펴보고 인내, 동기부여, 자신감 그리고 불안감과 같은 정의적인 요인과 이들 활동과의 관계를 분석할 것이다.

II. 이론적 배경과 선행 연구

1. 디지털 미디어를 이용한 교실 밖 영어학습

Rogers(2004)는 교실 밖의 영어학습 환경은 조직화되고 의도된 학교 기반의 학습과는 달리 일상생활에서 배울 수 있는 체계화·목적화되지 않은 광범위하고 중요한 일상의 영어학습이 가능하다고 주장하면서, 교실 밖에서 이루어지는 영어학습 활동이 학습자들에게 큰 역할을 하고 있으며 이러한 교실 밖의 학습 활동을 교실 내 학습 활동보다 더 흥미롭게 느낀다고 언급하였다.

Pearson(2004)은 뉴질랜드 대학의 아카데미 영어 수업에 등록한 ESL학생들을 대상으로 교실 밖 학습활동의 인식에 대한 설문조사를 실시하였고 그 결과 대부분의 학습자들은 그들의 영어학습 향상을 위해 인터넷, 라디오, TV, 영화, 신문, 다른 학습자들과의 교류 등을 통한 다양한 학습적 자원에 영향을 받는다고 보고하였다.

Lamb(2007)은 이들 학습활동이 영어학습자에게 언어 자원을 제공할 뿐만 아니라 긍정적인 학습자의 정체성 형성과 이로 인한 학습의 동기부여를 유지할 수 있는 다양한 기능들을 제공해 오고 있으며 학습자의 영어 능력과 동기부여는 그들의 교실 밖의 활동과 밀접한 관련이 있다고 주장하였다.

Murray(2008)는 영어학습자들의 성공적인 영어학습 경험은 특히 문화와 연계된 것으로 영화, TV프로그램, 음악, 소셜 등 교실 밖 영어학습 자료들이 학습자들의 학습 능력 배양에 있어 적극적으로 긍정적인 역할을 한다고 주장하였다.

Gao(2009) 또한 디지털 미디어를 이용한 교실 밖 영어학습은 학습자들에게 익숙하고 도움이 되는 학습 커뮤니티를 제공하며 더불어 학습자들의 자아 인식 향상과 자기표현의 장을 마련해줄 수 있다고 언급하였다.

II. 이론적 배경과 선행 연구

Lai와 Gu(2011)는 교실 밖의 영역에서 디지털 미디어 활용이 목표언어를 좀 더 효율적으로 습득하게 만드는 이유를 설문 조사하였으며, 그 결과 디지털 미디어 환경이 영어학습을 자유롭고 친밀하게 만드는 경향이 있으며, 흥미로운 시청각 자료와 환경이 목표 언어를 더 효율적으로 습득하게 만드는 원인이라고 보고하였다.

Sockett와 Tofoli(2012) 또한 Strasbourg 대학교의 225명 학생들을 대상으로 여가시간에 영어로 된 온라인 매체 사용에 대한 설문조사를 실시하였고 60% 이상의 응답자들은 영어로 된 영화, TV 시리즈를 영어 능력 향상을 위해 정기적으로 시청하는 것으로 나타났다고 보고 하였다.

Taher와 Tam(2012)는 터키인 영어학습자들은 교실 밖의 학습 환경에서 다양한 영어 자원을 이용할 수 있고, 영어학습의 효율성을 증대시킬 수 있는 사회적 상호 작용 환경에 충분히 노출될 수 있으며, 디지털 미디어 기술을 이용해 다양한 시청각적인 자원에 접근이 용이하여 실생활에서도 영어 습득이 가능함과 동시에 영어말하기 능력도 함양할 수 있다고 주장하였다.

종합하면, 교실 밖의 영어학습에 관한 연구들은 영어학습자들이 디지털 매체를 이용한 교실 밖의 다양한 학습 활동에 적극적으로 참여하고 있음을 보여주고 있다.

II. 이론적 배경과 선행 연구

2. 교실 밖 영어학습과 정의적 요인들

McCarthy(2010)는 바로 질의응답을 해야 하는 물리적 교실환경에서 보다 온라인에서 학습자들은 자신들의 학습 속도와 언어능력에 맞게 코멘트를 달고 응답을 할 수 있기에 자신감을 가지고 좀 더 유연한 상호 작용을 할 수 있다고 언급하면서 영어학습자들은 SNS상의 문법, 단어, 그리고 새로운 정보를 업데이트하며 서로 간의 격려를 통해 영어학습 동기를 고취할 수 있다고 주장하였다.

Baltner와 Fiori(2009)는 컴퓨터를 기반으로 한 SNS는 동시적·비동시적 상호작용을 할 수 있어 학습자들에게 심리적 안정감을 제공하고 그림, 비디오와 같은 자원을 제공하여 흥미를 북돋울 수 있기에 영어능력 향상에 도움을 줄 수 있으며, 온라인 커뮤니티에서 학습과정 및 결과를 많은 사용자와 함께 공유할 수 있기 때문에 Facebook과 같은 디지털 매체는 강력한 학습 도구일 뿐 아니라 의사소통적 도구로서 그 기능을 더욱 확장하고 있다고 주장하였다.

Akbari, Eghtesad과 Simons(2012)는 학생들이 SNS 사용 이후 SNS가 효과적인 공식적 교육 도구로서 매우 큰 잠재력을 가질 수 있음을 인식하게 되었고, SNS의 유용성에 대한 긍정적인 태도가 증가함과 동시에 SNS를 활용한 영어교육 참여에 흥미가 높아짐을 보고하였다.

II. 이론적 배경과 선행 연구

2. 교실 밖 영어학습과 정의적 요인들

AbuSa'aleek(2015)은 사우디아라비아Qassim 대학교 대학생 65명을 대상으로 SNS가 영어학습자의 영어 실력 향상에 도움이 되었는지를 언어 향상, 자신감, 동기, 그리고 태도 등 4가지 면에서 광범위하게 연구한 결과 SNS가 언어 능력을 향상시킬 뿐만 아니라 자신감 상승, 동기부여의 육성, 그리고 언어 상호 교류를 유발시킨다고 언급하였다.

Lee와Dressman(2018)은 한국인 영어학습자들을 대상으로 디지털 미디어 영어학습 활동과 WTC와의 관련성을 연구한 결과 디지털 미디어 영어학습 활동 중 문법 과업 수행과 WTC가 관련성이 가장 높으며, 디지털 미디어 영어학습 중 생산성 중심의 쓰기, 말하기 활동이 WTC 능력 지표를 예측하는 주요한 요인이라고 보고하였다.

Lee와Drajati(2019) 또한 183명의 인도네시아 영어학습자들을 대상으로 디지털 미디어 영어학습과 인내, 동기, 자신감 그리고 제2 언어 말하기에서의 불안감 등의 정의적 요소 및 WTC사이의 상관관계를 연구하였다. 연구 결과 인내, 자신감, 동기부여와 디지털 미디어 영어학습 중 생산적인 활동만이 L2학습자의WTC에 많은 영향을 주었으며 디지털 미디어 영어학습의 수용적인 활동, 말하기 불안감, 그리고 인구통계학적인 변인 등은L2학습자의WTC에 결정적 요소가 아니라는 사실을 지적하였다.

디지털 미디어를 사용하여 이루어지는 교실 밖 영어학습의 다양성이 학습자에게 영어 습득의 중요한 자원으로 불구하고 이와 관련된 국내의 연구는 많이 이루어지지 않았다. 따라서 디지털 미디어의 활용과 관련한 교실 밖 영어학습 활동과 그와 관련한 학습자의 정의적 요인의 관계에 대하여 살펴볼 필요가 있을 것으로 사려된다.

III. 연구 방법

1. 연구 대상

참여자	총 104명	남자 61명	여자 43명
디지털미디어 사용시간	평균 4.27시간		
영어 학습을 위한 디지털미디어 사용시간	평균 0.194시간		
가장 선호하는 영어 학습 도구	영어 TV시리즈		
영어 학습을 위해 사용하는 디지털 미디어 종류로	1. 태블릿	2. 휴대폰	3. 컴퓨터

2. 연구 도구

한국인 영어학습자들의 디지털 미디어를 이용한 교실 밖 영어학습 활동 현황 및 이들 활동과 학습자들의 정의적 요소와의 관계를 살펴보기 위해 디지털 미디어 영어학습 활동과 정의적 요인, 디지털 미디어 사용 정도에 대한 설문지를 구성하였다.

III. 연구 방법

1) 디지털미디어영어학습활동설문지

<수용적 활동>	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
7. 나는 온라인이나 TV에서 영어 뉴스 프로그램을 듣는다.	①	②	③	④	⑤
<생산적 활동>					
8. 나는 소셜미디어를 통해 다른 사람과 영어로 채팅을 한다.	①	②	③	④	⑤

2) 정의적 요인 설문지

<인내>	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
4. 나는 말할 때 열심히 말한다.	①	②	③	④	⑤
5. 나는 짧은 시간에 특정 아이디어나 프로젝트에 대한 생각에 사로잡히지만 시간이 지나면 흥미를 잃는다.	①	②	③	④	⑤
<중기>					
11. 나는 영어공부에 있어서는 많은 노력을 기울인다.	①	②	③	④	⑤
<자신감>					
17. 나는 영어 시간에 선생님이 하는 말을 대부분 이해할 수 있다고 자신한다.	①	②	③	④	⑤
<불안감>					
25. 나는 다른 학생들 앞에서 영어로 말할 때 긴장한다.	①	②	③	④	⑤

III. 연구 방법

3) 디지털 미디어 사용 설문지

2. 하루에 중 영어 학습을 위해 TV, 핸드폰 그리고 컴퓨터와 같은 디지털 미디어를 사용하는 시간은 몇 시간입니까? 사용 하지 않는다. □ 0 ~ 2시간 □ 3 ~ 5시간 □ 5 ~ 7시간 □ 7시간 이상 □
3. 영어 학습을 위해 사용하는(또는 사용할 예정인) 미디어는 어떤 종류입니까? (중복체크 가능) TV □ 컴퓨터 □ 핸드폰 □ 태블릿 □ 기타 □ (종류:)

3. 연구 절차 및 분석 방법

2020년 9월 2주 부터 3주간 관련 선행연구들을 분석하고 설문지 구성에 맞는 문항43개를 발췌하였으며 이들 문항은 한국인 영어 학습자들의 상황에 맞게 수정되었다. 영어 학습을 위한 디지털 미디어사용 설문지는 10월 첫 주 문항7개를 작성하여 완성하였다.

2020년 10월 2주, 한 광역시 소재 4년제 대학교 ‘해외 인턴십을 위한 영어회화’, “비즈니스 영어회화” 그리고 “글로벌어학교육 토익프로그램” 수업을 등록한 126명의 한국인 영어학습자들을 연구 대상으로 선정하고 담당 교수들에게 양해를 구한 후 학습자들에게 연구의 목적 및 방법, 그리고 연구 참여 동의를 확인한 뒤 연구를 시작하였다.

SPSS 25.0 프로그램을 사용하여 통계분석을 수행하였다. 연구 참여자들은 영어능력(TOEIC)을 기준으로 세 그룹으로 분류 되었으며 이들 그룹간에 교실 밖 디지털 미디어 영어 학습활동 및 정의적 요인에 차이가 있는지를 일원배치분산분석을 통해 비교하였다. 또한 피어슨(Pearson) 분석을 통해 정의적 요인들과 디지털미디어 영어학습활동간의 상관관계를 분석하였으며 디지털 미디어 영어 학습활동에 대한 정의적 요인들의 기

IV. 분석 결과

영어능력 기술통계

	참여자수	최소값	최대값	평균	표준편차
전체	104	250	890	657.21	129.992
상그룹	32	725	890	796.56	44.057
중그룹	42	590	720	669.76	39.043
하그룹	30	250	580	491.00	74.515

IV. 분석 결과

영어능력별 그룹간 일원배치분산 분석 결과

	제곱합	자유도	평균제곱	F	유의확률
그룹-간	1446801.852	2	728400.926	259.328	.000
그룹-내	283689.494	101	2808.807		
전체	1740491.346	103			

IV. 분석 결과

사후다중비교 분석 결과

그룹	그룹	평균차이	표준화오류	유의확률
상그룹	중그룹	126.801	12.436	.000
	하그룹	305.563	13.469	.000
중그룹	상그룹	-126.801	12.436	.000
	하그룹	178.762	12.669	.000
하그룹	상그룹	-305.563	13.469	.000
	하그룹	-178.762	12.669	.000

IV. 분석 결과

디지털 미디어 영어학습 활동 정도 기술통계

활동종류		참여자수	최소값	최대값	평균	표준편차
수용적	전체	104	8	31	17.75	5.004
	생산적		6	27	12.00	5.404
수용적	상그룹	32	9	31	18.94	5.054
	중그룹	42	8	28	17.40	5.022
	하그룹	30	9	24	16.97	4.860
생산적	상그룹	32	6	24	13.44	5.764
	중그룹	42	6	27	11.64	5.314
	하그룹	30	6	22	10.97	4.972

IV. 분석 결과

세 그룹간 일원배치분산 분석 결과

		제공합	자유도	평균제곱	F	유의확률
수용 적	그룹-간	68.539	2	34.270	1.378	.257
	그룹-내	2510.961	101	24.861		
	전체	2579.500	103			
생산 적	그룹-간	103.515	2	51.758	1.800	.171
	그룹-내	2904.485	101	28.757		
	전체	3008.000	103			

IV. 분석 결과

정의적 요인 기술통계

활동종류		참여자수	최소값	최대값	평균	표준편차
인내	전체	104	2	36	26.21	4.608
			15	40	27.14	5.608
자신감			9	39	23.74	5.963
불안감	전체	104	9	29	18.81	4.569
			13	32	24.91	3.658
인내	상그룹	32	13	32	24.91	3.658
	중그룹	42	2	34	25.95	5.165
	하그룹	30	21	36	27.97	4.271
동기	상그룹	32	15	40	29.88	5.701
	중그룹	42	15	38	25.38	5.342
	하그룹	30	17	36	26.70	4.865
자신감	상그룹	32	12	39	25.78	5.706
	중그룹	42	9	35	22.57	5.696
	하그룹	30	9	33	23.20	6.222
불안감	상그룹	32	13	29	20.56	4.690
	중그룹	42	11	27	18.74	4.073
	하그룹	30	9	27	17.03	4.529

IV. 분석 결과

세 그룹간 일원배치분산 분석 결과

		제공합	자유도	평균제곱	F	유의 확률
인내	그룹-간	149.759	2	74.878	3.712	.028
	그룹-내	2037.590	101	20.174		
	전체	2187.346	103			
동기	그룹-간	375.132	2	187.566	6.615	.002
	그룹-내	2863.705	101	28.354		
	전체	3238.837	103			
자신감	그룹-간	199.436	2	99.718	2.909	.059
	그룹-내	3462.554	101	34.233		
	전체	3661.990	103			
불안	그룹-간	193.193	2	96.597	4.985	.009
	그룹-내	1956.961	101	19.376		
	전체	2150.154	103			

IV. 분석 결과

상관관계

	수용적	생산적	전체	인내	동기	자신감	불안
수용적	1	.660**	.904**	-.007	.496**	.506**	-.080
생산적		1	.918**	-.184	.265**	.465**	.028
전체			1	-.108	.413**	.532**	-.026
인내				1	.013	-.035	-.082
동기					1	.650**	.024
자신감						1	-.119
불안							1

IV. 분석 결과

디지털 미디어 영어학습(통합) 활동 관련 다중회귀분석 결과

	변인	B	표준오차	베타(β)	t	p
모형1	상수	11.592	8.122		1.427	.157
	인내	-.192	.180	.093	1.065	.289
	동기	.201	.191	.119	1.056	.294
	자신감	.722	.181	.454	3.984	.000
	불안	.037	.188	.018	.197	.844
	영어점수	.000	.007	-.002	-.025	.980

IV. 분석 결과

수용적 영어학습 활동 관련 다중회귀분석 결과

	변인	B	표준오차	베타(β)	t	p
모형1	상수	5.387	4.263		1.264	.209
	인내	-.002	.095	-.002	-.022	.983
	동기	.265	.100	.297	2.653	.009
	자신감	.256	.095	.305	2.694	.008
	불안	-.059	.099	-.053	-.594	.554
	영어점수	.000	.004	-.009	.099	.921

IV. 분석 결과

생산적 영어학습 활동 관련 다중회귀분석 결과

변인	B	표준오차	베타(β)	t	p
모형1 상수	6.206	4.785		1.297	.198
인내	-.190	.106	-.162	-1.789	.077
동기	-.064	.112	-.067	-.572	.569
자신감	.466	.107	.514	4.363	.000
불안	.096	.111	.081	.864	.390
영어점수	-.001	.004	-.013	-.131	.896

V. 결론

영어능력의 높고 낮음에 관계없이 연구 참여자들은 유사한 정도로 디지털 미디어 영어학습의 수용적, 생산적 활동을 수행함을 알 수 있었고, 정의적 요인 중 ‘자신감’이 디지털 미디어 영어학습의 생산적, 수용적 활동 모두에게 영향을 미치는 주요한 요인

디지털 미디어를 통한 지식정보의 공유, 소통, 협업의 과정은 학습자의 영어능력과 함께 학습자가 지니는 정의적 요인의 중요성을 더욱 강조하고 있다. 이는 Lee(2019a)가 언급하였듯이 학습자가 지니고 있는 정의적 요인들이 비대면 학습환경에서 학습에 대한 흥미를 북돋우고, 영어능력과 상호작용하여 학습의 방향 및 결과 성취를 위한 다양한 노력을 견인하는 역할을 하기 때문이다.

본 연구는 디지털 미디어 영어학습 활동에 있어 정의적 요인 중 자신감의 중요성을 보고하고 있다. Bandura(1994)는 자신감이란 자기 통제적 신념을 형성하고 목표를 구체화시키며 결과성취를 위한 노력으로 정의하였다. 영어학습에서 자신감은 영어사용의 불안감을 통제하는 중요한 수단이 될 수 있으며, 학습과제나 난제에 부딪혔을 때 학습자들은 자신의 자신감에 따라 과제를 선택하고 성취범위를 결정할 수 있다(박근명, 2019; Schunk, 1983).

국내 영어학습자들을 대상으로 디지털 미디어 영어학습 현황과 정의적 요인간의 관계를 규명하는데 기여하였지만 다음과 같은 몇 가지 제한점을 가진다.

- 연구 참여자들의 디지털 미디어 영어학습 현황과 정의적 요인을 설문지를 사용하여 수집하였다.
- 특정 대학교 재학생들을 대상으로 연구를 수행하였기에 연구 결과를 일반화하는데 제한점이 있다.

다양한 디지털 자원이 공교육과는 무관하게 학습자에게 영어학습을 지원하고 있는 현시점에서 디지털 미디어는 강력한 학습 도구이자 영어 원어민들과 상호작용을 가능하게 하는 의사소통 도구로서 기능을 확장하고 있다. 따라서 디지털 미디어의 활용과 관련된 영어학습 활동에 관한 연구는 지속적으로 수행되어야 할 것이다.

실무용 한국어교재 개발 방안 연구

원해영
부산외국어대학교

- 차례 -

1. 연구목적 및 필요성
2. 선행연구 검토
3. 학습자 요구 분석
4. 비즈니스교재 분석
5. 비즈니스교재의 문제점
6. 실무용 한국어 교재 개발 방안
7. 마무리

1. 연구목적 및 필요성

• 연구목적

국내외 한국기업에 취직하고자 하는 외국인학습자를 대상으로, 한국회사에서 업무를 수행하고 한국인 상사 또는 동료와 한국어로 직무에 대해 논의할 수 있는 실무용 한국어교재 개발 방안을 연구하고자 함.

1. 연구목적 및 필요성

• 연구의 필요성(배경)

- 해외진출 한국기업의 꾸준한 증가
- 해외 한국기업의 인력 대부분이 현지인
(중간관리자+현장근로자)
- 한국인관리자와 현지인 현장근로자 사이의 의사조율(업무지시, 교육, 건의, 요청)을 담당할 수 있는 현지인 중간관리자를 위한 실무한국어교재 필요
- 실무 중심의 교재가 부족함

• 해외진출 한국기업의 현황

- 해외 이주 기업들은 최소한의 한국인관리자를 본국에서 파견하고 대부분의 노동인력을 현지에서 고용하고 있음
- 특히 중국과 베트남을 위시한 아세안지역은 한국기업들이 선호하는 지역으로 매년 해외 진출 기업이 늘어나고 있음
- 최근에는 근 24년 동안 해외진출기업 증가수 1위를 차지했던 나라가 중국에서 베트남으로 바뀌는 현상이 나타남



- 2017년 3월 8일자 한국경제 인터넷뉴스 :
 - 해외진출기업(1만1953) - 본사파견한국인력 : 6만 6166
 - 현지 채용 인력 : 338만 4281

- **산업통상자원부와 코트라(KOTRA)가 2019년 2월 19일 발표 내용.**
 - 세계 125개 코트라 무역관을 통해 1만2500여 개의 현지 진출 법인을 대상으로 실시한 '2018 해외진출기업 실태조사'에 따르면 매출실적 조사에 응한 현지법인 2513곳 중 46%가 앞으로 매출이 늘어날 것으로 내다봤다. 36%는 매출이 유지될 것으로 봤고 매출 감소를 예상한 법인은 18%에 그쳤다.
(2019년 02월 19일 20시 52분 | 세계일보 | 이우중 기자)
 - 고용현황 조사에 응답한 현지법인(3,502개사)은 **단순 생산직 중에 대부분 현지인력을 채용(99%)**하고 있었으며, 마케팅, 상품개발 등 관리직종에서는 **한국인 채용비중이 13%(12,200여 명)로 다소 높은 편으로** 나타났다.
- > 한국인관리자와 생산직 현지근로자들의 의사소통을 담당하고 있는 현지인 중간관리자 교육을 위한 실무 한국어 교육과정 및 교재 필요**

2. 선행연구 검토

- (1) 학습자 요구 및 교재 분석
 - 형희빈(2015), 취업목적 중국인을 위한 비즈니스 한국어 교재 구성에 관한 연구
 - 김도완(2016), 베트남 대학 한국학과 직업 목적 한국어 교재 개발을 위한 요구 분석
 - 함향·호길·전송철(2017), 중국인을 위한 비즈니스 한국어 교재 분석 연구.
 - 김현진 외 (2018), 세종학당 비즈니스 한국어 교재 개발 연구
 - 장성희(2019) 비즈니스 한국어 교재 개발 연구-베트남 내 전자기기 제조업 종사자의 요구분석을 바탕으로
 - 한상미,강승혜(2020), 베트남 학습자 대상 비즈니스 한국어 교육과정 개발을 위한 요구분석
- (2) 한국어교육내용 및 교육과정 연구
 - 김호정(2011), 태국 대학생을 위한 비즈니스 한국어 교육과정 설계 방안 연구
 - 이동희(2015), 실무 한국어 쓰기 교육과정 연구: 태국을 중심으로
 - 김중숙(2017), 직업목적 한국어교육과정 설계에 관한 연구
 - 박서윤(2017), 한국계 기업 취업을 위한 비즈니스 한국어교수요목 설계
 - 정윤아(2019), 비즈니스 한국어 교재의 주제 및 어휘 연구

(3) 현황 및 과제

- 이미혜(2003), 직업을 위한 한국어 교육 연구 : 교육 현황 및 '비즈니스 한국어' 개발 검토.
- 이미혜(2008) 국내 직업 목적 한국어 교육의 현황과 과제
- 안화현(2017) 태국대학교 '비즈니스 한국어' 교육 현황과 발전 방향(요구분석)

3. 학습자 요구 분석

• 베트남어권

- 베트남 현지 한국인직원, 베트남직원 요구조사

- 1) 한국인과 베트남의 주제에 대한 인식 차이 : 홍보, 시장 조사, 매출과 이익, 납품과 클레임, 계약 및 주문, 마케팅 전략, 윤리, 경영, 비즈니스 이메일, 재고 관리, 생산 관리' 등에서는 한국인에 비해 베트남인의 요구도는 비교적 낮았다
- 2) 한국인이 사고방식에 해당하는 비언어적 문화 영역과 관용어/ 비유어 등의 언어적 문화 영역에서 베트남인들의 요구도가 한국인 집단에 비해 낮았다
- 3) 비즈니스 관련 희망 교육내용으로 '어휘, 표현'에 대한 요구가 가장 많았고 '회화', '문화', '업무처리기술', '기타', '언어기능', '경제경영지식', '문법' 순으로 나타났다.(자유응답식)
- 4) '문화'관련 내용 또한 비즈니스 맥락에서 필요한 '직장 문화' 및 '직장 예절' 등

3. 학습자 요구 분석

- 중국어권

-대학생 대상(비즈니스 목적)

- 1) 비즈니스 한국어 교재에서 **적절한 비즈니스 실무 회화와 문서 작성 방법에 관한 내용과 직장생활과 직장문화에 관한 내용이 부족함**
- 2) 비즈니스 관련 직장에 대한 소개, 한국 사회 혹은 문화에 대한 소개, 비즈니스 관련 실무에 대한 설명 및 비즈니스 관련 직장에서 자주 사용되는 일상생활용어 등을 유용하다고 판단함
- 3) 비즈니스 용어 설명, 비즈니스 서식 작성 요령 과 관련 절차, 비즈니스 상담 회화, 취직 전 면접요령, 한국 비즈니스 관련 직장의 예절·생활과 문화, 비즈니스 관련 직장에서 자주 사용되는 일상생활 용어 등이 포함되어야 함
- 4) **비즈니스 현장에서 직접 사용할 수 있는 회화로 구성된 교재, 그리고 직장 취업 관련 내용이거나 직장의 일상생활 관련 내용이 풍부한 교재 희망**

- 중한 무역회사 재직 중인 비즈니스맨(중국인) 30명(남녀 각 15명) 대상 요구조사

- 1) 대부분의 재학생들이 듣기, 쓰기와 읽기를 가장 중요하다고 생각하는 반면 **직장인의 경우에는 말하기와 듣기가 중요하다고 판단**
- 2) 직장인들은 근무 연한에 따라 현재 처리하는 비즈니스 업무의 양과 종류가 차이가 있다고 생각함. 근무 연한 1년 이하의 직장인 응답자들은 보통 간단한 비즈니스 회화와 한국어를 자주 쓰지 않는 업무만 말는데 반해 근무 연한 1년 이상의 응답자들은 일반적인 비즈니스 업무를 담당함. 그리고 근무 연한이 3년 이상의 응답자들은 대부분 전문적인 비즈니스 업무 처리를 담당함.

3. 학습자 요구 분석

- 태국어권(비즈니스전공)

- 대학생, 취업자, 한국관리자, 교수

- 1) 비즈니스교과목 학습기간은 두 학기를 가르치는 것에 대한 요구가 가장 높음
- 2) 비즈니스교과목 담당교수는 회사 직장 경험이 있는 교수 그리고 한국인 교수와 태국인교수의 팀티칭에 대한 요구가 높음
- 3) '비즈니스 한국어' 교육내용 중요도는 '취업준비'(재학생, 졸업생, 교수) - '직장업무'(관리자) - '직장문화' - '직장 생활' 순으로 남(수요자에 따라 중요도가 다름)
- 4) 언어기능의 중요도에 대해서 '말하기'가 재학생은 63명(55.9%), 졸업생은 23명(56%), 교수는 5명(54.5%), 기업관리자는 '말하기'와 '듣기'가 각각 9명(47.4%)으로 가장 높게 나타났다. 듣기 중요도 높음

3. 학습자 요구 분석 정리

- 학습자(수요자) : 한국인관리자, 기취업자(졸업생), 학생, 교수

- 1) 취업준비와 관련된 내용이 포함되어야 함
- 2) 직장업무, 직장문화에 대한 내용을 중요하다고 판단함
- 3) 업무관련 어휘 표현에 대한 요구가 많음
- 4) 현장 중심의 회화, 직장의 일상생활 회화 중심의 내용 구성 원함
- 5) 업무전문성과 관련된 문서작성방법 소개
- 6) 말하기, 듣기 능력을 향상할 수 있는 내용 구성
- 7) 한국적 비언어적 사고방식에 대한 이해를 돕는 내용 구성
- 8) 비즈니스 학습에 대한 한국인과 현지인의 요구가 다름
(한국인은 전문업무- 현지인은 기초업무)

4. 비즈니스교재 분석

- 교재선정

- (가) 비즈니스한국어(연세대학교 한국어학당편, 2009)
- (나) 성공의지름길 비즈니스한국어(박창원외 2010)
- (다) 프로비즈니스한국어 1,2,3,4(김선정외 2014)
- (라) 바로 배워 바로 쓰는 비즈니스한국어1,2(세종학당재단 2019)

(가)교재 개요

- 한국회사 및 한국인이 다수 근무하는 환경에서의 외국인의 한국어능력 습득 목표
- 직장에서의 기본적 의사소통능력 + 이해능력
- 중급수준의 학습자 대상
- 일반목적 <연세한국어1, 2> 선행학습 권장

집필방향	한국어 비즈니스 활동에 도움을 주기 위함
학습대상	한국기업 또는 한국인과 일하는 중급 이상 학습자
구성표	총 30개 단원
단원구성	과-제목-학습목표-도입질문-회화-어휘문제-문법문제-듣기문제-읽기문제-과제와 활동
주제	업무관련 범위가 넓음(구직-면접-고객서비스, 계약, 광고, 홈쇼핑, 창업, 보험, 품질관리, 회식, 신제품개발...)
어휘	본문어휘 5~6, 관련어휘 10~15. 일상적 어휘가 많음
문법	단원별 2개 제시
문화	3회 제시(음식값, 선물, 미신터부) - 일반적 문화
총평	- 구체적인 업무내용보다 두루뭉실한 내용으로 구성 - 실무관련보다 일반적인 대화 상황 제시 - 제시된 주제들이 일관성이나 통일성을 찾기 어려움 - 어휘 문법이 주제와 관련성 부족 - 정확한 목표나 학습대상을 파악하기 어려움 - 단원의 첫 대화문 외 모든 영역이 연습문제로 제시됨.

(나)교재 개요

- 실용적 한국어 학습 목표
- 한국기업 취업자 및 취업준비생 대상
- 취업준비 + 취업 후 직무능력 향상 과제
- 다국적 인물 설정
- 말하기,듣기, 읽기, 쓰기, 어휘, 표현 포함

집필방향	직장생활에서 접할 수 있는 한국어 학습
학습대상	한국기업 취업의 외국인학습자
구성표	총 20과 단원
단원구성	단원표지(제목과 학습내용)-듣고 말하기-읽고 쓰기-어휘와 표현 문제-문법과 표현 문제
주제	취업준비부터 사무실 업무까지 순차적 배치(적성, 취직준비, 회사 조직과 업무, 일의 순서, 온라인업무, 출장)// 월급과 재테크, 무역, 벤처기업, 경제변화와 혁신
어휘	주제와 관련 있음. 연습문제나 덩어리표현은 관련 없음
문법	주제와 관련 없음
문화	1회 제시(직장윤리와 문화)
총평	- 목표어휘와 문법이 단원 각 영역에 충분히 활용되지 않음 - 단원간 내용이 중복됨 - 한 교재 내에 주제의 범위가 너무 포괄적임-개별취업준비부터 경제관련 전문 정보까지)

(다)교재 개요

- 상경계열 전공학습자, 한국기업 취업희망자, 무역업 희망자 대상
- 무역, 경영 중심, 기업문화 소개
- 총 4권 중 3, 4권이 비즈니스교재에 부합(1, 2권은 일상 주제로 구성)
- 읽기영역에 중점
- 고급 수준의 본문내용

집필방향	상경계열 전공준비서, 경영무역업무 지식 함양
학습대상	상경계열 학생 및 한국기업 취업준비생
구성표	총2권 각 권15단원
단원구성	(3,4권)제목-읽어봅시다-어휘-어휘문제-문법 및 표현-활동1,2
주제	1,2권-기초한국어(비즈니스 관련성 없음) 3,4권-경영무역실무(경영무역관련 주제 WTO, FTA, 아웃소싱, 브랜드정책, 임금피크제, 손익분기점, 납품, 벤치마킹, 재고관리, 리콜, 보상, 독과점 등)
어휘	주제관련 어휘 6~7 제시
문법	과당 3개, 주제 관련성 없음, 난이도 무시
문화	없음
총평	-1,2권은 직업실무와 관련성 없는 일반 한국어 -1,2권과 3,4권의 난이도 간극이 큼 -3,4권은 읽기활동으로 제한 -난이도 4,5등급으로 전체 내용이 어려움-특정 전공 대상, 전문성이 너무 강함) -한국기업문화에 대한 언급이 거의 없음

(라)교재 개요

- 해외거주자 대상, 해외 한국기업 취업 희망자
- 업무지식 및 직무능력 향상을 목표
- 일반 목적<세종한국어> 선행학습 필요
- 다양한 직책의 등장인물 설정 (본문활용 없음)
- 한국어와 영어 병기

집필방향	직장 내 기본한국어능력 향상, 기본업무지식 익힘
학습대상	취업 목적의 해외 한국어학습자
구성표	총2권 각 권 12과(과별 2파트)
단원구성	제목-학습목표-도입(그림+질문), -연습해봅시다-말해봅시다-해봅시다
주제	구직활동, 면접, 부서, 직급, 회의, 출장, 보고, 외근, 고객상담 등
어휘	단원별 12개(파트별 6개-주제와 관련성 없음)
문법	단원별 2개(파트별 1개-주제와 관련성 없음)
문화	없음
총평	-주제에 비해 제시어휘문법의 수준이 아주 낮음(2급 수준) -어휘가 주제와 연관성이 떨어짐 -주제는 교재의 목적과 부합하나 제시어휘나 문법은 주제를 뒷받침하지 못함

5. 기존 교재의 문제점

1. 교재 명칭의 문제
2. 주제의 광범위함
3. 특정(전공) 학습자 대상의 주제 편중
4. 학습 난이도가 일정하지 않음
5. 제시 어휘가 주제를 뒷받침하지 않음
6. 회화는 일반적인 내용으로 구성됨
7. 목표에 부합하지 않는 주제가 많음
8. 목표가 실무능력 향상인지 한국어능력향상인지 정확하지 않음
9. 듣기/말하기 영역에 대한 내용이 부족함
10. 직장문화에 대한 소개가 부족함
11. 문서작성과 관련된 사무업무, 실무에 대한 내용이 부족함

6. 실무용 교재 개발 방안

- 한국인관리자와 기취업자의 요구 적극적 반영
- 중간관리자용 현장 실무 중심의 내용 구성
- 교재유형 : 일반 사무직(공통)과 특수 사무직(개별)으로 이원화
- 교육내용 : 기초편(취업준비+한국기업에 대한 이해)과 심화편(직장업무 세분화)으로 이원화
- 대학 교육과정에 부합하는 교수요목 구성
- 중급 수준에 부합하는 어휘 및 문법 난이도 구성
- 실무 중심의 주제와 어휘에 중점을 둔 내용 구성
- 실무와 관련된 사무기술(문서 작성 및 기기 작동 등) 소개
- 듣기 영역을 강화한 내용 구성
- 다양한 한국기업의 문화 소개

7. 마무리

- 한국기업의 해외진출이 지속적으로 확대되고 있음
- 해외 한국기업의 현지인 비중이 높고 채용이 증가함
- 한국인과 현지인의 원활한 의사소통을 위해 현지인 중간관리자의 한국어능력 요구가 확대됨
- 현지인 중간관리자 양성을 위한 실무 한국어 교육과정 및 교재 필요
- 수요자인 한국인관리자와 현지인중간관리자의 요구 분석을 적극 반영한 실무 한국어교재가 필요함
- 실무 한국어교재 구성은 학습자의 국적보다 기업 및 업무의 특성을 반영하여야 함

Seeking for the potential of using Social Network Analysis via NVivo R 1

Chongwon Park

(Pukyong National University)

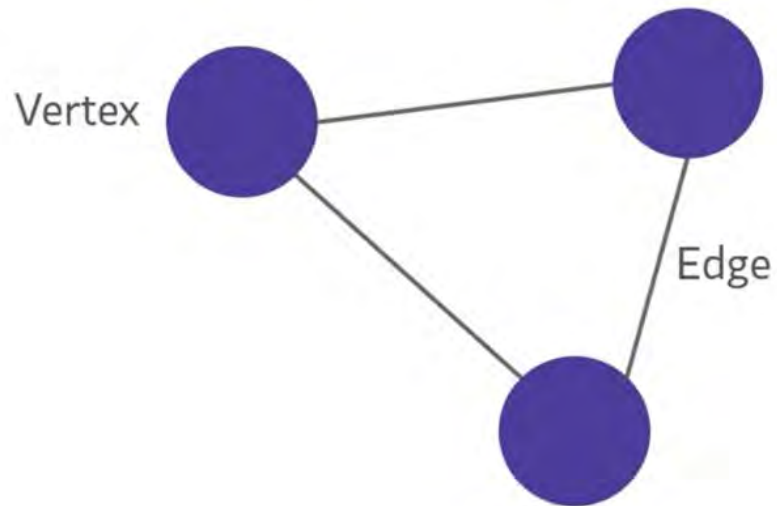
The presenter defines what social network analysis is and demonstrates how to conduct SNA studies via NVivo R1.

1. Defining Social Network Analysis

Social network analysis is a method used to map relationships and flows among people, organizations, groups, or other real-world entities. Sociogram and NVivo R1 show how social interdependence is identified as vulnerable people in the community or people who influence the flow of research information by displaying diagrams that help to visualize connections using sociogram to explore the network.

NVivo R1 can perform social network analysis on the population of cases and their relationships and create sociograms through various materials. NVivo R1 can handle various materials and automatically creates a sociogram when it receives Internet data from NCapture. To create an e-mail source diagram, you can import Outlook > data. In addition, creating a network that represents a case or relationship can generate a sociogram of any data.

What are Sociograms?



Sociogram is a visualization of social network analysis. The circle above is called the Vertex, and the Vertex represents a person, group, or other reality. The point between the vertex and the vertex is called the Edge, indicating the connection or interaction between the vertex. Vertex refers to Case Codes.

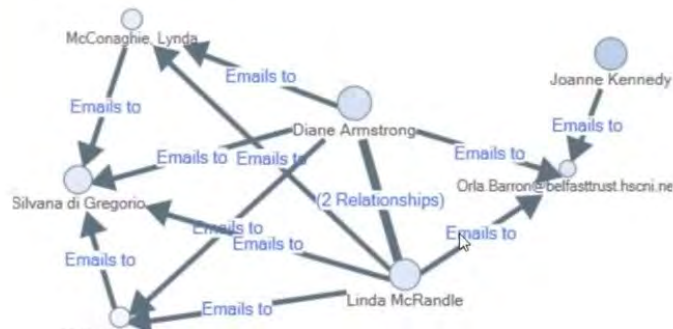
Sociograms in NVivo



There is a self-centered sociogram that can be created, which is a key vertex that has a relationship with another vertex. Ego visualized it as a self-centered sociogram. Ego is Maria, and it is possible to see all the cases

Ego visualized it as a self-centered sociogram. Ego is Maria, and it is possible to see all the cases associated with one particular case through the star-marked Ego. Unlike egocentric sociograms, network sociograms are represented by visualizing the entire or partial network and have no central point. All vertices make up an equal network. The sociogram visualizes how a group of cases are connected and isolated. The Twitter sociogram is automatically generated when Twitter material is called up to NVivo R1. The sociogram associated with the Vertex is the Twitter handle, and the Edge shows how a retweet and mention connect the user.

Sociograms in NVivo



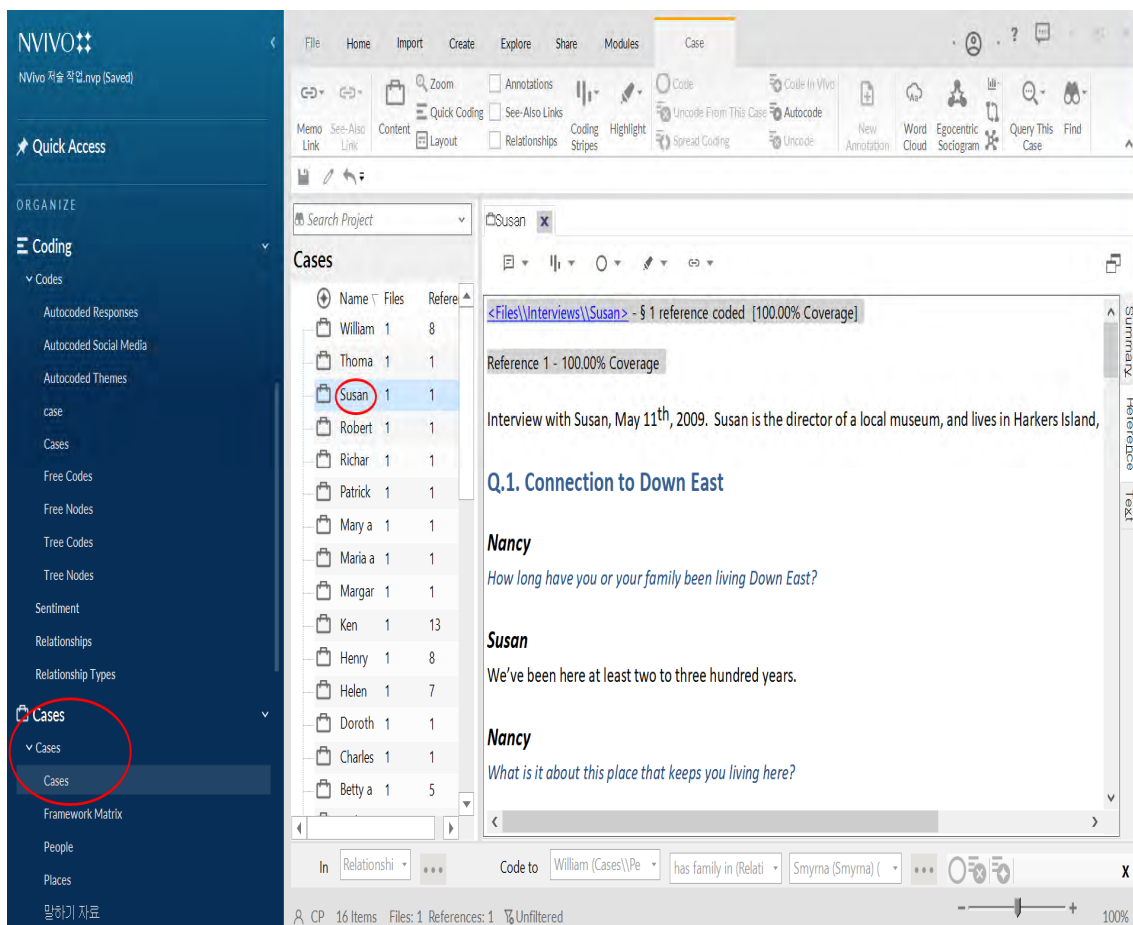
E-mail sociograms are automatically generated when researchers bring in Outlook e-mail data. The Vertex is the sender of the e-mail, and Edge is the one who receives the e-mail and answers.

I will demonstrate how to create sociograms and conduct social network analysis through NVivo R1. I will show you how to stack the basic blocks of NVivo R1 sociogram. NVivo R1 can make an example of anything an individual, organization, or researcher wants to create. Next, I create a case code, a relationship code that captures the associations between cases, and an edge. Sociogram helps researchers explore central measurements that make egocentric and network sociograms, bring in researchers' data, and visualize e-mail networks as well as Twitter sociograms.

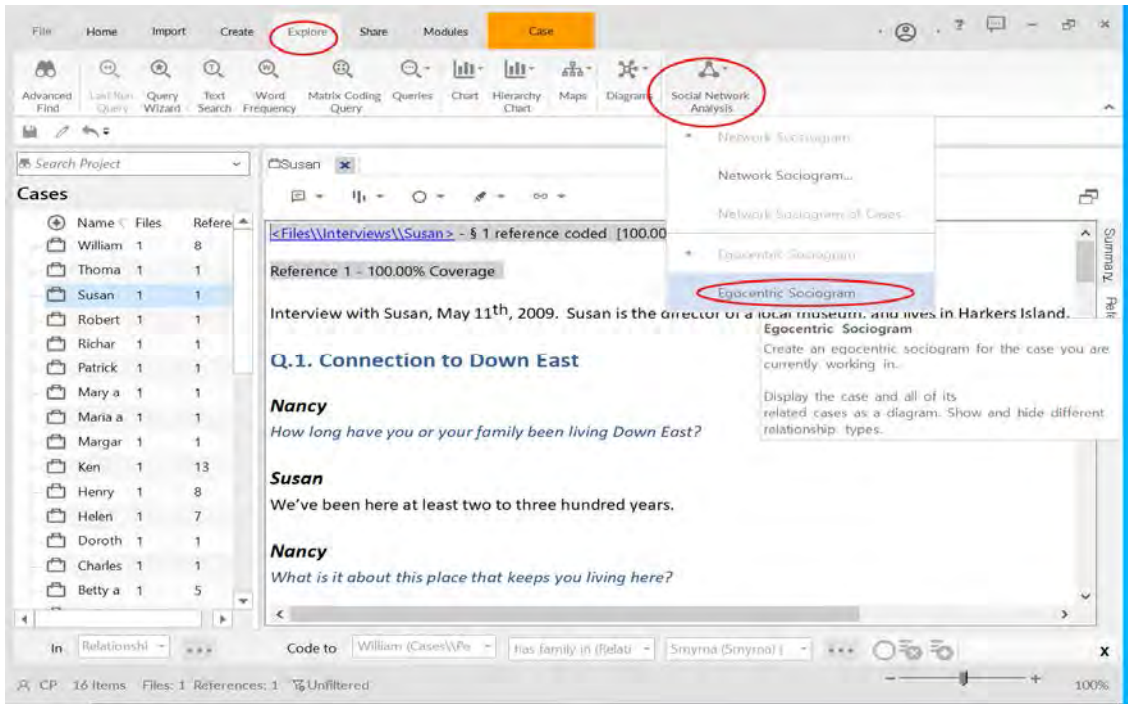
2. Sociogram

As mentioned earlier, in Twitter, the number of relationships among people can be used to create a sociogram based on relationships.

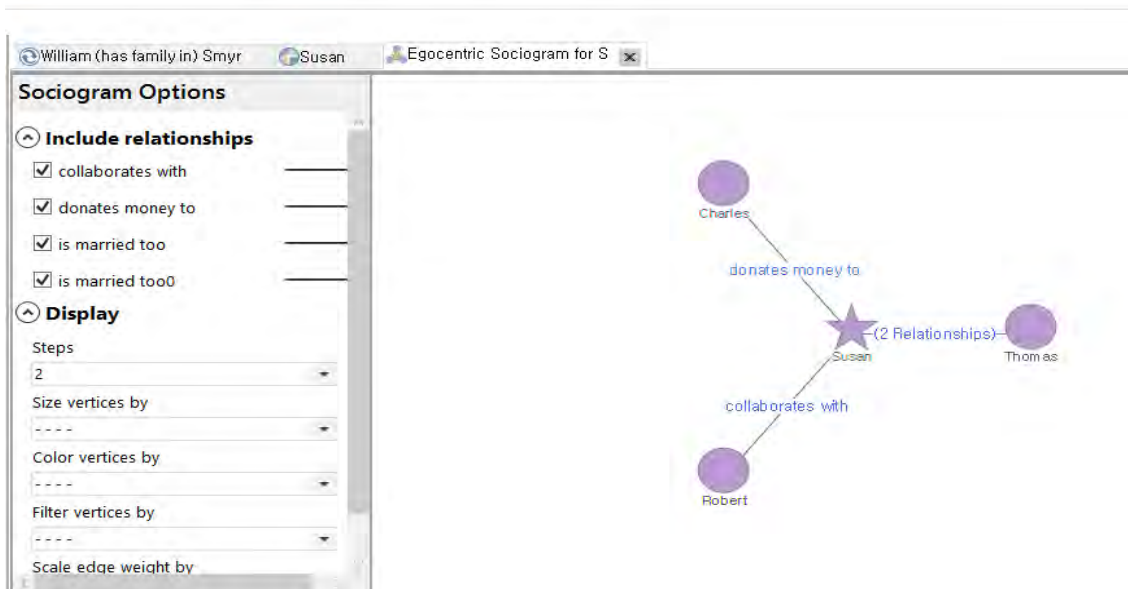
Let's make Susan's egocentric sociogram. Cases > Cases > Cases > Susan Click



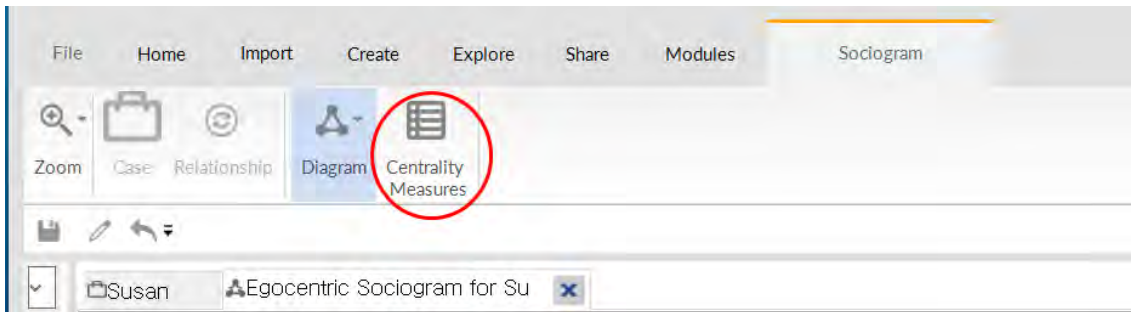
Explore > Social Network Analysis Click



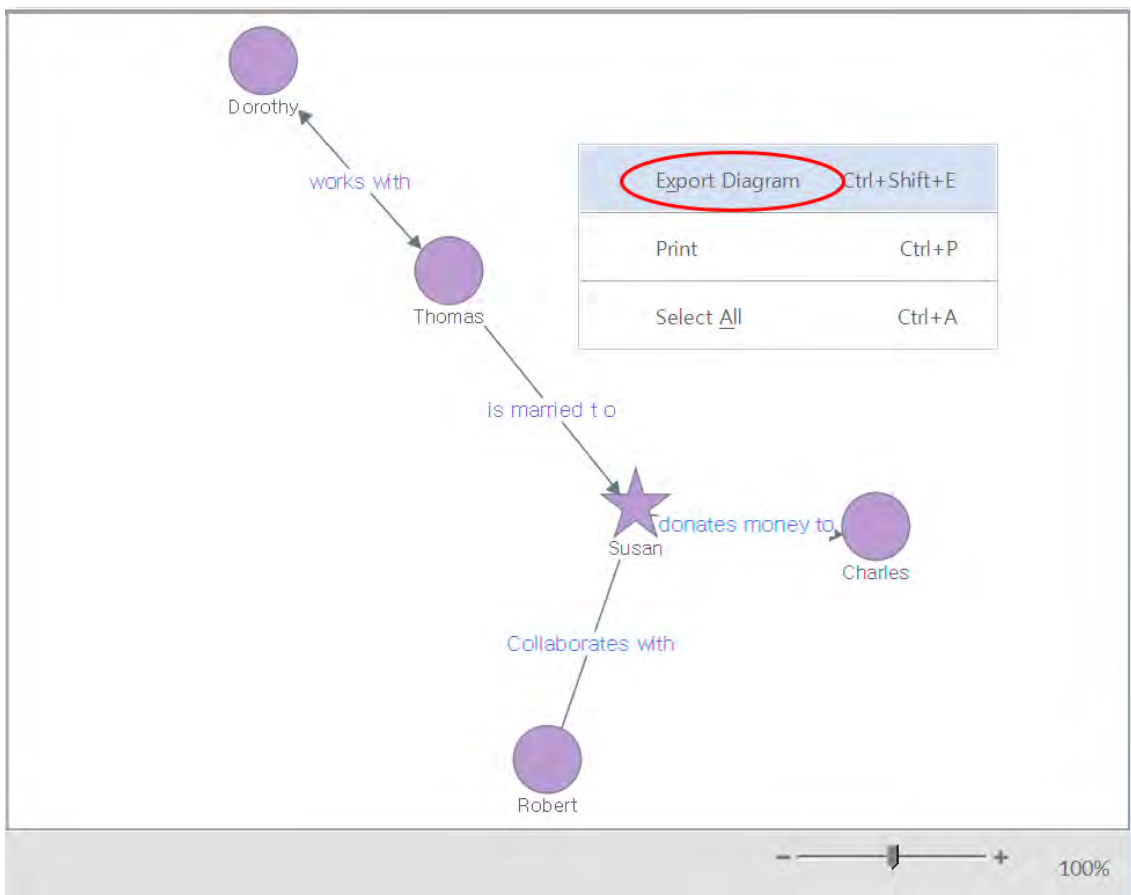
Susan is ego and indicates star.



The number of centrality measures can be measured and exported.



You can export to images or other social network analysis software can export to the Patrick edge list.



Cases > Charles, Dorothy, Robert, and Susan Select

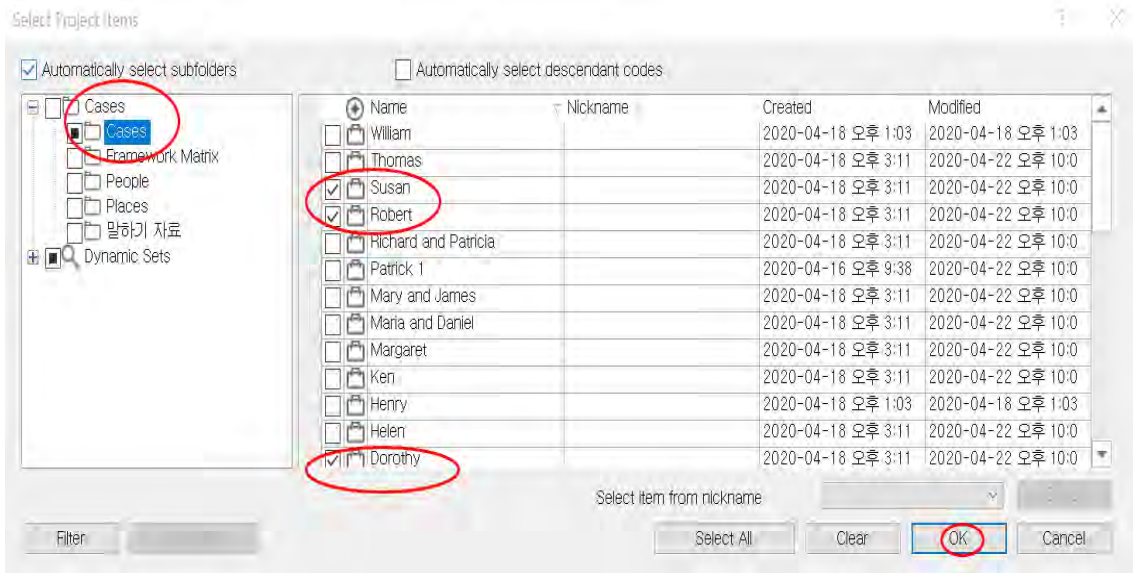
Name	Files	References	Modified on	Modified by	Classification
Barbara	1	1	2020-04-22 오후 2:47	CP	
Betty and Paul	1	5	2020-04-22 오후 10:08	CP	Case
Charles	1	1	2020-04-22 오후 10:08	CP	Case
Dorothy	1	1	2020-04-22 오후 10:08	CP	Case
Helen	1	7	2020-04-22 오후 10:08	CP	Case
Henry	1	8	2020-04-18 오후 1:03	CP	Person
Ken	1	13	2020-04-22 오후 10:08	CP	Case
Margaret	1	1	2020-04-22 오후 10:08	CP	Case
Maria and Daniel	1	1	2020-04-22 오후 10:08	CP	Case
Mary and James	1	1	2020-04-22 오후 10:08	CP	Case
Patrick 1	1	1	2020-04-22 오후 10:08	CP	Case
Richard and Patricia	1	1	2020-04-22 오후 10:08	CP	Case
Robert	1	1	2020-04-22 오후 10:08	CP	Case
Susan	1	1	2020-04-22 오후 10:08	CP	Case
Thomas	1	1	2020-04-22 오후 10:08	CP	Case
William	1	8	2020-04-18 오후 1:03	CP	Person

Social Network Analysis > Network Sociogram Click

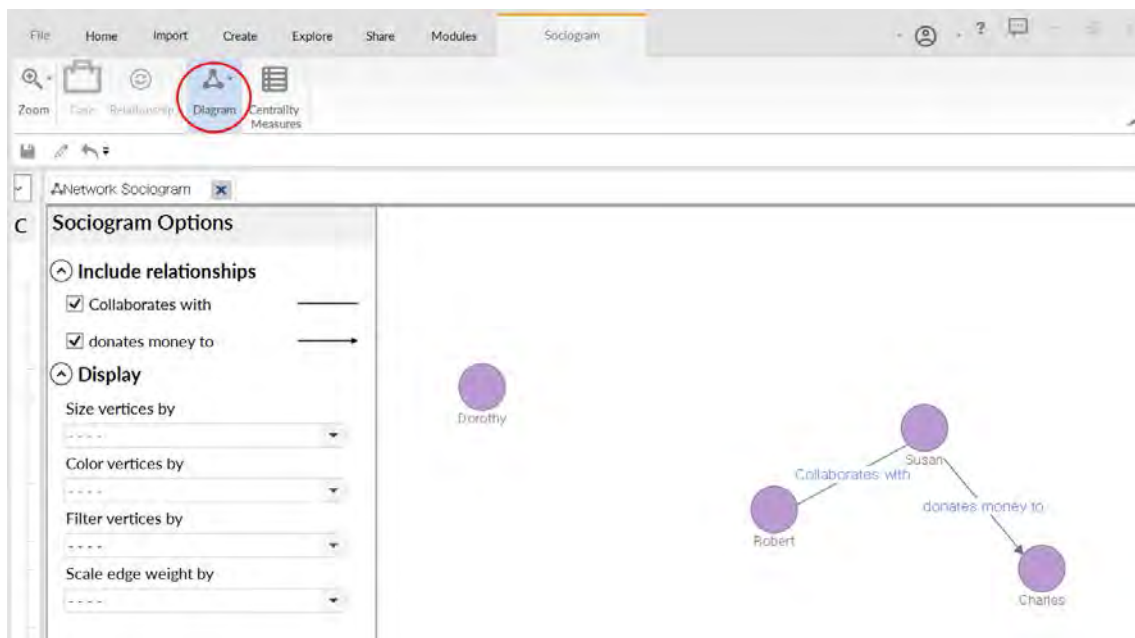
The screenshot shows the 'Social Network Analysis' menu with the following options: Network Sociogram, Network Sociogram..., Egocentric Sociogram, and Egocentric Sociogram. The 'Network Sociogram...' option is circled in red. Below the menu, the 'Cases' table is visible with 'Charles', 'Dorothy', 'Robert', and 'Susan' selected.

Name	Files	References	Modified on	Modified by	Classification
Barbara	1	1	2020-04-22 오후 2:47	CP	
Betty and Paul	1	5	2020-04-22 오후 10:08	CP	Case
Charles	1	1	2020-04-22 오후 10:08	CP	Case
Dorothy	1	1	2020-04-22 오후 10:08	CP	Case
Helen	1	7	2020-04-22 오후 10:08	CP	Case
Henry	1	8	2020-04-18 오후 1:03	CP	Person
Ken	1	13	2020-04-22 오후 10:08	CP	Case
Margaret	1	1	2020-04-22 오후 10:08	CP	Case
Maria and Daniel	1	1	2020-04-22 오후 10:08	CP	Case
Mary and James	1	1	2020-04-22 오후 10:08	CP	Case
Patrick 1	1	1	2020-04-22 오후 10:08	CP	Case
Richard and Patricia	1	1	2020-04-22 오후 10:08	CP	Case
Robert	1	1	2020-04-22 오후 10:08	CP	Case
Susan	1	1	2020-04-22 오후 10:08	CP	Case
Thomas	1	1	2020-04-22 오후 10:08	CP	Case
William	1	8	2020-04-18 오후 1:03	CP	Person

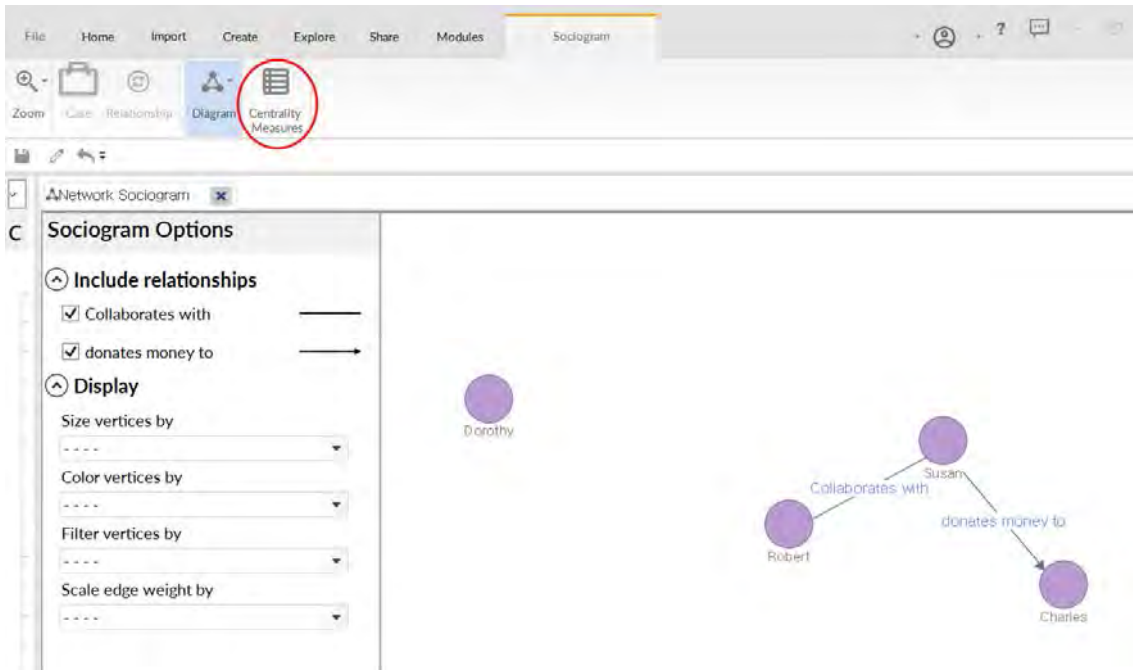
Cases > Cases > Charles, Robert, Dorothy and Susan checkmark > OK



The diagram can be viewed as follows.



Centrality Measures Click

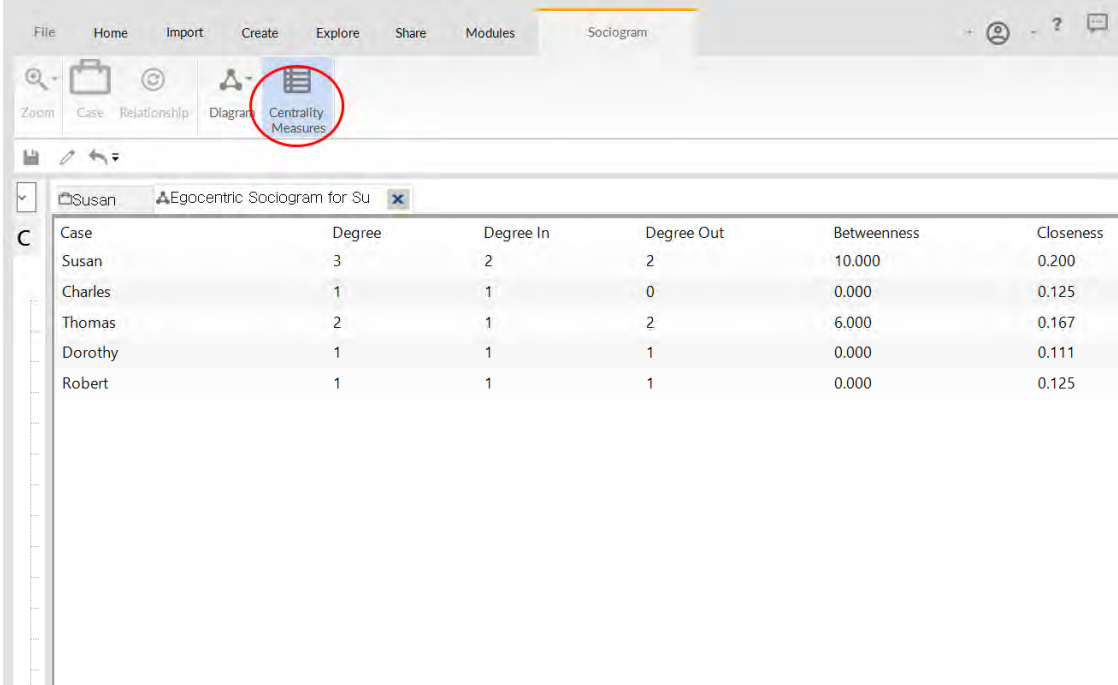


Centrality Measures Click

The screenshot shows the 'Sociogram' module interface. The 'Centrality Measures' button in the top toolbar is highlighted with a red circle. The main workspace displays a table of centrality measures for the nodes Dorothy, Susan, Charles, and Robert.

Case	Degree	Degree In	Degree Out	Betweenness	Closeness
Dorothy	0	0	0	0.000	0.000
Susan	2	1	2	2.000	0.500
Charles	1	1	0	0.000	0.333
Robert	1	1	1	0.000	0.333

The number of centrality, such as Degree, Degree In, Degree Out, and Betweenness is calculated.



The screenshot shows the NVivo software interface. The 'Centrality Measures' menu item is highlighted with a red circle. Below it, a table displays the results for an 'Egocentric Sociogram for Susan'. The table has columns for Case, Degree, Degree In, Degree Out, Betweenness, and Closeness.

Case	Degree	Degree In	Degree Out	Betweenness	Closeness
Susan	3	2	2	10.000	0.200
Charles	1	1	0	0.000	0.125
Thomas	2	1	2	6.000	0.167
Dorothy	1	1	1	0.000	0.111
Robert	1	1	1	0.000	0.125

Degree refers to the number of connections between participants.

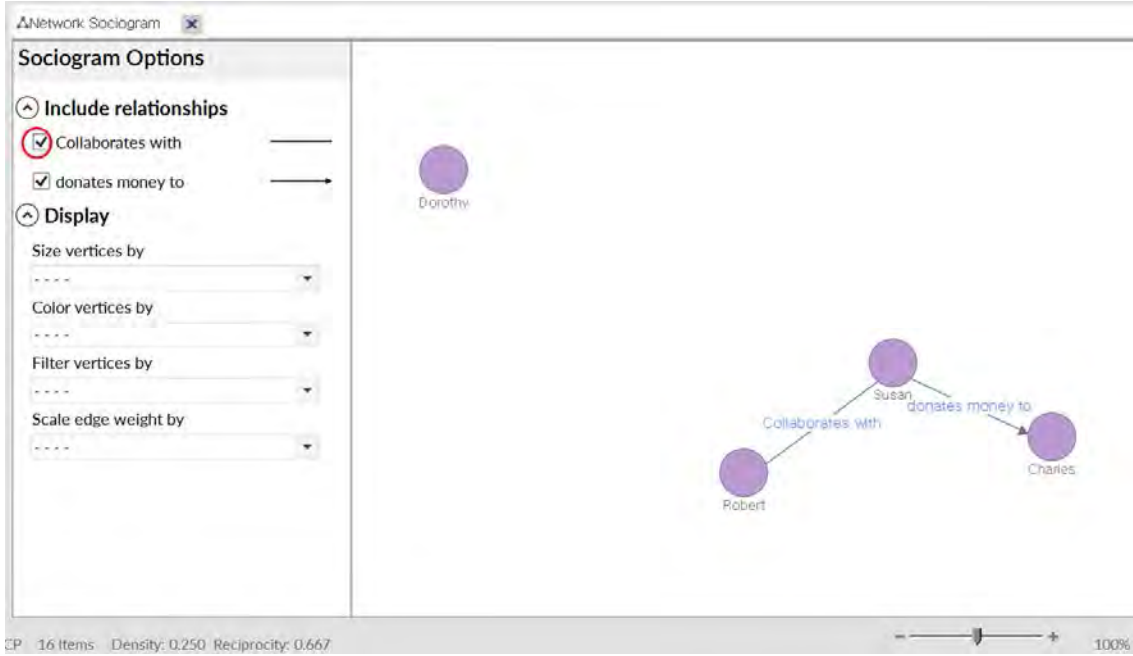
Degree in refers to the number of connections that go into the Vertex.

Degree out refers to the number of connections that go out of the Vertex.

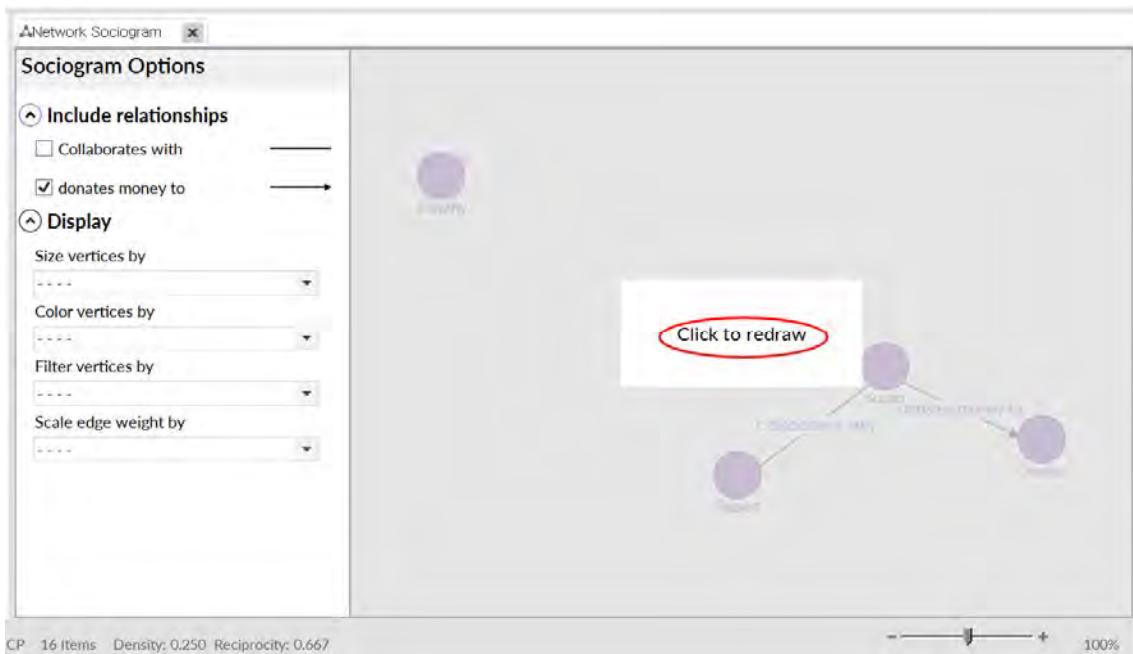
Betweenness is the number of vertices that connect the shortest distance from another vertex.

The closest distance between vertices and networks is the method of identifying hubs and data.

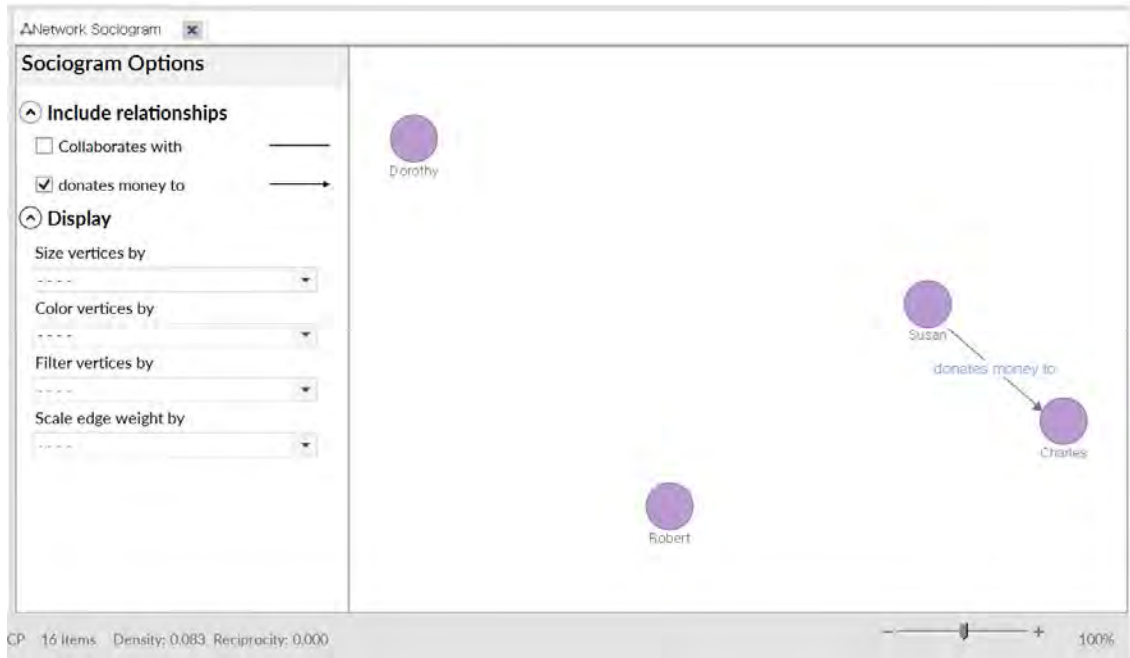
Diagrams have several sociogram options. Turn off Collaborates with and see if there are any changes in the network.



Click to redraw를 Click



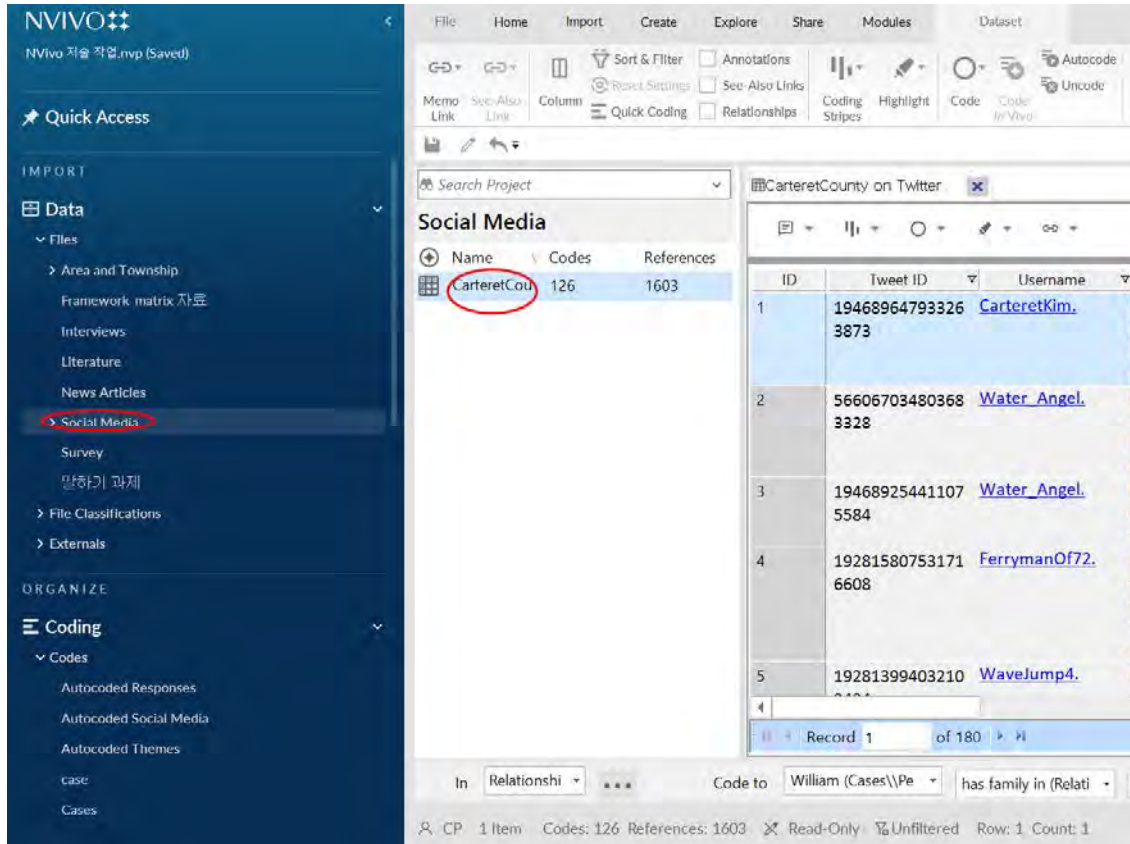
As shown below, the relationship between Susan and Robert appears in a disconnected form.



At the bottom of the display, you can adjust the Vertex's size, color, filter, and scale edge weight.

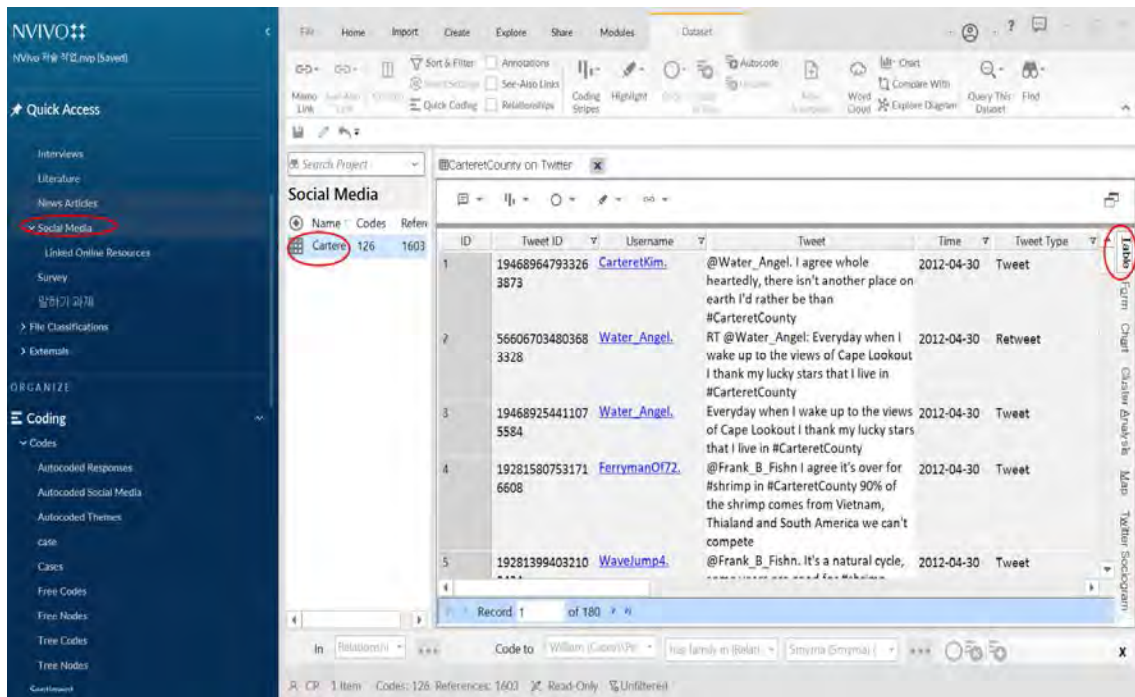
3. Twitter

File > Social Media > CarteretCounty Click

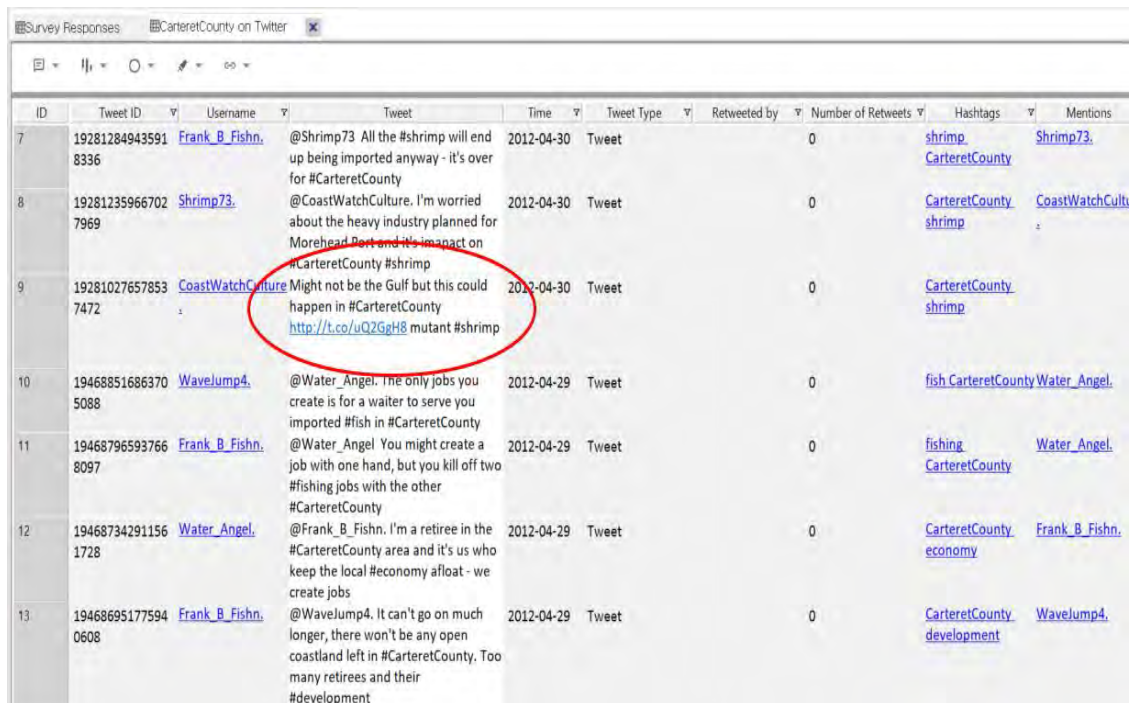


NVivo R1 can import social media data types such as Twitter, Facebook, and Youtube and create relationships between cases. In this way, it is possible to create a sociogram and obtain network metrics. Here, let's open the sample project's Twitter material and see what a typical set of data is like and find out each different perspective. NVivo R1 has charts, clustering, and maps, and I am going to add a Twitter sociogram to this, and I am going to consider the Tweeter set as a sort of sociogram.

Social Media > CartretCounty Click > Table



Each Tweet has several components: row and column. You can code Tweet contents.



To analyze unique ideas with Tweets, I select the Tweet material and code different topics or concepts. Code is a basket containing qualitative data that collects data related to the source, post, or Twitter.

ID	Tweet ID	Username	Tweet	Time	Tweet Type	Retweeted by	Number of Retweets	Hashtags
7	192812849435918336	Frank_B_Fishn.	@Shrimp73 All the #shrimp will end up being imported anyway - it's over for #CarteretCounty	2012-04-30	Tweet		0	shrimp CarteretCounty
8	192812359667027969	Shrimp73.	@CoastWatchCulture. I'm worried about the heavy industry planned for Morehead Port and it's impact on #CarteretCounty #shrimp	2012-04-30	Tweet		0	CarteretCounty shrimp
9	192810276578537472	CoastWatchCulture	Might not be the Gulf but this could happen in #CarteretCounty http://t.co/uQ2GgH8 mutant #shrimp	2012-04-30	Tweet		0	CarteretCounty shrimp
10	194688516863705088	WaveJump4.	@Water_Angel. The only jobs you create is for a waiter to serve you imported #fish in #CarteretCounty	2012-04-29	Tweet		0	fish CarteretCounty
11	194687965937668097	Frank_B_Fishn.	@Water_Angel You might create a job with one hand, but you kill off two #fishing jobs with the other #CarteretCounty	2012-04-29	Tweet		0	fishing CarteretCounty
12	194687342911561728	Water_Angel.	@Frank_B_Fishn. I'm a retiree in the #CarteretCounty area and it's us who keep the local #economy afloat - we create jobs	2012-04-29	Tweet		0	CarteretCounty economy
13	194686951775940608	Frank_B_Fishn.	@WaveJump4. It can't go on much longer, there won't be any open coastland left in #CarteretCounty. Too many retirees and their #development	2012-04-29	Tweet		0	CarteretCounty development

Inverse triangle click on the top left zero > Code Selection Click,

ID	Tweet ID	Username	Tweet	Time	Tweet Type	Retweeted by	Number of Retweets	Hashtags
7	192812849435918336	Frank_B_Fishn.	@Shrimp73 All the #shrimp will end up being imported anyway - it's over for #CarteretCounty	2012-04-30	Tweet		0	shrimp CarteretCounty
8	192812359667027969	Shrimp73.	@CoastWatchCulture. I'm worried about the heavy industry planned for Morehead Port and it's impact on #CarteretCounty #shrimp	2012-04-30	Tweet		0	CarteretCounty shrimp
9	192810276578537472	CoastWatchCulture	Might not be the Gulf but this could happen in #CarteretCounty http://t.co/uQ2GgH8 mutant #shrimp	2012-04-30	Tweet		0	CarteretCounty shrimp
10	194688516863705088	WaveJump4.	@Water_Angel. The only jobs you create is for a waiter to serve you imported #fish in #CarteretCounty	2012-04-29	Tweet		0	fish CarteretCounty
11	194687965937668097	Frank_B_Fishn.	@Water_Angel You might create a job with one hand, but you kill off two #fishing jobs with the other #CarteretCounty	2012-04-29	Tweet		0	fishing CarteretCounty
12	194687342911561728	Water_Angel.	@Frank_B_Fishn. I'm a retiree in the #CarteretCounty area and it's us who keep the local #economy afloat - we create jobs	2012-04-29	Tweet		0	CarteretCounty economy

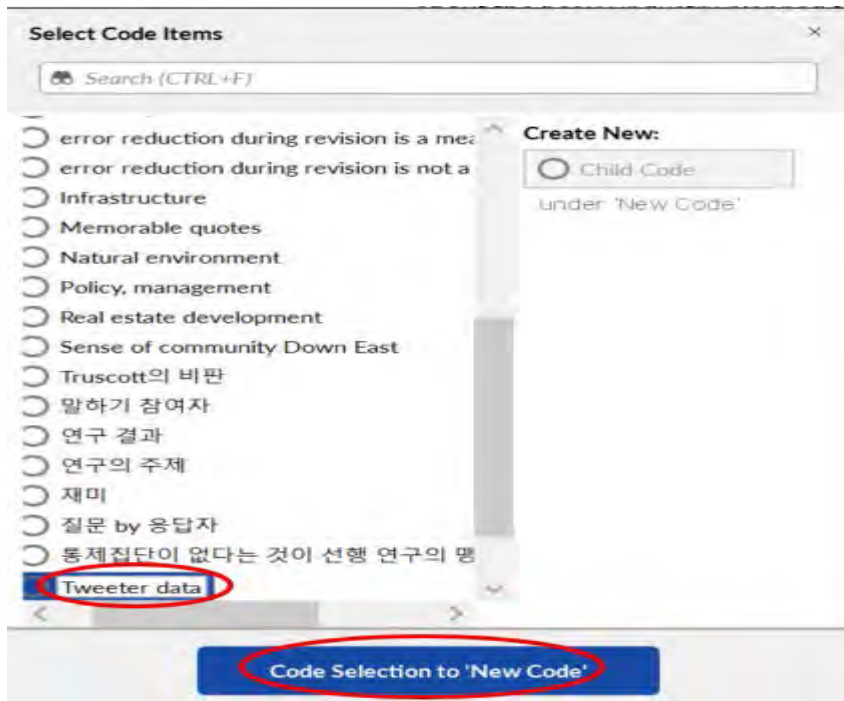
or next to the appropriate data > Right-click on the mouse > Code Selection
Select

ID	Tweet ID	Username	Tweet	Time	Tweet Type	Retweeted by	Number of Retweets	Hashtags
7	192812849435918336	Frank_B_Fishn.	@Shrimp73 All the #shrimp will end up being imported anyway - it's over for #CarteretCounty	2012-04-30	Tweet		0	#shrimp #CarteretCounty
8	192812359667027969	Shrimp73.	@CoastWatchCulture. I'm worried about the heavy industry planned for Morehead Port and it's impact on #CarteretCounty #shrimp	2012-04-30	Tweet		0	#CarteretCounty #shrimp
9	192810276578537472	CoastWatchCulture	Might not be the Gulf but this could happen in #CarteretCounty http://t.co/uQ...	2012-04-30	Tweet		0	#CarteretCounty #shrimp
10	194688516863705088	WaveJump4.	@Water_Angel create is for a v imported #fish		weet		0	fish #CarteretCounty
11	194687965937668097	Frank_B_Fishn.	@Water_Angel job with one h #fishing jobs w #CarteretCoun		weet		0	fishing #CarteretCounty
12	194687342911561728	Water_Angel.	@Frank_B_Fish #CarteretCoun keep the local f create jobs		weet		0	#CarteretCounty #economy
13	194686951775940608	Frank_B_Fishn.	@WaveJump4 longer, there w coastland left i many retirees i #development		weet		0	#CarteretCounty #development

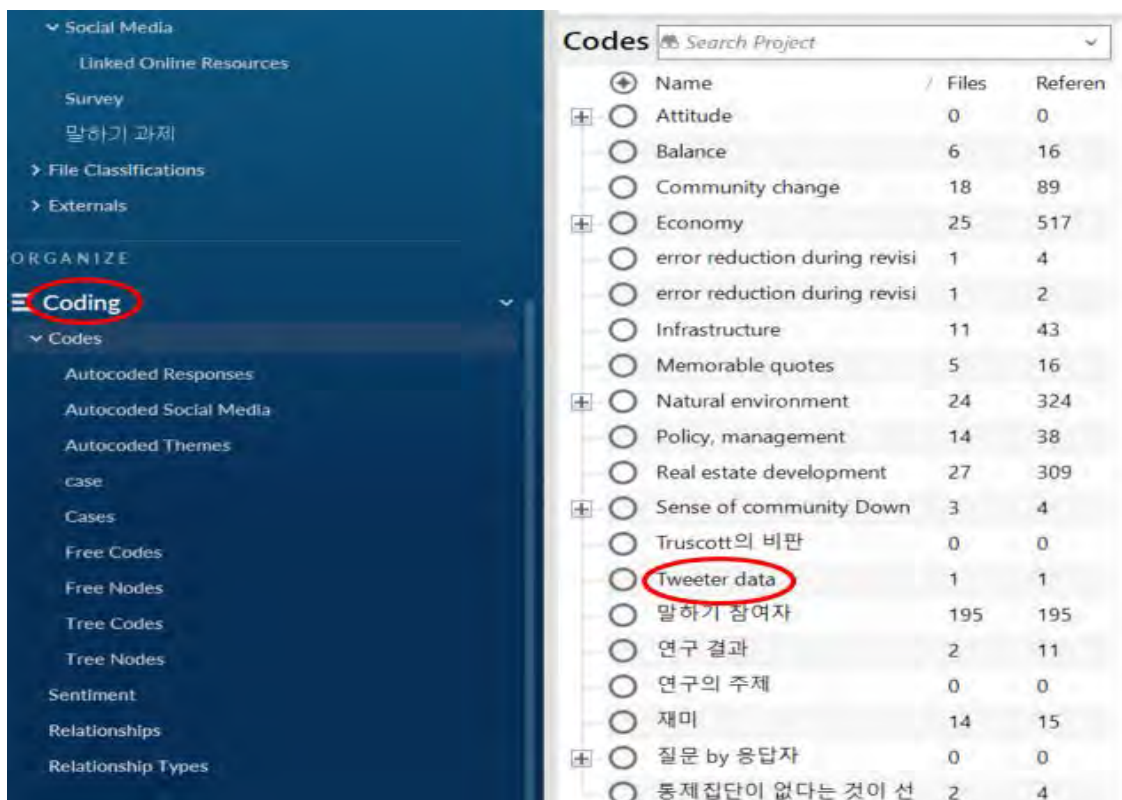
Top-Level Code Click

The image shows the 'Select Code Items' dialog box in NVivo. It features a search bar at the top with the placeholder text 'Search (CTRL+F)'. Below the search bar, a tree view shows the 'Codes' folder expanded. Under 'Codes', several sub-folders are listed, including 'Autocoded Responses', 'Autocoded Social Media', 'Autocoded Themes', 'case', 'Cases', 'Free Codes', 'Free Nodes', 'Tree Codes', and 'Tree Nodes'. At the bottom of the tree view, several radio buttons are visible, including 'Attitude', 'Balance', 'Community change', 'Economy', and 'error reduction during revision i'. To the right of the tree view, a 'Create New:' section contains a radio button next to 'Top-Level Code', which is circled in red. At the bottom of the dialog box, there is a 'Code Selection' button.

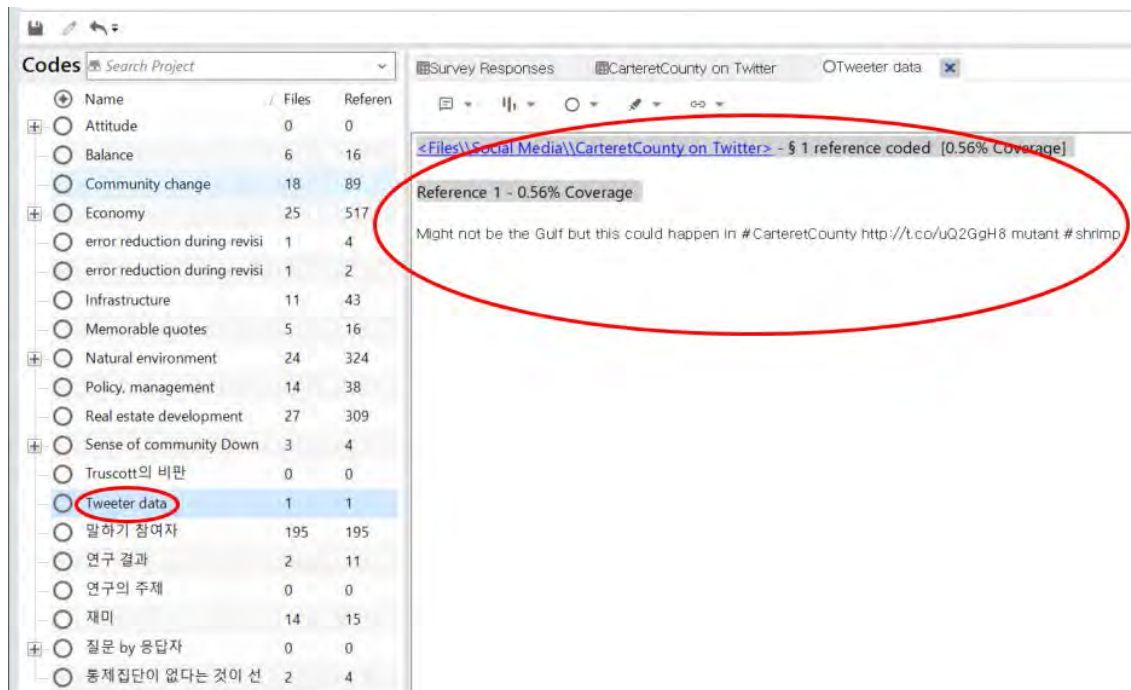
Twitter data Type > Code Selection to 'New Code' Click



Coded content can be found in Codes.



Codes > Tweeter data Click



Click on the link in the above material to view the relevant materials and videos.

The screenshot shows a news article from Al Jazeera. The main headline is 'Gulf Seafood Deformities Raise Questions Among Scientists And Fisherman (VIDEO)'. Below the headline is a video player showing a close-up of a deformed shrimp being held by a person. The article text below the video reads: 'While the true extent of the 1989 Exxon Valdez oil spill was not known for about 4 years, as Al Jazeera notes in the video above, the repercussions of BP's 2010 Deepwater Horizon spill in the Gulf of Mexico may become apparent more quickly.' To the right of the article is a sidebar with related news items, including '눈 노화, 10일 선명한 눈' and 'Republica Americans In Scathin'.



1. 서론

1. COVID-19 발발에 따른 비대면 수업으로의 전환
2. 실제 교육환경에 COVID-19이 미치는 영향에 대한 연구 필요
3. 온라인 수업환경과 수업운영에 관한 교육의 주체인 학생들의 반응과 인식이 추후 비대면 강의 운영에 미칠 영향에 대한 분석 필요

III. 선행연구

- Abbad & Jaber(2014): 바레인 대학생들의 온라인 학습시스템에 대한 인식 조사를 수행, 학생들은 온라인 학습시스템에 대해 대체로 긍정적인 인식을 지니 것으로 나타났다.
- Fedynich, Bradley & Bradley(2015): 미국 대학에서 249명의 대학생들을 대상으로 온라인 학습에 대한 인식을 조사한 결과, 학생들과 교수자 사이의 상호작용이 수업 만족도에 가장 큰 영향을 끼치며, 교수자 역할의 중요성
- Harry & Newline(2006): 시간제(part-time) 대학생인 217명을 대상으로, 50% 이상의 학생들이 대면수업의 질이 높고, 시간을 더욱 균형 있게 사용할 수 있다는 긍정적인 인식을 보임.
- Dobba, Waid, del Carmen(2009): 대면 수업을 신청한 대학생 180명+ 온라인 수업을 신청한 100명의 학생 대상. 학생들 모두 대면 수업이 온라인 수업보다 학습이 쉬우며, 대면 수업 시 더 많은 내용을 습득한다고 응답.

III. 선행연구

- 이은진(2013): 115명의 한국 대학생 대상 온라인 수업에 대한 인식 조사. '사회적 실재감의 부족', '학습 부담' 과 '교수자와의 제한적 상호작용' 등에 대한 불안 수준이 낮은 집단인 수업에 대한 전반적인 만족도가 높음.
- 이쌍철, 김정아(2018): 학습자 요인 중 수강 이유 및 자발성, 정기적 수강 그리고 교수자 요인 중 교과 교사 및 관리교사의 지원, 시스템 관리자 등이 온라인 수업의 만족에 유의미한 영향을 미침.

VI. 연구 방법: 연구참여자

1. 연구 참여자: 부산과 안동 소재 2개의 대학교 초·중등 영어 교육과 학부생 110명

<표 1> 연구 참여자 정보

수강강좌	학년	참여자 수(N=110)	
		남	여
영어 쓰기	1학년	7	12
제2외국어 습득론	2학년	19	27
영어 교재연구 및 방법론	3학년	4	11
영어교과논리및논술	4학년	9	21
합계	110명	39명	71명

VI. 연구 방법: 연구 도구 및 절차



설문지조사: 비대면 온라인 강의 이후 온라인 수업에 영향을 미치는 주요 요인들에 대한 폐쇄형 문항(26문항)과 온라인수업 진행에 대한 의견피력과 개선방안을 조사하는 개방형 문항들(8문항).



세부 문항: 강의 전반에 대한 문항(1-4번), 학습자 요인(5-9번), 교수자 요인(10-16번), 시스템 요인(17-19번), 영어학습 효과(19-26번)로 구성



이상철, 김정안(2018)의 연구에서 사용된 검사 도구를 참조하여 본 연구의 목적에 맞게 수정 및 활용



5점 척도의 Likert 방식

VI. 연구 방법: 분석방법

- google-forms를 통해 수집된 각 문항의 측정치들을 엑셀 프로그램으로 전
환
- 연구질문 1, 2, 3: SPSS를 활용하여 먼저 요인 분석(Factor
Analysis)을 통해 총 26문항에서 공통적인 요인들을 분석한 후, 상관관계 분
석과 일원 배치 분산분석(one-way ANOVA)을 시행, 사후검사는 학
년 간 평균의 차이가 통계적으로 유의 할 경우, Tukey를 사용연구
- 질문 4: 개방형 질문들에 대한 학생 교사들의 응답을 질적으로 분석

V. 결과 및 논의: 요인분석

- 요인분석 결과: 학생자 요인, 교수자
요인, 시스템 요인, 학생효과 요인 도출

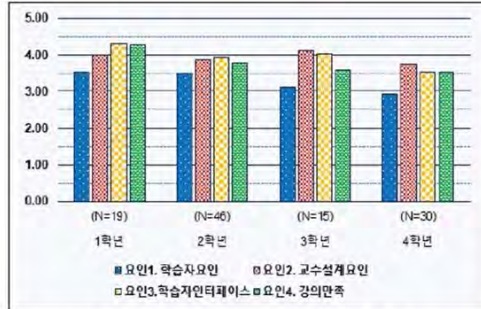
<표 4> 설문조사 분석 결과 (N=110)

요인	#	문항	M	SD
요인1 학생자 요인 (14문항)	3	나는 온라인 수업을 선호한다.	3.19	1.252
	4	나는 온라인 수업을 통한 학습 성취에 만족한다.	3.18	1.119
	5	나는 온라인 수업에 적극 참여한다.	3.86	.829
	6	비대면 수업은 학습 동기에 도움이 된다.	2.83	1.000
	7	비대면 수업은 자기 주도적 학습에 도움이 된다.	3.35	1.230
	8	비대면 수업은 학습의 집중에 도움이 된다.	3.02	1.188
	9	비대면 수업은 다른 학생들과의 상호작용에 도움이 된다.	2.20	1.021
요인2 교수자 요인 (7문항)	20	비대면 수업은 영어 읽기 학습에 도움이 된다.	3.30	1.036
	21	비대면 수업은 영어 듣기 학습에 도움이 된다.	3.41	1.069
	22	비대면 수업은 영어 말하기 학습에 도움이 된다.	2.83	1.156
	23	비대면 수업은 영어 쓰기 학습에 도움이 된다.	3.32	1.108
	24	비대면 수업은 전반적인 영어 실력 향상에 도움이 된다.	3.25	1.033
	25	비대면 수업은 영어교과 내용이해에 도움이 된다.	3.50	1.064
	26	비대면 수업은 영어교과 흥미도에 도움이 된다.	3.32	1.065
요인3 시스템 요인 (3문항)	10	교수의 수업 요구 사항이 명확하다.	3.83	.876
	11	교수는 학습자에게 적절한 정보와 피드백을 제공한다.	3.87	.990
	12	교수는 학생들의 지원요청에 대한 빠른 반응을 한다.	4.01	.994
	13	교수는 의사소통과 상호작용에 적극적이다.	3.96	.917
	14	교수는 과제에 대한 적절한 피드백을 제공한다.	3.83	1.003
	15	교수는 비대면 수업을 통해 교수의 전문성을 잘 전달한다.	3.93	.885
	16	교수의 비대면 수업의 수업 내용 전달은 이해하기 쉽다.	3.78	.913
요인4 학습효과 요인 (2문항)	17	비대면 수업의 시스템에 대한 학생들의 접근이 쉽다.	3.75	1.065
	18	비대면 수업의 시스템에 대한 학생들의 편의성이 높다.	3.85	1.082
	19	비대면 수업을 위한 환경(컴퓨터, 인터넷 등)이 잘 준비되어 있다.	4.08	.940
요인5 학습효과 요인 (2문항)	1	비대면 영어관련 수업 강의를 전반적으로 나는 만족한다.	3.76	.877
	2	나는 비대면 영어과목 수업의 질에 대해 만족한다.	3.79	.899

V. 결과 및 논의: 학년별 주요 요인들에 대한 인식 차이

<표 5> 학년별 주요 요인들에 대한 인식 차이 분석 결과(N=110)

	1학년	2학년	3학년	4학년	F	Sig.	사후 분석
	(N=19) M(SD)	(N=46) M(SD)	(N=15) M(SD)	(N=30) M(SD)			
요인1. 학습자 요인 (14문항)	3.52 (.755)	3.50 (.891)	3.13 (.855)	2.93 (.931)	1.855	.142	--
요인2. 교수자 요인 (7문항)	4.00 (.680)	3.86 (.797)	4.11 (.630)	3.74 (.902)	.903	.443	--
요인3. 시스템 요인 (3문항)	4.30 (.542)	3.92 (.946)	4.02 (.885)	3.54 (1.060)	2.801*	.044	1학년> 4학년
요인4. 학습효과 (2문항)	4.29 (.673)	3.79 (.827)	3.60 (.687)	3.52 (.978)	3.619*	.016	1학년> 4학년



<그림 1> 학년별 요인별 중요도 인식 차이

V. 결과 및 논의: 학생 교사들의 콘텐츠 선호도

<표 6> 콘텐츠 선호도

종류	1학년		2학년		3학년		4학년		전체	
	N	%	N	%	N	%	N	%	N	%
직접제작 영상	13	68.5	30	65.2	10	66.7	16	53.3	69	62.7
실시간 강의	2	10.55	8	17.4	1	6.7	14	46.7	25	22.7
과제 실행	2	10.55	5	10.9	2	13.3	0	0	9	8.2
유튜브 /동영상	2	10.55	3	6.5	2	13.3	0	0	7	6.4
합계	19	100	46	100	15	100	30	100	110	100

<표 7> 콘텐츠 선호 이유

콘텐츠 종류	선호도 이유
1. 교수자가 제작한 수업 영상	-시간 제약에서 벗어난다. -복습이 가능하다. -자기 일정에 따라 유동적으로 계획할 수 있고 편하다. -이해되지 않는 부분을 다시 들을 수 있다. -대면수업과 가장 유사하게 들을 수 있다. -자기 주도적 학습이 가능하다. -저장할 수 있고 효율적이다.
2. 실시간 강의	-그나마 집중을 할 수 있다. -교수자의 설명을 따라가기 쉽다. -실시간으로 소통할 수 있다. -실시간 피드백이 가능하다. -강제성으로 학습 동기와 규칙적인 생활에 도움이 된다. -현장감이 있다. -상호작용이 이루어질 수 있다. -가장 학교의 현장과 유사하다.
3. 과제 수행	-시간과 관계없이 학습할 수 있다. -과제를 수행하면 반강제적으로 강의를 듣게 된다. -규칙적인 학습이 가능하다. -자기 주도적 학습이 가능하다. -편리하다.
4. 유튜브 등 동영상	-딱딱한 교과의 내용보다 유튜브 영상을 통해 이해를 높인다. -내용이 재미있고 더욱 집중하게 한다.

V. 결과 및 논의: 비대면 온라인 수업의 장·단점 및 개선 방향

<표 8> 비대면 수업의 장·단점

	비대면 수업의 장점	비대면 수업의 단점 및 문제점
학습자유인	-녹음이 가능함 -이해하지 못하는 부분을 재학습 가능 -자기 주도적 학습이 가능함 -수업에 대한 부담이 적음 -개인에게 유리한 시간대에 학습 가능 -효율적인 시간 관리 가능함 -편안한 복장 편한 자세로 수업 가능 -이동시간이 없기에 개인적인 학습 시간을 많이 확보할 수 있음	-집중력 저하 -나태해짐 -동기부여가 떨어짐 -흥미 유발이 되지 않음 -학생들과의 상호작용이 힘들고 적응 -수업 마감 시간 지키기 어려움 -수업시간을 놓침 -공부하는 경소의 문제점 -학습효과 떨어짐
시스템요인	-컴퓨터만 있으면 학습할 수 있음 -인터넷에 접속만으로 출석가능 -결근이 영구적임 -시간과 공간의 제약울 덜 받음	-시스템 결근이 힘들다 (예상치 못한 오류, 와이파이 문제, 팀즈 프로그램 오류, 프로그램에 익숙지 않음, 진행속도가 느림, 인터넷 연결) -실시간 수업 시 자료 가독성이 떨어짐 -수업 진행이 힘들 -실시간 수업 상 적극적으로 참여하는 학생도 드물 -의사소통이 원활하지 않음 -조별 과제, 그룹 활동 시 소통이 어려움 -의사소통과 전달력이 낮음 -피드백이 느리고 원활하지 않음 -과제 양이 많음(수업과 과제를 동시에 수행하기 힘들) -과제를 하기 위한 학습
교수자요인		
기타	감염의 우려로부터 예방될 수 있음	

<표 9> 비대면 온라인 강의 개선을 위한 제안 사항

#	제안 방안
1	적절한 과제의 양
2	무리한 그룹 활동 및 조별 과제 지양
3	온라인 수업의 방식에 대한 다양성 추구
4	원활한 피드백과 상호작용을 할 수 있는 수업 방안 모색
5	실시간 원격수업 진행
6	시험과 평가에 대한 체계적이고 다각적인 구성 및 진행
7	전자 칠판을 활용한 시각적인 학습 자료 제시로 강의의 질과 학생의 이해도 향상
8	토론식 수업을 원활하게 진행하기 위한 프로그램 모색

VI. 결론 및 시사점

■ 연구 질문 1 & 2: 학습자 요인, 교수자 요인, 시스템 요인, 영어학습 효과 등 4가지 요인들이 도출된 요인 모형에서의 상대적 중요도를 나타냄. 세부적으로 학습자 요인에는 적극적으로 참여도, 자기 주도적 학습이 교수자 요인에는 교수의 빠른 피드백, 시스템 요인으로는 시스템 환경(저속 응답성, 편의성, 사용 편리성) 등이 온라인 수업에 영향을 미치는 주요인들로 파악

■ 연구 질문 3: 온라인 수업에서 중요하게 느끼는 요인들을 한꺼번에 분석한 결과, 시스템 요인과 학습효과 부분에서 통계학적 차이를 보이고 그하녀으로 올라갈수록 학습 시스템과 전체적인 강의 만족도 하모에서 나온 퍼그인 보이는데, 이는 저학년 학생들보다 고학년 학생들이 온라인 강의 방식의 한계를 느끼고 학습효과에 대한 부정적이 이식이 강화되고 있음을 보여주는 결과

VI. 결론 및 시사점

- 연구지점 4: 온라인 수업의 개선방안으로는 과제량의 조정, 조별 활동과 조별 과제 활동 지양, 전자 칠판 등을 활용한 시각적 학습 자료의 활용으로 강의 전달력과 학습효과 향상, 교수자의 즉각적인 피드백과 상호작용 개선, 시험 및 평가에 대한 체계적이고 다각적인 접근이 필요하다는 방안이 제시



Zoom 1

초청특강

<https://pusan.zoom.us/j/83772894412?pwd=UnozVzILYzEzRDF1TTdoc0laWHZxQT09>

회의 ID: 837 7289 4412

암호: 123456

인공지능 시대에 언어학의 발전 방안

- 한국언어과학회 2021년 동계 학술대회 -

2021년 2월 18일(목)

정덕길(동의대학교 산업/CT기술공학 전공)

E-mail : dgjung@deu.ac.kr

전화(핸드폰) : 010-6280-6563 전화(연구실) : 051-890-2351

언어 공학(대학 교육의 변화)

"변해야 산다"...4차산업혁명 '간판' 바꾼 대학들, 논쟁 사이언스인(2017), <https://www.yna.co.kr/view/AKR20170701057700004>

한국외대 영어학과 → ELLT학과

- 한국외대는 2018년부터 '영어학과'의 명칭을 'ELLT학과'(English Linguistics & Language Technology)로 변경

- 시용 영어교육과 이론 영어학 중심으로 이뤄진 교육과정엔 언어공학 분야를 도입해 인문학적 어문학 교육의 한계 극복

- 데이터가 핵심인 언어학의 연구방법은 4차산업의 화두인 인공지능과 연계

▪ 학생들의 IT 기업 취업 기회 확대

경희대는 소프트웨어융합 단과대 신설

- 단순히 학과명만 바꿀 게 아니라 컴퓨터 공학으로 분류되는 언어공학(컴퓨터 언어 처리법 등)을 배우는 교육과정 추가

성균관대학

아주경제(윤상민, 2019), <https://www.ajunews.com/view/20190901134835846>

- AI는 미래 언어 : 전공 언어 동시에 구사하는 이중 언어자

▪ MIT·스탠퍼드대 벤치마킹에 AI교육 대중화에 주력

▪ 어문학과는 언어학에 AI 도입하고 예술 분야도 SW·IT 기술 접목에 응함

정보 기반 인문언어학

언어정보연구(연세대) 인문학컴퓨터사업단

<https://ils.yonsei.ac.kr/%EC%B2%B0%ED%95%9B%EC%97%B0%EA%B5%AC%EC%B4%B0%ED%95%9C%EA%B5%AD%EC%B2%AC%EC%97%B5%EB%B8%AB/>

인문학컴퓨터사업단

- 4차 산업혁명 시대의 정보 기반 인문언어학을 선도

- 대규모 말뭉치와 아카이브, 빅데이터를 기반으로 언어정보 구축과 활용, 언어 습득과 교육, 번역과 언어 대조 분석이라는 세 가지의 전문적, 세부적 주제를 설정

- 인공지능 비서, 인공지능 스피커, 챗봇 등의 애플리케이션이 대중화되면서 다양한 사용역(register)의 언어 자료 연구 필요성

- 학사자 언어 빅데이터 기반의 언어 교수 및 학습 연구를 중앙적으로 하여 한국어 및 영어 학사자의 중간어어를 연구하고 한국어와 영어 자동 문형 패턴 분석 도구 및 자동 평가 시스템 개발을 위한 기초적 자원을 개발하고 보급함으로써 한국어와 영어 교사의 양적 질적 제고

- (1단계) 빅데이터 및 데이터베이스 구축을 위한 기초 연구를 정리하고 실제로 구축, (2단계) 챗봇 개발, 어휘망 검색서비스를 제공하고 자동 분석과 평가 도구 등을 실제 교육 현장에 적용

전산언어학의 방법론과 응용분야

전산언어학, http://www.aistudy.com/linguistics/computational_hausser.htm#_bookmark_190bb78

전산언어학의 기초 : Roland Hausser 지음, 장석진, 이기용, 홍은, 최기성, 이민영, 윤애선, 민미선, 류병택, 윤광의 옮김, 한국문화사, 2002 (원서 : Foundations of Computational Linguistics, Springer, 2001), Page 11~28

전산언어학의 여러 가지 실용적 과제

- 텍스트 데이터베이스의 색인과 검색
- 기계번역
- 자동 텍스트 생성
- 자동 텍스트 검사 : 철자검사기, 통사구조 검사기
- 자동 내용분석
- 자동 학습 보조 시스템
- 자동 대화 정보 시스템 : 열차 운행시간에 대한 자동 안내서비스, 의료 데이터베이스를 기초로 한 질의와 저장, 자동 세금상담 등

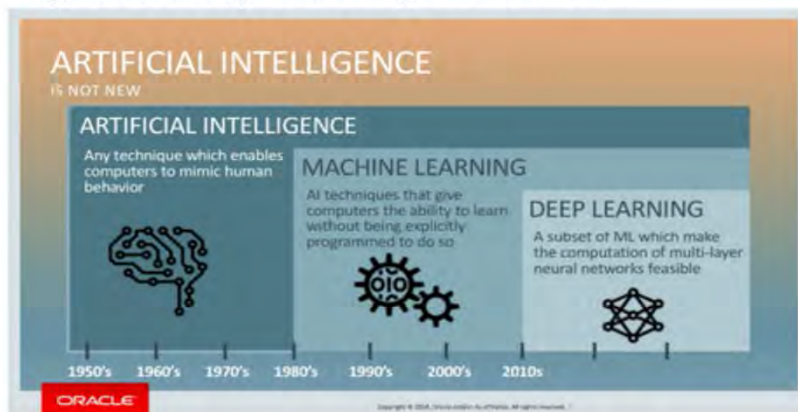
인공지능의 4가지 유형

[레벨1] 단순한 제어 프로그램 단순한 제어 프로그램 탑재 제품을 마케팅적으로 '인공지능 탑재'라 광고 (세탁/청소)	[레벨2] 패턴이 다양한 고전적 인공지능 입력과 출력 관계를 맺는 방법의 수가 극단적으로 많고 세련된 경우 (탐색/추론, 지식베이스 활용)
[레벨3] 머신러닝을 받아들인 인공지능 데이터를 바탕으로 학습되는 기계학습 알고리즘 이용 (레벨2+기계학습→레벨3)	[레벨4] 딥러닝을 받아들인 인공지능 기계학습을 할 때 입력값의 특징 (feature)을 사람이 입력하지 않고 기계가 직접 학습

[출처: 인공지능과 딥러닝, 마쓰오 유카카 (박기연 옮김), 동아엠앤비(2015)]

What's the Difference Between AI, Machine Learning, and Deep Learning?

Peter Jeffcock(2018)



[출처: <https://blogs.oracle.com/bigdata/difference-ai-machine-learning-deep-learning>]

AI, ML, DL

What Is The Difference Between Artificial Intelligence And Machine Learning?
(Bernard Marr , 2016)

Machine Learning has certainly been seized as an opportunity by marketers. After AI has been around for so long, it's possible that it started to be seen as something that's in some way "old hat" even before its potential has ever truly been achieved. There have been a few false starts along the road to the "AI revolution", and the term Machine Learning certainly gives marketers something new, shiny and, importantly, firmly grounded in the here-and-now, to offer.

[출처]: <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#339d6852742b>

What Is The Difference Between Deep Learning, Machine Learning and AI?
(Bernard Marr , 2016)

Deep Learning is used by Google in its voice and image recognition algorithms, by Netflix and Amazon to decide what you want to watch or buy next, and by researchers at MIT to predict the future. The ever-growing industry which has established itself to sell these tools is always keen to talk about how revolutionary this all is. But what exactly is it? And is it just another fad being used to push "old fashioned" AI on us, under a sexy new label?

[출처]: <https://www.forbes.com/sites/bernardmarr/2016/12/08/what-is-the-difference-between-deep-learning-machine-learning-and-ai/#1667db626cfa>

Artificial Intelligence vs. Machine Learning vs. Deep Learning(1/3)

Artem Opperma(2019)

[출처]: <https://www.deeplearning-academy.com/p/ai-wiki-machine-learning-vs-deep-learning>

Machine Learning incorporates "classical" algorithms for various kinds of tasks such as clustering, regression or classification. **Machine Learning algorithms must be trained on data.** The more data you provide to your algorithm, the better it gets.

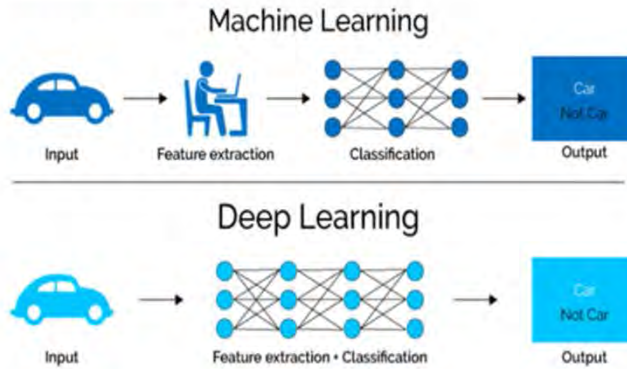
Machine Learning is a **pretty old field** and incorporates methods and algorithms that have been **around for dozens of years**, some of them since as early as the sixties.

Artificial Intelligence vs. Machine Learning vs. Deep Learning(2/3)

Why is Deep Learning better than Machine Learning?

- Feature Extraction

The first advantage of Deep Learning over machine learning is the needlessness of the so-called Feature Extraction.



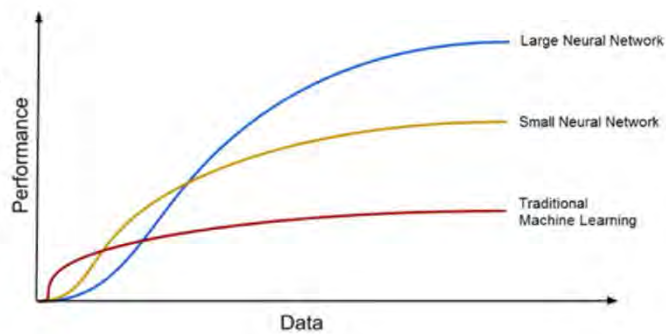
Artificial Intelligence vs. Machine Learning vs. Deep Learning(3/3)

- The Era of Big Data

The "Big Data Era" of technology will provide huge amounts of opportunities for new innovations in deep learning.

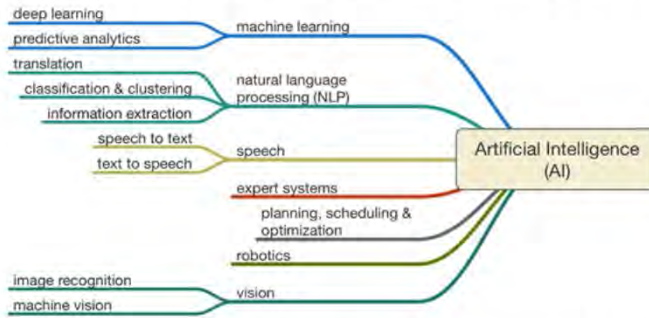
As per Andrew Ng, the chief scientist of China's major search engine Baidu and one of the leaders of the Google Brain Project,

"The analogy to deep learning is that the rocket engine is the deep learning models and the fuel is the huge amounts of data we can feed to these algorithms."



A.I. ARCHITECTURE INTELLIGENCE

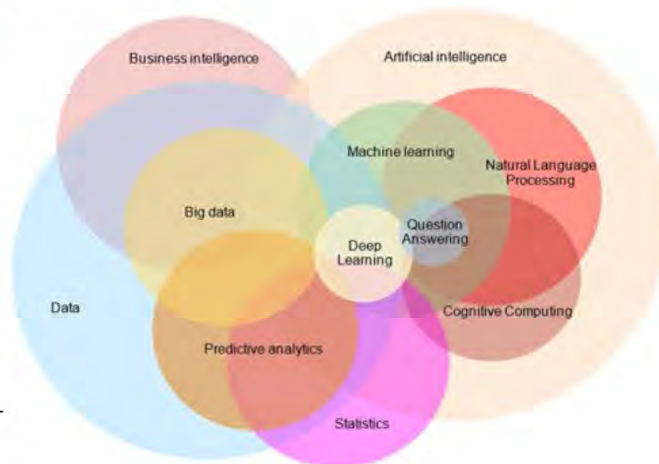
What's required for a machine to be intelligent



[참조]: Architecture and Urban Planning in the age of Artificial Intelligence in an article by Ron Beqiri(2016), <http://futurearchitectureplatform.org/news/28/ai-architecture-intelligence/>

ARTIFICIAL INTELLIGENCE TO DIGEST BIG DATA

AI Venn Diagram – one take on classification of AI and related terminology according to the Big Data Dictionary



[참조]: Why Finance professionals need to know about AI—and where they can learn, Paul Bard(2017)

<https://www.ibm.com/blogs/business-analytics/why-finance-professionals-need-artificial-intelligence/>

자연어처리 기반의 인공지능 기술 현황(1/3)

최미환(2017), TTA Journal Vol.169, 2017

- 음성 인터페이스 기술

1) 자연어 음성 인터페이스 기술

- 음성인식 기술 : 자연어 음성모델링 기술, 의미기반 언어모델링 기술, 환경에 가변하는 음성인식 기술, 핵심어 추출 기술 등
- 언어 이력 서비스 : 자유 발화를 인식하여 대화형태의 언어 학습 시스템을 이용해 영어 사교육비 문제 및 다문화 가족의 언어소통 문제 해결 기대

2) 자연어 대화처리 기술

- 입력 대화체 문장에 대해 발화자의 의도를 이해하기 위한 대화 이해 기술, 자연스러운 대화 응답을 위한 대화모델링 기술, 외국어교육 등에 적용할 수 있는 대화 오류 수정 기술 등
- 대화처리 기술을 외국어 교육에 접목하여 자동 교육 방식을 제공하려는 연구가 활발히 진행 중이며 향후 대화식 외국어 교육 피드백에 대한 연구 필요

자연어처리 기반의 인공지능 기술 현황(2/3)

자연어 음성인식 기술을 이용한 외국어 학습서비스



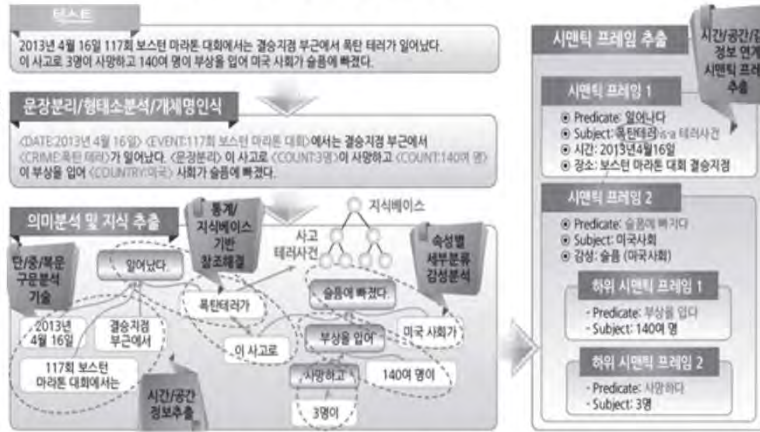
자동통역 기술

- 다국어 언어분석 기술
 - 다국어 문법 문장에 대해 컴퓨터가 이해할 수 있는 단위로 언어학적으로 분석하는 기술
 - 명사 분석, 구문 분석 및 의미 분석 기술은 자연어 처리의 기반 기술
 - 규칙 기반 방법과 통계 기반 방법으로 구분
- 지식추출 및 학습 기술
 - 언어 분석과 생성을 위한 지식추출
 - 자동번역을 위한 단어 및 구 단위 번역 지식학습
- 하이브리드 자동번역 기술
 - 복수개의 자동번역 엔진에서 최적의 번역결과를 선택하는 기술
- 다국어 음성 자동통역 기술
 - 사람의 말을 인식하고 다국어로 불역하는 기술
 - 텍스트, 제스처, 음성 등을 활용하는 멀티모달 인터페이스 기술과 연계 필요

자연어처리 기반의 인공지능 기술 현황(3/3)

자연어 질의응답 기술

- 자연어 질의응답 및 이해 기술
- 정답 후보 생성 기술
- 언어이맥을 위한 구조 의미 결합 모델 기술
- 개방형 지식 추출 기술
- 지식베이스 연동 기술



< 자연어 질의응답을 위한 자연어처리의 예 >

국내 NLP 시장 현황(1/2)

출처: 자연어처리(NLP) 무엇인가... 그 기술과 시장은, 최창현, 인공지능신문(2020), <http://www.aitime.kr/news/articleView.html?idxno=15036>

말뭉치(말뭉들, 글뭉들) 데이터

- 일본: 150억 어절
- 중국: 300~800억 어절
- 미국: 300억 어절 구축

21세기 세종 계획

- 1차 세종계획(1998 ~ 2007)
세종 문어 말뭉치(약 3700만 어절), 형태 분석 말뭉치(약 1000만 어절), 의미 분석 말뭉치(1,000만 어절), 구문 분석 말뭉치(약 45만 어절), 워시 말뭉치(80만), 형태 분석 말뭉치(80만 어절)
- 2차 세종계획(2018 ~ 2022, 총 175억 원): 155억 어절 규모의 말뭉치 구축 계획

한국어 언어모델 코BERT(KorBERT)

- 과거 저투본의 소프트웨어 분야의 국가 핵심기술 개발영역에 포함
과거에 영서서자도려 프로젝트 엑소브레이크인 사어, 한국자등신언
언(ETRI), 2019)

- (1) 구글의 언어표현 방법을 기반으로 더 많은 한국어 데이터를 넣어 만든 언어모델 (2) 한국어의 고차원 특성을까지 반영해 만든 언어모델

딥러닝 기반 형태소분석기 '카이(khacii)'

- ('카이') 2018년 말부터 오픈소스로 제공
- 딥러닝을 통해 학습한 데이터를 활용해 형태소를 분석하는 모델
- 콘텐츠 신경망(CNN, Convolutional Neural Network)을 이용해 음절기반으로 형태소를 분석하는 방법 채택

국내 NLP 시장 현황(2/2)

SQuAD(The Stanford Question Answering Dataset)

- 국내 딥러닝 기반 QA(Question Answering) 검색 기술을 개발하는 AI 스타트업 포티투마루(대표 김도환)
- (2019년) 스탠포드 대학(Leland Stanford Junior University)에서 주관하는 세계 최대의 기계 독해 경진대회인 SQuAD 2.0에서 스타트업으로는 전 세계 최초로 3위 기록

코퀴드 2.0(KorQuAD 2.0)

- 국내 처음이자 언어 AI 개발 위한 핵심 인공지능 학제 융합 데이터셋 '코퀴드 2.0' 무료 공개
- 한국어 표준데이터를 7만 개에서 10만 개로 확대하고 다단계에서 차원이 답변이 가능한 시퀀스 개발할 수 있도록 데이터셋 강화한 버전

네이버

- 네이버 검색이 자연어 처리 분야에서 검색 이용자의 의도를 더욱 잘 이해하는 검색으로 진화
- 모바일 상에서의 검색이 일상화된 이용자를 위해 AI 기술 기반 검색어 고징 시스템인 'AIQSpell' 개발
- 세계 최고 권위의 자연어처리 분야 학회(EMNLP-IJCNLP 2019)에서, 다국어 읽기 이해도를 위한 제한된 데이터 학습으로 총 6만 4천 개 이상의 단어를 학습한 'AIQSpell'이 1등 기록을 세우며, 이는 자연어 처리 분야 연구자들에게 큰 자극제가 될 것으로 기대된다.

PROCESSING & UNDERSTANDING TEXT(1/4)

[출처]: A Practitioner's Guide to Natural Language Processing (Part I) — Processing & Understanding Text, Dipanjan (DJ) Sarkar(2018)

<https://towardsdatascience.com/a-practitioners-guide-to-natural-language-processing-part-i-processing-understanding-text-9f4abfd13a72>

➤ Text Analytics with Python: A Practitioner's Guide to Natural Language Processing

2nd ed. Edition, Apress(2019), by Dipanjan Sarkar (Author)

- Processing & Understanding Text
- Feature Engineering & Text Representation
- Supervised Learning Models for Text Data
- Unsupervised Learning Models for Text Data

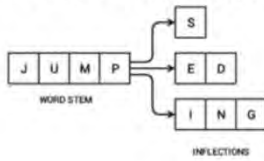
➤ Standard NLP Workflow



PROCESSING & UNDERSTANDING TEXT(2/4)

Stemming

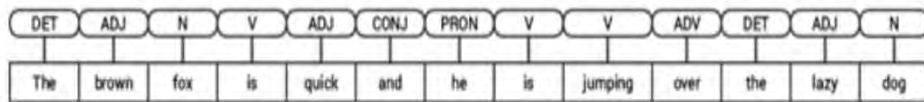
- Word stem and its inflections (Source: Text Analytics with Python, Apress/Springer 2016)



Understanding Language Syntax and Structure

- Sentence : hierarchical structure
 • Sentence → clauses → phrases → words

1) Tagging Parts of Speech : POS tagging for a sentence

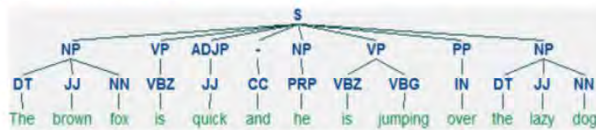


PROCESSING & UNDERSTANDING TEXT(3/4)

Understanding Language Syntax and Structure

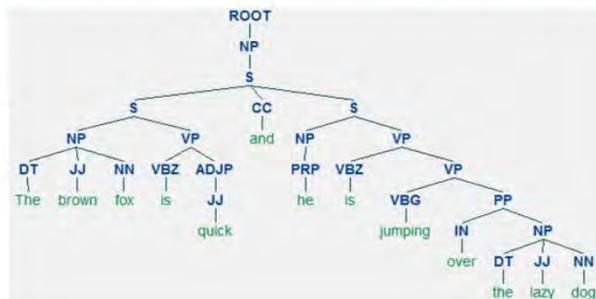
2) Shallow Parsing or Chunking

An example of shallow parsing depicting higher level phrase annotations



3) Constituency Parsing

An example of constituency parsing showing a nested hierarchical structure



PROCESSING & UNDERSTANDING TEXT(4/4)

Named Entity Recognition / Sentiment Analysis with AFINN Lexicon

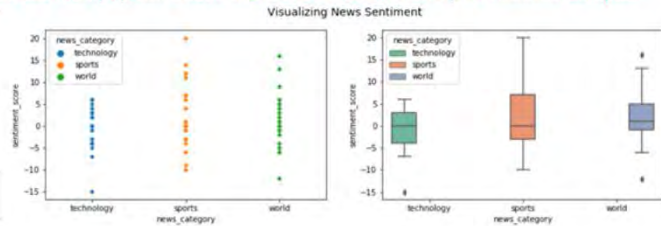
Visualizing named entities in a news article with spaCy

Spacy had two types of English dependency parsers based on what language models you use.

US GPE unveils world's most powerful supercomputer, beats China GPE . The US GPE has unveiled the world's most powerful supercomputer called 'Summit', beating the previous record-holder China GPE 's Sunway TaihuLight one . With a peak performance of 200,000 CARDINAL trillion calculations per second ORDINAL , it is over twice as fast as Sunway TaihuLight one , which is capable of 93,000 CARDINAL trillion calculations per second. Summit has 4,608 CARDINAL servers, which reportedly take up the size of two CARDINAL tennis courts.

The AFINN lexicon is perhaps one of the simplest and most popular lexicons that can be used extensively for sentiment analysis.

```
f, (ax1, ax2) = plt.subplots(1, 2, figsize=(14, 4))
sp = sns.stripplot(x='news_category', y='sentiment_score',
                  hue='news_category', data=df, ax=ax1)
bp = sns.boxplot(x='news_category', y='sentiment_score',
                hue='news_category', data=df, palette='set2', ax=ax2)
t = f.suptitle('Visualizing News Sentiment', fontsize=14)
nlp_strategy_26.py hosted with ❤ by GitHub
```



인공지능 기술과 자연어 처리(1/2)

인공지능(A.I) 기술과 자연어 처리(Natural Language Processing), SD 아카데미 공신님(2018), <https://m.blog.naver.com/sundooedu/221235729489>

자연어 서비스

- 구글**
 - 구글 음성 인식
 - 집주인의 음성을 알아 들어 지시를 수행하는 가정용 스마트 스피커
 - 구글 어시스턴트
 - 사용자 가 대화의 맥락과 내용을 이해한 후 실시간으로 답변을 지어내 주는 머신러닝 기술을 활용
 - 워드랩
 - 카메라로 비추는 언어를 즉시 번역에 주는 앱 서비스

IBM

- 인공지능 플랫폼 왓슨(Watson) API
 - 번역, 문맥 분석, 이미지와 음성을 인식하는 기능을 제공
- 왓슨 개발자 클라우드(Watson Developer Cloud)
 - 언어 구조를 인식하고 애플리케이션으로 통해 음성 명령을 인식할 수 있게 하는 스피치 투 텍스트(Speech to text) 서비스와 텍스트를 음성으로 변환 시켜주는 기능을 머신러닝 기술로 구현

마이크로소프트(MS)

- 코타나(2014년)
 - 인공지능 음성 비서 서비스
- 스키이프 트랜스레이터(Skype Translator)
 - 실시간 언어통역 번역 서비스

인공지능 기술과 자연어 처리(2/2)

자연어 처리 기술과 챗봇

- 챗봇 프레임워크
 - MS사의 봇 프레임워크(Bot Framework)
 - IBM의 Watson conversation
 - 구글의 API.ai
 - 아마존의 렉스(Lex)
- 챗 프로그램
 - 슬랙, MS 메신저
- 음성 AI 비서
 - 애플의 시리(Siri)
 - 아마존의 알렉사(Alexa)

자연어 처리 기술의 미래

- 자연어 처리 활용 분야
 - 텍스트 분석, 기계 번역과 언어 모델, 질문응답 시스템
 - 음성인식, 영상처리 기술과 결합된 개인비서나 이미지 캡션 생성 기술
 - 자연어 처리 기술은
 - 기계 번역 QA 시스템
 - 감성분석(sentiment analysis)
- 자연어 처리 시장 전망(시장조사 기관 Tractica 결과 인용)
 - (2015년) 약 3000억 원
 - (2024년) 약 2조 3천억 원

데이터 사이언스 / 기계학습 과 딥러닝(1/3)

[출처: <http://www.saltflux.com/service/dataScience.do?menuNumber=4>]

데이터 사이언스란?

데이터 수집, 크래이셔, 통계 분석과 기계학습 등의 다양한 기술과 지식을 활용하여 복잡한 데이터로부터 인사이트를 얻거나 지능화된 시스템을 구현하기 위한 모든 업무를 총칭

기계학습 과 딥러닝(학습단계/예측단계)



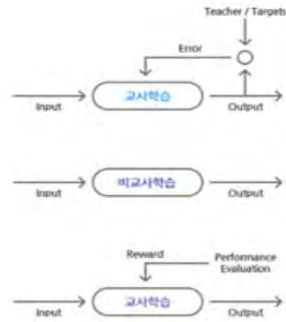
데이터 사이언스 / 기계학습 과 딥러닝(2/3)

기계학습 과
딥러닝

- 1) 교사 학습
- 2) 비교사 학습
- 3) 강화 학습

레이블이 있는 학습 데이터의 사용

- 예제가 매우 많은 경우 효과적인 학습 가능
- 분류(classification) : 레이블이 이산적인 경우
- 회귀(regression) : 레이블이 연속적 값을 가진 경우



레이블 없이 데이터 내의 패턴, 특성, 구조를 학습

- 학습 데이터는 체계의 특성만으로 구성
- dimension reduction, clustering 등

시스템 동작 결과 적절성 피드백, 보상 학습

- 환경 내에서 보상(reward)이 최대화 되는 일련의 행동을 수행
- action selection, planning, policy learning

데이터 사이언스 / 기계학습 과 딥러닝(3/3)

챗봇의 주요 기능

특보 대화 엔진 Talkbot - 대화 처리 구조



자연어 처리의 단계별 핵심 기술

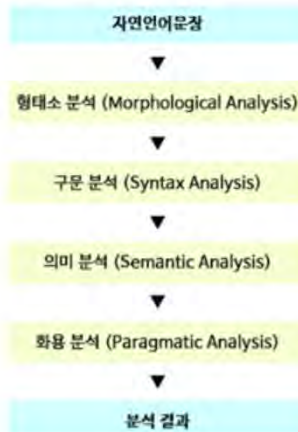
자연어 처리의 핵심 기술

- ① 형태소 분석 기술
- ② 구문 분석 기술
- ③ 의미 분석 기술
- ④ 담화 분석 기술
- ⑤ 단어 및 문장 생성 기술

[출처: 인공지능(AI) 기술과 자연어 처리(Natural Language Processing)]

SD아카데미 공식블로그(2018)

<https://m.blog.naver.com/sundooedu/221235729489>

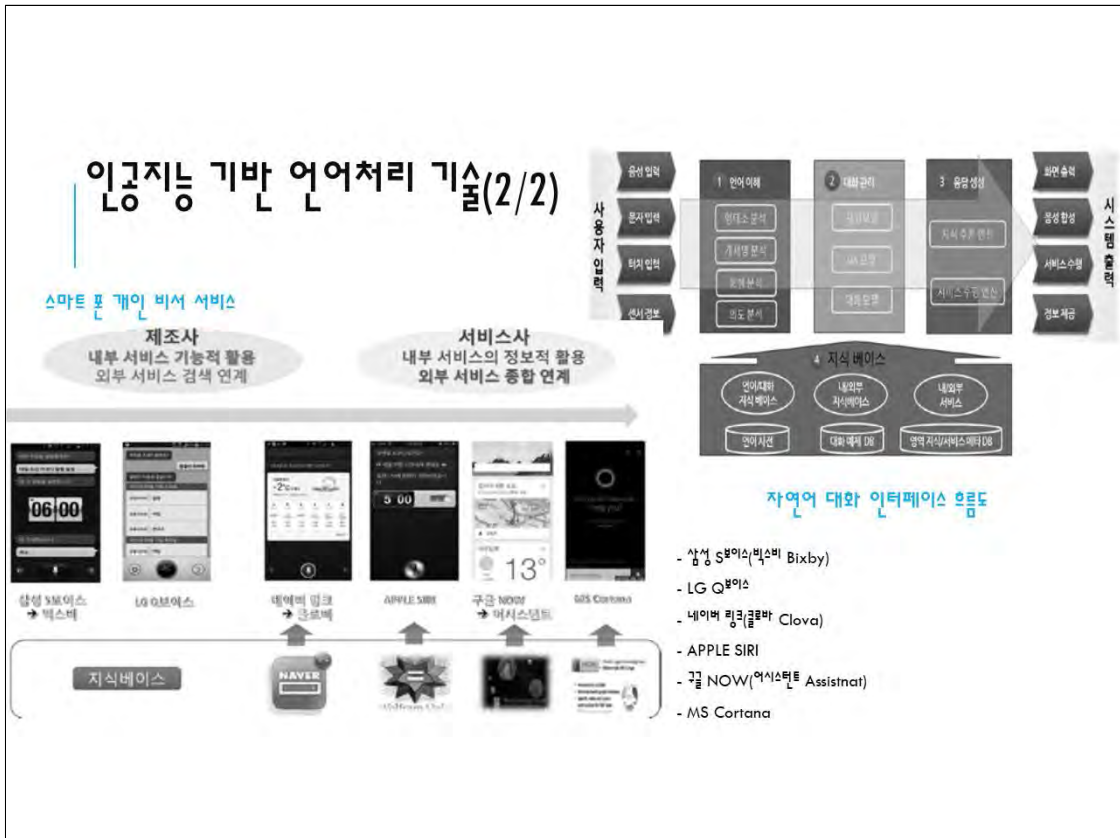


KoNLPy : 파이썬 (Python) 한국어 NLP

<http://konlpy-ko.readthedocs.io/ko/v0.4.3/>

➤ KoNLPy에서 지원하는 형태소분석기 :

- 꼬꼬마 (Kkma)
- 한나눔(Hannanum)
- 자바 한국어 형태소 분석기 (Komoran)
- 오픈한글 프로젝트 (Mecab)
- Twitter 등



자연어 처리 시장

음성인식부터 인공지능까지, 무한한 가능성의 자연어 처리 시장, 김지은, KOTRA 해외시장 뉴스(2019)
<http://news.kotra.or.kr/user/globalAIBbs/kotraneews/album/781/globalBbsDataAllView.do?dataIdx=176188&column=title&search=&searchAreaCd=&searchNationCd=101001&searchTradeCd=&searchStartDate=&searchEndDate=&searchCategoryIdx=&searchIndustryCatelIdx=&page=10&row=10>

자연어 처리 기술 사용 분야

- 맞춤법 검사 및 번역
- 인공지능 음성인식 스피커
- 정서 분석

자연어 처리 시장 현황

- (시장 규모) 2017년 약 32억 달러에서 2025년 약 432억 달러로 확대될 전망 (statista, 2019)
- (인공지능 음성인식 스피커 시장) 2017년 약 11억 1,000만 달러로 2024년까지 약 7천억 달러 시장으로 성장 전망
- 최근 다국어 지원은 필수로 하는 소비자가 증가하면서 스피커 시장과 자연어 처리 시장이 동반 상승 예상

자연어 처리 기술 보유 기업

- Apple 음성인식 시스템인 Siri
 - 사용자가 음성명령을 할 때 마다 반복적으로 사용하는 단어와 문장을 프로파일링하여 사용자가 이메일 메시지 확인과 같은 음성명령을 내릴 시 기존에 구축해놓은 음성 정보와 대조해 사용자의 신원을 확인한 후 해당 정보 제공
- 자연어 처리 시스템 Amazon Comprehend
 - 고객이 기업에 보내 이메일 제품 후기 소셜 네트워크 서비스에 게재된 기업 평가 등을 분석하고, 언어 뒤에 숨겨진 고객의 심리 상태를 통찰해 기업이 제공하는 서비스 및 제품에 대한 고객의 만족도가 긍정적인지 부정적인지를 식별
- IBM 자연어 처리 소프트웨어인 Watson Analytics
 - Amazon Comprehend와 비슷한 기본으로 문서가 내포하고 있는 전반적인 정서 및 감정 또는 텍스트에 포함된 키워드와 관련하여 의도된 정서 및 감정을 파악하여 심층 분석 가능

딥러닝 음성합성 기술(1/2)

머니브레인, <http://moneybrain.ai/>

인공지능 영어회화 Speak Now

- 서비스 개요 : 가상의 인공지능과 대화하며 학습할 수 있는 서비스
- 주요 기술 : 음성인식, 음성합성, 영상합성, 대화 시나리오, 학습 콘텐츠, 오픈트 텍스트, 레베퍼가 학습추진
- 활용 방안 : 스마트폰을 이용한 편리한 인공지능 영어회화 서비스 제공

인공지능 영어회화 Speak Now

최 인공지능과 영어로 대화하고 실시간으로 피드백 받는 인공지능 학습 과정

- 인공지능 영어학습
- 실시간 발음교정
- 수준별, 상황별 시나리오

다보기



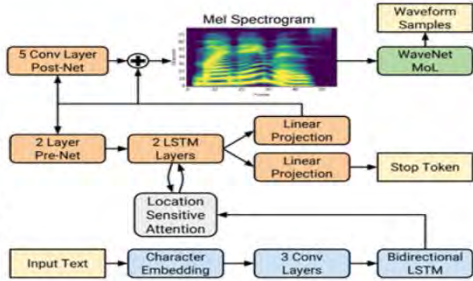
자연어 대화 처리 엔진의 처리절차




딥러닝 음성합성 기술(2/2)

인공 신경망/딥러닝 기술을 활용한 음성합성 기술

- Sequence-to-Sequence (with attention) 기반 모델
- 텍스트 입력 문자를 특정한 목소리로 변환
- 입력 문자를 알파벳 단위, 하위 자모 단위로 쪼개어 각 단위의 임베딩(word embedding)을 학습하고 인코더와 디코더(encoder/decoder)로 구성된 신경망을 거쳐 음성 스펙트로그램을 학습





대화 인터페이스, 챗봇, 자연어처리(1/2)

서경연(서경대, 2017) <http://sigai.or.kr/workshop/AI-for-everyone/2017/slides/> 대화_인터페이스_연구에_관련된_자연어_처리와_인공지능_기술_이야기.pdf

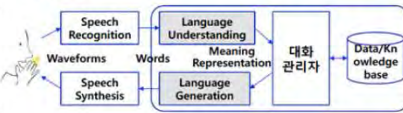
음성 대화형 intelligent assistant



음성 대화 시스템

- 음성 인식, 대화 처리, 음성 합성
- 대화 처리 (대화 시스템)
 - 언어 이해: 사용자의 발화를 언어처리를 통하여 분석하여 사용자의 의도를 표현하는 의미표현을 생성
 - 대화 관리: 대화 흐름과 상황을 고려해서 사용자 발화의 의도에 대한 최선의 대화 전략을 결정하여 다음 시스템 발화에 해당하는 의미표현을 생성
 - 언어 생성: 입력된 의미 표현으로부터 시스템이 발화할 문장을 생성

음성 대화 시스템 구성도



대화 인터페이스, 챗봇, 자연어처리(2/2)

지능형 로봇의 멀티모달 상호작용

멀티모달 대화 인터페이스 (Multimodal Dialogue Interface)

- 음성 인식: 마이크로 입력된 음성 시그널을 문자열로 변환하는 기술
- 음성 합성: 문자열을 스피커로 출력할 수 있는 음성 시그널로 변환하는 기술
- 영상 인식: 카메라로 입력된 화상에서 물체를 인식할 수 있는 기술
- 상황정보: 사용자의 위치정보, 시간정보, 주변 정보, 화면 정보 등을 반영

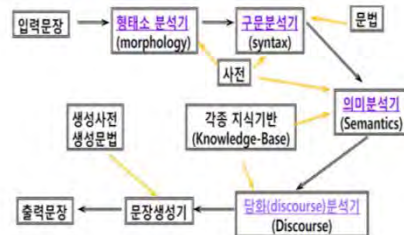
멀티모달 상호작용 대화의 예

노인: "약자 위의 저 컵 가져고 와." (손으로 가리키며)
 로봇: "약을 드시려고 하십니까?"
 노인: "그래"
 로봇: "(컵을 가리려 약자에 갔으나, 컵에 물이 없음을 확인한 후)
 컵에 물이 없습니다. 떠 오겠습니다."
 노인: "그래라." 또는 "아니야, 물은 여기 있으니까 컵만 가져고 와."



- ✓ "약자 위의 저 컵"을 실세계 모델에서 표현된 객체인 '컵101'로 매핑
- ✓ 사용자의 발화를 분석하고, 사용자 모델을 이용하여 약 먹음 시간임을 간파
- ✓ "약을 드시려고 하십니까"라는 문장을 생성하여 노인을 쳐다보면서 발화

자연어처리 시스템의 구성도



AI 음성 비서

[출처: 나만의 자동차 AI 음성비서를 소개합니다, <https://blog.hyundai-mnsoft.com/1606> (현대엠엔소프트 공식 기업 블로그, 2019)

자동차 AI 음성 비서

1. 현대자동차 '카카오이'
2. 르노삼성자동차 '기가지니'
3. 제네시스 브랜드 '알렉사'
4. 벤츠 MBUX
5. BMW '인텔리전트 개인 비서'
6. 현대엠엔소프트 '매피(Mappy)'

현대엠엔소프트 '매피(Mappy)'



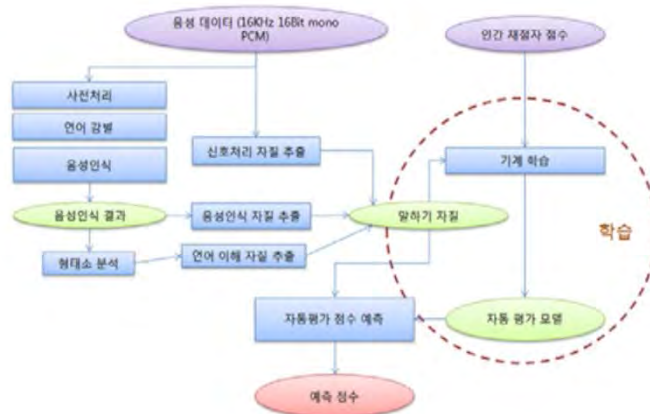
지금 바로 '매피'라고 불러보시고, 다양한 기능들을 이용해보세요.



영어 말하기 자동채점 시스템(1/2)

Shin, Dongkwang, Park, Yonghyo, Park, Taejoon, & Yim, Su Yon. (2015). The present and future of an automated scoring program for speaking skills of English. Multimedia-Assisted, Language Learning, 18(1), 93-114.

한국형 영어 말하기 자동채점 시스템 구성도



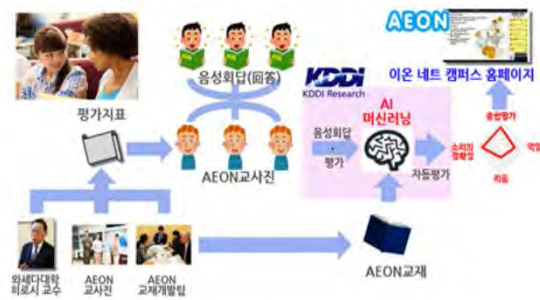
영어 말하기 자동채점 시스템(2/2)

AI가 영어 회화·발음 능력 자동으로 평가한다, 최창현, 인공지능신문(2018), <http://www.aitime.kr/news/articleView.html?idxno=12924>

AI 일반인 영어 사용자를 위한 발음 자동 평가 시스템

- 일본의 영어전문교육기관인 이온(AEON)과 정보 통신을 중심으로 한 전자, 전자, 사인화 등에 관한 조사 연구 컨설팅 기업인 KDDI 종합연구소(KDDI Research, Inc.)는 인공지능(AI)으로 일반인 영어 회화 능력을 자동으로 판정하는 시스템을 개발

- 이온 네트 캠퍼스내에서 학생들이 말하는 음성을 수록하고 그에 대한 AI가 자동 평가하는 '발음 진단' 킷(초로서 2019년 1월부터 서비스 제공



AI 통번역

네이버 번역기 papago

<https://papago.naver.com>

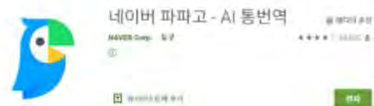
<https://web.archive.org/web/20170304083251/http://linspace.naver.com/nmt/>

<https://play.google.com/store/apps/details?id=com.naver.jobs.translator&hl=ko>

- 제공 기능 : (1) 텍스트 번역 (2) 웹사이트 번역

- 제공 외국어 : 15개

한국어	영어	일본어
한국어(번체)	중국어(번체)	스페인어
프랑스어	독일어	러시아어
포르투갈어	이탈리아어	베트남어
태국어	인도네시아어	힌디어



▶ 주요 기능

텍스트 번역, 이미지 번역, 음성 번역, 오프라인 번역, 대화 번역, 필기 번역, 사진, 클립board 회화

시스트란 번역기

<http://www.systranbox.co.kr/>

시스트란의 언어처리 엔진의 작동 원리

챗봇/AI 스피커

Digital First

"정보 검색 시대"

Web

Mobile First

"모바일에서의 User Network"

SNS

Now

"모바일에서 AI로 진입하기 위한 준비"

Chat bot & AI Speaker

***AI First**

"기계가 사람의 생활 방식에 적용하는 시기"

?

(AI 서비스와 UX의 역할, 라이프타임 챗봇/AI 스피커 사업소개서, 2018, https://www.slideshare.net/Rightbrain_lab/rightbrain-ai-ux-ai)

챗봇/AI Speaker 서비스의 필요성

- 챗봇/AI 스피커 서비스의 성장
- 금융/커머스 중심의 챗봇 서비스 성장
- 음악/날씨/스마트홈 중심의 AI 스피커 성장
- 대화형 사용자 인터페이스
 - 텍스트 기반 대화형 인터페이스
 - 음성 기반 대화형 인터페이스

제공 목적

- 온라인 업무 처리 붐
- 이벤트/마케팅 붐
- 디지털 상담 붐
- 업무 생산성 붐

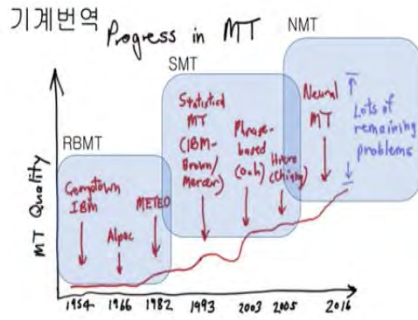
구현 방식

- 단답형
- 시나리오형
- 서비스 연동형

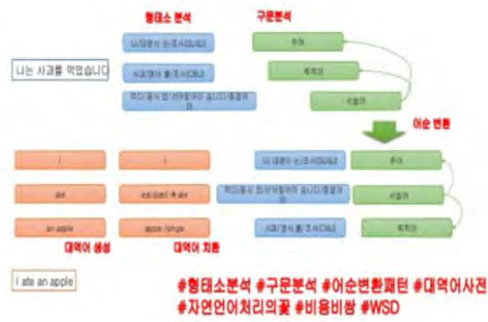
시스템 구성

GNMT(GOOGLE'S NEURAL MACHINE TRANSLATION SYSTEM) (1/2)

GNMT 100 알아보는 신경망 기반 기계번역, 2명일, 2016년, <https://www.slideshare.net/ByeongilKo/gnmt-69817390>



Rule-based Machine Translation



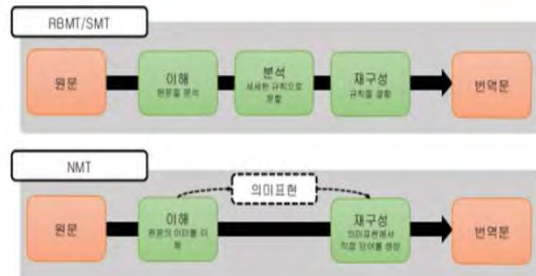
GNMT(GOOGLE'S NEURAL MACHINE TRANSLATION SYSTEM) (2/2)

Statistical Machine Translation



NMT 번역

- 1) Word Embedding
- 2) Recurrent Neural Network
- 3) Encode Decode Model



COMPUTATIONAL LINGUISTICS

Ling 1330/2330 Computational Linguistics
Na-Rae Han, Fall 2019, University of Pittsburgh
<http://www.pitt.edu/~naraehan/ling1330/>

□ Machine translation

- Classical MT

Direct translation

Word-to-word translation

Transfer approaches

Uses syntactic and possibly semantic transfers

Based on **transfer rules**: hand-crafted per source-target language pair

Interlingua approaches

- Statistical MT

The three classic architectures focus on the appropriate representations to use symbolic approach

In **statistical MT**, however, the focus is more on the result

- Neural machine translation

Encoder-decoder model

The topics include: spell-checking, machine translation, part-of-speech tagging, parsing, document classification, and corpus building and exploration.

- NLTK & NLP:

• Natural Language Toolkit (NLTK) Book, Python3 Edition

• FSNLP Ch.4 Corpus-Based Work

Foundations of Statistical Natural Language Processing

Christopher D. Manning and Hinrich Schütze

• Speech and Language Processing 3rd Ed.

- Machine Translation

<http://www.pitt.edu/~naraehan/ling1330/Lecture25.pdf>

NATURAL LANGUAGE PROCESSING WITH DEEP LEARNING(1/2)

CS224d: Deep Learning for Natural Language Processing, Winter 2017, <http://cs224d.stanford.edu/syllabus.html>

CS224n: Natural Language Processing with Deep Learning, Stanford / Winter 2020, <http://web.stanford.edu/class/cs224n/>

What is Natural Language Processing (NLP)?

- Natural language processing is a field at the intersection of computer science,
- artificial intelligence
- and linguistics.

NLP in Industry

- Search (written and spoken)
- Online advertisement
- Automated/assisted translation
- Sentiment analysis for marketing or finance/trading
- Speech recognition
- Automating customer support

NLP Applications

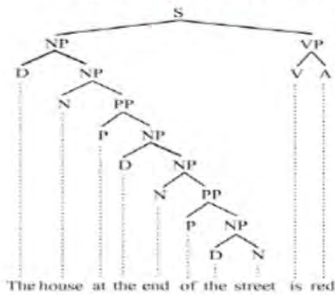
- Spell checking, keyword search, finding synonyms
- Extracting information from websites such as
- product price, dates, location, people or company names
- Classifying, reading level of school texts, positive/negative sentiment of longer documents
- Machine translation
- Spoken dialog systems
- Complex question answering

NATURAL LANGUAGE PROCESSING WITH DEEP LEARNING(2/2)

Representations at NLP Levels: Syntax

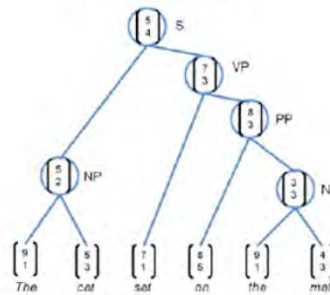
• Traditional: Phrases

Discrete categories like NP, VP



• DL:

- Every word and every phrase is a vector
- a neural network combines two vectors into one vector
- Socher et al. 2011



NLTK(NATURAL LANGUAGE TOOLKIT)

<http://www.nltk.org/>

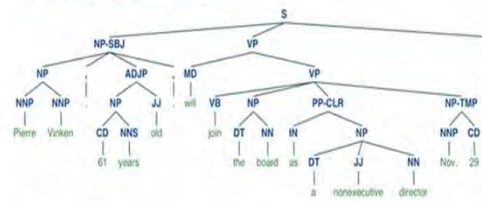
Tokenize and tag some text

```
>>> import nltk
>>> sentence = """At eight o'clock on Thursday morning
... Arthur didn't feel very good."""
>>> tokens = nltk.word_tokenize(sentence)
>>> tokens
['At', 'eight', "o'clock", 'on', 'Thursday', 'morning',
'Arthur', 'did', "n't", 'feel', 'very', 'good', '.']
>>> tagged = nltk.pos_tag(tokens)
>>> tagged[0:6]
[('At', 'IN'), ('eight', 'CD'), ('o'clock', 'JJ'), ('on', 'IN'),
('Thursday', 'NNP'), ('morning', 'NN')]
```

Identify named entities

```
>>> entities = nltk.chunk.ne_chunk(tagged)
>>> entities
Tree('S', [Tree('IN', ('At', 'IN')), Tree('CD', ('eight', 'CD')), Tree('JJ', ('o'clock', 'JJ')),
Tree('IN', ('on', 'IN')), Tree('NNP', ('Thursday', 'NNP')), Tree('NN', ('morning', 'NN')),
Tree('PERSON', [Tree('NNP', ('Arthur', 'NNP'))]),
Tree('VBD', ('did', 'VBD')), Tree('RB', ('n't', 'RB')), Tree('VB', ('feel', 'VB')),
Tree('RB', ('very', 'RB')), Tree('JJ', ('good', 'JJ')), Tree('.', '.')])
```

Display a parse tree



Installing NLTK Setting up a Python Environment(Windows)

- 1) Install Python 3.7: <http://www.python.org/downloads/>
- 2) Install Numpy : <https://www.scipy.org/scipylib/download.html>
- 3) Install NLTK: <http://pypi.python.org/pypi/nltk>
- 4) Test installation: Start>Python37, then type import nltk

영어학습을 위한 인공지능 채보 활용 및 제작

저자 김혜영, 신동광, 이장호, 김영우, 양혜진, 발행일 2019-06-30, ISBN 978-89-254-1388-4(03370), 페이지 246

제1부 AI 채보의 이해와 활용

제1장 AI 채보의 특성과 현황

- 영어 채보의 현황: 텍스트 채보에서 스마트스피커까지

제2장 영어 텍스트 채보의 이해와 활용

- 텍스트 채보의 활용 분석: Mitsuku vs. Cleverbot

제3장 영어 음성 채보의 이해와 활용

- AI 스피커의 활용 분석: Google Assistant와 Alexa

제2부 AI 채보의 제작

제4장 AI 채보 빌더와 간단한 채보 제작

- 단비(danbee)를 이용한 간단한 기관 소개 채보 만들기

제5장 Dialogflow를 이용한 채보 제작

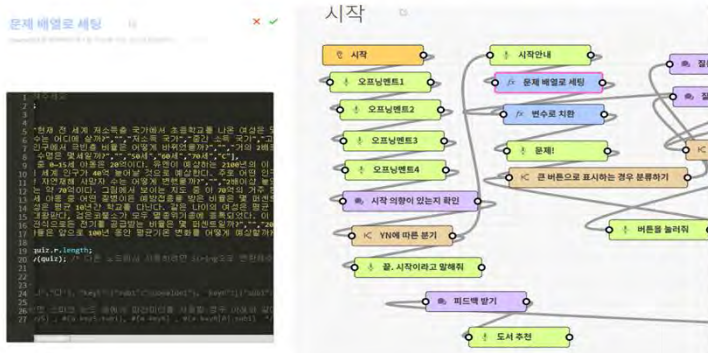
제6장 AI 채보를 활용한 영어학습 과업 설계 I

제7장 AI 채보를 활용한 영어학습 과업 설계 II

단비(DANBEE.AI) 채보

<https://danbee.ai/landing.html>
<https://blog.naver.com/danbeeai>

FunctionNode에 담겨있는 퀴즈 데이터 SET



인공지능 채보와 자연어처리 취창업과정

한국인공지능아카데미

- 교육명칭
인공지능 채보와 자연어처리 취창업과정
- 교육기간
2019.8.29 ~ 11.22 (452 시간)
- 교육시간
월요일 09:00 ~ 18:00
- 교육내용
파이썬 / 웹 / 데이터스지 / 딥러닝 자연어처리 / 딥러닝 이미지처리 / 채보

교육역량 단계	과정명	세부내용
기초과정	파이썬	OT(1시간) / 챗봇 개론
		개발환경 설정환경 세팅
		파이썬
		배열 행렬 다루기(numpy)
		엑셀파일 다루기(pandas)
		데이터 시각화(matplotlib,seaborn)
		웹 프로그래밍
		광규표형식
		자연어처리 기초
		파이썬 프로젝트
심화과정	챗봇	챗봇 서버
		서버용라이언트 OpenAPI Restful
		챗봇 case / 챗봇 시나리오 / 대화디자인
	딥러닝 자연어처리	챗봇 페르소나 / intent, entities, context
		챗봇 프로젝트
		음성 챗봇
		웹 음성 챗봇 프로젝트
	딥러닝 기초	
	딥러닝 이미지처리	
	딥러닝 자연어처리	
	딥러닝 챗봇 프로젝트	

딥러닝 채보(재직자 교육)

NCIA(차세대융합기술진흥협회) 교육센터, http://edu.ncia.kr/ncia/aducation.application.php?ep_no=338

딥러닝 채보

재직자 직무능력향상 교육
교육 일정: 2018년 6월~12월

교육목표:

- RNN과 텍스트표현을 역등한 기본 기술 이해
- Word2Vec, Doc2Vec 알고리즘 습득하고 해당 알고리즘 접목할 수 있는 수준의 코드 작성
- 자연어 처리에 필요한 개념 이해 및 간단한 채보 구성
- 파이썬 자연어 처리 패키지(NLTK) 모듈을 통해 자연어 처리에 대한 개념 정리

교육내용

일차	과목	내용
1일차	자연어 처리 입문	챗봇 구현에 필요한 자연어 처리의 모든 개념을 학습한다. - 텍스트마이닝, 한국어 코퍼스, 단어 추출, tokenizer
2일차	챗봇 알고리즘	챗봇 구현에 필수적인 다양한 알고리즘에 대해 학습한다. - word2vec, doc2vec, t-sne, vector Indexer
3일차	챗봇 프로젝트	챗봇 프로젝트를 직접 구현해 보고, 추가적인 알고리즘에 대해 학습한다. - Auto encoder, CNN, RNN, seq2seq, 챗봇 프로젝트

인공지능 기술을 활용한 영어 교사 전문성 신장 특별 연수 프로그램

프로그램 개요

- 일시: 2019년 1월 16일(수)~17일(목) 9:30시~18시
- 장소: COEX HALL A, B홀
- 주최: 국제언어교육연구회, 교육부 인가 비영리 공익법인)
- 주관: (주)Exporum, 글로벌영어교육연구소, 전국외국어교육협회
- 후원: 한국외국어교육학회, 전국외국어교육협회, 글로벌영어교육연구소, 지능정보산업협회 등
- 연수 주제: 인공지능 기술을 활용한 영어 교사 전문성 신장 특별 연수회(안)

주요 프로그램 Topics

- 서울시교육청 인공지능 활용 영어교육 정책 안내
- 인공지능형 영어학습용 로봇 작동원리 및 영어수업에의 적용 방안- 인공지능 기술의 학교 영어수업에의 활용방안
- TTS(ReadSpeaker)/STT(Dragon Speaking) 기술을 활용한 영어수업 방안
- AI 스피커 인공지능 기술의 영어수업에의 활용방안
- (GenieTutor) 영어 말하기 학습을 위한 음성인식 및 발음 평가 기술

인공지능어어공학

세종대학교 인공지능어어공학 대학원 학과과정

전공 소개

인공지능어어공학과 학과과정은 이가 중심 인터페이스 지능 인공지능 언어 처리 기술, 빅데이터 분석, 딥러닝, 자연어 처리 기술, 그리고 대량의 데이터를 검색하기 위한 웹 기반 정보 검색 기술의 고도에 중점을 두고 있다. 이를 위해 인간의 뇌에 대한 전방적인 지식을 기초하여 Text Mining, Deep Learning, 컴퓨터어어학의 공학에 언어학 지식을 접목하고 인공 지능을 통사론 의미론 화용론 언어학의 기반을 형성하는 지식체계에서 IT의 접근 지식을 접목하여 융합 교과목을 신설한다.

- 세종대학교 컴퓨터공학과 영어영문학과, 교육학과 교수진이 다음 학기부터 세종대 일반대학원에 석사과정으로 인공지능어어공학과를 신설(2018학년도)

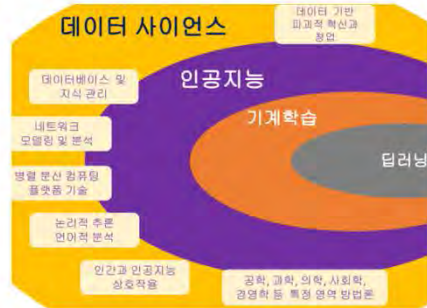
교육과정

교과목명	대상과정	미수구분	학점
역사논문연구1	석사	심화	3.0
역사논문연구2	석사	심화	3.0
고급딥러닝	석사	핵심	3.0
딥러닝개론	석사	핵심	3.0
기계학습개론	석사	핵심	3.0
텍스트마이닝개론	석사	핵심	3.0
언어데이터분석이론과방법	석사	핵심	3.0

데이터사이언스대학원(서울대)

<http://gsds.snu.ac.kr/>

데이터사이언스와 인공지능의 관계



고급과정(주요 교과목)

· **적고(기초)**
· **적고(고급)** (적고, 기초, 공통, 계산, 분석, 응용)

- **적고(기초)**
 - 데이터사이언스를 위한 프로그래밍
 - 데이터사이언스를 위한 통계
- **적고(계산)**
 - 머신러닝 및 딥러닝 기초
 - 머신러닝 기반 음성인식 및 영상 인식
 - 영상 비데이터 분석 방법론
 - 텍스트 및 자연어 비데이터 분석 방법론
 - 심층강화학습을 이용한 비데이터 응용

AI 프로그래밍 언어

함수 언어 / 논리 언어

- LISP : Functional Language
- **PROLOG** : Logic Language
 - 자연어 처리(기계 번역)

Script 언어

- JavaScript : TensorFlow.js
- **R** : 데이터 사이언스

절차형(Procedural) 언어

- C/C++, Rust
- JAVA

ML(DL) 라이브러리

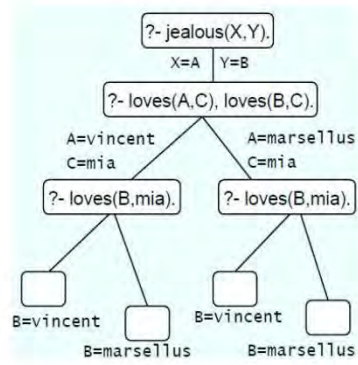
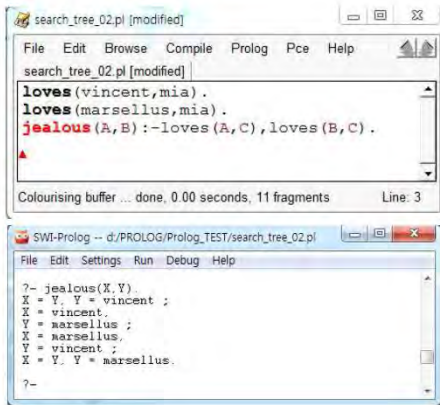
- **Tensorflow**
- PyTorch
- Chainer
- Apache MXNet
- Theano

Python

- 파이썬 라이브러리
 - NumPy : Tensorflow 표준 API
 - Pansa : R 의 데이터 프레임
 - NLTK, SpaCy : 자연어 처리(NLP)

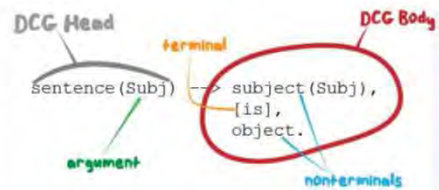
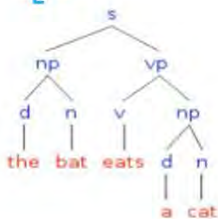
PROLOG(1/2) : 논리 프로그래밍과 탐색 트리

인공지능, 정덕기/김성희, 도서출판 지남(2018)



PROLOG(2/2) : 자연어 처리

구문 트리

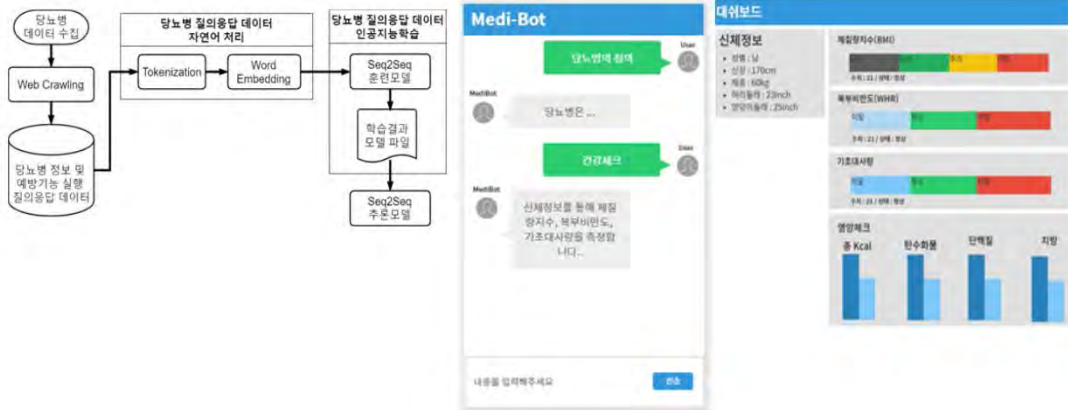


Prolog

- Definite Clause Grammar(DCG)
- 2005년에는 NASA 에 의해 International Space Station 에서 자연 언어 (natural language) 인터페이스를 프로그래밍 하는데 사용
- 2011년에는 IBM사의 쿼리 슈퍼 컴퓨터인 Watson 이 Prolog를 코딩됨

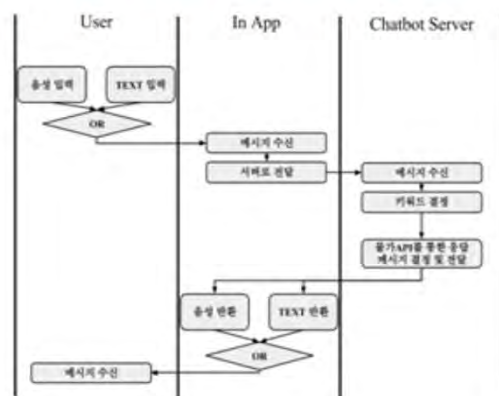
만성질환 정보 제공 및 예방 관리를 위한 인공지능 기반 챗봇 시스템

공학석사학위논문, 장재홍(지도교수: 정덕기)동의대학교학위논문(2019)



공공데이터와 AI 챗봇을 이용한 물가 음성안내 앱 서비스

이재선, 강경돈, 박백어, 정덕기(동의대학교), 한국경제통신학회 춘계 학술대회(2018)

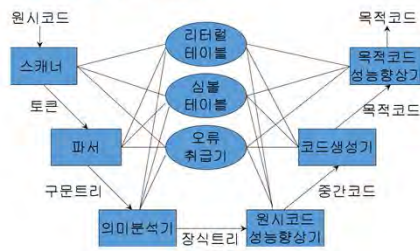


COMPIER

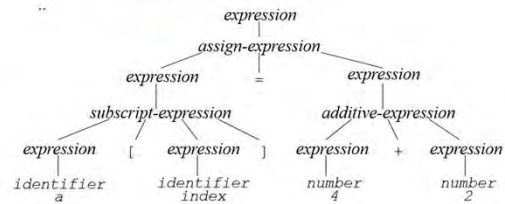
컴파일러 제작: 원리와 실제 (Compiler Construction: Principles and Practice, K.C. Louden, 1997), 정덕길 외(번역), 영남출판사(2002)



컴파일 단계(Phases)



파서(Parser)



인공지능 교재

Artificial Intelligence: A Modern Approach

Third Edition, Stuart J. Russell and Peter Norvig, 1151 pp, Copyright 2010, 2003, 1995 by Pearson Education, Inc.

