

# 2022년 한국생성문법학회-한국언어정보학회 가을 공동학술대회 발표논문집

---

## “화행: 통사와 의미의 만남”

일시:	2022년 11월 12일(토) 9:20 ~ 17:00
진행방식:	비대면 방식 (실시간 Zoom 접속)
주최/주관:	한국생성문법학회 한국언어정보학회



# 차 례

학술대회 조직위원회.....	2
학술대회 프로그램.....	4

## 초청 발표 초록

### 초청발표 1

임동식 (홍익대).....	6
Korean sentential particle <i>-ney</i> : its semantics, pragmatics, and theoretical implications	

### 초청발표 2

Ricardo Etxepare (Iker lab-CNRS) & Myriam Uribe-Etxebarria (UPV/EHU) .....	6
Disentangling the Underlying Structures of Constituent Negation	

## Session I

박종언 (한성대).....	10
Meta <i>Why</i> Questions and Speech Act Projections	

이주원 (전주대).....	24
A New Type of English Secondary Predicate Construction	

## Session II

노강산 (고려대).....	34
Korean EFL Learners' Underuse of the Verb-Particle Constructions	

이지은, 이규민, 송상현 (고려대).....	48
딤러닝 한국어 모델은 관형사절 속 공백을 포착하는가?	

## Session III

박명관 (동국대).....82  
On the necessity of the plural marker '-tul' in Korean

왕건민 (서강대).....94  
'NP2+무섭다' 구문의 의미와 사건구조: -어휘 분해 이론에 의한 접근

박혁재, 김광섭 (한국외대).....106  
현대중국어 비부정 不, 没(有)의 문법 기능에 관한 연구

## Session IV

송영숙 (경희대), 정유남 (중앙대)..... 115  
국어사전 용례에서의 윤리성 자동 판별 문제

강아름 (충남대)..... 132  
The illocutionary degree modifier 'wancen' at the discourse level

황동진, 신운섭, 이지은, 송상헌 (고려대)..... 153  
Intervention effects in WHY questions as presupposition violation: Evidence from an ERP study

## 초청 발표 #2

Ricardo Etxepare (Iker lab-CNRS) & Myriam Uribe-Etxebarria (UPV/EHU) ..... 169  
Disentangling the Underlying Structures of Constituent Negation

## 학술대회 조직위원회

<b>대회장</b>	이광호	(한국생성문법학회 회장)
	조세연	(한국언어정보학회 회장)
<b>조직위원장</b>	김광섭	(한국외대)
	위혜경	(단국대)
<b>조직위원</b>	임창국	(중앙대)
	최재훈	(대구대)
	임동식	(홍익대)
	강아름	(충남대)
<b>운영위원</b>	김한샘	(연세대)

# 한국생성문법학회-한국언어정보학회 가을 공동학술대회

2022년 11월 12일 (토)

09:20 – 09:30	<p>인사말 이광호 (경상대) <a href="https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFBtUjF1O0kxZUZwUT09">https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFBtUjF1O0kxZUZwUT09</a></p>	
09:30 – 10:50	<p>초청 발표 #1 발표자: 임동식 (홍익대) 주 제 : Korean sentential particle <i>-ney</i>: its semantics, pragmatics, and theoretical implications 사회자: 이주원 (전주대) <a href="https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFBtUjF1O0kxZUZwUT09">https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFBtUjF1O0kxZUZwUT09</a></p>	
10:50 – 11:00	휴식	
11:00 – 12:00	<p><b>Session I</b> 사회자: 임창국 (중앙대) <a href="https://cau.zoom.us/j/84441700735?pwd=SE1ZN0xUTFgzanl5TjlwOnNqUGVYUT09">https://cau.zoom.us/j/84441700735?pwd=SE1ZN0xUTFgzanl5TjlwOnNqUGVYUT09</a></p>	<p><b>Session II</b> 사회자: 김정수 (경희대) <a href="https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFBtUjF1O0kxZUZwUT09">https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFBtUjF1O0kxZUZwUT09</a></p>
11:00-11:30	<p>발표자: 박종언 (한성대)   Meta <i>Why</i> Questions and Speech Act Projections 토론자: 김선웅 (광운대)</p>	<p>발표자: 노강산 (고려대)   Korean EFL Learners' Underuse of the Verb-Particle Constructions 토론자: 황주현 (한양대)</p>
11:30 – 12:00	<p>발표자: 이주원 (전주대)   A New Type of English Secondary Predicate Construction 토론자: 박상희 (덕성여대)</p>	<p>발표자: 이지은, 이규민, 송상현(고려대)   딥러닝 한국어 모델은 관형사절 속 공백을 포착하는가? 토론자: 한수미 (한림대)</p>
12:00 – 12:30	<p>생성문법학회 확대이사회 <a href="https://cau.zoom.us/j/84441700735?pwd=SE1ZN0xUTFgzanl5TjlwOnNqUGVYUT09">https://cau.zoom.us/j/84441700735?pwd=SE1ZN0xUTFgzanl5TjlwOnNqUGVYUT09</a></p>	점심 휴식
12:30 – 13:30	점심 휴식	

13:30 – 15:50	<b>Session III</b> 사회자: 최재훈 (대구대) <a href="https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09">https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09</a>	<b>Session IV</b> 사회자: 남윤주 (한양대) <a href="https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFhtUjF1Q0kxZUZwUT09">https://cnu-ac-kr.zoom.us/j/86142744533?pwd=azNpN3FwOEc5ZFhtUjF1Q0kxZUZwUT09</a>
13:30 – 14:00	발표자: 박명관 (동국대)   On the necessity of the plural marker '-tul' in Korean  토론자: 문귀선 (한성대)	발표자: 송영숙 (경희대), 정유남 (중앙대)   국어사전 용례에서의 윤리성 자동 판별 문제  토론자: 이상아 (서울대)
14:00 – 14:30	발표자: 왕건민 (서강대)   'NP2+무섭다' 구문의 의미와 사건구조: -어휘 분해 이론에 의한 접근  토론자: 명정희 (서강대)	발표자: 강아름 (충남대)   The illocutionary degree modifier 'wancen' at the discourse level  토론자: 이지은 (고려대)
14:30-14:50	휴식	
14:50 – 15:20	발표자: 박혁재, 김광섭 (한국외대)   현대중국어 비부정 不, 没(有)의 문법 기능에 관한 연구  토론자: 이슬기 (협성대)	발표자: 황동진, 신운섭, 이지은, 송상헌 (고려대)   Evaluating the World Knowledge of Language Models: English Backshift Phenomenon  토론자: 이용훈 (충남대)
15:20 – 16:20	<b>초청 발표 #2</b> 발표자: Ricardo Etxepare (Iker lab-CNRS) & Myriam Uribe-Etxebarria (UPV/EHU) 주 제 : Disentangling the Underlying Structures of Constituent Negation 사회자: 박명관 (동국대) <a href="https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09">https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09</a>	
16:20-16:50	<b>연구윤리교육</b> 이해윤 (한국외대) <a href="https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09">https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09</a>	
16:50 – 17:00	<b>맺음말</b> 조세연 (강원대) <a href="https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09">https://us06web.zoom.us/j/88154720066?pwd=RmlnazRkQOWNrUUhYMVVlQlJFMlBFdz09</a>	

## 초청 발표 초록

- 초청발표 1

Korean sentential particle *-ney*: its semantics, pragmatics, and theoretical implications

Dongsik Lim (Hongik University)

(with Semoon Hoe and Yugyeong Park)

Korean sentential particle *-ney* has received much attention from previous literature about evidentials, due to its implication similar to that of direct evidential, its mirativity, and its epistemic inference. To explain these various aspects of *-ney*, we propose that *-ney* requires the speaker's directly perceived evidence, just like another widely studied evidential *-te-*, but it also has the restriction that the time when the evidence is acquired (or learning time: LT) should overlap with the utterance time (UT) (in contrast with *-te-*, where LT should precede UT, as discussed in previous literature). To derive other aspects of *-ney* from this proposal, we also propose that, depending on the context, including the expected information state of the addressee, either the prejacent of *-ney*, or the information state of the speaker indicated by *-ney*, can be the main point of the utterance with *-ney* (see also two possible main points in the utterance with attitude predicates: Simons 2007, a.o.). In this presentation we try to show that our proposal can explain unique behaviors of *-ney* discussed in previous studies, and that previous proposals on *-ney* can be easily derived from our proposal. We also discuss the theoretical implications of our proposal, specifically with respect to the typology of evidentials.

- 초청발표 2

Disentangling the underlying structures of Constituent Negation

Ricardo Etxepare (Iker lab-CNRS) & Myriam Uribe-Etxebarria (UPV/EHU)

The study of quantification has played a crucial role in our understanding of the syntax-semantics interface and the architecture of grammar. In this paper we focus on some sequences formed by Negation and a Quantifier Phrase, which are usually characterized as involving Constituent Negation, and investigate the syntactic and semantic structure that underlies these sequences. The Spanish example in (1)

illustrates the type of sequence under analysis.

(1) No todo el mundo está contento

Neg all the world is happy

“Not everyone is happy”

We discuss previous analyses of this type of structure (a.o. Lasnik 1972, Sánchez López 1999, Collins 2017) and present a new alternative proposal. More specifically, we will show that there are at least three different structures underlying Negation + QP sequences and that languages vary regarding which one(s) they choose.

**i. Direct Merge.** Negation merges directly with the QP. Restricted to a limited set of Qs.

(2) [<sub>QP</sub> **NO** [<sub>QP</sub> ...Q... ]]

**ii. Association with Focus.** Neg and the QP are born separately. They surface together due to the movement of the QP to the Spec of a FocP which is close to a NegP located in a high position of the left periphery.

(3) [<sub>NegP</sub> **NO** [<sub>FoCP</sub> **QP** **Foc<sup>o</sup>** [<sub>TP</sub> **QP** ]]]

**iii. Mixed strategy.** The QP has a Foc and a Neg feature and has to move to the specifier of a sentential FocP.

(4) [<sub>FoCP</sub> [<sub>NegP/QP</sub> **NO** [<sub>FoCP</sub> **Foc<sup>o</sup>** [**QP**]]]i **Foc<sup>o</sup>** [<sub>TP</sub> **QP**i ]]]

Our hypothesis that both (2) and (3) are necessary to account for the syntactic behavior of *Neg+QP* sequences goes against much of the literature on constituent negation, which takes either (2) (or a variant of (2)) or (3) to be the only available strategy. Thus, (3) is reminiscent of Lasnik’s (1972) treatment of constituent negation in English as being external to the QP and located in a pre-sentential S



position. Collins and Postal (2014) and Collins (2017, 2020) argue in favor of (2) as the proper analysis of *Neg+QP* sequences in English. We will argue that these proposals are insufficient to capture the properties of *Neg+QP* sequences in Spanish, and show that the two structures in (2) and (3) are necessary to capture the behavior of the sequences under analysis in this language. Further, we will show that the limited cases of *Neg+QP* found in Basque can only be accounted for under the strategy described in (3). Finally, we will give evidence from some other Romance varieties (Portuguese and Paduan) for the existence of the mixed strategy, which seems to be dependent on the presence of a focus feature in the QP which may force pied-piping of the whole structure to a preverbal focus position.

**발 표 문**

# Meta *Why* Questions and Speech Act Projections<sup>1)</sup>

Jong Un Park (Hansung University)

jupark@hansung.ac.kr

## 1. Introduction

### Core data

- It is well known that *why* questions like (1B) solicits information about a reason or purpose for which Jenny came back from the school earlier than usual.

(1) A: Jenny-ka onul hakkyo-eyse ilccik wa-ss-e.  
 J.-Nom today school-from early return-Past-Decl  
 ‘Jenny returned from the school early today.’

B: Way?

why

- (i) ‘For what reason did Jenny return from the school early today?’ [Reason]
- (ii) ‘For what purpose did Jenny return from the school early today?’ [Purpose]

- By contrast, *why* questions like (2B) do not seem to convey the same meaning as the question in (1B): rather than requesting the information about a reason or purpose for Ellie’s being accepted to the school, the speaker B asks the interlocutor A why (s)he raised his/her question in (2A).

(2) A: Ellie-ka hakkyo-ey hapkyekhay-ss-ni?  
 E.-Nom school-to get.accepted-Past-Int  
 ‘Did Ellie get accepted to the school?’

B: Way?

why

‘Why are you asking that question?’

- To the best of my knowledge, the second type of fragmentary *why* questions have not been touched on in the literature on Korean linguistics; only the English equivalent has been discussed in a few recent works (e.g., Ginzburg 2012; Woods & Vicente 2021).

### Some immediate questions

The question in (2B), apparently similar to the one in (1B), raises a couple of inquiries:

- (3) a. How does the distribution of question (2B) differ from question (1B)?  
 b. Is the *wh*-word in (2B) a sentential unit with invisible structure or non-sentential unit whose meaning arises due to pragmatic principles?  
 c. If it is sentential, what does the structure for question (2B) look like?, etc.

### Goals of this paper

- To demonstrate that *wh*-questions involving only *way* ‘why’ given in (2) is a novel type of fragmentary questions distinguished from *wh*-questions of the sort in (1) in terms of their semantic meaning and syntactic distribution.

---

1) This work was supported by the National Research Foundation (2022S1A5A2A01045589).

- To argue that the *wh*-word in (2) that I will call ‘meta(communicative) *why* question’ (in the sense of Ginzburg (2012) and Woods & Vicente (2021)) is not a nonsentential unit conveying a pragmatic kind of meaning, but a sentential unit with part of its syntactic structure deleted (à la Merchant (2004) for fragment answers).
- To put forth a clause structure for Korean meta *why* questions, which supports the claim that there are speech act projections in the clause structure (e.g., Speas & Tenny 2003; Haegeman & Hill 2013; Yim 2016; Hong 2018; Park 2021, 2022, *a.o.*)

#### □ Roadmap of the paper

- Section 2 discusses more properties of what I call ‘meta *why* questions’ in comparison with typical fragmentary *why* questions.
- Section 3 demonstrates that the meta *why* question is not a nonsentential unit that yields pragmatic meanings but a sentential unit derived with part of its structure deleted.
- Section 4 discusses, under the ellipsis view, what the silent structure of the meta *why* question looks like and where the remnant *why* appears.
- Section 5 draws the conclusions.

## 2. More properties of meta *why* questions

### □ Syntactic distribution of meta *why* questions

In addition to the meaning difference shown in Sec 1, meta *why* questions and regular *why* questions are not identical in their distribution.

#1. Only interrogatives can be an antecedent for meta *why* questions, but this restriction doesn’t apply to typical fragmentary *why* questions.

(4) Declarative clause as an antecedent (No)

A: Ellie-ka hakkyo-ey hapkyekhay-ss-e.  
 E.-Nom school-to get.accepted-Past-Decl  
 ‘Ellie got accepted to the school.’

B: Way?

Why

≠ ‘Why are you asking that question?’ [Meta *why*]  
 = ‘Why did she get accepted to the school?’ [Regular *why*]

#2. Only true types of interrogatives can be an antecedent for meta *why* questions.<sup>2)</sup>

- Rhetorical questions cannot be an antecedent for meta *why*-questions.

(5) Rhetorical question as an antecedent (No)

A: Coyonghi mos hay? (Mom to her naughty boy)  
 be.quiet not do

---

2) Woods & Vicente (2021) show that an echo question also fails to license a meta *why*-question in English, as in (i). However, it’s not easy to test whether the same type of question can precede a meta *why*-question in Korean, as the latter does not involve overt *wh*-movement.

(i) A: He said WHAT?

B: #Why?

Intended: ‘Why are you asking that?’

Woods & Vicente (2021: 3)

Lit. 'Can't you be quiet?'

B: Way?

Why

≠ 'Why are you asking that question?' [Meta *why*]

= 'Why should I be quiet?' [Regular *why*]

- Declarative questions, another atypical kind of questions, don't seem to license meta *why* questions in Korean, though not quite as clearly attestable as in English (see (7) below).

(6) Declarative question as an antecedent (No)

A: Ellie-ka yeki-ey iss-ta?  
E.-Nom here-Loc exist-Decl

B: #Way?

Why

Intended: 'Why are you asking that question?' [Meta *why*]

(7) English equivalent

A: Sally is here? [Declarative question]

B: #Why

Intended: 'Why are you asking that question?' [Meta *why*]

Woods & Vicente (2021: 3)

#3. Even interrogatives, when embedded, can't serve as an antecedent.

(8) A: Halmeni-kkeyse [Ellie-ka hakkyo-ey hapkyekhass-nunci] kungkumhayha-si-n-ta.  
grandma-Nom(Hon) E.-Nom school-to got.accepted-Int wonder-Hon-Pres-Decl  
'Grandma wonders [whether Ellie got accepted to the school].'

B: Way?

Why

≠ 'Why are you asking me that?' [Meta *why*]

= 'Why is she wondering about that?' [Regular *why*]

#### Summary

- This section has shown that meta *why*-questions are not identical to typical fragmentary *why*-questions in their syntactic distribution—esp., as to which clauses can be an antecedent.
  - ✓ Only true interrogatives in root context can be an antecedent that feeds meta *why* questions.
  - ✓ In contrast, not only declaratives but also various types of interrogative clauses can license typical fragmentary *why*-questions, whether or not appearing in root or embedded context.
- These differences in the syntactic distribution, as well as the semantic meaning, lead us to conclude that meta *why*-questions cannot be treated on a par with typical *why* questions in Korean (cf. Woods & Vicente (2021) for the English counterpart).

### 3. Do Korean meta *why*-questions have a silent syntactic structure?

#### Question about the status of way 'why' in meta *why*-questions

The next question to ask is ...

- whether way 'why' in meta *why*-questions is a 'nonsentential' unit that doesn't accompany any

syntactic structure at all or a ‘sentential’ unit with invisible syntactic structure.

#### □ **Base-generation approach vs. deletion approach to fragments**

Over the last 20 years there have been two competing views on fragmentary questions or answers, namely the base-generation approach and the ellipsis approach.

- Base-generation approach: Fragments do not make up a sentential unit at all, and their meanings are obtained by pragmatic principles depending on context (e.g., Stainton 2006, *a.o.*).
- Ellipsis approach: Fragments are base-generated with a fully-blown syntactic structure, some of which is supposed to be deleted after a remnant has moved to the periphery (e.g., Stanley 2000; Merchant 2004; B.-S. Park 2005; Weir 2014, *a.o.*).

Similarly to Woods & Vicente’s (2021) view on the English equivalent, this paper argues for the so-called ‘ellipsis’ approach to Korean meta *why* questions and claims ...

- that they start out as a fully-blown syntactic structure; and
- that the *wh*-word *way* ‘why’ is the remnant derived with part of the syntactic structure deleted.

#### □ **Immediate evidence against the base-generation/pragmatic approach**

- One might argue that meta *why* questions are nothing but a typical fragmentary *why* question, and an extraordinary meaning from them is what is called a ‘conversational implicature’ arisen from the Cooperative Principle and the conversational maxims (e.g., Lee 1998).
- Under this pragmatic view, it is predicted that the meta *why* question should be available even when its potential antecedent appears in the embedded context.

⊞ Recall, however, that the prediction turns out to be false, as shown in (8) in Sec 2: unlike the prediction, the embedded question can’t be a target of a meta *why* question.

#### □ **Meta *why* questions are an instance of sluicing?**

We are left with the ellipsis view, and the next possible question is whether meta *why* questions behave similarly to ‘sluicing,’ a construction that has been shown to support the existence of a silent structure (e.g., Merchant 2001).

- The idea of assimilating the meta *why* question to sluicing like (9) appears to be plausible since a remnant that survives in both constructions is a *wh*-constituent.

#### (9) Sluicing constructions in English

- a. Jack bought something, but I don’t know **what** ~~{he bought #}~~.
- b. Peter was talking with someone, but I don’t know **(with) who** ~~{he was talking #}~~.
- c. They want to hire a senior, but I don’t know **why** ~~{they want to hire a senior #}~~.

- To verify this idea, it’s necessary to examine whether meta *why* questions pattern with *why* sluicing cases with respect to some properties such as island sensitivity and scope ambiguity.

⊞ However, it doesn’t seem easy to directly compare these two constructions since *way* ‘why’ in the meta *why*-questions is not part of its antecedent.

But as shown in Woods & Vicente (2021) for English meta *why* questions, ‘*why* stripping’ like (10) can be an alternative for verifying the presence of a silent structure in the Korean equivalent.

(10) *Why* stripping in English

A: Sally is here.

B: Why Sally? [= Why is Sally, specifically, here (as opposed to some other person?)]

Woods & Vicente (2021: 6)

- Yoshida et al. (2015) suggest a couple of diagnostics, such as variable binding effects, scope effects and island effects, etc., all of which demonstrate that sluicing differs from stripping.
- Adopting Yoshida et al.’s suggestions, Woods & Vicente (2021) compare what they call ‘meta *why* fragments’ like (11) with *why* stripping and *why* sluicing in terms of those diagnostics.
- Observe that ‘meta *why* fragments’ are different from ‘meta *why* questions,’ in that while only a *wh*-word survives in the latter, a *wh*-word and a non-*wh*-constituent survive in the former.

(11) Meta *why* fragment in English

A: Is Sally here?

B: Why Sally? [= Why are you asking this about Sally specifically?]

By adopting the same logic and diagnostics as those in Woods & Vicente, let’s see whether Korean meta *why* fragments like (12) are closer to *why* stripping in (13) or *why* sluicing in (14).

(12) Meta *why* fragment in Korean

A: Ellie-ka            yeki-ey            iss-ni?  
E.-Nom            here-Loc            exist-Int  
‘Is Ellie here?’

B: Way Ellie?

why E.            [= Why are you asking this about Ellie specifically?]

(13) *Why* stripping in Korean

A: Ellie-ka            yeki-ey            iss-e.  
E.-Nom            here-Loc            exist-Decl  
‘Ellie is here.’

B: Way Ellie?

why E.            [= Why is Ellie, specifically, here (as opposed to some other person)?]

(14) *Why* sluicing in Korean

Ellie-ka    Jenny-lul            cohaha-nuntey,    na-nun    way-inc-i-lul            molun-ta.  
E.-Nom    J.-Acc            like-but            I-Top    why-Int-Acc            not.know-Decl  
‘Ellie likes Jenny, but I don’t know why’

## #1. Variable binding effects

### English

- As shown in Yoshida et al., a pronominal variable contained in remnants in the *why* stripping construction in English can be bound, as in (15), and this kind of ‘connectivity effect’ can be taken to be a strong support for the ellipsis view (cf. Merchant 2001).

(15) Variable binding effect in English *why* strippingA: No linguist<sub>i</sub> here recommended her<sub>i</sub> own book.B: Why her<sub>i</sub> own book?

Woods &amp; Vicente (2021: 9)

Korean

- Likewise, the same connectivity effect can be found in the Korean *why* stripping construction, as in (16); what's more, it's also observed in the Korean *why* fragment construction, as in (17).

(16) Variable binding effect in Korean *why* stripping

A: yeki-uy [amwu enehakca]<sub>i</sub>-to caki<sub>i</sub>-uy chayk-ul chwuchenha-ci ahn-ass-e.  
 here-Gen any linguist-even self-Gen book-Acc recommend-Link not-Past-Decl  
 'No linguist<sub>i</sub> here recommended his<sub>i</sub> own book.'

B: Way ku<sub>i</sub>-uy chak-ul?  
 why he-Gen book-Acc  
 'Why his<sub>i</sub> own book?'

(17) Variable binding effect in Korean meta *why* fragment

A: yeki-uy [etten enehakca]<sub>i</sub>-ka caki<sub>i</sub>-uy chayk-ul chwuchenha-ss-ni?  
 here-Gen any linguist-Nom self-Gen book-Acc recommend-Past-Int  
 'Did any linguist<sub>i</sub> here recommend his<sub>i</sub> own book?'

B: Way caki<sub>i</sub>-uy chak-ul?  
 why self-Gen book-Acc

[= Why are you asking me about [each of those linguists]' own book specifically?']

- Notice that this variable binding effect can't apply to *why* sluicing constructions, since the remnant doesn't accompany any other constituent than a *wh*-phrase.

**#2. Island effects**English

- Yoshida et al. show that the remnants in the *why* stripping construction in English do not exhibit an island effect, as in (18).

(18) Island effect in *why* stripping (NO)

A: Katie loves girls who learn Italian.

B: Why Italian?

- As reported in Woods & Vicente (2021), by contrast, the sluice in the *why* sluicing construction is sensitive to an island constraint (esp., the complex NP constraint), as in (19).<sup>3)</sup>

(19) Island effect in *why* sluicing (YES)

Katie loves a girl who is learning Italian, but I don't know why.

[= Why Katie loves a girl who is learning Italian]

[≠ Why that girl is learning Italian]

- More importantly, as given in (20), the remnants in the meta *why* fragment in English are just

3) Note that other sluices than *why* in English have been shown to be insensitive to island constraints in the literature (e.g., Merchant 2001; Fox & Lasnik 2003, *a.o.*).



as immune to the same constraint as those in the *why* stripping construction.

(20) Island effect in meta *why* fragment (NO)

A: Who loves the girl who is learning Italian?

B: Why Italian? [= Why are you asking me about Italian specifically?]

Woods & Vicente (2021: 10)

### Korean

- Just like the English equivalent, the *why* stripping construction in Korean is not subject to the island constraint, as in (21).

(21) Island effect in *why* stripping (NO)

A: Jenny-ka	hankwuke-lul	paywu-nun	ai-tul-ul	cohahay.
J.-Nom	Korean-Acc	learn-Rel	chil-Pl-Acc	like.Decl

'Jenny likes kids who learn Korean.'

B: Way hankwuke?

why Korean

'Why Korean?'

- However, the *why* sluicing construction in Korean shows island effects, as in (22).

(22) Island effect in *why* sluicing (Yes)

Jenny-ka	hankwuke-lul	paywu-nun	ai-ul	cohaha-nuntey.
J.-Nom	Korean-Acc	learn-Rel	child-Acc	like-but

na-nun way-inci-lul molu-keyss-e.

I-Top why-Int-Acc not.know-Mod-Decl

'Jenny likes a kid who learns Korean, but I don't know why.'

[= Why Jenny likes a kid who learns Korean.]

[≠ Why that kid is learning Korean.]

- The meta *why* fragment in Korean turns out to pattern with the *why* stripping, in that it may violate the same island constraint, as in (23).

(23) Island effect in meta *why* fragment (NO)

A: Nwu-ka	hankwuke-lul	paywu-nun	ai-tul-ul	cohaha-ni?
J.-Nom	Korean-Acc	learn-Rel	chil-Pl-Acc	like-Int

'Who likes a kid who learns Korean?'

B: Way hankwuke?

why Korean [= Why are you asking me about Korean specifically?]

### #3. Scope ambiguity

#### English

- As shown in Woods & Vicente, the *wh*-word *why* always takes wide scope over other quantifier in *why* stripping in English, as in (24).

(24) Scope ambiguity (NO) A:

Everyone hates John. B:

Why John?

A: Because they say he's rude. (why > every)

A': #Tom said he's rude and Mary said he's mean and James said... (\*every > why)

- However, *why* in the *why* sluicing construction scopally interacts with other quantifier in the sentence, taking either wide or narrow scope with respect to the latter, as in (25).

(25) Scope ambiguity in *why* sluicing (OK)

Everybody hates John, but I don't know why. (*why* > every, every > *why*)

- The *wh*-phrase in the meta *why* fragment doesn't exhibit scope ambiguity, as in (26), just like the *why* stripping construction.

(26) Scope ambiguity in *why* sluicing (NO)

A: Does everyone hate John?

B: Why John? [= Why are you asking me about John specifically?]

A: Because you know him well and might be able to confirm what I'm hearing about him.

(*why* > every)

A': #Because Tom said he's rude and Mary said he's mean and James said... (every > *why*)

Woods & Vicente (2021: 11)

### Korean

- Just as in the case of English equivalent, the *wh*-word *why* always invariably takes scope over other quantifier in the Korean *why* stripping construction, as in (27).

(27) Scope ambiguity in *why* stripping (NO)

A: Motun salam-tul-i John-ul silhehay.

every person-PI-Nom J.-Acc hate.Decl

'Everyone hates John.'

B: Way John-ul?

why J.-Acc

'Why John?'

A: waynayhamyen salamtul-i ku-ka mwulyeyha-tay.

because people-Nom he-Nom be.rude-Evid

'Because they said he's rude.' (*why* > every)

A': #Tom-un ku-ka mwulyeyha-tay, kuliko Mary-nun ku-ka piyelha-tay...

T.-Top he-Nom be.rude-Evid and M.-Top he-Nom be.mean-Evid

'Tom said he's rude, Mary said he's mean, and ...' (\*every > *why*)

- On the other hand, the *wh*-sluice in the Korean sluicing construction exhibits scope ambiguity with respect to other quantifier, as in (28).

(28) Scope ambiguity in *why* sluicing (YES)

Motun salam-tul-i John-ul silheha-nuntey, na-nun way-inci molu-keyss-e.

every person-PI-Nom J.-Acc hate.but I-Top why-Int not.know-Mod-Decl

'Everyone hates John, but I don't know why.' (*why* > every, every > *why*)

- But the meta *why* fragment in Korean patterns with the *why* stripping in that there is no scope ambiguity, forcing *way* 'why' to take wide scope.

(29) Scope ambiguity in meta *why* fragment (NO)

A: Motun salam-tul-i John-ul silhehay?

- every person-PI-Nom J.-Acc hate.Int  
 ‘Does everyone hate John?’
- B: Way John-ul?  
 why J.-Acc  
 ‘Why John?’ [= why are you asking me about John specifically?]
- A: waynayhamyen ne-nun ku-lul cal al-ko, nay-ka ku-eytayhay  
 because you-Top he-Acc well know-and I-Nom  
 tulun-kes-ul hwakinhaycwul-swu.iss-ul.kes.kat-ase.  
 hear-thing-Acc confirm-can-might-Decl.  
 ‘Because you know him well and you might be able to confirm what I heard.’ (why > every)
- A’: #Tom-un ku-ka mwulyeyha-tay, kuliko Mary-nun ku-ka piyelha-tay...  
 T.-Top he-Nom be.rude-Evid and M.-Top he-Nom be.mean-Evid  
 ‘Tom said he’s rude, Mary said he’s mean, and ...’ (\*every > why)

#### □ Summary

The results of the application of these diagnostics to three apparently similar constructions can be summarized as follows:

Table 1. Test results for *why* stripping, *why* sluicing and meta *why* fragment in English & Korean

Languages	Constructions	Variable binding	Island effects	Scope ambiguity
English	<i>why</i> stripping	OK	NO	NO
	<i>why</i> sluicing	Unattested	OK	OK
	meta <i>why</i> fragment	OK	NO	NO
Korean	<i>way</i> stripping	OK	NO	NO
	<i>way</i> sluicing	Unattested	OK	OK
	meta <i>way</i> fragment	OK	NO	NO

What these results suggest is ...

- Just as shown in Woods & Vicente (2021) for English, meta *why* fragments in Korean are closer to *why* stripping than *why* sluicing.
- Then, given that *why* stripping has been shown to have a silent structure, as in Yoshida et al. (2015), meta *why* fragments can be said to accompany an invisible structure as well.
- Recall that meta *why* fragments are distinguished from meta *why* questions only in terms of whether a *wh*-remnant co-occurs with another constituent; as such, we can conclude that there is a silent structure in the Korean meta *why* questions as well.

#### 4. What does the clause structure look like?

We have seen that meta *why* questions accompany invisible structure, and we should answer the following two related questions about their clause structure.

- What does the clause structure of meta *why* questions look like?
- Where does the *wh*-remnant *way* ‘why’ appear in the structure?

#### □ Clause structure for meta *why* questions in Korean

Let’s begin with the first question.

- Notice first that meta *why* fragments appear to show contradictory properties.
  - ✓ The availability of bound variable readings suggests that the co-occurring element with *way* ‘why’ in the remnants starts out somewhere inside an invisible clause structure.
  - ✓ By contrast, the lack of island effects and the scope rigidity point to an opposite conclusion that the *wh*-word in the meta *why* fragment is generated in a fairly high position.

#### Yoshida et al.’s (2015) structure for English *why* stripping

- Recall, as shown in Sec 4, that meta *why* fragments behave similarly to *why* stripping in both English and Korean, and to capture these apparently contradictory properties in the English *why* stripping construction, Yoshida et al. (2015) suggest the structure given below.

☞ There are two CP layers on the top of TP, and while *why* is directly merged to the Spec-CP<sub>1</sub>, the co-occurring element with *why* moves to the Spec-CP<sub>2</sub> before TP is deleted.

#### (30) Structure for English *why* stripping

[CP<sub>1</sub> Why [C<sub>1</sub>] [CP<sub>2</sub> [DP Sally] [C<sub>2</sub>] [~~TP-TP is here~~]]]

↑ \_\_\_\_\_ |

Yoshida et al. (2015: 355)

#### Woods & Vicente’s (2021) structure for English meta *why* questions

Given the similar behavior between the *why* stripping and the meta *why* fragment/question, one could adopt the same structure for the meta *why* question.

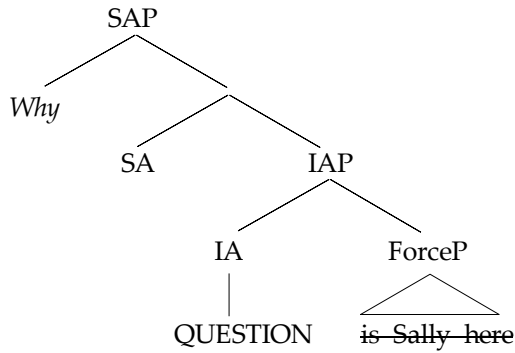
- The similarity between the *why* stripping and the meta *why* fragment can be readily explained with the structure of the sort in (30).
- Nonetheless, the structure seems to have some limitations in capturing the syntactic distribution and properties of meta *why* questions at issue.

☞ More than anything else, it is unclear how to account for the fact that a meta *why* question must take an interrogative as its antecedent.

Thus, Woods & Vicente (2021) propose the structure in (31) for English meta *why* questions.

- They suggest that there are two speech act related projections above Force Phrase (ForceP): i.e., Speech Act Phrase (SAP) and Illocutionary Act Phrase (IAP).
- They assume SAP and IAP above ForceP in order to capture the important properties of meta *why* questions.
  - ✓ First, IAP is a projection for an illocutionary act, especially corresponding to the speaker’s intention, and the QUESTION operator is postulated under IA<sup>0</sup>, which explains why the meta *why* question always takes an interrogative as its antecedent.
  - ✓ Second, SAP (in the sense of Speas & Tenny (2003)) is assumed above IAP, which hosts *why* in its specifier position, and it enables us to capture the invariant wide scope of *why*.

(31) Structure for English meta *why* questions



Woods & Vicente (2021: 14)

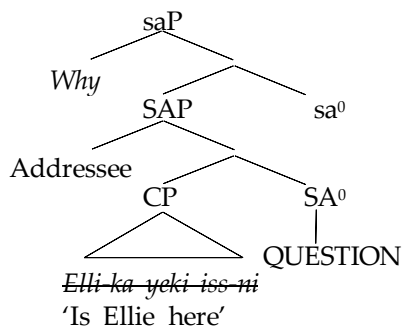
Suggested structure for Korean meta *why* questions

Modifying Woods & Vicente's structure of English meta *why* questions, we suggest the structure in (32) for the Korean equivalent, along with a couple of assumptions listed below.

- First, note that unlike Woods & Vicente's structure, there are two different kinds of speech act projections—i.e., saP for a speaker projected higher than SAP for an addressee (cf. Yim 2016; Hong 2018; Park 2021).
- Second, Spec-saP is assumed to be a base-generation position for *way* 'why', while Spec-SAP is reserved for (an) addressee(s), which may be occupied by a(n exclamative) vocative.
- Furthermore, we assume with Woods & Vicente that there is the QUESTION operator selecting for a proposition as its complement.

⊞ However, slightly departing from them, we argue that it is under SA<sup>0</sup> that hosts the operator, and this idea is consonant with the recent view that the head SA<sup>0</sup> encodes various illocutionary forces, signaled by intonation in Korean (Hong 2018; Park to appear).

(32) Structure for Korean meta *why* questions



□ **Benefits of the proposed structure**

With the proposed structure for Korean meta *why* questions in (32), we can explain not only their syntactic distribution but also important structural properties.

#1. Syntactic distribution explained

- First, the presence of the QUESTION operator under SA<sup>0</sup> enables us to explain why the meta *why* question in Korean always requires an interrogative as its antecedent (e.g., (2), (4)~(6)).
- Second, the QUESTION operator under SA<sup>0</sup> readily predicts that even an interrogative, if

embedded, would fail to license meta *why* questions; this is because the head SA<sup>0</sup> encodes an illocutionary force, which is generally taken to be a ‘root’ phenomenon (e.g., (8)).

## #2. Structure-dependent properties revisited

- It has been shown that meta *why* fragments, which are akin to meta *why* questions, do not exhibit island effects nor narrow scope below other quantifiers.
- These two structure-dependent properties can be readily explained if *way* ‘why’ is directly merged to Spec-saP, as in (32), preempting any chance to scopally interact with a quantifier.

### □ More predictions by the proposed structure

#### #1. Relative ordering between *way* ‘why’ and a vocative

- Under the proposed structure in (32), the Spec-SAP position reserved for an addressee is predicted to be filled with an addressee-related constituent such as a vocative.
- In particular, the vocative is expected to follow ‘why,’ not the other way around, because the Spec-SAP position it occupies is structurally lower than the Spec-saP position.
- Indeed, this prediction is borne out by the meta *why* question with an exclamative vocative, as in (33).

(33) A: Ellie-ka hakkyo-ey hapkyekhay-ss-ni?  
 E.-Nom school-to get.accepted-Past-Int  
 ‘Did Ellie get accepted to the school?’

B: Way inma?

why dude [= ‘Why are you dude asking that question?’]

B’: \*Inma way?

dude why [= (Intended) ‘Why are you dude asking that question?’]

#### #2. Lack of meta *why* readings with *etten iywulo* ‘for what reason’

- Ko (2005) provides an insightful discussion of the syntactic distribution and structural position of *way* ‘why’ and *etten iywulo* ‘for what reason’
- She convincingly shows, based on the contrast in scopal behavior, that while *way* ‘why’ is merged directly to CP, *etten iyulo* ‘for what reason’ starts out in the TP domain.<sup>4)</sup>
- Assuming her analysis to be correct, the current analysis predicts that *etten iyulo* would fail to cause the same interpretation as the meta *why* question.

⊞ Indeed, this prediction turns out to be true, as shown in (34).

4) Ko (2005) observes that *way* ‘why’ always takes wide scope over negation, as in (i), while *etten iyulo* ‘for what reason’ exhibits scope ambiguity with respect to negation, as in (ii). Based on this, she argues that the contrast between the two types of *wh*-adjunct is due to the difference in their structural position.

(i) John-un way Mary-lul cohaha-ci-anh-ni?

John-Top why Mary-Acc like-CI-not-Q

‘What is the reason x such that John does not like Mary?’ (Reason >> Not)

\*‘What is not the reason x such that John likes Mary for x?’ (Reason << Not)

Ko (2005: 893)

(ii) John-un etten iyu-lo Mary-lul cohaha-ci-anh-ni?

John-Top what reason-for Mary-Acc like-CI-not-Q

‘What is the reason x such that John does not like Mary?’ (Reason >> Not)

(?)‘What is not the reason x s.t. John likes Mary for that reason?’ (Reason << Not)

Ko (2005: 894, fn. 27)

- (34) A: Ellie-ka hakkyo-ey hapkyekhay-ss-ni?  
 E.-Nom school-to get.accepted-Past-Int  
 ‘Did Ellie get accepted to the school?’  
 B: #Etten iyulo?  
 what reason [= (Intended) ‘Why are you asking that question?’]

## 5. Concluding remarks

### □ Conclusions

To summarize our discussion so far ...

- It has been shown that there are a novel type of *why* questions, which we call ‘meta *why* questions,’ and they should be distinguished from typical *why* questions.
- This paper has investigated the syntactic distribution and structural properties of the meta *why* questions: in brief, they can only target a genuine interrogative in root context.
- The *wh*-remnant in the meta *why* questions is not a non-constituent unit but a constituent unit with an invisible structure, in favor of the so-called ellipsis approach.
- It has been suggested that there are two different speech act related projections above CP, namely saP and SAP, and that the *wh*-remnant is directly merged to Spec-saP, with the QUESTION operator under SA<sup>0</sup>.
- The distribution and structure-dependent properties of the meta *why* questions have been captured by the proposed structure, while a couple of correct predictions also made by it.

## References

- 이한규. 1997. 한국어 담화 표지어 ‘왜’. 『담화학인지』 4(1): 1-20.
- Fox, D. & H. Lasnik 2003. Successive-cyclic movement and island repair: The difference between sluicing and VP-ellipsis,” *Linguistic Inquiry* 34: 143-154.
- Ginzburg, J. 2012. *The interactive stance: Meaning for conversation*. Oxford: Oxford University Press.
- Haegeman, L. and V. Hill. 2013. “The syntacticization of discourse.” In A. Folli, C. Sevdali and R. Truswell (eds.), *Syntax and Its Limits*, Vol. 48, 370-390. Oxford University Press.
- Hong, Y.-T. 2018. Embeddability of sentence end particles in Korean and speech act projections. *Studies in Generative Grammar*, 28(3): 527-552.
- Ko, H. 2005. Syntax of *why*-in-situ: Merge into [Spec, CP] in the overt syntax. *Natural Language Linguistic Theory* 23: 867-916.
- Merchant, J. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27(6): 661-738.
- Park, B.-S. 2005. Island-insensitive fragment answers in Korean. *Proceedings of the 24th West Coast Conference on Formal Linguistics*, 317-325.
- Park, J. U. 2021. On the syntax of agreement in the left periphery with special reference to honorific agreement. *Language and Linguistics* 93: 49-80.
- Park, J. U. to appear. Pseudo-root imperatives and speech act projections. *Language and Information Society* 47.
- Speas, P. & C. Tenny. 2003. Configurational properties of point of view roles. In D. Sciullo & A. Maria (eds.), *Asymmetry in grammar*, 315-344. John Benjamins.
- Stainton, R. J. 2006. *Words and thoughts: Subsentences, ellipsis, and the philosophy of language*. Oxford: Oxford University Press.
- Stanley, J. 2000. Context and Logical Form. *Linguistics and Philosophy* 23(4): 391-434. doi:10.1023/A:1005599312747.
- Weir, A. 2014. Fragments and clausal ellipsis. Doctoral dissertation, University of Massachusetts, Amherst.
- Woods, R. & L. Vicente. 2021. Metacommunicative-*why* fragments as probes into the grammar of the speech act layer. *Glossa: a journal of general linguistics* 6(1): 84. 1-32.

- Yim, C. 2016. The unembeddable allocutive marker *-yo* in Korean and the syntax-pragmatics interface. *Studies in Generative Grammar*, 26(4): 395-406.
- Yoshida, M., C. Nakao & I. Ortega-Santos. 2015. The syntax of *why*-stripping. *Natural Language and Linguistic Theory* 33: 323-370.



# A New Type of English Secondary Predicate Construction

---

## Juwon Lee

Jeonju University, Republic of Korea

juwonlee@jj.ac.kr

<https://sites.google.com/view/juwonlee19>

한국생성문법학회, 한국언어정보학회  
가을 온라인 공동학술대회

2022.11.12 (토)

1

## 1. Introduction (1/3)

- **Topic:**

English sentences like *John **built** the house **big***

- **Question:**

Are they a **resultative** or **depictive** construction?

- **A proposal:**

They would constitute a **third category** distinct from **resultative** or **depictive constructions**.

## 1. Introduction (2/3)

- The secondary predicates can serve as a **resultative predicate** as in (1a) or **depictive predicate** as in (1b)

- (1) a. Bill hammered **the metal** flat.  
 b. Bill ate **the fish** raw.

- The sentences like (2) have not been studied in the literature to my best knowledge.
- **Boas (2003)** provides a comprehensive classification of English resultative constructions extracted from **the BNC** (British National Corpus), but sentences like (2) are not reported in the study.

- (2) a. Tom drilled **a hole** deep.  
 b. Tom built **a house** big.

3

## 1. Introduction (3/3)

- *Build-big* constructions:

- (2) a. Tom drilled **a hole** deep.  
 b. Tom built **a house** big.

- The sentences in (2) appear to have **the same structure** as that of the resultative and depictive constructions: roughly, [NP + V + NP + AP].
- These commonalities between (1) and (2) suggest that they all belong to the same category (namely, **secondary predicate construction**).
- **A question** is whether the sentences in (2) are **resultative** or **depictive**.
- They have some **properties** of **resultative** or **depictive constructions**.

## 2. Corpus Data (1/4)

- *Build-big* constructions in **the Web**:

(3) a. And I *carved a hole big enough*.

(<https://www.bruwriter.com/the-little-green-penguin.html>)

b. ... and they *carved a ship big enough for fifty penguins*.

(<https://www.bruwriter.com/the-little-green-penguin.html>)

c. It's a consequence of having *invented a technology powerful enough to alter an entire species*.

(<https://www.media.mit.edu/articles/gene-drives-could-stop-the-world-s-oldest-problems-kevin-esvelt-wants-to-make-sure-they-don-t-start-any/>)

d. Only when you have *digged a hole deep enough for Shihong* that would make his escape impossible, you should contact me.

(<https://www.wuxiaworld.co/LEGEND-OF-THE-GOLDEN-TIGRESS/1939972.html>)

e. I *brewed the coffee strong*, the way I like it in order to get the full flavor of the coffee and to see if it had a bitter flavor to it.

([https://www.walmart.com/reviews/product/23555013?reviews\\_limit=10&reviewId=56944877](https://www.walmart.com/reviews/product/23555013?reviews_limit=10&reviewId=56944877))

5

## 2. Corpus Data (2/4)

- *Build-big* constructions in **the COCA**:

(4) a. ? Garfield: I don't think they've ever *baked a cake big enough*. I love to get cake-faced on my birthdays.

b. When setting concrete pier forms in the ground, *dig the holes large enough to allow room for side-to-side adjustment*.

c. They haven't *invented the hat big enough to pull that rabbit out of*

d. I design illusions and *construct the apparatus necessary for performing them*.

e. Brew the coffee strong please.

f. ...and she figures they must be *baking the pie fresh*, because it's taking forever.

g. ...an English man who was *building the buggies cheaper than anyone else*.

## 2. Corpus Data (3/4)

- *Build-big* constructions in **the iWeb**:
- (5)
- a. *Dig a hole* larger and deeper than the container, add a handful of bonemeal and some rotted manure and mix into the soil.
  - b. *Dig a hole* large enough to hold the plants root system.
  - c. Use a trowel to *dig a hole* big enough to accommodate the roots of the plant.
  - d. Mourinho is likely to make multiple defensive signings this summer as he looks to *build a squad* capable of challenging for the Premier League title.

7

## 2. Corpus Data (4/4)

- **Three patterns** of *build-big* constructions:
- (6)
- a. V NP AP
  - b. V NP AP *enough (to V)*
  - c. V NP A-er (*than ...*)
- **V**: *drill, carve, dig, build, invent, brew, bake, construct* (**creation verbs**)
  - **NP**: *a hole, a house, a ship, a technology, the coffee, a cake, the holes, the hat, the apparatus, the pie, the buggies, a squad* (**product**)
  - **A**: *deep, big, powerful, strong, large, necessary, fresh, cheaper, larger and deeper, large, capable* (**states**)

### 3. Property #1: Change of State (1/2)

- Resultative constructions: change of state
- (7) a. He hammered **the metal** flat.  $\Rightarrow$  The metal became flat.  
 b. He wiped **the table** clean.  $\Rightarrow$  The table became clean.  
 c. He followed Lassie free of his captors.  $\Rightarrow$  He became free of his captors.
- Depictive: no change of state
- (8) a. Bill ate **the fish** raw.  $\nRightarrow$  The fish became raw.  
 b. Bill ate **the fish** naked.  $\nRightarrow$  Bill became naked.
- BB: no change of state
- (9) a. Tom drilled **a hole** deep.  $\nRightarrow$  A hole became deep.  
 b. Tom built **a house** big.  $\nRightarrow$  A house became big.

9

### 3. Property #1: Change of State (2/2)

- Resultative constructions: change of state
- (10) a. **The ice** froze solid.  $\nRightarrow$  The ice became solid.  
 b. **The water** froze solid.  $\nRightarrow$  The water became solid.
- Are BB constructions like those in (10)?
  - But a fundamental difference:
    - Resultative: **the same entity**: water (not solid) = ice (solid)
    - “Bring the water” = “Bring the ice”
    - “The president was poor at that time.”
  - BB construction: **not the same entity**: ground  $\neq$  hole (deep)

### 3. Property #2: [V N A] vs. [V A N] (1/1)

- Resultative constructions : V N A  $\neq$  V A N

(11) a. He hammered **the metal** flat.  $\neq$  He hammered the flat **metal**.  
 b. He wiped **the table** clean.  $\neq$  He wiped the clean **table**.

- Depictive: V N A = V A N

(12) a. **Bill ate the fish** raw. = Bill ate the raw **fish**.  
 b. **Bill ate the fish** naked. = **Bill**, who was naked, ate the fish.

- BB: V N A = V A N

(13) a. Tom drilled **a hole** deep. = Tom drilled a deep **hole**.  
 b. Tom built **a house** big. = Tom built a big **house**.

11

### 3. Property #3: AP only at the end of event (1/2)

- Resultative constructions: AP = only at the end of event

(14) a. He hammered **the metal** flat.  
 $\Rightarrow$  The metal was flat only at the end of event.  
 b. He wiped **the table** clean.  
 $\Rightarrow$  The table was clean only at the end of event.

- Depictive: AP  $\neq$  only at the end of event

(15) a. **Bill ate the fish** raw.  
 $\neq$  The fish was raw only at the end of event.  
 b. **Bill ate the fish** naked  
 $\neq$  Bill was naked only at the end of event.

### 3. Property #3: AP only at the end of event (2/2)

- BB: AP = only at the end of event
- (16) a. Tom drilled **a hole** deep.  
 ⇒ The hole was deep only at the end of event.
- b. Tom built **a house** big.  
 ⇒ The house was big only at the end of event.

13

### 3. Property #4: [A N] Creation (1/1)

- Resultative constructions: [A N] Creation
- (17) a. He hammered **the metal** flat. ⇒ The **flat metal** became to exist.  
 b. He wiped **the table** clean. ⇒ The **clean table** became to exist.
- Depictive: No [A N] Creation
- (18) a. Bill ate **the fish** raw. ⇏ The **raw fish** became to exist.  
 b. Bill ate the fish naked. ⇏ **Naked Bill** became to exist.
- BB: [A N] Creation
- (19) a. Tom drilled **a hole** deep. ⇒ A **deep hole** became to exist.  
 b. Tom built **a house** big. ⇒ A **big house** became to exist.

### 3. Property 5: Object existence (1/2)

- Resultative constructions: Object existence

- (20) a. He hammered **the metal** flat. ⇒ **The metal** existed during the hammering event.
- b. He wiped **the table** clean. ⇒ **The table** existed during the wiping event.

- Depictive: Object existence

- (21) a. Bill ate **the fish** raw. ⇒ **The fish** existed during the eating event.
- b. Bill ate the fish naked. ⇒ **Bill** existed during the eating event.

- BB: No object existence

- (22) a. Tom drilled **a hole** deep. ⇏ **The hole** existed during the drilling event.
- b. Tom built **a house** big. ⇏ **The house** existed during the building event.

15

### 3. Property 5: Object existence (2/2)

- Resultative constructions: Object existence

- (23) a. **The ice** froze solid. ⇏ **The ice** existed during the freezing event.
- b. **The water** froze solid. ⇏ **The water** existed during the freezing event.

- But in (20) **the same entity** is referred to as ‘the ice’ or ‘the water.’ That is, the entity existed during the freezing event.



## 4. Pseudo-resultative construction (1/1)

- Pseudo-resultative constructions (Levinson 2010):

- (24) a. Mary **braided** her hair tight. ⇒ tight **braid**  
 b. Mary **piled** her cushions high. ⇒ high **pile**  
 c. Mary **sliced** the bread thin. ⇒ thin **slice**

- Pseudo-resultative predicates do not modify any ‘word’ in the syntax. They modify the **root** of the verb (Levinson 2010).
- But the secondary predicates of BB constructions modify **the objects**.
- The roots of the creation verbs like *drill* or *build* cannot be modified by the secondary predicates.

17

## 5. Conclusion: Properties (1/3)

	Resultative	BB	Depictive
1. Change of state	1	0	0
2. [V N A] = [V A N]	0	1	1
3. AP only at the end of event	1	1	0
4. [A N] Creation	1	1	0
5. Object existence	1	0	1

## 5. Conclusion: Aspectual differences (2/3)

- Resultative constructions: *Bill hammered **the metal** flat.*



- Depictive constructions: *Bill ate **the fish** raw.*



- BB constructions: *Tom built **a house** big.*



19

## 5. Conclusion: A third category (3/3)

- BB constructions share **some properties** with **resultative** or **depictive constructions**.
- BB constructions would constitute **a third category** distinct from **resultative** or **depictive constructions**.

# Korean EFL Learners' Underuse of the Verb-Particle Construction

Noh Kang San  
Korea University



---

Introduction

---

Theoretical Background

---

Methodology

---

Results

---

Conclusion



# 1. Introduction

4

The goal of the present study is to detect Korean EFL learners' underuse of the English verb-particle constructions (phrasal verbs).

- (1) a. Do you **exercise** a lot? (single-word verb)
- b. Do you **work out** a lot? (verb-particle construction)

## Research Questions

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

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

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

## 2. Theoretical Background

## What are Verb-Particle Constructions?

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

(Biber et al., 2002)

take (lexical verb) + out (adverbial particle)



## What is Underuse Phenomenon?

### Dagut and Laufer (1985)

Hebrew EFL learners turned out to prefer single-word verbs over verb-particle constructions.

(2) We didn't believe that John could ever \_\_\_\_ his friends.

(let down, solve, disappoint, carry on)



## 3. Methodology

### Distinctive Collexeme Analysis

: It is applied to identify a lexeme that shows a strong attraction to one member of the pair as opposed to the other.

Thus, it can be applied to *grammatical alternation*.

→ An alternation concerning verb-particle constructions

## The Alternation between [V - OBJ - Prt] and [V - Prt - OBJ]

: The case of verb-particle constructions can be the instance of an alternation.

- (3) a. John **picked up** the book. (V - Prt - OBJ)  
 b. John **picked** the book **up**. (V - OBJ - Prt)

(Gries and Stefanowitsch 2004: 110)



## How to Perform Distinctive Collexeme Analysis

The two-by-two contingency table is needed. expected frequency

observed frequency

	make up	¬ make up	Row Totals
[V - Prt - OBJ]	652 (603)	2,668	3,320
[V - OBJ - Prt]	18 (66)	348	366
Column Totals	670	3,016	3,686





## Learner Corpus Data

Corpus	Words
Gachon Learner Corpus (GLC)	2,507,899
Yonsei English Learner Corpus (YELC)	1,082,295
<b>Sum</b>	<b>3,590,194</b>



## The List of Verb-Particle Constructions

The Top 100 Phrasal Verb Lemmas in the BNC

(Gardner and Davies 2007)

From 100 phrasal verb lemmas, 10 verb-particle lemmas were removed.

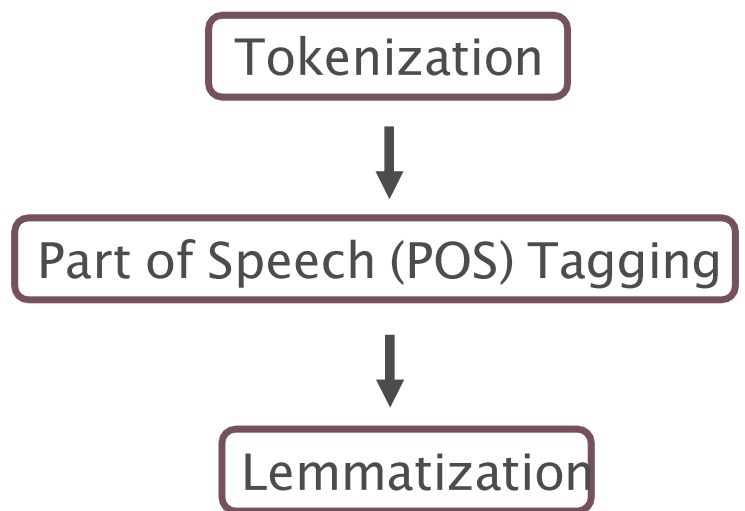
go on ↔ continue

sit down ↔ ?



## Extracting Verb-Particle Lemmas

The selected 90 verb-particle lemmas were extracted from the GLC and YELC by using NLTK (Natural Language Tool Kit).

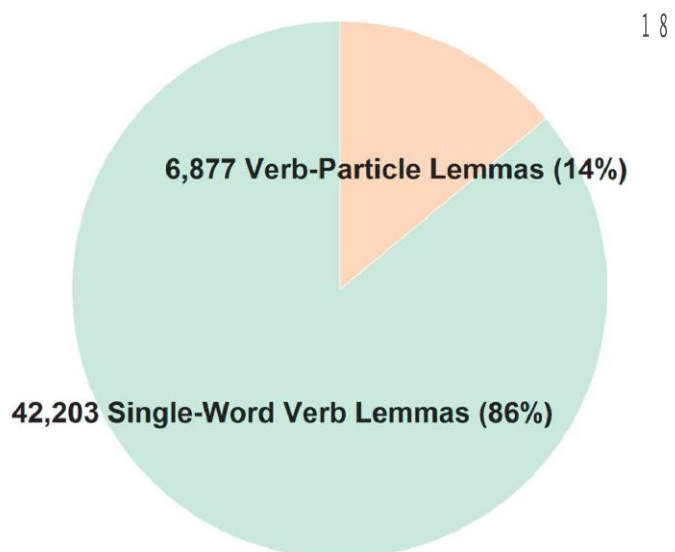


## Extracting Verb-Particle Lemmas from Learner Corpora

- 16
- |  |                                     |
|--|-------------------------------------|
| 1) Verb + Particle (+ X)               | ex) pick up (+ something)           |
| 2) Verb + X + Particle                 | ex) pick it up                      |
| 3) Verb + X + X + Particle             | ex) pick your coat up               |
| 4) Verb + X + X + X + Particle         | ex) pick your blue coat up          |
| 5) Verb + X + X + X + X + Particle     | ex) pick your big blue coat up      |
| 6) Verb + X + X + X + X + X + Particle | ex) pick your very big blue coat up |
- The list shows six patterns for verb-particle constructions, each with an example. The patterns range from a simple verb and particle to a verb followed by up to six 'X' placeholders and a particle. The examples use the verb 'pick' and the particle 'up' with various objects and modifiers. A horizontal line is positioned above the list, and a vertical grey bar with a plus sign is on the right side.

## 4. Results

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



The Distribution of Verb-Particle and Single-Word Lemmas from the GLC and YELC

From particle lemmas, only 17 of them were more frequently used than their single-word counterparts.

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

19

17 Verb-Particle Lemmas More Used than Single-Word Counterparts

20

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

Type	Count (%)
[V - Prt - OBJ]	3,320 (90.1)
[V - OBJ - Prt]	366 (9.9)
Sum	3,686

## The Prediction of Distinctive

21

### Collexeme Analysis

The [V - Prt - OBJ] Construction ↔ Figurative Verb-Particle Lemmas

ex) give up

The [V - OBJ - Prt] Construction ↔ Literal Verb-Particle Lemmas

ex) give back

## The Results of Distinctive Collexeme Analysis

22

V - Prt - OBJ	p.value	V - Prt - OBJ	p.value
make up	4.20E-12	go over	0.3186
take out	7.53E-06	pick out	0.3186
give up	1.35E-05	turn down	0.3336
go through	2.26E-05	carry out	0.3436
break down	9.03E-05	carry on	0.4157
find out	0.0001918	hold on	0.4254504
put on	0.0008229	bring in	0.4575
get on	0.003829	hold up	0.4575
bring about	0.004728	turn over	0.4575
break up	0.02502	pick up	0.464
bring up	0.0281	set out	0.5064
take on	0.05177	hold out	0.5064
set up	0.06304	break off	0.5064
look up	0.1027	bring down	0.6386
get through	0.1027	take up	0.7269
put off	0.1365	set about	0.7398
give in	0.1468	take over	0.787
come through	0.1832	give out	0.787
put in	0.1886	turn off	0.9223

V - OBJ - Prt	p.value
give back	2.34E-124
put back	1.67E-09
get back	1.82E-08
take down	0.0001794
make out	0.0005689
bring back	0.001569
put up	0.002993
take off	0.1381
put down	0.1536
hold back	0.1751
put out	0.1913
take back	0.1913
set off	0.3133
take in	0.4773
bring out	0.9305

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

Type	Number of Lemmas (%)	Number of Tokens (%)	Type-Token Ratio
Literal	41 (46)	2,982 (43)	1.4
Figurative	48 (53)	3,761 (55)	1.3
Completive	1 (1)	134 (2)	0.7
Sum	90	6,877	-

## 5. Conclusion

## Results

Verb-particle constructions were less frequently used than their single-word counterparts.

The [V - OBJ - Prt] construction was underused compared to the [V - Prt - OBJ] construction.

No specific preference for figurative verb-particle lemmas was observed.



## Implications

1. Analyzed the two representative Korean EFL learner corpora (3,590,194 words in total)
2. Applied distinctive collexeme analysis to show which construction is being underused compared to the other



## References

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

Dagut, M. and B, Laufer. (1985). Avoidance of Phrasal verbs: A Case for Contrastive Analysis. *Studies in Second Language Acquisition* 7, 73–79.

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

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

Thank you for listening!



## 딥러닝 한국어 모델은 관형사절 속 공백을 포착하는가?

이지은\*·이규민 2)·송상현3)

### 차 례

1. 서론
2. 목표 구문의 선택 및 기술
3. 실험 방법
  - 3.1 딥러닝 한국어 모델의 학습
  - 3.2 추론 결과의 분석
4. 실험 결과
  - 4.1 목표 구문의 학습 여부
  - 4.2 개별 데이터의 분석 및 예시
5. 논의점
6. 결론

### □ 국문 요약

본고는 딥러닝 한국어 모델이 공백이 있는 관계 관형사절(공백 관계절)과 그 밖

\* 고려대학교 BK21 언어학교육연구팀 연구교수

2) 고려대학교 언어정보연구소 연구원

3) 고려대학교 언어학과 부교수

의 관형사절(비공백 관계절)을 구별하는지 살펴본 연구이다. 궁극적으로는 딥러닝 한국어 모델이 관형사절 속 공백을 포착하는지 살펴보려고 했다. 연구 결과는 다음과 같다. 첫째, 딥러닝 한국어 모델이 공백 관계절과 비공백 관계절을 최소 89%에서 최대 100%의 정확도로 분류하였다. 둘째, 히트맵 분석 결과 공백 관계절의 경우 머리명사보다 관계절 속 서술어에 더 많은 어텐션(Attention)이 놓였다. 이러한 결과들은 딥러닝 한국어 모델이 관계절 속 공백을 포착한다는 직접적인 단서는 되지 못하므로 후속 연구가 필요하다.

주제어: 한국어 딥러닝 모델, 공백, 관계절, 히트맵, 어텐션

## 1. 서론

본고는 딥러닝 한국어 모델이 공백이 있는 관계 관형사절을 학습할 수 있는지를 분석하여 궁극적으로 딥러닝 한국어 모델이 관계절 속 공백을 포착할 수 있는지 살펴보는 것을 목적으로 한다. 더불어 딥러닝 한국어 모델이 무엇을 기준으로 공백이 있는 관계 관형사절을 그 밖의 관형사절과 구별하는지 살펴보려고 한다.

한국어의 관형사절은 크게 관계 관형사절과 동격 관형사절로 나뉜다. 전자의 경우 관형사절 안에 공백 성분이 있는데, 해당 성분은 머리명사와 공지시의 관계를 갖는다. 각각의 예를 살펴보자.

(1) 가. 영희가 읽은 책

나. 철수가 온 사실

먼저, 관계 관형사절의 예인 (1가)에서 관형사절 속 서술어 '읽다'는 두 자리 서술어로 행동주와 대상을 논항으로 요구하는데 관형사절 안에 행동주만 '영희'로 명시되어 있다. 대상 역할의 목적어는 삭제되어 있는데 관형사절 속 서술어 '읽다' 및 머리명사 '책'이 제공하는 구조적, 의미적, 화용적 정보에 의해 삭제된 목적어가 '책'임을 알 수 있다. 다음으로 동격

관형사절의 예인 (1나)를 살펴보면 (1가)에서와 달리 관형사절의 서술어가 요구하는 성분이 관형절 속에 모두 나타나 있다. 관형사절 속 서술어인 '오다'는 한 자리 서술어로 행동주인 '철수'가 주어로 나타나고 있다. 이때 관형사절은 머리명사인 '사실'과 동격의 관계를 갖는다. 즉, (1가)와 (1나)는 관형사절 속에 머리명사와 공지시적인 성분이 삭제되어 있는지의 여부로 구별된다.

그러나 관계 관형사절과 동격 관형사절 사이에는 이 둘 중 하나에 속하는 것으로 보기 어려운 예들이 연속성을 이루며 무수히 존재한다. 아래 예를 살펴보자.

(2) 가. 영희가 읽은 책

나. 걷는 운동

다. 생선이 타는 냄새

라. 철수가 온 사실

위의 예에서 (2나)는 관형절 속에 생략된 성분이 있다는 점에서 앞서 살펴본 관계 관형절의 예인 (2가)와 유사하다. 관형절 속 서술어인 '걷다'는 한 자리 서술어로 행동주를 요구하는데 관형절 안에 행동주에 해당하는 성분이 나타나 있지 않다. 그러나, 관계 관형사절과 달리 머리명사 '운동'이 관형사절 속 공백 성분과 공지시적이지 않다. 같은 맥락에서 (2다)는 동격 관형사절의 예인 (2라)와 유사한데, (2라)에서와 같이 머리명사가 관형사절 속 서술어에 대해 주어나 목적어와 같은 문법적 기능을 가지지 않는다. 그러나, 동격 관형사절과 달리 관형사절이 머리명사와 동격의 관계에 있는 것으로 보기 어렵다. 전형적인 관계 관형사절과 동격 관형사절 모두에서 조금씩 벗어나 있는 (2나, 다)와 같은 예에서 (2가)를 구별하기 위해서는 머리명사의 의미적 특성과 함께 관형사절 속 서술어가 요구하는 논항의 개수 및 의미역 정보가 요구된다.

(2가)와 같은 공백이 있는 관계 관형사절을 구별해 내는 데에는 머리명사와 관형사절 속 서술어의 의미/구조적 특성과 함께 의미적/화용적 지식이 요구되기도 한다. 아래 예는 관계 관형사절의 예인데 머리명사가 관형사절 속 서술어의 논항 지위를 갖지 않는다. 관형사절 속 서술어 '먹다'와 '사다'는 두 자리 서술어로 행동주와 대상을 요구하는데 머리명사 '식당'과 '잔돈'은 각각 장소와 수단을 나타내는 부가어의 역할을 한다. 따라서, (3)에 제시된 예를 관계 관형사절로 인식하려면 '먹다'라는 사건과 '사다'라는 사건에 '식당'과 '잔돈'이 관련될 수 있음을 알아야 한다. 의미적 혹은 화용적 지식이 개입되는 것이다.

(3) 가. 철수가 먹은 식당

나. 담배를 산 잔돈

같은 맥락에서 아래 예를 살펴볼 수 있다. (4)에 제시된 관형사절의 예는 머리명사가 관형사절 안에 공지시 명사구를 갖는 예로 볼 수도 있고 관형사절 안에서 주제어로 기능하는 것으로 분석될 수도 있다.

(4) 가. 귀가가 늦어지는 교외생활

나. 취직이 어려운 물리학

(4) 를 관형사절 속에 공지시 명사구가 없는 특이한 관형사절의 예로 보든 공지시 명사구가 주제어인 관계 관형사절로 분석하든 두 경우 모두 관형사절의 성립에 화용적 지식이 개입된다. 직장이 도심에서 먼 경우 일 이 끝난 후 집에 가기 위해 길에서 보내는 시간이 길어지고 결과적으로 귀가 늦어질 수 있다. 또한 물리학의 경우 직접적으로 관련된 직장의 폭이 크지 않을 수 있다. 이러한 화용적 지식이 화자와 청자 사이에 공유되는 경우 (4가, 나)가 화자에 의해 발화되고 청자에 의해 수용될 것이다. 본고

에서는 (4)에서와 같이 머리명사가 관형사절 안에서 주제어로만 환원되는 경우는 공백이 있는 관계 관형사절로 보지 않는다. 이는 연재훈(2012) 등의 입장을 따른 것으로 (4)의 경우 (3)과 달리 머리명사가 관형사절 속 서술어와 의미적 관련성을 지니지 않는다는 점에서 구별된다.

위에서 살펴본 바를 정리하면, 한국어 관형사절의 예에서 공백이 있는 관계 관형사절을 식별해 내기 위해서는 머리명사의 의미적 특성, 관형사절 속 서술어의 자릿수와 논항의 의미역 정보 및 화용적 지식이 필요한 것을 알 수 있다. 이를 염두에 두고 본고는 아래 질문에 대한 답을 찾아보고자 한다.

첫째, 딥러닝 한국어 모델이 공백이 있는 관계 관형사절을 학습할 수 있는가?

둘째, 공백이 있는 관계 관형사절의 학습이 가능하다면 학습은 무엇을 바탕으로 이루어지는가?

셋째, 딥러닝 한국어 모델이 관형사절 속 공백을 포착한다고 볼 만한 직접적인 증거가 있는가?

이를 위해 먼저 딥러닝 한국어 모델에게 공백이 있는 관계 관형사절을 학습시킨 후 해당 모델이 공백이 있는 관계 관형사절을 다른 관형사절과 구별하는지를 살펴본다. 이후에 딥러닝 한국어 모델이 무엇을 기준으로 공백이 있는 관계 관형사절을 판별했는지를 살펴보기 위해 히트맵 자료를 분석한다. 마지막으로 결과를 종합적으로 검토하여 딥러닝 한국어 모델이 관형사절 속 공백의 존재를 포착한다고 볼 수 있는지 살핀다.

이어지는 논의는 다음과 같다. 2장에서는 딥러닝 한국어 모델을 학습시킬 목표 구문을 선택하고 해당 구문의 특성을 선행연구를 중심으로 기술한다. 3장과 4장에서는 각각 실험 방법 및 실험 결과를 기술한다. 5장에서는 앞서 제시한 세 가지 질문을 중심으로 실험 결과의 함의점을 다

룬다. 6장에서는 연구를 정리하고 본고의 한계 및 남은 문제를 언급한다.

## 2. 목표 구문의 선택 및 기술

본장에서는 딥러닝 한국어 모델에게 학습시킬 목표 구문인 공백이 있는 관계 관형사절에 대해 기술하고자 한다. 본고에서 공백이 있는 관계 관형사절의 예로 다룬 것은 (5가, 나)이다.

- (5) 가. 영희가 읽은 책
- 나. 철수가 먹은 식당
- 다. 귀가가 늦어지는 교외생활
- 라. 걷는 운동
- 마. 생선이 타는 냄새
- 바. 철수가 온 사실

이어지는 논의에서는 관형사절의 하위 범주에 대해 알아본 후 관계 관형사절을 식별하는 기준을 제시한 후 본고에서 (5라-마)와 같은 예를 공백이 있는 관계 관형사절로 분류하지 않은 이유에 대해 설명하고자 한다.

먼저, 한국어 관형사절의 하위 범주에 대해 살펴보겠다. 선행 연구들에서 관형사절을 분류한 방식을 살펴보면 (6)과 같다.

- (6) 가. 명사구 수식 구성
- 나. 관계 관형사절, 연계절, 동격 관형사절
- 다. 공백이 있는 관계 관형사절, 공백이 없는 관계 관형사절,  
        동격 관형사절

(6가)는 한국어의 관형사절을 통사적으로 구별되는 하위 범주를 가지

지 않는 단일한 명사구 수식 구성으로 보는 입장이다. 이러한 입장은 Matsumoto (1988) 및 Comrie (1998)의 범언어적 연구 결과에 기반한다. 해당 연구들에서는 관계 관형사절이 동격 관형사절과 구별되지 않는 언어의 예 중 하나로 한국어를 들고 있다. 문숙영(2012)에서도 한국어에 공백이 있는 관계 관형사절의 예가 있음을 인정하면서도 단지 공백의 유무만으로 관계 관형사절과 동격 관형사절이 구별되지 않는 점을 지적한 바 있다. (6나)는 한국어의 관형사절의 하위 범주로 관계 관형사절, 연계절, 동격 관형사절을 상정한다. 각 하위 범주의 전형적인 예는 (5가), (5마), (5바)와 같다. 이러한 입장을 취하는 연구에는 이흥식(1990), 김지은(2002), 홍윤기(2010) 외 다수가 있다.<sup>1)</sup> 마지막으로 (6다)는 한국어의 관형사절의 하위 범주로 관계 관형사절과 동격 관형사절을 설정한다. 다른 연구들에서 연계절로 다루어진 것은 관계 관형사절의 일종으로 보고 공백이 없는 관계 관형사절로 분류하였다. 이는 연계절 구성에서도 관형사절이 머리명사의 외연을 좁히는 기능을 하는 것으로 보았기 때문이다. 관련 연구에는 박형진(2015)가 있다.

본고에서는 아래 (7)을 기준으로 공백이 있는 관형사절을 식별하였다.

- (7) 가. 관형사절의 내용이 머리명사에 정보를 제공함으로써  
 머리명사의 외연을 좁히거나 머리명사에 추가적인 정보를  
 더함.  
 나. 머리명사와 공지시 관계에 있는 명사구가 관형사절에 안에  
 공백으로 존재함.  
 다. 관형사절이 관형사형 어미에 의해 표시되어 있음.

---

1) 본고에서 사용한 술어인 '연계절'은 김지은(2002)에서 '연계관형절'로 부른 것에 기반한 것이다. 동일한 구성을 이흥식(1990)은 '의사관계절'로, 홍윤기(2010)은 '연결관형사형'로 분류하였다.

라. 머리명사와 공지시 관계에 있는 명사구가 관형사절 안에서  
주제어 이외의 문법 기능을 가짐.

먼저, (7가)는 관계 관형사절의 기능에 대한 것이다. 관형사절에 기술된 내용이 머리명사의 외연을 좁히는 경우는 제한적 관계 관형사절로 볼 수 있다. 또한 관형사절에 기술된 내용이 머리명사에 추가적인 정보를 더하는 경우는 비제한적 관형사절로 볼 수 있다. (7나, 다)는 한국어에서 관계 관형사절이 입력되는 일반적인 방식을 기술하고 있다. 즉, 관형사절 속 서술어가 관형사형 어미 '-(으)ㄴ'이나 '-(으)ㄹ'로 표시되어 있고, 머리명사와 공지시 관계에 있는 명사구가 관형사절에서 삭제된다.

위에서 살펴본 관계 관형사절의 식별 기준 중 (8가-다)는 공백이 있는 관계 관형사절을 식별하는 일반적인 기준으로 볼 수 있다. (8가-다)의 기준을 적용하면 아래 예 중 (8가, 나)가 공백이 있는 관형사절로 식별되고, (8마, 바)는 그 밖의 관형사절로 식별된다. (8다, 라)의 경우는 연구자들마다 입장이 다를 수 있다. 본고는 (8라)의 기준을 더해 (8가, 나)만을 공백이 있는 관형사절로 분류하였다.

- (8) 가. 영희가 읽은 책
- 나. 철수가 먹은 식당
- 다. 귀가가 늦어지는 교외생활
- 라. 걷는 운동
- 마. 생선이 타는 냄새
- 바. 철수가 온 사실

먼저, 앞 장에서 살펴본 것과 같이 (8가, 나)의 경우 머리명사가 관형사절 안에서 각각 논항 혹은 부가어로 기능한다. 아래 예문이 보여 주듯이 (8가)의 경우 머리명사 '책'이 관형사절 안에서 목적어로 기능하고, (8나)



의 경우 머리명사 '식당'이 부가어로 기능한다.

(9) 가. 영희가 책을 읽었다.

나. 철수가 식당에서 점심을 먹었다.

이와 달리, (8마, 바)는 머리명사가 관형사절 속에서 논항이나 부가어로 기능하지 않으므로 관계사절 안에 머리명사와 공지시적인 공백이 존재하는 것으로 보기 어렵다. (8다, 라)의 경우는 해당 관형사절의 대당 문장을 설정할 수 있는지에 대해 의견이 갈린다. (8다, 라)를 각각 아래 예문 속 주제어 및 부가어가 관계화된 것으로 볼 경우 (8다, 라) 역시 공백이 있는 관계 관형사절로 볼 수 있다.

(10) 교외생활은 귀가가 늦다.

(11) ?운동으로 걷는다. (박형진 2015: 27)

본고에서는 (8다, 라)를 (8가, 나)의 부류에 포함시키지 않았는데 이유는 다음과 같다. 첫째, (8다)의 경우 (8나)와 달리 머리명사가 관형사절 속 서술어와 어떠한 의미적 관련을 가지지 않는다. 즉, (8나)의 경우 머리명사 '식당'은 관형사절 속 서술어 '먹다'의 논항은 아니지만 의미적으로 관련이 있는 성분인 반면, (8다)의 경우 머리명사 '교외생활'과 관형사절 속 서술어 '늦다'는 이러한 관련성을 보이지 않는다. 아래 (9가)와 (9나)의 경우 머리명사가 관계사절에 속 서술어와 의미적 관련이 있는 것으로 볼 수 있다. 둘째, (8라)의 경우 대당문이 없는 예로 보는 것이 타당해 보인다.<sup>2)</sup> 가령, 박형진(2015)에서는 (8라)의 대당 문장이 (11)과 같이 설정될 수도 있으나 자격을 나타내는 명사구의 경우 (12가)에서와 같이

2) 대당 문장을 설정하기 어려운 관계 관형사절의 예는 문숙영(2012), 박형진(2015), 박소영(2021) 등을 참고할 수 있다.

관계화가 불가능함을 볼 때 (8라)를 공백이 있는 관계 관형사절로 보기 어렵다고 하였다.

- (12) 가. \*회의에 참석한 대표  
 나. 회의에 대표로 참석했다.  
 (박형진 2015: 27)

셋째, 딥러닝 한국어 모델의 학습을 용이하게 하고자 하는 기술적인 측면을 고려하였다. 즉, 공백이 있는 관계 관형사절로 보기에 논란이 없는 예만 포함시킴으로써 딥러닝 한국어 모델의 학습 가능성을 높이고자 하였다.

지금까지 선행연구를 참고하여 관형사절을 공백이 있는 관계 관형사절과 그 밖의 관형사절로 분류하고 각각의 예를 살펴보았다. 이어지는 논의에서는 용어를 간명히 하고자 아래와 같이 용어를 사용하도록 하겠다.

- (13) 가. 관형사절 → 관계절  
 나. 공백이 있는 관계 관형사절 → 공백 관계절  
 다. 그 밖의 관형사절 → 비공백 관계절

### 3. 실험 방법

#### 3.1 딥러닝 한국어 모델의 학습

딥러닝을 포함한 기계학습은 크게 지도학습과 비지도학습 방법으로 이루어진다. 이 중 지도학습은 데이터와 해당 데이터에 대한, 개발 목적에 부합하는 라벨을 같이 제시하여 기계로 하여금 데이터의 패턴을 파악해

새로운 데이터에 대해서도 올바른 라벨을 부착할 수 있도록 훈련시키는 경우를 지칭한다. 본 연구에서는 문장의 형태로 된 데이터에 공백 유무의 라벨을 붙일 수 있기에, 공백 유무 정보가 담긴 다량의 문장 데이터를 학습시켜 새로운 문장이 들어왔을 때 공백 유무를 분별할 수 있는 지도학습 방법을 사용하였다. 지도학습을 통한 분류에는 구체적인 통계 분포를 기반으로 한 여러 알고리즘이 사용되었으나, 최근에는 구체적 통계 분포까지도 기계가 스스로 파악하도록 하는 딥러닝 방식이 실현되었고, 사람의 언어 처리에 이 방법이 이론적 단순성에 비해 높은 성능을 보이는 점이 경험적으로 입증되어 학계 및 산업에서 널리 사용되고 있다. 이에 본 연구에서도 딥러닝 방법론을 적용하여 지도학습을 진행하였다.

구체적으로, 본 연구에서는 BERT 모델(Devlin 외, 2019)을 기반으로 한국어 형태론을 고려해 학습한 한국어 BERT 모델인 KR-BERT(Lee 외, 2020)를 공백 여부를 사람이 부착한 데이터로 파인 튜닝(fine-tuning)하여 딥러닝이 공백 여부를 분류할 수 있는지를 검토하였다. 일반적인 딥러닝 방법으로 학습을 진행하면 학습에 사용된 데이터에 대해서만 작업을 수행할 수 있다. 허나, 최근에는 큰 규모의 데이터를 통해 일반적인 언어 모델을 학습하고, 이를 구체적인 작업을 수행할 수 있도록 비교적 소량의 데이터로 추가적으로 학습시켜 범용성을 높이는 방법이 사용된다. 이렇게 추가적으로 학습하는 과정을 파인튜닝이라 한다. 이러한 기법은 이미 훈련되어 공개된 모델을 파인튜닝한다면 기존의 방법에 비해 적은 비용으로 구체적인 모델 구축이 가능하면서 동시에 더욱 높은 성능의 모델을 획득할 수 있어 자주 사용되는 방법이다. 본 연구에서는 Python 스크립트를 직접 구축해, HuggingFace Hub에 공개된 KR-BERT-char16424 모델을 PyTorch와 HuggingFace Transformers 패키지를 사용하여 파인튜닝과 기계 추론을 진행하였다.<sup>3)</sup>

3) 모델 정보는 이 웹페이지를 참고 : <https://huggingface.co/snunlp/KR-BERT-char16424>

위와 같이 사용한 데이터에는 비공백구문의 예시에 비해 공백구문이 매우 많다는 특징이 있다. 본 연구에서는 공백 유무를 분류하는 과업을 수행하여야 하는데, 이러한 분류 과업을 위해 모델을 훈련시킬 때에는 분류하는 카테고리의 비율을 일정하게 유지하는 것이 좋다. 이에 따라 본 연구에서는 숫자가 적은 비공백구문의 숫자(총 1980개)만큼 공백구문 데이터를 임의 추출하여 1:1 비율로 훈련 데이터셋을 구성하였다.

일반적으로 기계학습을 훈련할 때는 학습데이터를 분할하여, 각각을 훈련과 테스트 용으로 사용한다. 이는 모델의 훈련과 성능 평가에 정확히 동일한 데이터를 사용하면 모델이 일반적인 데이터가 아니라 주어진 데이터에만 높은 성능을 보이는 결과를 낼 수 있기 때문이다. 이를 방지하기 위해 일반적으로 데이터를 임의 분할하여, 모델이 훈련에 등장하지 않았지만 같은 패턴을 지닐 것으로 예상해야 할 데이터를 잘 파악하는지 판단한다. 그러나 본 실험의 학습용 데이터셋은 이렇게 분할하기에는 숫자가 많지 않다. 또한, 비공백구문의 개수로 인해 상대적으로 소량의 데이터를 사용하기 때문에, 임의 추출을 거쳤다고 해도 모델의 결과가 해당 데이터에서만 학습이 가능할 수도 있다는 가능성을 배제할 수 없다.

따라서, 본 실험에서는 1980개의 비공백구문은 공통으로 하되, 1:1 비율을 맞추는 공백구문 데이터를 서로 겹치지 않게 10번 추출하였다. 이를 통해 공백구문 데이터와 비공백구문 데이터가 서로 1:1 비율을 이루는 데이터셋 10개를 준비하였다. 이 중 한 개의 데이터셋은 모델의 성능을 정성적으로 분석하기 위하여 별도로 구분하였다. 나머지 9개의 데이터셋은 동일한 조건 하에 각각의 파인튜닝에 활용되었다. 각 데이터셋은 9:1 비율로 임의 분할되어 훈련 및 테스트에 사용되었다.

모든 파인튜닝은 전체 데이터를 4번씩 볼 수 있도록 에폭을 4로 설정하였다. 한 에폭의 훈련이 완료된 경우 0.1만큼 나누어 둔 테스트용 데이터로 중간 성능을 평가하였다. 성능 평가에는 카테고리의 비율을 고려하지 않고 정답을 맞춘 비율을 나타내는 정확도와 카테고리의 비율을 고려

하여 모델의 성능을 평가하는 F1 점수를 사용하였다. 기계학습은 모델이 데이터에 대해 예측을 진행하고, 이러한 예측값이 얼마나 실제와 어긋나는지를 손실값(loss)으로 계산해, 이를 활용해 학습을 진행한다. 테스트 과정에서 이러한 손실값을 사용하면 모델이 얼마나 실제와 가까운 결과를 내는지 가늠할 수 있다. 따라서, 9개의 파인튜닝 모두 4번의 에폭을 모두 실행하나, 최종적으로는 테스트 과정의 평균 손실값을 기반으로 가장 손실이 적은 모델을 저장하였다. 이러한 과정을 통해 모델이 주어진 데이터 셋에만 치중하여 일반적인 데이터를 처리하지 못하는 과적합 문제를 피하고자 하였다.

### 3.2 추론 결과의 분석

다량의 데이터에서 어떠한 패턴을 찾아내 이를 기반으로 예측을 진행하는 기계학습은 사람이 이해하는 것과 다른 방식의 패턴을 찾아낸다. 이러한 문제로 인해 기계학습 모델, 특히 패턴에 대한 통계학적 가정을 상정하지 않는 딥러닝 모델은 그 예측이 정확하든 부정확하든 구체적으로 어떻게 그러한 예측이 나오게 되었는지 파악할 수 없다. 흔히 블랙박스 문제라고 불리는 이러한 문제로 인해 본 실험의 모델이 공백 구문 여부를 학습하는지에 상관 없이 왜 특정 사례를 공백 구문인지 아닌지로 구분하는 이유는 구체적으로 파악할 수 없다.

하지만 본 연구에서 사용하는 모델의 기반이 되는 BERT 모델을 비롯한 최근의 딥러닝 한국어 모델은 셀프-어텐션(Self-Attention)을 기반으로 하는 Transformer(Vaswani 외, 2017) 모델을 활용한다. 기존에는 자연어 텍스트를 토큰이 차례대로 등장하는 데이터로 파악하여, 이러한 토큰 하나하나를 순차적인 입력으로 받아 처리하는 순환신경망(Recurrent Neural Network)을 기반으로 처리하였지만, 이러한 모델은 학습 과정에서 멀리 떨어진 토큰 정보를 반영하지 못한다는 단점이 있다. 이러한 단점

을 극복하기 위해 단순히 토큰의 나열을 순차적으로 입력할 뿐만 아니라, 텍스트에서 어느 토큰이 중요한지에 대한 정보도 같이 전달하여 중요한 토큰 정보를 집중하여 처리에 참고할 수 있도록 하는 어텐션 기법이 제시되어 기계번역이나 문서 요약 등 장문을 이해하고 새로운 텍스트를 생성하는 작업에서 활용되었다(Bahdanau 외, 2015; Luong 외, 2015). Transformer는 이러한 어텐션 만을 사용하여 자연어 이해 및 생성 처리를 구현한 것으로, BERT는 이 중 '자연어 이해' 부분을 가져와 거대한 규모로 훈련시킨 사전훈련 모델이다.

어텐션 기법을 사용한다고 해서 훈련 텍스트에서 어느 토큰에 집중하여야 하는지 사람이 알려주지는 않는다. 어텐션 기법에서 어느 토큰에 집중할 것인가 또한 기계가 직접 판단하기 때문에, 이러한 어텐션의 결과 어느 토큰에 가중치가 얼마나 주어졌는지를 알면 이를 통해 모델이 어느 토큰을 근거로 그러한 예측을 내어 놓았는지 어느정도 가늠할 수 있다. 물론 어텐션 기법이 이러한 용도를 위해 설계된 것이 아니기 때문에 어텐션의 결과로 부여된 가중치인 어텐션 점수만으로 모델의 작동을 '설명explain'할 수 없다는 견해가 있다(e.g., Jain & Wallace, 2019). 하지만 Wiegrefe & Pinter(2019)는 모델의 작동을 '설명explain'한다는 것을 '투명성transparency', '설명가능성explainability', '해석가능성interpretability'로 나누어 본다면, 어텐션 점수는 이 중 최소한 투명성을 드러낸다는 것을 실험을 통해 입증해 보였다. 이를 종합적으로 고려한다면, 어텐션 점수를 통해 모델의 작동 원리 전체를 해석 및 설명할 수는 없지만, 적어도 모델이 무엇을 중요하게 보는지는 가늠해 볼 수 있음은 분명하다.

따라서, 본 연구에서는 전술한 방법으로 파인튜닝한 모델을 통해 공백 및 비공백 구문의 예측을 진행하고, 이 예측의 어텐션 점수를 사용하여 모델이 각 데이터의 어느 부분을 중요하게 보았는지를 가늠해 본다. 이를 통해 파인튜닝을 통해 공백 구문 여부를 학습시킬 수 있는가를 넘어, 학습이

가능하다면 모델은 무엇을 근거로 판단을 내렸는지를 추측한다. 이러한 분석은 BERT 등의 언어모델을 통한 언어 현상 분석 및 언어 모델의 추론 능력 평가에서 정성적 분석의 방법으로 많이 차용되었다(박진호 2018, Chaves & Richter 2021, Park & Song 2021).

이 때, BERT의 작동 원리로 인해 부가적인 처리가 이루어졌다. 언어 모델은 텍스트 전체를 입력으로 받을 때 보다 이를 토큰으로 분절하여 입력으로 받을 때 더욱 용이한 처리를 할 수 있다. 이 때, 전통적으로 어절 및 실질형태소를 기반으로 분절한 경우, 영어의 접미사와 같이 반복되어 등장하는 형태소도 어간에 따라 다른 토큰으로 처리되어 따로 학습되어야 하므로 효율성이 떨어진다. 또한, 해당 접미사를 사용하지만 학습시에 등장하지 않은 단어를 처리하는데 어려움을 겪게 된다. 이러한 문제를 해결해 학습에 등장하지 않은 단어를 포함해 토큰을 효율적으로 처리하기 위해서 BERT는 단어를 더욱 작은 단위로 쪼갬다. 이렇게 단어를 분절하는 근거는 해당 언어의 형태론이 아닌 데이터 처리의 효율성이다. 따라서, 가령 영어에서 이러한 토큰화의 결과 동사의 과거시제를 나타내는 접미사 '-ed'가 분절될 수 있지만, 이는 모델이 '-ed'를 접미사라고 판단한 것이 아니라, 데이터 처리에서 '-ed'의 꼴이 매우 많이 등장하기 때문에 이를 별도로 처리하는 것이 효율적이라 판단하였기 때문이다. 그 결과 영어 BERT의 경우 'pre-', '-ed', '-ing'와 같은, 자주 등장하지만 한정되어 있는 영어의 접사를 비교적 잘 분리하나, 'juxtaposition'과 같은 저빈도 어휘는 'ju + xt + ap + osition'과 같이 분리하여 이해한다.<sup>4)</sup>

한자어가 많은 한국어는 같은 형태의 글자가 반복되는 경우가 많다. 즉, '언어'라는 단어는 '언'과 '어'라는 두 개의 한자가 결합된 것이기 때문에, 가령 '어'라는 한자가 '한국어', '영어', '프랑스어' 등 다른 단어에서도 빈번하게 등장한다. 이러한 경향으로 인해 한국어 BERT 모델은 비록 자

4) 해당 결과는 HuggingFace Hub의 bert-large-uncased 모델의 토큰나이저가 처리한 결과이다.

주 등장하는 어휘일지라도 이러한 어휘를 분절하여 이해하는 경우가 많다. 예를 들어, 본 연구에서 사용한 KR-BERT는 '기차'라는 자주 사용되는 어휘 또한 '기 + 차'로 이해한다. 이는 형태론적으로 한자 하나 하나를 형태소로 인식했다기보다, KR-BERT의 훈련에 사용된 데이터에서 '차'를 분절하는 것이 더욱 효율적으로 데이터를 처리할 수 있다고 모델이 판단하였기 때문이다. 이는 '기관차'를 '기 + 관 + 차'가 아닌 '기관 + 차'로 이해한다는 점을 보면 분명히 할 수 있다.

이러한 과정은 분명 언어 모델을 통한 자연어의 전산 처리에는 이점이지만, 언어 모델의 결과를 통해 언어 현상을 보고자 하는 본 연구에서는 장애물로 작동한다. 이는 가령 '역을 떠난 기차를 바라보았다'라는 문장을 판단한 근거를 어텐션 점수로 확인할 때, 모델은 '기차를'을 하나의 토큰으로 처리하여 어텐션 점수로 주는 것이 아니라, '기차를'을 '기', '차', '를'의 세 토큰으로 나누고, 각각의 토큰에 대해 별개의 어텐션 점수를 부여하기 때문이다. 이렇게 나누어진 토큰의 어텐션 점수를 분석한다 하더라도 각각의 토큰이 언어학에서 바라보는 단위가 아니기에 이로부터 언어학적으로 유의미한 함의를 도출하기는 힘들다.

이에, 본 연구에서는 어절을 기반으로, 한 어절을 구성한 토큰의 어텐션 점수를 평균으로 통합하였다. 예를 들어 위의 사례에서 '기', '차', '를'이 각각 0.03, 0.01, 0.05 만큼의 어텐션 점수를 받았다고 가정한다면, 이 세 토큰은 하나의 어절을 구성하였기에 '기차를'이라는 어절에 0.03 만큼의 어텐션 점수가 주어졌다고 계산하였다. 평균을 통해 점수를 합산하는 것은 어절을 구성하는 문자의 길이에 따라 점수가 높아지는 것을 방지하기 위함이다. 이러한 토큰화를 사용하는 모델을 통해 한국어 언어 현상을 분석하는 연구에서는 주로 이러한 토큰화를 고려하여 설계한 데이터셋에 대해 분석을 진행하기 때문에, 이러한 처리를 진행하는 사례를 찾을 수는 없었다. 다만 본 연구에서 이러한 방법으로 통합한 어절의 어텐션 점수가 어느정도 일정한 패턴을 보였음을 근거로 위 방법을 통해 분석을 진행한



다.

## 4. 실험 결과

### 4.1 목표 구문의 학습 여부

본 연구에서는 3장에서 제시한 방법으로 서로 다른 데이터셋을 사용한 9번의 파인튜닝을 진행하였다. 각각의 파인튜닝은 사전 훈련 모델에서 서로를 참고하지 않고 독립적으로 진행되었다. 각각의 파인튜닝에서는 4번의 에폭이 진행되었고, 한 번의 에폭이 끝날 때 마다 정확도, F1, 평균 손실값으로 훈련된 모델의 성능을 측정하였다. 이 때, 정확도와 F1은 0에서 1사이의 값으로, 모든 경우에 정답일 경우 1이다. 손실값은 구체적으로 교차엔트로피손실값(Cross-Entropy Loss)이 사용되었으며, 최대값은 무한이고 최저값은 0이다. 손실값은 클 수록 모델의 예측이 실제와 다르다는 뜻이므로 0에 가까울 수록 훈련이 잘 된 것이라 볼 수 있다. 9번의 파인튜닝 중 첫번째의 경우 최고 0.89의 정확도와 0.88의 F1, 그리고 0.34의 최저 평균손실값을 2번째 에폭에서 달성하였다. 이 경우를 제외한 모든 경우는 0.97에서 1 사이의 정확도와 F1을 달성하였으며, 0.1 이하의 평균손실값을 달성하였다. 8번의 파인튜닝에서 훈련에 사용된 데이터에 한하여 모두 정답을 맞춘 점을 미루어 본다면 BERT 모델에게 공백 구문 여부의 분류는 매우 쉬운 과업이었을 것이라 판단된다. 첫번째 파인튜닝에서 유독 낮은 성능이 나온 이유는 우연히 구축된 데이터셋의 특성이 반영된 것으로 판단된다. 허나, 이 경우에도 0.9에 근접하는 점수는 BERT 모델을 통해 수행하는 텍스트 분류 과업에서 높은 점수로 인식되는 만큼 9번의 파인튜닝 모두 성공적으로 공백 여부를 판단하는 법을 학습하였다. 기계학습을 진행할 때에는 실제 세계에 존재하는 전체 데이터 중 일부를 샘플링하여 훈련을 진행하기 때문에, 기계가 이러한 샘플에 과

도하게 훈련되어 과적합될 가능성이 있다. 이러한 과적합은 성능 평가 단계에서 손실값의 추이를 통해 파악할 수 있다. 성능 평가에서는 훈련에 사용되지 않은 데이터를 사용하기 때문에, 훈련 중의 손실값이 작아지더라도 성능 평가 단계의 손실값이 다시 커진다면 모델이 훈련에 사용된 데이터에만 치중하여 성능 평가 데이터에서 조차 제대로 된 예측을 하지 못하기 시작했기에, 이를 과적합이라 판단할 수 있다. 전술한 성능은 이러한 방식으로 과적합이 이루어지기 이전의 결과이며, 모델 또한 손실값을 기준으로 과적합되지 않은 경우를 저장하여 활용하였다. 따라서, 최소한 파인튜닝 과정 자체에서의 과적합은 이루어지지 않았다.

## 4.2 개별 데이터의 분석 및 예시

본장에서는 본 연구에서 얻은 결과 중 학습이 가장 잘 이루어진 데이터 세트에 대해 기술한다. 해당 세트는 공백 관계절 1981개와 비공백 관계절 1980개에 대한 딥러닝 한국어 모델의 분류 결과를 담고 있다. 이 중 딥러닝 한국어 모델이 옳게 분류한 경우는 총 3,704개로 이 중 공백 관계절과 비공백 관계절을 각각 임의로 400개씩 추출하여 살펴보았다. 다음으로 딥러닝 한국어 모델이 잘못 분류한 경우는 총 257개인데, 공백 관계절을 비공백 관계절로 분류한 경우가 225개였고, 비공백 관계절을 공백 관계절로 분류한 경우는 모두 32개로 해당 사례들을 모두 살펴보았다.

먼저, 각 예시들은 관계절 속 서술어와 머리명사 중 무엇이 어텐션 점수가 높은지를 기준으로 아래와 같이 분류하였다.

- (14) 가. 관계절 속 서술어
  - 나. 머리명사
  - 다. 관계절 속 서술어&머리명사

(14가)는 관계절 속 서술어의 어텐션 점수가 더 높은 경우이고, (14나)는 머리명사의 어텐션 점수가 더 높은 경우이다. (14다)는 관계절 속 서술어와 머리명사의 어텐션 점수가 같은 경우이다. 데이터별 각 유형에 해당하는 예문의 개수는 아래 표1과 같다.

표1: 데이터별 어텐션의 위치

공백 관계절을 공백 관계절로 분류	
관계절 속 서술어	301
머리명사	93
관계절 속 서술어&머리명사	6
총합	400
비공백 관계절을 비공백 관계절로 분류	
관계절 속 서술어	179
머리명사	174
관계절 속 서술어&머리명사	47
총합	400
공백 관계절을 비공백 관계절로 분류	
관계절 속 서술어	109
머리명사	107
관계절 속 서술어&머리명사	9
총합	225
비공백 관계절을 공백 관계절로 분류	
관계절 속 서술어	18
머리명사	12
관계절 속 서술어&머리명사	2
총합	32

딥러닝 한국어 모델이 어떠한 기제로 특정 예를 특정 범주로 분류했는지는 정확히 알 수 없다. 그러나 히트맵의 결과를 통해 아래와 같이 몇 가지 추정을 할 수 있다.

- (15) 가. 공백 관계절의 경우 관계절 속 서술어 정보에 더 많이 의존하는 경향이 있다.  
 나. 비공백 관계절의 경우 관계절 속 서술어와 머리명사 모두에 의존하는 경향이 있다.  
 다. 분류가 어려운 경우 비공백 관계절을 고르는 경향이 있다.

첫째, 공백 관계절의 경우 관계절 속 서술어 정보에 더 많이 의존하는 경향이 있어 보인다. 이러한 경향은 딥러닝 한국어 모델이 분류 작업을 성공적으로 완수한 경우 더 뚜렷하게 나타났다. 먼저, 성공적으로 공백 관계절로 분류한 400개의 사례의 경우, 관계절 속 서술어에 더 많은 어텐션이 간 경우가 301개였고, 머리명사에 더 많은 어텐션이 간 경우는 93개로 관계절 속 서술어 정보에 의존하는 경향이 분명해 보인다. 실패한 사례 32개, 즉, 원래 비공백 관계절인데 이를 공백 관계절로 분류한 경우, 관계절 속 서술어에 더 많은 어텐션이 놓인 경우가 18개였고, 머리명사에 더 많은 어텐션이 놓인 경우가 12개였다. 이때에도 관계절 속 서술어에 더 많은 어텐션이 놓인 사례가 더 많으나 그 경향성은 성공 사례에서만 뚜렷하지는 않았다. 둘째, 비공백 관계절의 경우 관계절 속 서술어와 머리명사 모두에 의존하는 것으로 보인다. 이러한 경향은 딥러닝 한국어 모델이 성공적으로 분류 작업을 완수한 경우와 그렇지 못한 경우 모두에서 비슷하게 나타났다. 성공적으로 비공백 관계절로 분류한 400개의 예에서, 관계절 속 서술어에 더 많은 어텐션이 놓인 경우는 179개였고, 머리명사 더 많은 어텐션이 놓인 경우는 174개였다. 실패한 사례 225개, 즉, 원래 공백 관계절인데 이를 비공백 관계절로 분류한 경우에는 관계절 속 서술어에 더 많은 어텐션이 놓인 경우가 109개였고, 머리 명사에 더 많은 어텐션이 놓인 경우는 105개였다. 셋째, 분류가 어려운 경우 비공백 관계절을 고르는 경향이 있어 보인다. 먼저, 딥러닝 한국어 모델이 분류에 실패한 사례는 모두 257개인데 이 중 공백 관계절을 비공백 관계절로 분류한 경우가 225개이고, 비공백 관계절을 공백 관계절로 분류한 경우는 32개였다. 비공백 관계절로 잘못 분류한 경우가 많았다는 것은 몇 가지 함축을 갖는다. 먼저, 비공백 관계절은 공백 관계절과 달리 이질적인 특성을 가진다. 2장에서 살펴본 바와 같이 해당 범주는 연구자에 따라 연계 관형사절을 포함한 하위범주로 구별되기도 한다. 동일한 것을 다른 측면에서 보면 이는 공백 관계절이 딥러닝 한국어 모델이 분류해 낼 수 있는 범주임을

보여 주는 것이기도 하다. 마지막으로 실패한 사례와 관련하여 특기할 만한 사항은 성공한 사례와 달리 실패한 사례의 경우 하나의 예문 안에 관계절이 1개 이상 포함된 예가 발견된다는 것이다.<sup>5)</sup> 이러한 예는 공백 관계절을 비공백 관계절로 분류한 경우나 비공백 관계절을 공백 관계절로 분류한 경우 모두에서 관찰되었다.

이어지는 절에서는 표2에 제시한 각 유형의 사례들을 구체적으로 살펴보고자 한다. 먼저, 공백 관계절을 성공적으로 분류한 예는 (16)과 같다. 각 예에서 밑줄로 표시한 부분은 관계절 속 서술어와 머리명사 중 어텐션이 많이 놓인 것을 의미한다. 관계절 속 서술어와 머리명사에 동일한 어텐션이 놓인 경우에는 둘 모두에 밑줄로 표시하였다. 각 예의 어텐션 데이터는 표2-표7에 제시되어 있다.

- (16) 가. 갓 들어온 신입 경리 사원이  
 나. 그 이름을 알 사람이 없느니라,  
 다. 기모노 하니까 생각나는 영화가

표2: 성공적인 공백 관계절의 분류 (관계절 속 서술어)

ID	21557.00							평균(표준편차)
데이터	갓	들어온	신입	경리	사원이	:	:	
어텐션 점수	0.09	0.20	0.12	0.08	0.09	0.06	0.06	0.10(0.04)
(점수 평균)/표준편차	-0.31	2.48	0.56	-0.44	-0.18	-1.01	-1.03	

5) 하나의 예문 안에 관계절이 한 개 이상 포함된 경우 원자료의 분류 레이블링을 기준으로 관계절을 식별했다. 가령, 원자료의 레이블링이 공백 관계절로 되어 있는 경우 예문 안에 비공백 관계절이 있더라도 공백 관계절을 기준으로 관계절 속 서술어와 머리명사의 어텐션 정도를 비교하였다.

표3: 성공적인 공백 관계 관계절의 분류 (머리명사)

ID	59914.00						평균(표준편차)
데이터	그	이름	알	사람이	없느니라	,	
어텐션 점수	0.06	0.07	0.16	0.27	0.10	0.03	0.11(0.07)
(점수-평균)/표준편차	-0.67	-0.61	0.75	2.28	-0.16	-1.12	

표4: 성공적인 공백 관계절의 분류 (관계절 속 서술어&amp;머리명사)

ID	17520.00				평균(표준편차)
데이터	기모노	아니까	생각나는	영화가	
어텐션 점수	0.01	0.01	0.02	0.02	0.08(0.22)
(점수-평균)/표준편차	-0.31	-0.29	-0.27	-0.27	

다음으로, 성공적인 비공백 관계절의 예는 (17)과 같다. 각 예의 어텐션 데이터는 표5-표7에 제시되어 있다.

- (17) 가. 뭔가 풀리는 느낌 같은 것이 들어.  
 나. 근데 어쩌다가 실패하는 경우가 있어.  
 다. 눈이 진짜 창밖이 안 보일 정도로 막 내리는 거야.

표5: 성공적인 비공백 관계절의 분류 (관계절 속 서술어)

ID	119.00							평균(표준편차)
데이터	뭔가	풀리는	느낌	같은	것이	들어		
어텐션 점수	0.11	0.15	0.07	0.06	0.05	0.15	0.15	0.01(0.04)
(점수-평균)/표준편차	0.06	0.79	-1.04	-1.12	-1.45	1.00	0.89	

표6: 성공적인 비공백 관계절의 분류 (머리명사)

ID	11.00											
데이터	근데	어쩌다가	실패하는	경우가	있어							평균(표준편차)
어텐션 점수	0.04	0.04	0.05	0.60	0.02	0.02						
(점수-평균)/표준편차	-0.33	-0.30	-0.23	3.15	-0.42	-0.42						0.09(0.16)

표7: 성공적인 비공백 관계절의 분류 (관계절 속 서술어&amp;머리명사)

ID	272.00												
데이터	눈이	진짜	장박이	안	보일	정도로	막	내리는	거야				평균(표준편차)
어텐션 점수	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.92			
(점수-평균)/표준편차	-0.27	-0.27	-0.25	-0.24	-0.26	-0.26	-0.26	-0.26	-0.26	3.87			0.06(0.22)

다음으로, 공백 관계절을 비공백 관계절로 잘못 분류한 예는 (#)와 같다. 각 예의 어텐션 데이터는 표8-표10에 제시되어 있다.

- (18) 가. 그냥 계속 흐린 날씨였는데,  
 나. 느낌으로 알 수 있는 그런 사랑을 했으면 좋겠어,  
 다. 스몰라렉 선수 표정을 봐도 지금 많이 지쳐있는 표정이네요.

표8: 공백 관계절을 비공백 관계절로 잘못 분류 (관계절 속 서술어)

ID	28156.00						
데이터	그냥	계속	흐린	날씨였는데			
어텐션 점수	0.01	0.14	0.30	0.11	0.03		평균(표준편차)
(점수-평균)/표준편차	-1.19	-0.01	1.40	-0.31	-0.98		

평균/표준편차										
---------	--	--	--	--	--	--	--	--	--	--

표9: 공백 관계절을 비공백 관계절로 잘못 분류 (머리명사)

ID	20480.00										
데이터	느낌으로	알	수	있는	그런	사랑을	했으면	좋겠어			평균(표준편차)
어텐션 점수	0.02	0.10	0.12	0.13	0.09	0.20	0.08	0.04	0.04		
(점수 평균) 표준편차	-1.00	0.46	0.73	0.93	0.29	2.40	-0.02	-0.72	-0.60		0.08(0.05)

표10: 공백 관계절을 비공백 관계절로 잘못 분류 (관계절 속 서술어&amp;머리명사)

ID	46602.00										
데이터	스몰라렉	선수	표정	봐도	지금	많이	지쳐	있는	표정이	네요	평균(표준편차)
어텐션 점수	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.85	
(점수 평균) 표준편차	-0.25	-0.23	-0.25	-0.24	-0.27	-0.21	-0.24	-0.24	-0.24	4.12	0.06(0.19)

다음으로, 비공백 관계절을 공백 관계절로 잘못 분류한 예는 (19)와 같다. 각 예의 어텐션 데이터는 표11-표13에 제시되어 있다.

(19) 가. 온다 오는 길이라매.

나. 놀 계획 짤려고 그러지 너.

다. 예~또 다른 얘기 사관학교 간 이야길 해 드리겠습니다.

표11: 비공백 관계절을 공백 관계절로 잘못 분류한 예(관계절 속 서술어)



ID	17252.00				
데이터	준다	오른	같이때매		
어텐션 점수	0.10	0.25	0.14	0.14	평균(표준편차)
(점수- 평균)/ 표준편차	-0.40	1.06	-0.07	-0.04	0.14(0.10)

표12: 비공백 관계절을 공백 관계절로 잘못 분류한 예(머리명사)

ID							
데이터	늘	계획	짧려고	그러지	너		
어텐션 점수	0.05	0.18	0.15	0.11	0.13	0.11	평균(표준편차)
(점수- 평균)/ 표준편차	-1.72	1.41	0.55	-0.29	0.03	-0.25	0.13(0.04)

표13: 비공백 관계절을 공백 관계절로 잘못 분류한 예(관계절 속  
서술어&머리명사)

ID	2637											
데이터	예	~	또	다른	애기	사관	간	이야	해	드리		평균(표준편차)
어텐션 점수	0.00	0.0	0.0	0.0	0.0	0.01	0.02	0.02	0.01	0.01	0.80	
(점수- 평균)/ 표준편차	-0.31	-0.4	-0.2	-0.2	-0.2	-0.25	-0.23	-0.23	-0.26	-0.25	4.00	0.06(0.18)

## 5. 논의점

앞 장에서는 히트맵 분석 결과를 바탕으로 이루어진 세 가지 관찰을

제시하였다. 즉, 공백 관계절의 경우 관계절 속 서술어 정보에 더 많이 의존했다. 또한, 비공백 관계절의 경우 관계절 속 서술어와 머리명사 모두에 의존했다. 마지막으로, 분류가 어려운 경우 비공백 관계절로 분류하는 경향이 있었다. 이러한 관찰은 1장에서 제시한 세 가지 질문 중 공백 관계절의 학습이 가능한지, 또한 공백 관계절의 학습이 무엇을 바탕으로 이루어지는지에 대한 잠정적인 답을 제시한다. 본장에서는 1장에서 제시한 마지막 질문인 딥러닝 한국어 모델이 관계절 속 공백을 포착하는가에 대한 질문을 다루어 보고자 한다. 결론부터 말하자면, 본 연구는 딥러닝 한국어 모델이 관계절 속 공백을 포착한다는 직접적 증거를 찾지 못했다.

먼저, 공백 관계절의 경우 비공백 관계절과 달리 서술어에 어텐션이 있는 경향이 있었는데 이는 공백 관계절이 딥러닝 한국어 모델에게 구별되는 구성임을 말해 준다. 4장에서 살펴본 바와 같이 성공 사례 중 400개의 공백 관계절의 경우 관계절 속 서술어에 더 많은 어텐션이 간 경우가 301개였다. 반면, 비공백 관계절의 경우 관계절 속 서술어와 머리명사 모두에 의존하는 것으로 보인다 성공 사례 중 400개의 비공백 관계절의 경우 관계절 속 서술어에 더 많은 어텐션이 놓인 경우는 179개였고, 머리명사에 더 많은 어텐션이 놓인 경우는 174개였다. 이러한 경향이 나타나지 않은 것은 관계절이 한 개의 형용사로 이루어진 경우이다. 해당 경우에도 공백 관계절의 경우 관계절 속 서술어에 더 많은 어텐션이 놓인 경우가 머리명사에 더 많은 어텐션이 놓인 경우보다 많았으나 그 차이가 크지 않았다. 그러나 공백 관계절의 경우 서술어에 가장 큰 어텐션이 놓이는 경향이 있다는 것이 한국어 딥러닝 모델이 공백을 포착한다는 직접적인 증거를 제시하는 것은 아니다.

다음으로 원자료에서 공백 관계절로 분류한 것을 딥러닝 한국어 모델이 비공백 관계절로 분류한 경우 중 모델의 판단이 맞는 것으로 보이는 사례가 다수 있었는데 이는 딥러닝 한국어 모델이 공백 관계절의 패턴을 일관성 있게 인식하고 있다는 추정을 하게 한다. 아래 (20)이 해당 예이

다.

- (20) 가. 자신이 생각하는 최적의 거리가 유지되지 않으면  
 나. 그런 문제들도 지적되는 부분이 사실이거든요?  
 다. 뭐 살 시간도 없고,

위의 예에서 밑줄 그은 부분은 머리명사가 관계절 안에서 어떤 문법 역할을 갖는 것으로 보기 어렵다. (20가)의 경우 (20가)에서 '최적의 거리'가 관계화된 것으로 볼 수도 있겠지만 (20나)가 비문이 되는 것을 볼 때 (20가)는 비공백 관계절로 처리하는 편이 합당해 보인다.

- (21) 가. 자신이 (그것을) 최적의 거리로 생각한다.  
 나. 내가 영희를 반장으로 생각한다.  
 나'. \*내가 영희를 생각하는 반장

그러나 딥러닝 한국어 모델이 공백 관계절의 패턴을 일관성 있게 인식하고 있다고 해도 해당 사실이 한국어 딥러닝 모델이 한국어 관계절 속에서 공백을 포착한다는 직접적인 증거는 되지 못한다.

본 연구에서 추가적으로 진행한 실험의 결과 역시 딥러닝 한국어 모델이 공백을 포착하지 못한다는 것을 제시하지는 않는다. 본고에서는 본 실험을 마친 후 딥러닝 한국어 모델이 공백을 포착하는지 확인해 보고자 비논항 성분이 머리명사인 관계절과 선행연구에서 비공백 관계절로 분류한 예 몇 가지에 대한 분류 실험을 진행하였는데 결과는 다음과 같다. 먼저, 비논항 성분이 머리명사인 관계절의 경우 머리명사가 (22가-다)에서와 같이 장소를 가리키거나 (22라-바)에서와 같이 도구를 가리키는 경우에는 성공적으로 공백 관계절로 분류하였고, 머리명사가 (23가-다)에서와 같이 시간을 가리키는 경우에는 비공백 관계절로 잘못 분류하였다.

## (22) 공백 관계절로 분류한 예

- 가. 나는 민지가 점심을 먹은 식당에 갔었는데.  
 나. 어제 진수가 책을 산 서점에 가서 나도 책 한권을 사 왔어.  
 다. 내가 어제 쿠키를 산 커피숍은 아주 유명한 곳이다.  
 라. 너가 함께 영화를 본 친구한테 전화해 봐.  
 마. 은지가 사과를 깎은 칼이 탁자에 놓여 있다.  
 바. 나는 동생이 창문을 닦은 천을 세탁기에 넣었다.

## (23) 비공백 관계절로 분류한 예

- 가. 내가 헤지랑 커피를 마신 일요일에 너는 뭐 했어?  
 나. 지혜가 과제를 하던 그날 저녁에 태풍이 몰아쳤다.  
 다. 지희가 빨래를 널고 있던 오후에 갑자기 비가 내렸다.

위의 예 중 (22)에서는 관계절 속 서술어에 더 많은 어텐션이 놓였고, (23)의 경우에는 (23가, 나)의 경우에는 관계절 속 서술어에, (23다)에서는 머리명사에 더 많은 어텐션이 놓였다. 이러한 예들의 경우 머리명사가 관계절 속의 논항이 아니고, 관계절 속 논항 자리는 이미 채워져 있다. 그럼에도 (22)를 공백 관계절로 분류했다는 것은 딥러닝 한국어 모델이 머리명사에 중점을 두고 예측했거나 관계절 속 서술어와 함께 나타나는 비논항 성분에 대한 정보를 가지고 있을 가능성을 제시한다. 첫 번째 가능성은 해당 예에서 동사에 더 많은 어텐션이 놓였으므로 설득력이 높지 않다. 두 번째 가능성의 경우 왜 시간 정보만 파악하지 못했는지에 대해 답하기 어렵다.

다음으로 선행연구에서 비공백 관계절로 분류된 것을 비공백 관계절로 분류한 경우는 (24)와 같고, 공백 관계절로 분류한 경우는 (25)와 같다. 비공백 관계절로 성공적으로 분류한 (24)에서, (24가-다)의 경우 머리명사에 더 많은 어텐션이 놓였고, (24라-마)의 경우 관계절 속 서술어와 머리명사에 동일한 어텐션이 놓였다. 나머지는 관계절 속 서술어에 더 많은

어텐션이 높았다. 비공백 관계절을 공백 관계절로 분류한 (25)에서는 (25가)를 제외하면 모두 관계절 속 서술어에 더 많은 어텐션이 높았다.

(24) 비공백 관계절로 분류

가. 나는 걷는 운동이 좋더라고.

나. 아마도 가까운 미래에 갈 수 있을 것 같아.

다. 주말에 화장실에 가기 어려운 연속극을 보고 있는데 전화가 오는 거야.

라. 인간은 모두 살 권리가 있다.

마. 현대인은 귀가가 늦어지는 교외생활을 피할 수 없다.

바. 어디서 타는 냄새 나는 것 같은데?

사. 그가 바라보던 눈빛이 잊혀지지 않는다.

아. 이제 야채 삶은 물에다가 파스타를 넣어 주세요.

자. 그냥 순종하는 마음으로 살려고.

차. 이제 슬슬 싸울 준비를 좀 해 볼까?

카. 오늘은 얼굴이 예뻐지는 세안법을 알려 드리겠습니다.

(25) 공백 관계절로 분류

가. 우리는 비참한 땅에 들어가게 되었다.

나. 좀 서운한 얼굴로 앉아 있는 거야.

다. 개는 그냥 단순한 부적응증이야.

라. 그러니까 유일한 결점이 드러나 버렸다고?

마. 그건 정말 새빨간 거짓말이라니깐.

바. 어제 타고 남은 재까지 갖다 버렸어.

사. 누가 오바이트하지 않는 술을 만들어 줬으면 좋겠다.

아. 그러게 애초에 취직이 어려운 물리학을 왜 해가지고는.

위에 제시한 예들은 딥러닝 한국어 모델이 공백 관계절과 비공백 관계

절을 일관성있게 분류해 내지 못하는 것처럼 보일 수 있다. 그러나 두 가지 점은 생각해 볼 만하다. 첫째, 2장에서 살펴본 바와 같이 위에 제시된 예를 공백 관계절로 보아야 하는지는 사람에게도 판단하기 어렵거나 의견이 갈릴 수 있는 문제일 수 있다. 둘째, 위의 예에서도 공백 관계절의 경우 관계절 속 서술어에 더 많은 어텐션이 놓이는 경향성이 포착된다. 그러나 본고에서 추가적으로 살펴본 위의 예들 역시 딥러닝 한국어 모델이 관계절 속 공백을 포착한다는 직접적인 증거를 제시해 주지 않는다. 해당 문제는 가능성을 열어 두고 계속 살펴봐야 할 것이다.

마지막으로 과적합의 문제를 살펴보고자 한다. 본 연구에서 진행한 9번의 파인튜닝 중 첫번째를 제외한 모든 경우에서 최대 0.97에서 1사이의 정확도와 F1을 달성하였다. 이러한 수치는 매우 높은 수치로, 특히 점수가 1인 경우는 훈련된 모델의 평가에 사용된 데이터를 모두 제대로 예측하였다는 뜻이다. 이렇게 높은 점수를 근거로 BERT모델이 공백 유무를 잘 판단한다고 보았지만, 실제로 모델이 이러한 과업을 잘 수행하는지에 대해서는 추가적인 검토가 필요하다.

우선 파인튜닝 과정에서 성능 평가에 사용한 데이터는 1980개의 비공백절과 1:1의 비율이 되도록 구축한 데이터셋의 10%이므로 총 396개이다. 이러한 학습을 진행하는데 필요한 데이터 양이 절대적으로 정해진 것은 아니다. 3,564개의 데이터로 훈련을 진행하고 396개의 데이터로 성능을 평가하는 것은 일반적으로 진행되는 기계학습, 특히 딥러닝에 비해 적은 양이다.

앞에서는 이렇게 훈련된 데이터셋이 훈련 및 테스트에 사용하지 않은 10번째 데이터셋에서도 높은 성능의 분류를 해 낸다는 것을 근거로 각각의 파인튜닝이 주어진 데이터셋에만 훈련되는 과적합을 진행하지는 않았다고 판단하였다. 이는 특히 연구자가 새로이 생성하여 분류하게 한 문장도 어느정도 잘 분류하는 점을 보았을 때 더욱 두드러진다. 하지만 과도하게 높은 성능을 보았을 때 본 연구의 모델이 과적합이 되었을 가능성을

완전히 배제하기는 조심스럽다.

이를 확인하기 위해서는 결국 더 많은 데이터가 필요하다. 따라서, 더욱 큰 말뭉치에서 많은 데이터를 추출하거나, 템플릿을 기반으로 수가 부족한 비공백 데이터를 증폭시키는 등의 방법으로 공백 구문과 비공백 구문 사이의 비율을 유지하면서 전체적으로 양이 많은 데이터를 구축할 필요가 있다. 이러한 데이터에 대해 BERT 모델의 파인튜닝과 더불어 사전 훈련을 진행하지 않는 Support Vector Machine, 순환신경망 등의 방법으로 훈련을 진행해 이 결과를 비교한다면 공백 여부 분류가 사전훈련을 통해 얻을 수 있는 어느 정도의 통사적 패턴을 동원해야만 해결할 수 있는 문제인지, 그러한 패턴 없이 해결할 수 있는 문제인지 파악할 수 있을 것이다. 다만 현재 데이터 양으로는 상기한 기계학습 방법론으로 훈련하기에 충분한 양이라 볼 수 없으므로 이러한 연구는 데이터를 더욱 구축한 다음에 진행할 수 있다.

이러한 한계에도 불구하고 본 연구는 자연어 텍스트의 순서를 넘어 계층적 구조도 어느정도 포착하는 언어 모델을 활용하여 공백 유무 판단을 시도하였다는 점에서 의의가 있다. 특히, 어텐션 점수를 기반으로 한 분석은 파인튜닝이 진행된 모델이 임의의 어절이 아닌 어느정도 일관된 통사적 요소에 집중한다는 것을 보여주었다. 이러한 모델의 성능이 과적합의 결과라면, 어텐션 점수 또한 사람이 이해하는 통사적 요소와 관련이 없어 임의로 부여된 것과 같이 보일 수 있다. 이를 고려한다면, 비록 과적합의 가능성이 있더라도, 본 연구는 BERT 언어 모델이 통사적 패턴을 기반으로 공백 여부를 판단한다는 가능성만은 분명하게 암시한다.

## 6. 결론

본고는 딥러닝 한국어 모델이 공백 관계절과 비공백 관계절을 구별할 수 있는지 살펴보았다. 이를 위해 말뭉치에서 추출한 공백 관계절과 비공

백 관계절을 . 연구 결과는 다음과 같다. 첫째, 딥러닝 한국어 모델이 공백 관계절과 비공백 관계절을 최소 89%에서 최대 100%의 정확도로 분류하였다. 둘째, 히트맵 분석 결과 공백 관계절의 경우 두 유형의 분류 시 관계절 속 서술어에 더 많은 주위가 기울여지는 것으로 보인다. 셋째, 비논항 성분이 머리명사이고 관계절 속 서술어의 논항이 모두 채워져 있는 경우에도 해당 예를 어느 정도 공백이 있는 관계절로 분류해 냈다. 이러한 결과들은 딥러닝 한국어 모델이 관형사절 속 공백을 포착하지 못한다는 직접적인 단서를 제공하지는 않는다. 따라서 해당 주제에 대한 지속적인 연구가 요구된다.

#### □ 참고 문헌

- 김지은(2002), 관형절의 한 유형에 대한 연구, 애산학보 27, 애산학회, 157-185.
- 박형진(2015), 한국어 관계절의 두 유형, 우리말연구 43, 우리말학회, 5-30.
- 박소영(2021), 절 구성이 아닌 관형절의 존재에 대하여, 국어학 98, 국어학회, 113-150.
- 박진호(2018), 딥러닝을 이용한 유사표현 연구 방법론: '오히려', '도리어', '차라리'의 사례, 한국문법교육학회 학술발표논문집, 2018(1), 한국문법교육학회, 1-17.
- 문숙영(2012), 유형론적 관점에서 본 한국어 관계절의 몇 문제, 개신어문연구 35, 개신어문학회, 31-68.
- 이홍식(1990), 현대국어 관형절 연구, 국어연구 98호, 국어연구회.
- 연재훈(2012), 유형론적 관점의 한국어 관계절 연구, 국어학 63, 국어학회, 413-457.
- 홍윤기(2010), 문법 교육 항목으로서의 새로운 관형사절 도입에



- 대하여, *이중언어학* 42, 이중언어학회, 331-361.
- Bahdanau, D., K. Cho, and Y. Bengio. 2015. Neural machine translation by jointly learning to align and translate. arXiv:1409.0473.
- Chaves, R. P. and S. N. Richter. 2021. Look at that! BERT can be easily distracted from paying attention to morphosyntax. *Proceedings of the Society for Computation in Linguistics 2021*.
- Comrie, B. .1998. Attributive clauses in Asian languages: Towards an areal typology. In W. Boeder, C. Schroeder, K. H. Wagner & W. Wildgen (Edn.), *Sprache in Raum und Zeit: In memoriam Johannes Bechert*, Band, 2, 51-60. Tbingen: Gunter Narr.
- Devlin, J., M-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171-4186.
- Jain, S. and B. C., Wallace. 2019. Attention is not explanation. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.
- Lee, S., H. Jang, Y. Baik, S. Park, and H. Shin. 2019. KR-BERT: A small-scale Korean-specific language model. arXiv:2008.03979.
- Luong, T., H., Pham, and C. D. Manning. 2015. Effective approaches to attention-based neural machine

- translation. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.
- Matsumoto, Y. 1988. Semantic and pragmatics of noun-modifying constructions in Japanese. Proceedings of the Annual Meeting of the Berkeley Linguistic Society, 14, 166-175.
- Park, K. and S. Song. 2021. A deep learning-based understanding of nativelikeness: A linguistic perspective. Korean Journal of English Language and Linguistics 21, 487-509.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. 2017. Attention is all you need. Advances in Neural Information Processing Systems, 30.
- Wiegrefe, S. and Y. Pinter. 2019. Attention is not not explanation. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).

## On the necessity of the plural marker ‘-tul’ in Korean

**Myung-Kwan Park**  
(Dongguk University)

### 1. Introduction

Borer (2005, 94) notes that as in (1a-c), Armenian can have either a classifier or a plural marker, but not both.

- (1) a. Yergu had hovanoc  
two CL umbrella  
b. Yergu hovanoc-ner  
two umbrella-PL  
c. \*Yergu had hovanoc-ner  
two CL umbrella-PL

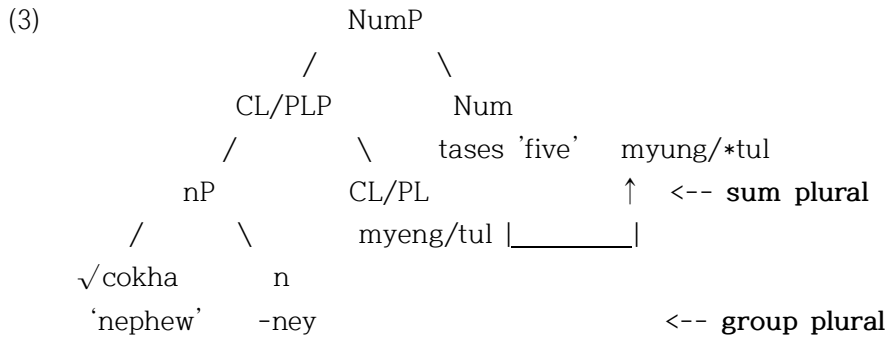
Based on this paradigm concerning the distribution of classifier and plural marker, Borer (2005) suggests that plural markers in non-classifier languages function in a parallel fashion to classifiers in that they create count nouns from mass/kind ones. Borer thus goes on to submit that plural markers and classifiers compete for the same syntactic head position; a plural marker cannot occur if a classifier does, and vice versa. Simply speaking, they are in complementary distribution.

Borer’s suggestion is intriguing, in that it unifies the roles of the two categories, though it seems to need a careful scrutiny to assess its cross-linguistic validity. Park (2022) in fact embraces Borer’s suggestion to argue that the distribution of classifier and plural marker in Korean can be captured by it, which in turn renders support to it. Particularly, the following paradigm of examples in Korean make a case for the incompatibility of plural marker and classifier:

- |  |   |
|--|---|
| <p>(2) a. haksayng(tul) tases myeng<br/>student-PL five CL</p> <p>b. koyangi(??tul) tases mali<br/>cat-PL five CL</p> <p>c. sakwa(*tul) tases kay<br/>apple-PL five CL</p> | <p>d. cha(*tul) tases tay<br/>car-Pl five CL</p> <p>e. congi(*tul) tases cang<br/>paper-Pl five CL</p> <p>f. mul(*tul) tases can<br/>water-Pl five CL</p> |
|--|---|
- Park (2022, 151)

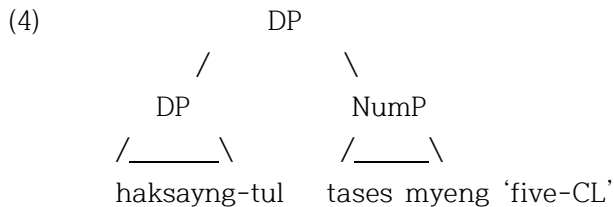
Admittedly, there is an exception to their incompatibility as in (2a), where a plural marker on [+human] nouns can optionally occur with a classifier (furthermore, there is a speaker variation regarding the (in)compatibility of a plural marker on [+animate] nouns as in (2b) with the following (numeral) classifier). But other examples than (2a) (or maybe (2b)) display the complementary distribution of the

plural marker ‘-tul’ with a (numeral) classifier, which leads to the postulation of one function category for them. Park thus proposes the following structure of nominals in Korean:



Once again, the upshot of Park’s (2022) proposal in (3) is that a classifier and a plural marker have the same grammatical function, syntactically competing for the unifying category, CL/PL. In addition, Park also proposes a functional category that additional group-denoting plural markers like ‘-ney’ occupy; this is served by the *n* head that converts a root head to a noun category.

Evidently, Park’s (2022) proposed structure cannot account for the optional realization of the plural marker on human nouns with a classifier as in (2a). To accommodate cases like (2a), Park suggests that human nouns on top of definite pronouns like *ku-tul tases myeng* ‘he/him-PL five-CL’ have not the structure in (3), but the structure in (4):



According to Park (2022), *haksayng-tul* ‘student-PL’ in (4) has an appositive relation with the following NumP, thus the plural marker and the classifier being part of the separate nominal phrases, in keeping with Park’s initial hypothesis that they are in complementary distribution.

## 2. Assessing the validity of Park's (2022) incompatibility of '-tul' with a classifier

We acknowledge that as shown in (2), there is a human vs. non-human noun distinction in regard to plural marker – classifier incompatibility. But this distinction does not always obtain. As noted by Park and Park (2020), when modified by a relative clause (RC), nouns tend to be neutralized in the human vs. noun distinction. Like that on a bare (i.e., not modified by a RC) human noun, the plural marker on a human noun like *haksayng* 'student' modified by a RC in (6) is optionally realized with the numeral classifier. Likewise, unlike that on a bare non-human noun, the plural marker on a non-human noun like *chayk* 'book' in (5) also optionally comes with the numeral classifier.

(5) RC > non-human noun(-PL) > numeral classifier

[<sub>RC</sub> poko siph-un] chayk(-tul) (#) ney kweni-i<sup>1)</sup>  
 read want book-PL four CL-NOM  
 'four books that (I) want to read'

(6) RC > human noun(-PL) > numeral classifier

[<sub>RC</sub> myentamha-yss-te-n] haksayng(-tul) (#) ney myengi-i<sup>2)</sup>  
 interviewed student-PL four CL-NOM  
 'four students that (I) interviewed'

Second, the human vs. non-human noun distinction in light of plural marker and classifier incompatibility does not hold, either, when the numeral classifier occurs as a pre-nominal, as follows:

(7) (ku/i/ce) sey kwen-uy chayk(-tul)-i  
 that/this/yonder three CL-GEN book-PL-NOM  
 'three volumes/copies of those/these books/books yonder'

(8) (ku/i/ce) sey myeng(-uy) haksayng(-tul)-i  
 that/this/yonder three CL-GEN student-PL-NOM  
 'those/these three students/students yonder'

Regardless of the presence of the demonstrative that makes the following nominals definite, when followed by a pre-nominal numeral classifier, either a human noun in (8) or a non-human noun in (7) can come with the optional plural marker.

Third, the crucial thrust of Park's (2022) proposal incorporated in the

---

1)

(i) . . . [제가 본] 책들 두 권을 추천드립니다 . . .

(ii) . . . [<오마이뉴스>에서 발간한] 책들 두 권을 받아갈 . . .

2)

(i) [확진자와 접촉한] 학생 3명이 추가로 확인됐다

(ii) [비슷한 학업 수준의 동일 교과목을 희망하는] 학생 3명이 . . .

structure (3) is that a classifier as a functional category directly performs an individuating or dividing function on a non-human noun as well as a human noun. It is doubtful that this proposal is justified, given that *mass nouns* like mwul ‘water’ or rice ‘rice’ as in (9a) and (9b) are resistant to such a function. Furthermore, a Case marker can intervene between a host noun and a numeral classifier. In this regard, Korean classifiers are more reasonably assimilated to unit nouns like ‘block’ and ‘drop’ in English as in (10a) and (10b).

- (9) a. mwul(-ul) sey thong  
       water-ACC three bucket  
       ‘three buckets of water’  
    b. ssal(-ul) yel thol  
       rice-ACC ten grain  
       ‘ten grains of rice’  
 (10) a. three drops of water  
       b. four blocks of ice

Fourth, when a nominal and a numeral classifier form a constituent separated from the anaphoric definite marker ‘ku’<sup>3)</sup> by an intervening pause or prosodic boundary (#), they display the expected human vs. non-human distinction as in (12b) vs. (11b). The plural marker on a non-human noun is degraded with the following numeral classifier as in (12b), but that on a human noun can be optionally realized as in (11b). In a parallel fashion, when the definite marker and nominal sequence is separated from a numeral classifier by a pause, the plural marker is obligatorily realized on a human noun as in (12a). However, it is optionally realized on a non-human noun as in (11a).

- (11) definite marker > non-human noun(-PL) > numeral classifier  
    a. [ku chayk(-tul)] # ney kwen-i  
       that book-PL four CL-NOM  
       ‘four volumes of those books’  
    b. ku # [chayk(??-tul) ney kwen]-i  
       those four volumes of books’  
 (12) demonstrative > human noun(-PL) > numeral classifier  
    a. [ku haksayng\*(-tul)] # ney myeng-i  
       that student-PL four CL-NOM  
       ‘those four students’  
    b. ku # [haksayng(-tul) ney myeng]-i

---

3) Other definite markers such as demonstratives and possessives in analogous contexts will be discussed in the next section.

‘those four students’

The (a)-examples of (11) and (12) can be re-constructed by the adding Nom/Acc Case-marker after a noun, instead of a pause, as in (13)-(14). In (14) where the plural numeral classifier follows, the plural marker on a human noun is obligatory, as can be found in the contrast between (14a) and (14b). However, the plural marker on a non-human noun is optional, but the two forms (i.e., without and with the plural marker) yield different interpretations. (13a) is interpreted as ‘three copies of that book’, while (13b) is interpreted as ‘three volumes of those books’, which may be copies of the same book or different books. The contrast in grammaticality between (11a) and (12a) indicates that while the definite marker ‘ku’ yields a definite entity reading both for the following human or the non-human nouns, the numeral classifier particularly with the following non-human evidently does not perform an individuating/dividing operation on them. Rather, it denotes different copies of the same book entity.

(13) a. ku chayk-ul sey kwen<sup>4)</sup>

that book-ACC three volume  
‘three copies of the book’

b. ku chayk-tul-ul sey kwen  
-PL

(14) a. \*ku haksayng-ul sey myeng vs. ku haksayng-ul han myeng

that student-ACC three CL one  
‘those three students’

b. ku haksayng-tul-ul sey myeng

What has been learned from the discussions in this section is that the Korean plural marker is in principle optionally realized on a noun (though it will be argued in the section that the optionality of the plural marker is only apparent in Korean), regardless of whether it is human or non-human, though in some restricted structural contexts the marker needs to be added on a human noun to express a plurality of entities that the resulting plural form denotes. This indicates that contrary to Park’s (2022) claim, a plural marker and a classifier do not compete for the same functional category: they have the different grammatical functions. The former is a pluralizer of entities, whereas the latter is characterized as denoting a quantity of the host noun that it associates with. The optionality of the plural marker in Korean implies that its realization as a syntactic head is not warranted: nor is its semantic function operative. In addition, as seen in this section, classifiers in Korean do not directly perform their general function of

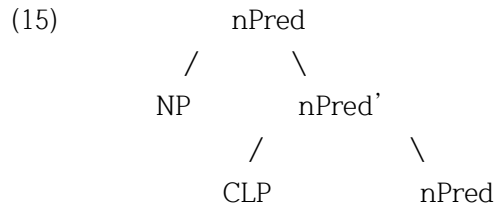
---

4)

(i) 그 고양이를 여러 마리 사가지고 찾집으로

individuation/division on nouns that are interpreted as kind/mass nouns. Rather, preceded by numerals, they serve as a predicate or modifier, describing the quantities of the nouns that they associate with.

Accommodating the properties described in the preceding paragraph, we propose the following structure that the host NP builds with the immediately following numeral classifier; this structure is adapted from that in Park (2008) and Park and Park (2019):



The host NP and the classifier phrase (which contains a numeral) are in nominal small clause relation. The classifier phrase serves as a predicate, being predicated of the host NP in terms of quantity. Assumed to be a modifier, the plural marker undergoes Late Merge, attaching to the host NP in a later point of derivation.<sup>5)</sup> Also assumed to be a modifier, the demonstratives/possessives are subject to Late Merge, attaching to the host NP or to the nominal small clause (nPredP). In the former case, the demonstratives/possessives modify the denotation of the host NP, whereas in the latter, they modifies that of the numeral classifier.<sup>6)</sup>

The structure of a noun with a numeral proposed in (15) has several advantages. First, the host NP and the numeral classifier are separate constituents and assume different grammatical functions. Thus, as mentioned above, the former can be modified by the plural marker and demonstratives/possessives, independently of the latter. The former can be modified by a relative clause (RC), as seen before. (5) and (6) are repeated below as (16a) and (16b):

- (16) a. RC > non-human noun(-PL) > numeral classifier
- |                               |             |     |      |                       |
|-------------------------------|-------------|-----|------|-----------------------|
| [ <sub>RC</sub> poko siph-un] | chayk(-tul) | (#) | ney  | kweni-i <sup>7)</sup> |
| read want                     | book-PL     |     | four | CL-NOM                |

5) The plural marker cannot be attached after a numeral classifier since it is semantically redundant. The numeral classifier denotes one or more than one, which the plural marker does.

6) The numeral classifier phrase in the nominal small clause structure of (15) is not a genuine predicate. As a semantic head of the nominal small clause, it functions as a referring expression, denoting a numeral-expressed number of entity/entities. Hence, it can be modified by a demonstrative/possessive.

7)

(i) . . . [제가 본] 책들 두 권을 추천드립니다 . . .

(ii) . . . [<오마이뉴스>에서 발간한] 책들 두 권을 받아갈 . . .



- ‘four books that (I) want to read’  
 b. RC > human noun(-PL) > numeral classifier  
 [rc myentamha-yss-te-n] haksayng(-tul) (#) ney myengi-i<sup>8)</sup>  
 interviewed student-PL four CL-NOM  
 ‘four students that (I) interviewed’

The host NP can also be elided as follows, independently of a numeral classifier:

- (17) Cheli-ka chayk sey kwen-ul sass-ko,  
 Cheli-NOM book three CL-ACC bought  
 Yengi-nun [ e ] ney kwen-ul sassta.  
 Yengi-TOP four CL-ACC bought  
 ‘Cheli bought three volumes of books and Yengi bought four.’

The structure in (15) also accounts for other variant forms of host NP – numeral classifier form. The host NP can move independently from the nPredP and pick up a Case/case particle outside it, producing the following form:

- (18) chayk-i/ul sey kwen  
 book-NOM/ACC three CL  
 ‘three volumes of books’

Alternatively, a numeral classifier phrase can be inverted with the host NP and attached with a Genitive Case particle, producing the following variant form:

- (19) sey kwen-uy chayk-i/ul  
 three CL-GEN book-NOM/ACC  
 ‘three volumes of books’

Besides, the predicate status of a numeral classifier in the nominal structure of (15) can be confirmed in the fact that it cannot be elided as on (20a) and (20b) in the same way that a predicate in the verbal small clause as in (21) cannot be.

- (20) a. Cheli-ka [chayk sey kwen]-ul sassta.  
 Cheli-NOM book three CL-ACC bought  
 Yengi-to [chayk \*(sey kwen)]-ul sassta.  
 -also  
 ‘Cheli bought three volumes of books, and Yengi did too.’

---

8)

(i) [확진자와 접촉한] 학생 3명이 추가로 확인됐다

(ii) [비슷한 학업 수준의 동일 교과목을 희망하는] 학생 3명이 . . .

- b. chelika chaykul sey kwen sassta. yengito chaykul \*(sey kwen) sassta.  
 (21) Cheli-ka caki chinkwu-lul chencay-lo sayngkakhanta.  
       -NOM his own friend-ACC genius-as think  
 Yengi-to caki chinkwu-lul \*(chencay-lo) sayngkakhanta.  
       -also  
 ‘Cheli thinks his own friend to be a genius, and Yengi does too.’

Furthermore, as in (21a-b), a numeral classifier phrase cannot be relativized, which shows that it is assimilated to a predicate nominal in (22):

- (21) a. \*[[Cheli-ka ecey [chayk [e]]-ul kwuiphan] [sey kwen]-i  
       Cheli-NOM yesterday book -ACC purchased three CL-NOM  
       wuphyen-ulo tochakhayssta.  
       mail-by arrived  
       ‘Three volumes of books that Cheli bought arrived by mail.’  
 b. \*[[Cheli-ka ecey chayk-ul [e] kwuiphan] [sey kwen]-i  
       -ACC -NOM  
       wuphyenulo tochakhayssta.  
 (22) \*[[Cheli-ka [e] toy-lyeko nolyekha-n] [uysa]]  
       Cheli-NOM became tried doctor  
       ‘The doctor that Cheli made efforts to become.’

To summarize, we have argued that the correct structure for nominals composed of a noun and a numeral classifier is as in (15). A noun projects a noun phrase accompanied by a demonstrative/possessive and the plural marker inserted via Late Merge, and the noun phrase enters into subject-predicate relation with the following numeral classifier phrase, forming a nominal small clause (nPredP) structure. With this structure in place, in the next section we are going to investigate the reasons for the insertion of the plural marker in a point of derivation building a nominal small structure in Korean.

### 3. The Korean plural marker is not in free variation

The insertion of the plural marker in Korean crucially hinges on the structural contexts and the interpretations of the host nouns in such contexts that come with numeral classifiers. In the following discussion we investigate this issue considering each of the structural contexts where the host nouns are situated.

#### 3. 1. When a host noun/NP is bare

The cases where a host noun/NP is bare (i.e., not modified by anything nor

accompanied by a Case/case particle) are as follows:

- (23) a. haksayng(-tul) sey    myeng-i/ul  
          student-PL        three CL-NOM/ACC  
          ‘three students’  
       b. chayk(\*-tul) sey    kwen-i/ul  
          book-PL        three CL-NOM/ACC  
          ‘three books’

We can characterize bare nouns/NPs in this structural context as follows

- (24) When host nouns followed by a numeral classifier are bare:  
       host human nouns: optional ‘tul’    vs.    host non-human nouns: (\*tul)

While host non-human nouns of bare forms strongly prefer to be interpreted as kind/mass, host human nouns of bare forms prefer to be interpreted as kind, but they can also be interpreted as individuated/singular human entities. In the latter interpretation, the plural marker ‘tul’ needs to be added on bare host human nouns before a classifier preceded/modified by the number 2 or more than 2. To repeat, bare host non-human nouns denote a kind/mass, which is fine with the following plural-denoting numeral classifier. However, when bare host human nouns denote singular human entities, they need to be supported by the plural marker before being quantified using the following plural-denoting numeral classifier.

### 3. 2 When a host noun/NP is not bare (i.e., Case/case-marked)

The examples with apparently the same structure as in (23) but with the Case marker on nouns/NPs will be as follows:

- (23) a. haksayng(-tul)-ul sey    myeng-i/ul  
          student-PL-ACC    three CL-NOM/ACC  
          ‘three students’  
       b. chayk(-tul)-ul sey    kwen-i/ul  
          book-PL-ACC    three CL-NOM/ACC  
          ‘three books’

These examples in question are described as follows

- (26) When a host noun is Case/case-marked:  
       Both human and non-human nouns: optional ‘tul’

When a non-human noun as well as a human noun is attached with a Case-marker, it is ambiguously interpreted, either denoting a kind or an entity. In other words, when a non-human noun is Case/case-marked, unlike a bare (i.e., non Case/case-marked) one it is interpreted either a kind or an entity. Having such an entity reading, the Case/case-marked non-human noun when followed by a plural-denoting numeral classifier needs to be supported by the plural marker.

### 3. 3. When a RC or demonstrative/possessive precedes either a bare or Case-marked noun/NP followed by a numeral classifier

When modified by a RC, both a Case-marked non-human noun and a bare non-human noun can denote an entity as well as a kind as in (28). Thus, when followed by a plural numeral classifier, the plural marker on both nouns enforce the entity readings of them:

(27) a. RC > non-human noun(-PL) > numeral classifier

[<sub>RC</sub> poko siph-un] chayk(-tul) (#) ney kweni-i<sup>9)</sup>  
read want book-PL four CL-NOM

‘four books that (I) want to read’

b. RC > non-human noun(-PL)-NOM > numeral classifier

[<sub>RC</sub> poko siph-un] chayk(-tul)-i ney kweni-i

(28) a. RC > human noun(-PL) > numeral classifier

[<sub>RC</sub> myentamha-yss-te-n] haksayng(-tul) (#) ney myengi-i<sup>10)</sup>  
interviewed student-PL four CL-NOM

‘four students that (I) interviewed’

b. RC > human noun(-PL)-NOM > numeral classifier

[<sub>RC</sub> myentamha-yss-te-n] haksayng(-tul)-i ney myengi-i

Recall that whether or not modified by a RC, human nouns can have an entity reading, hence being attached with the plural marker.

Likewise, when preceded by a demonstrative like the proximate ‘i’ or the distal ‘ce’, a bare non-human noun tends to be supported by the plural marker since being modified by the demonstrative it is not bare any longer. In other words, being preceded by the demonstrative it can denote an entity as well as a kind, hence being supported by the plural marker that enforces its entity reading, as in

9)

(i) . . . [제가 본] 책들 두 권을 추천드립니다 . . .

(ii) . . . [<오마이뉴스>에서 발간한] 책들 두 권을 받아갈 . . .

10)

(i) [확진자와 접촉한] 학생 3명이 추가로 확인됐다

(ii) [비슷한 학업 수준의 동일 교과목을 희망하는] 학생 3명이 . . .

(29b).<sup>11)</sup>

- (29) a. i/ce            haksayng(-tul) sey    myeng-i/ul  
           this/yonder student-PL    three    CL-NOM/ACC  
           'these/those three students'
- b. i/ce            chayk(-tul) sey    kwen-i/ul  
           this/yonder book-PL    three CL-NOM/ACC  
           'these/those three books'

A different picture emerges when a demonstrative modifies a non-bare Case-marked noun. In this case, demonstrative such as the proximate 'i' and the distal 'ce' can be ambiguously interpreted; as 'this' and 'that', on the one hand, and 'on this side' and 'yonder', on the other hand. On the former interpretations, the human noun in (30a) needs to come with the plural marker when followed by the plural-denoting numeral classifier. But the non-human noun in (30b) can come without the plural marker even when followed by the plural-denoting numeral classifier because the reading available here is 'four copies of this/that book'. On the latter interpretations, the human and the non-human nouns without the plural marker are interpreted as denoting a situation-sensitive kind or group, thus meaning 'a group of students/books on this side/yonder', hence being compatible with the following plural-denoting numeral classifier. The same pattern of interpretations obtain with the pronominal possessive and the Genitive Case form of proper noun as in (30c):<sup>12)</sup>

- (30) a. i/ce            haksayng(-tul)-ul sey    myeng-i/ul  
           this/yonder student-PL-ACC three CL-NOM/ACC  
           'these/those three students'
- b. i/ce            chayk(-tul)-ul sey    kwen-i/ul  
           this/yonder book-PL-ACC three CL-NOM/ACC  
           'these/those three books'
- c. nay/kimkyoswu-uy haksayng(-tul)-ul sey    myeng-i/ul  
           my/Kim Prof.-GEN student-PL-ACC three    CL-NOM/ACC  
           'my/Prof. Kim's three students'

By contrast, unlike deictic demonstratives such as 'i' and 'ce' the demonstrative 'ku' (which was dubbed 'the anaphoric definite marker' in the previous section) that tends to be used as anaphoric (i.e., referring to an

11) See the previous section for the discussion of the anaphoric definite marker 'ku' in the same structural context as in (29).

12)

(i) 저 아이를 세 명 키우나 봐요.

antecedent in a text) does not allow a situation-dependent kind/group reading. Thus, followed by a plural-denoting numeral classifier, the Case-marked human and non-human nouns need to be supported by the plural marker. Of course, the non-human noun without the plural marker allows a plurality of the same token; for example, four copies of the same book in (31b).

- (31) a. ku haksayng\*(-tul) sey myeng-i/ul  
 that student-PL three CL-NOM/ACC  
 ‘these/those three students’  
 b. ku chayk\*(-tul) sey kwen-i/ul  
 that book-PL three CL-NOM/ACC  
 ‘these/those three books’

### 3. 4. When a numeral classifier precedes a host noun

When preceded and modified/quantified by a numeral classifier, Case/case marked human and non-human nouns can come with the plural marker, as in (32):

- (32) a. ney myeng(?-uy) haksayng(-tul)-ul  
 four CL-GEN student-PL-ACC  
 ‘four students’  
 b. sey kwen-uy chayk(-tul)-ul  
 three volume-GEN book-PL-ACC  
 ‘three volumes of books’

These cases represent the optionality of the plural marker. Regardless of the presence of the plural marker, both human and non-human nouns denote individual entities. To repeat, the pre-nominal numeral classifier results in predicting the entity reading of the following noun.

### References

- Borer, Hagit. 2005. *Structuring sense* I. Oxford: Oxford University Press.  
 Park, So-Young. 2008. *Functional categories: the syntax of DP and DegP in Korean*. Ph. D. dissertation. University of Southern California.  
 Park, So-Young. 2022. Two types of plurals and numeral classifiers in classifier languages: the case of Korean. *Journal of East Asian Linguistics* 31:139-177.  
 Park, Jong-Un, and Myung-Kwan Park. 2020. Syntactic structure, plurality and anaphoric interpretations of Korean numeral classifier phrases. *Language and Linguistics* 88:107-134.

## ‘NP2+무섭다’ 구문의 의미와 사건구조:

### -어휘 분해 이론에 의한 접근-

왕건민

#### 1. 서론

이 논문의 목적은 한국어 심리형용사의 특성을 고려해 어휘 분해 이론(lexical decomposition)에 따라 ‘NP2+무섭다’<sup>1)</sup> 구문의 의미 표상과 사건구조(event structure)에 대해 설명하는 데 있다. 심리술어(psychological predicate)는 인간의 심리적인 상태를 표현하는 용어로서 심리술어 구문에서 쓰일 때 경험주(experiencer)를 반드시 필요로 한다. 경험주가 없는 심리형용사 구문에서 서술어는 성상형용사(state-property adjective)로 취급되기 쉽다.

유현경(1998)은 동일한 형태의 형용사가 경험주 논항의 존재 유무에 따라 심리형용사와 성상형용사로 각각 쓰일 수 있다고 설명하였다. 심리형용사로 분류되는 ‘무섭다’는 (1가)에서 심리형용사로, (1나)에서 성상형용사로 간주하였다.

(1) 가. 나는 호랑이가 무섭다.      나. 호랑이가 무섭다.

NP1 + NP2 + A

NP2 + A

김정남(2005)은 주관형용사가 객관형용사로, 객관형용사가 주관형용사로 쓰이는 많은 형용사들이 서로 겹쳐서 교집합 이루는 것을 설명하는 데 난점이 있다고 지적하고 형용사를 새롭게 ‘상관적 형용사(성질)’과 ‘비상관적 형용사(평가)’로 분류하였다. 이런 분류는 형용사가 서로 겹쳐서 교집합을 이루는 문제점을 해소하는 면에서 의미가 있다.<sup>2)</sup>

하지만 유현경(2015)은 형용사의 의미적 특성이 통사구조와 어떻게 연관되는지에 관해서 김정남이 충분히 설명하지 않았다고 지적하였다. 유현경, 김정남 등 형용사의 분류에 관한 선행연구에서는 주로 주관과 객관의 분류 방식이 타당한지를 논의했으나, ‘무섭다’ 어휘 자체를 성상형용사로 볼 수 있는 자격이 있는지 논의하지 않았다.<sup>3)</sup>

- 1) 본 연구에서는 ‘NP1+NP2+무섭다’가 쓰이는 구문, ‘NP2+무섭다’가 쓰이는 구문으로 나누어 그중에 두 번째 구문의 의미와 사건구조를 집중적으로 다룬다. 여기서 NP1은 경험주, NP2는 자극(stimulus)의 의미역에 해당된다. 심리술어 구문에는 경험주 논항과 함께 자극 논항이 많이 등장한다. 자극 논항은 흔히 경험주 논항으로 하여금 심리적 사건을 경험하게 하는 원인을 제공한다. 심리술어 구문에서 자극 논항의 의미역은 원인(cause), 영향주(effecter) 등으로 기술되기도 한다. 본고에서 자극 논항은 대상(theme) 논항과 구별된다. ‘자극’과 ‘대상’의 구별은 남승호(2007:99)에서 참고.
- 2) 김정남(2005:103)은 형용사 자체에 객관적이거나 주관적인 속성이 들어 있는 것이 아니라 실제로 그 형용사가 서술어로 나타난 구문에 따라 그 형용사의 의미가 달라진다고 본다. 이에 따르면 그렇게 되면 형용사를 객관형용사와 주관형용사로 분류하는 것 자체가 중요하지 않다. 형용사를 어휘적인 차원에서 객관형용사와 주관형용사로 나눌 것이 아니라 동일한 어떤 형용사가 실제 구문에서 객관적 용법으로 사용되었는지 주관적 용법으로 사용되었는지를 판별하는 것이 더 합리적인 방법이라고 하였다. 하지만 본고에서는 형용사의 의미적 특성이 형용사 구문의 통사 구조와 구문의 의미 표상에 영향을 끼치기 때문에 이런 특성을 우선 고려해야 한다고 본다.
- 3) 국립국어원 표준국어대사전에 제시된 정의와 예시에서 확인되듯이, 성상형용사는 객관성 형용사와 주관성 형용사를 모두 포함한다(표준국어대사전에서 성상형용사는 ‘사물의 성질이나 상태를 나타내는 형용사’로 기술한다. 예: ‘달다’, ‘붓다’, ‘아프다’, ‘기쁘다’ 등). 기존의 입장은 성상형용사 부류 안에 있는 주관성 형용사가 속성을 표현할 수 있는 기능을 갖는다. 하지만 주관성 형용사가 과연 대상의 속성을 직접 표현할 수 있는지 따져볼 필요가 있다. 본고에서는 성상형용사 분류 안에 있는 주관성 형용사 어휘 자체가 속성을 직접적으로 표현하는 기능이 없다고 본다. 속성을 나타내는 기능은 주관

정연주(2011:216)에서 이런 술어는 외적인 대상에 대한 심리적 판단을 나타내며 여러 사람에 의해 공유되어 왔다고 말했다. 이런 해석에 대해 두 가지 의문이 남는다. 일단 ‘NP2+무섭다’ 구문에서 여러 사람에 의해 공유될 수 있는 통사적인 표상을 어떻게 포착할 수 있을까? 두 번째, 이 구문의 서술어는 과연 외적인 대상에 대한 심리적 판단인지 자극체로 인한 심리 상태의 표현인지 확인할 수 있을까?

이런 의문점에서 출발하여, 심리형용사인데도 불구하고 성상형용사로 전용되는 대표적인 단어 ‘무섭다’를 선택하여 ‘NP2+무섭다’ 구문을 분석하고자 한다. 특히 ‘NP2’와 서술어의 의미적인 관계를 통해 ‘NP2+무섭다’ 구문의 의미 표상을 어떻게 갖게 되는지 설명하고자 한다. 본고의 구성은 다음과 같다. 먼저 2장에서 심리형용사 구문의 인칭제약이라는 특징과 함께 ‘무섭다’를 성상형용사로 삼기에는 적절하지 않다는 것을 설명한다. 3장에서 인간의 심리 상태를 객관화하는 장치인 ‘-어 하다’와 사동화 장치인 ‘-게 하다’ 통합구성을 통사부에 도입하여 이를 도입하는 이유와 ‘NP2+무섭다’ 구문의 도출 과정을 논의한다. 4장에서 어휘 분해 이론을 기반으로 ‘NP2+무섭다’ 구문에 대한 사건구조와 ‘NP2+무섭다’의 사역 의미에 대해 기술적인 설명을 시도해보고자 한다. 이어 언어 유형론적인 관점에서 심리형용사 구문에서 나타난 사역 의미에 대해 살펴본다.

## 2. ‘무섭다’ 구문의 통사적 실현 양상

2장에서는 먼저 한국어 심리형용사 구문의 통사적 특성에 대해 살펴보고, 이어서 심리형용사 ‘무섭다’ 구문의 실현 양상을 확인한다. 심리형용사를 서술어로 사용하는 영어 심리술어 구문에서, 주어가 1인칭일 때와 3인칭일 때 형용사의 형태 변화는 없다. (2)는 심리형용사로 분류되는 bored 구문인데, 이 구문에서 경험주 논항이 1인칭 ‘I’이든 3인칭 ‘Sam’이든 수용 가능하다. 서술어의 형태에 변화가 없기 때문에, 서술어의 어휘 의미는 하나의 단일한 표현으로 이해된다.

(2) 가. I am bored with school. (나는 학교가 지루하다.)

나. Sam is bored with school. (샘은 학교에 싫증이 났다.) (박재희 2014:3)

하지만 한국어 심리형용사 구문에는 경험주의 인칭제약(person constraint)이 있다. 1인칭에서 3인칭으로 바뀌게 되면, (3다)의 ‘싫다→싫어하다’, (3다')의 ‘그립다→그리워하다’처럼 심리형용사의 형태 변화가 일어난다. (3가, 3가')는 서술어가 심리형용사로 분류되는 ‘싫다’, ‘그립다’ 구문이며, 경험주 논항이 1인칭 ‘나’를 허용한다. 반면에 (3나, 3나')에서 경험주 3인칭 ‘철수’, ‘그’는 문장에서 매우 어색하다.<sup>4)</sup> 영어 심리술어 구문과 달리 한국어 심리형용사의 구문에서 서술어의 형태에 변화가 생겼으며, 이 때문에 서술어의 어휘 의미가 달라진다.<sup>5)</sup>

(3) 가. 나는 호랑이가 싫다. / 가'. 나는 고향 집이 그리워.

나. \*철수는 호랑이가 싫다. / 나'. \*그는 고향 집이 그리워.

다. 철수는 호랑이를 싫어하다. / 다'. 그는 고향 집을 그리워해.

성 형용사가 구성하는 구문에 내포된 의미이다.

- 4) 심리형용사 구문에 3인칭이 나타날 수 있는 경우는 화용론 측면에서 수용성을 전제한다. 화용론 측면에서는 경험주를 문장 내부의 3인칭으로 보는 것이 아니라 문장 바깥의 발화자 1인칭으로 해석하기 때문에 인칭 제약을 준수한다고 본다.
- 5) 본고에서는 ‘그립다’가 ‘그리워하다’ 형태로 변경될 때 의미도 달라진다고 본다. ‘그리워하다’의 의미는 ‘그립다’의 의미에 사건 의미 BECOME이 부가된 것이다.



(4)는 'NP1+NP2+A' 구문인데 이 구문에는 서술어 형용사 앞에 명사구 두 개가 실현되어 있다. 이들 구문에서 표면적 통사 구조가 같아 보이지만 실제로 경험주의 유무에 따라 문장의 수용성도 달라진다.

(4) 가. 나는 오늘 같은 날씨가 좋다. / 가'. 날씨가 좋다.

나. 나는 그 개가 부럽다. / 나'. \*그 개가 부럽다. (김세중 1989:153)

(4가, 4나)에서 NP1 '나'가 존재하지 않는다면 정상형용사 '좋다' 구문은 아무 영향 없이 여전히 정문인 반면에, 심리형용사 '부럽다' 구문에 영향을 받아 비문이 된다. 정상형용사 '좋다' 구문에서 NP2는 A가 직접 서술하는 대상이다. 만약 정상형용사 구문에서 A가 필수적으로 요구하는 논항의 개수가 하나라면 그 논항은 NP2이다. 하지만 (4나)에서 '개'는 A가 서술하는 대상이 아니라 마음의 상태를 일으키는 자극에 해당되어 A가 요구하는 논항은 자극 외에 경험주 논항까지 총 두 개다. 심리형용사 '부럽다'는 경험주 '나'의 심리 상태를 표현하는 용어로서 NP1이 없어도 서술하는 대상이 존재하지 않기 때문에 (4나')는 비문이 된다.<sup>6)</sup> 따라서 심리형용사 구문에서 경험주 논항이 필수적인 존재이다.

한국어 심리형용사 구문에 두 번째 현저한 특징은 바로 NP2가 심리술어에 의하여 직접 서술하는 대상이 아니라는 데 있다.<sup>7)</sup> 심리형용사 구문에서 '-시-'의 호응(concord) 현상을 통해 이런 특징을 확인할 수 있다. 심리형용사 '그립다' 구문 (5가, 5가')에서 주체 존재 '-시-'는 'NP1+NP2+그립다' 구문의 NP2 '할머니'뿐만 아니라 'NP2+그립다' 구문의 NP2 '할머니'와도 호응하지 않는다. 하지만 정상형용사 구문 (5나, 5나')에서 주체 존재 '-시-'는 'NP1+NP2+화려하다' 구문의 NP2 '김 선생님'뿐만 아니라 'NP2+화려하다' 구문의 NP2 '김 선생님'과도 호응한다. '김 선생님'은 정상형용사 '화려하다'가 직접 서술하는 대상이다. 즉 '화려하다'는 김 선생님을 대한 외모(속성)를 표현한다. 이와 달리 '할머니'와 심리형용사 '그립다'와의 관계는 그렇지 않다. 만약 '할머니'가 '그립다'에 의해 직접 서술하는 대상이라면 심리형용사 '그립다'는 주체 존재 '-시-'와 결합해야 한다. 하지만 위에서 분석했듯이, 심리형용사 '그립다'는 주체 존재 '-시-'와 결합하지 못한다.

(5) 가. \*나는 그 할머니가 그리우셔. / 가'. \*그 할머니는 그리우셔.

나. 나는 김 선생님이 화려하셔. / 나'. 김 선생님은 화려하셔. (정연주 2011:214,224)

지금까지 정상형용사 구문과의 비교를 통해서 심리형용사 구문의 통사적 특성을 확인했다. 비록 대부분 선행연구에서 심리형용사는 정상형용사로 점용된다고 간주해 왔지만 여전히 정상형용사와 같은 종류의 부류로 볼 수 없다. Locke(2017:268-272)는 속성에 대해 두 가지 종류를 나눠서 설명했다. 첫 번째 속성(primary quality)은 어떤 상황에서도 대상과 분리되지 않는 속성이다. 두 번째 속성(secondary quality)은 사람한테 작용해야 알 수 있는 속성이다. 우리는 두 번째 속성에 대해 일

- 6) 비록 심리형용사 가운데 'NP2+A' 구문을 형성할 수 있는 심리형용사도 존재하지만 (5나')에 확인했듯이 이것은 심리형용사 고유의 특징이 아님을 알 수 있다. 즉 모든 심리형용사가 'NP2+A' 구문을 형성할 수 있는 것이 아니다.
- 7) 서술어(predicate)는 주어의 동작이나 속성, 상태 따위를 풀이하는 기능을 하는 문장 성분이다. 'NP2+A' 구문에서 NP2와 서술어의 사이에는 다양한 의미적 관계가 존재한다. 예를 들자면 '철수가 예쁘다' 예문에서 서술어 '예쁘다'는 '철수'를 직접 표현하는 것이다. 즉 '철수의 모습'의 속성을 서술하는 것이다. '철수'의 의미역은 대상이다. 이와 달리 '그 할머니는 그립다' 예문에서 '할머니'와 '그립다'의 의미적인 관계는 '대상+속성 표현'이라는 의미 관계가 아니다. 'NP2+A' 구문에서 NP2와 A의 의미적인 관계에 대해 본고에서는 사건의 관점에서 구체적으로 설명하고자 한다.

반적으로 주관적인 느낌을 객관적인 상태로 받아들인다. 하지만 이것은 심리형용사가 구성된 구문을 통해 구현되는 속성을 설명할 수 있을 뿐, 심리형용사 자체가 속성을 표현하는 기능을 가진다고 설명할 수 없다. 또한 정상형용사와 성상형용사로 전용된 심리형용사의 본질에는 차이가 있어서 완전히 같은 분류로 귀속시키기 어렵다. 이에 대해 (6)에서 그 차이를 살펴보자.

(6) 가. 영희는 머리가 좋고, 마음씨가 착하고, 몸이 건강하다.

나. \*영희는 영리하고 궁금하다.<sup>8)</sup>

(김세중 1989:154)

(6가)에서 보듯이 정상형용사 ‘좋다, 착하다, 건강하다’ 이들과 병렬되는 것은 문제가 없지만, (6나)처럼 성상형용사 ‘영리하다’와 성상형용사로 전용된 ‘궁금하다’가 병렬되는 것에는 문제가 있다. 이는 두 어휘 부류가 전혀 다른 종류의 부류임을 다시 보여 주는 것이다.

앞에서 심리형용사 구문에서 인칭제약, 주어와 술어의 의미적인 관계, ‘-시-’의 호응 현상 등 특징에 대해 살펴보았다. 인칭제약은 한국어 심리형용사 구문에서 반드시 고려해야 한다는 전제 하에 이제부터 심리형용사 ‘무섭다’ 구문의 경험주의 인칭제약이 어떻게 실현되는지 살펴보자.

(7) 나는/ \*너는/ \*철수는 호랑이가 무섭다.

심리형용사 ‘무섭다’ 구문 (7)에서 경험주 1인칭 ‘나’일 때 매우 자연스럽게 경험주 2인칭 ‘너’ 또는 3인칭 ‘철수’일 때 문장의 수용성이 많이 떨어진다. (7)을 통해서 인칭제약은 ‘NP1+NP2+무섭다’ 구문에 잘 반영되어 있는 것으로 확인된다. 이어 심리형용사 ‘무섭다’ 구문에서 NP2와 심리형용사의 의미적 관계를 살펴보자.

(8) 가. 나는 호랑이가 무섭다. / 가'. 호랑이가 무섭다.

나. 나는 선생님이 무섭다. / 나'. 선생님이 무섭다.

다. 나는 할아버지가 무섭다. / 다'. 할아버지가 무섭다.

(8가~다)는 심리형용사 ‘무섭다’로 구성된 예문인데, ‘무섭다’가 요구하는 논항은 NP1 ‘나’와 NP2 ‘호랑이, 선생님, 할아버지’이다. 심리형용사 구문에서 NP1이 구문에서 없으면 ‘무섭다’의 서술 기능이 발휘되지 않는다. 물론, 언뜻 보기에 심리형용사 ‘무섭다’는 여전히 (8가~다)와 같은 ‘NP2+무섭다’ 구문을 형성할 수 있다. 하지만 이것은 ‘NP1+NP2+무섭다’ 구문에서 NP1이 삭제되고, ‘NP2+무섭다’ 구문만으로 성립한다는 뜻이 아니다. 앞에도 언급했지만 NP2는 심리형용사가 직접 서술하는 대상이 아니며, ‘NP2+무섭다’ 구문이 성립하기 위해서는 이들 사이에 성립하는 의미적 관계에서 그 합리성을 확인할 수밖에 없다.

이번에는 ‘무섭다’ 구문에서 ‘-시-’의 호응 현상을 살펴보겠다. ‘-시-’가 없을 때 (8)과 같이 모두 성립 가능하다. 하지만 (9)에 확인하듯이 NP2가 ‘할아버지, 선생님’일 때 ‘NP1+NP2+무섭다’ 구문이든 ‘NP2+무섭다’ 구문이든, ‘무섭다’가 ‘-시-’와 결합할 수 없다.

(9) 가. \*나는 할아버지가 제일 무서우시다. / 가'. \*할아버지가 제일 무서우시다.

나. \*내가 선생님이 무서우시다. / 나'. \*선생님이 무서우시다.

8) 박재희(2017)는 ‘궁금하다’가 심리형용사이지만 ‘그것이 궁금하다’ 예문에서 ‘궁금하다’는 성상형용사

로 전용되어 해석될 여지가 충분하다고 언급하였다.

다. \*나는 할아버지께서 무서우시다./ 다'. \*할아버지께서 무서우시다. (문법성판단 말뭉치)

NP2는 ‘무섭다’가 직접 서술하는 대상이 아니다. 그런데도 (8가'~8다')와 같은 ‘NP2+무섭다’ 구문이 여전히 성립하는 것은 ‘NP2+A’ 구문의 문장성분(constituent of sentence)끼리 형성하는 의미적 관계가 표면의 실현 형태처럼 단순하지 않기 때문이다. 따라서 ‘NP2+A’ 구문을 분석할 때 A는 일률적으로 성상형용사로 처리되면 (9가'~9다')와 같은 언어 현상을 설명할 수 없다. 적어도 우선 A의 기본적 의미가 무엇인지, 이 구문과 관련된 통사적 현상을 확인하고, 다음에 NP2와 A의 관계를 분석하는 것이 보다 합리적인 설명 방법처럼 보인다.

정리해보자면, 심리형용사 ‘무섭다’는 일반 성상형용사와 비교할 때 많은 차이가 존재하기 때문에 성상형용사로 볼 수 없다. 심리형용사 자체가 속성을 표현할 수 없다. 심리형용사를 성상형용사로 보는 것은 심리형용사와 심리형용사 구문 자체에 대한 인식이 부족하다는 것을 의미한다. 본고에서는 ‘NP2+무섭다’ 구문을 분석할 때 여전히 심리형용사의 의미 특성과 함께 고려해야 한다는 입장에서 ‘NP2+무섭다’ 구문에는 경험주가 존재한다고 주장한다. 심층 구조에서의 경험주는 pro로 표시한다.

(11) NP2+pro+무섭다.

하지만 pro는 어디서 오는가? 왜 심층 구조에는 존재하고 표면 구조에 안 나타나는가? NP2는 원래 NP1 뒤에 있는데 지금 문두 위치에 오는 것에 대해 어떻게 이해하는지, ‘NP2’는 ‘무섭다’의 서술하는 대상이 아니라면 ‘무섭다’와의 의미적인 관계를 어떻게 봐야 하는지, 기존에 ‘외부 대상의 속성을 표현하는 용법’으로 이해해 온 견해에 대해 어떻게 설명해야 하는지 의문이 남는다. 이에 대해 다음 장에서 자세히 살펴본다.

### 3. ‘NP2+무섭다’ 구문의 도출 과정

3장에서는 ‘NP2+무섭다’ 구문의 합리성과 도출 과정을 설명하기 위해 인간의 심리 상태를 객관화하는 장치인 ‘-어 하다’ 구성과 사동 장치인 ‘-게 하다’ 구성을 통사부에 도입하여 이에 대해 설명하고자 한다.

‘NP2+무섭다’ 구문이 외부 대상의 속성에 대해 표현하는 기능을 가지려면, 외부 대상을 여러 사람이 경험해야 한다. 따라서 경험주로서의 pro는 ‘NP2+pro+무섭다’의 형식 안에, 잠재적으로 존재한다. 예를 들어 설명하자면 ‘호랑이가 무섭다’ 문장의 심층형은 ‘호랑이가 pro 무섭다’로 표상된다. 심층구조에 경험주의 존재가 있어서 ‘무섭다’는 ‘호랑이’ 자체의 속성이라기보다는 잠재된 경험주인 ‘pro’를 무섭게 느끼게 한다는 뜻으로 해석된다. 이런 의미 분석을 감안하여 ‘호랑이가 무섭다’는 ‘호랑이가 사람을 무섭게 한다’ 문장의 의미로 해석할 수 있다. 따라서 이 구문은 사역 의미를 잠재적으로 취하는 것으로 볼 수도 있다.<sup>9)</sup> 하지만 ‘NP2+pro+무섭다’ 구문에서 의미적으로 볼 때 경험주는 인칭제약이 없다. 즉 pro는 ‘여러 사람’에 해당된다. 그 이유는 주관적인 느낌이 객관적인 상태로 받아들여지기 위해서는 여러 사람의 경험이 필요하기 때문이다. 경험주 1인칭 ‘나’뿐만 아니라 ‘나’ 외에 많은 사람이 이 사건(event)에 참여해야 한다.

반면에 심리형용사 ‘무섭다’는 인칭제약의 특성을 갖고 있다. 즉 심리형용사 ‘무섭다’ 구문은 1인칭

9) 심리형용사 구문이 사역 의미의 특징을 갖는 것은 한국어의 개별적인 특징이 아니다. 유형론 관점에

서 보편성을 가지고 있다. 이에 대해서는 4장에서 다룬다.

경험주만 허용한다. 이것은 심리형용사 ‘무섭다’가 요구하는 경험주가 ‘NP2+pro+무섭다’ 구문이 요구하는 경험주와 모순된다는 것을 의미한다. 이런 모순된 경험주의 문제를 해결하기 위해, 또한 ‘NP2+무섭다’ 구문의 특징과 심리형용사 ‘무섭다’의 특징을 동시에 만족하는 차원에서 본고에서는 ‘NP2+pro+무섭다’ 구문의 통사적인 표상은 ‘-어 하다’와 관련성이 있다고 본다. 한국어 심리형용사의 내재 감정을 나타내는 기능이 여러 사람의 공유에 의하여 외현된다는 방법은 구문의 시제가 변화하지 않은 이상 외현화한 장치가 필요하다. 한국어 통합 구성 ‘-어 하다’는 주로 내적 심리 상태를 표현하는 심리형용사와 결합하여 내적 심리 상태의 외현화를 지시하는 것으로 알려져 있다.<sup>10)</sup> 이상 분석에 따라 인간의 심리 상태를 객관화는 과정은 다음과 같이 도식화할 수 있다.

(12) 인간 심리 상태를 객관화하는 과정:

A는 NP2를 무서워한다.<sup>11)</sup>

B는 NP2를 무서워한다.

C는 NP2를 무서워한다.

.....

A+B+C+.... NP2를 무서워한다.> NP1(여러 사람)은 NP2를 무서워한다.

(12)에서 A라는 사람이 NP2에 의하여 일어나는 ‘무섭다’라는 내재적인 심리 상태가 ‘-어 하다’를 통해서 외현화된다. 이어 B라는 사람이 자기의 심리 상태를 객관화한다. A, B, C 뒤에 D, F 등 존재하는 것으로 가정하고 이들을 원소로 본다면, ‘여러 사람’은 원소들의 한 집합으로 이해될 수 있다. 이 과정을 통해 여러 사람의 내재적인 심리 상태가 외현화되었다. 따라서 여러 사람에 의해 공유될 수 있는 심리형용사 구문이, 외부 대상의 속성을 표현하는 것으로 보게 되는 이유를 논리적으로 확인할 수 있다. 이런 심리형용사를 객관화하는 과정은 ‘NP2+무섭다’ 구문을 분석할 때 필수적인 전제로 고려해야 한다. 한편 (13) 예문을 통해 심리형용사 객관화의 중요성을 확인할 수도 있다.

(13) 가. 나는 뱀이 무섭다.

가'. \*모든 사람이 내가 뱀이 무섭다고 안다.

나. 나는 뱀을 무서워하다.

나'. 모든 사람이 내가 뱀을 무서워한다고 안다.<sup>12)</sup>

(13)은 심리형용사 ‘무섭다’를 중심으로 구성된 예문이다. (13가)에서 ‘나’의 감정이나 심리 상태는 경험주 ‘나’만 알 수 있는 관점에서 쓰이는 것이고, (13나)는 (13가)의 외현화된 표현이다. 경험주 ‘나’의 감정을 표현하는 구문에서 왜 ‘무섭다’가 ‘-어 하다’와 결합하는지에 대해 의문이 생길 수 있다. 외현화된 ‘나’의 감정을 보여줘야 다른 사람 모두가 ‘나’의 심리 상태를 알 수 있다. (13가)이 비문인 이유는 ‘모든 사람’이 내재화된 ‘나’의 ‘무섭다’ 심리 상태를 알 수 없기 때문이다. ‘-어 하다’라

10)현대국어의 ‘-어 하다’가 가지는 통사적 특징에 대해 최태영(1973)은 ‘심리적, 주관적 형용사에 통합하여 행동성을 높이다’라고 하였다. ‘-어 하다’의 의미 기능에 대해 최정진(2005)은 ‘형태적 표출’이라고 보았다. 즉, 이들은 심리 상태를 표현하는 형용사의 의미 및 개념을 겉으로 드러나는 특정한 형태에 의해서 파악한다.

11)‘무서워하다’는 ‘무섭--+어 하다’의 구성으로 이루어진 본용언과 보조용언의 결합이다. 원칙적으로는 띄어 써야 하지만, 심리형용사 ‘무섭다’에 ‘하다’가 결합하여 이를 동사화한 하나의 단어로 파악할 수 있는 것이므로, 앞말에 붙여 적도록 하고 있다.

12) (13가',13나') 예문은 문법성판단 말뭉치에서 가져왔다.

는 심리형용사를 객관화하는 통사적인 장치에 의하여야만 ‘모든 사람’이 ‘나’의 심리를 인지할 수 있다. (13나)의 문장 자체가 수용성이 있다는 것은 아마도 이런 이유로 가능한 것이다.

한편 위와 같은 입장을 동의한다고 해도 ‘NP1+NP2+무서워한다’와 ‘NP2+pro+무섭다’는 무슨 관계가 있을까라는 질문이 있을 수 있다. 본고에서는 사역 의미를 가진 ‘NP2+pro+무섭다’ 구문이 ‘NP1+NP2+무서워하다’의 사동문에서 도출된 것으로 보고자 한다. 한국어의 전형적인 사동은 장형사동과 단형사동이 있는데<sup>13)</sup> 형용사의 경우는 주로 단형사동 방식으로 사동이 이루어지지만 ‘NP1+NP2+A(어근)+-어 하다’ 구문의 사동문은 이론적으로 ‘-게 하다’ 사동 장치에 의하여 실현된다.<sup>14)</sup> 인간 심리 상태를 객관화한 구문의 사동문은 다음과 같다.

(14) NP2 + NP1 + A<sub>심리</sub>(어근) + -어 하 + -게 하다.

하지만 유념할 점은 ‘무섭다’ 구문이 사동화될 때의 표면 형태가 ‘무서워하게 하다’로 실현되지 않고 ‘무섭게 하다’로 실현된다는 것이다. 이것에 대해 본고에서 다음 방식으로 설명한다. 아시다시피 형용사는 주로 단형사동 방식으로 사동이 이루어지나 장형사동은 형용사와 결합이 불가능하다. 예를 들어, 형용사 ‘넓다’ 경우는 (15가)의 ‘넓히다’와 같은 단형사동이 가능하지만 (15나)처럼 장형사동 ‘넓게 하다’는 불가능하다.

(15) 가. 일꾼들이 길을 넓히다.

나. (?)일꾼들이 길을 넓게 한다.

다. 일꾼들이 길을 넓어지게 한다.

(서정수 1996:1086)

서정수(1996:1086)는 ‘넓다’와 같은 형용사가 ‘넓어지게 하다’와 같은 방식으로 사동화를 설명하는데 이는 다음과 같이 분석된다. 형용사 ‘넓다’는 (16가)와 같이 동사화를 거쳐서 사동화를 하면 자연스럽게 되는 것이다. 이와 비슷한 방식으로 ‘넓다’가 단형사동의 사동화되는 과정을 분석하여 본다면 (16나)와 같이 될 것이다. (16)와 같은 분석과 해석에서 단형사동과 장형사동은 어느 경우나 나란히 동사화를 거쳐 사동화를 이루게 되는 것이 더 타당해 보인다. 그렇게 보아야만 단형사동과 장형사동의 경우는 형평성을 이룰 수 있다.

(16) 가. 넓 + -어 지 + -게 하다 → 넓어지게 하다

형용사 동사화 사동화 동사성

나. 넓 + ∅ + 히 → 넓히다

형용사 동사화 사동화 동사성

본고에서는 이런 분석을 받아들이는 입장에서 ‘무섭다’의 사동문을 분석한다. 심리형용사 ‘무섭다’ 경우는 단형사동이 존재하지 않고 장형사동문만 존재한다. 하지만 ‘무섭다’의 장형사동문은 심리형용

13)한국어 사동에 대한 연구에서는 접사 ‘-이-, -히-, -리-, -기-’ 등에 의하여 형태적으로 파생된 단형사동과 ‘-게 하다’에 의한 통사적 구성으로 실현되는 장형사동이라는 사동의 두 가지 형태에 대한 통사적인 고찰이 주된 관심사였다.

14)사동을 ‘시킴’의 의미와 관련하여서만 이해한다면, 형용사에서 파생된 사동사가 쓰인 문장을 사동문으로 보기 어려울 수도 있다. 그러나 ‘사동’이 나타내는 의미의 핵심은 ‘원인-결과’의 복합 사태이다. 이러한 관점에서 ‘사동’을 이해한다면 형용사에서 파생된 사동사가 쓰인 문장이 ‘시킴’의 의미를 지니지 않더라도 충분히 사동의 범위 내에 들어올 수 있다(남기심 외 (2019:465) 참고).

사가 직접 ‘-게 하다’와 결합한 것이다. 비동사성용언이 사동화하기 전에 동사화를 해야 사동화할 수 있다는 점을 감안하여 본고에서는 ‘무섭게 하다’와 같은 장형사동에 대해 다음(17)과 같이 설명한다.

(17) 호랑이는 사람을 무섭- + ∅ + -게 하다.

형용사    동사화    사동화

(17)에서 심리형용사 ‘무섭다’는 ‘-게 하다’와 결합하기 전에 먼저 동사화하고 나서 ‘-게 하다’ 장형사동 장치와 결합한다. 심리형용사 ‘무섭다’에 동사화를 만들어주는 장치는 ‘-어 하다’이고 이는 영형태 ‘∅’로 이 구문에서 존재한다. 이상 논의를 종합하여 ‘NP2+무섭다’ 구문의 도출 과정에 대해 ‘호랑이가 무섭다’로 설명하면 다음(18)과 같다.

(18) ‘호랑이가 무섭다’ 구문 도출 과정:

가. 나는 호랑이가 무섭다. (A<sub>심리</sub>: 무섭다)~

나. 여러 사람들이 호랑이를 무서워한다. (외현화/객관화 적용:-어 하다)~

다. 호랑이가 사람을 무섭게 한다/\*무서워하게 한다. (사동 적용:-게 하다)~

라. 호랑이가 무섭다. (A<sub>심리</sub>, 무섭- +∅(-어 하다)+∅(-게 하다))

(18)은 ‘NP2+무섭다’ 구문 도출 과정에 대해 분석하였다. 이에 의하여 ‘호랑이가 무섭다’ 구문은 ‘나는 호랑이가 무섭다’ 구문의 NP1 ‘나’가 삭제된 이후 형성된 것이 아님을 알 수 있다. ‘NP2+무섭다’ 구문에서 NP2와 ‘무섭다’의 의미적인 관계를 포착하려면 단순히 ‘NP2+A’ 구문의 통사구조에 의하여 확인하기 어렵다. 왜냐하면 A에 따라 NP2의 의미역이 다르고 NP2와 용언이 형성하는 의미 관계도 다르다. 보통 정상형용사 구문에서 NP2는 대상(theme)이고, 이와 달리 ‘NP2+무섭다’의 ‘NP2’는 사동주(causer)로서, 사건이나 상태의 발생을 촉진하는 사람이나 사물 등 결과의 원인이 된다. ‘NP2+무섭다’ 구문의 서술어는 사건의 관점에서 분석하자면 심리형용사 ‘무섭다’만으로 이룬 것이 아니라 영 형태의 의미 요소를 포함하는 ‘무섭-+∅(어 하다)+-게 하다’로 이룬다. 심리형용사 ‘무섭다’는 이 서술어 사건의 구성요소일 뿐이다. ‘∅(-어 하다)+∅(-게 하다)’는 실제 표면 형태에서 외면적으로 구현되지 않은 부분으로 존재하고, 이 부분은 통사부에서 잠재적인 의미기능을 담당하고 있다. NP2는 ‘무섭다’와의 의미적인 관계는 결국 ‘원인-결과’의 관계를 귀납할 수 있다.

#### 4. ‘NP2+무섭다’ 구문의 사건구조 분석

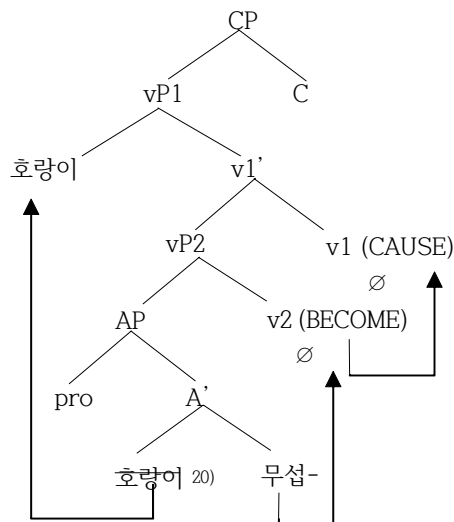
4장에서는 어휘 분해 이론<sup>15)</sup>에 따라 ‘NP2+무섭다’ 구문의 사건구조<sup>16)</sup>를 통해 ‘무섭다’와 ‘NP2’의 의미적인 관계에 대한 기술적인 설명을 시도해보고자 한다.

15) 어휘 분해 이론은 초기 생성의미론(Generative Semantics) 연구에서 유래하였고, McCawley(1968), Ross(1972) 등을 위시한 지지파와 Fodor(1970) 등 반대파 학자들 사이에서 폭넓은 반향을 일으켰다. 1980년대 후반 어휘 분해 이론은 지배결속이론 및 최소주의이론 연구 가운데 함께 발전하였다 (Huang 2017 참고).

16) 사건구조(event structure)는 서술어(사건)와 주어(참여자)의 관계를 말하는 것이고 이들이 형성한 관계는 의미 구조 관계에 해당된다.

어휘 분해는 ‘서술어 분해’라고도 한다. 이는 서술어가 서술하는 사건을 관련된 여러 의미 요소로 분해하는 방법이다. 어휘 분해 이론의 관점에서 서술어의 핵심 부분은 총 두 가지를 포함한다. 하나는 기초적으로 어휘적 의미를 담당하는 부분이며 어근(root)이라고 부르기도 한다. 또 하나는 사건의 미(eventualities)를 나타내는 부분인데, Vendler(1967)는 이를 크게 ‘상태(state)’, 행위(activity), 완성(accomplishment), 달성(achievement)이라는 총 네 가지 부류로 구분한다.<sup>17)</sup> 사건의미를 기술하는 것은 일반적으로 사건술어(eventuality predicates)인 경동사(light verbs)<sup>18)</sup> BE, DO, BECOME, CAUSE로 표현한다. 전자의 어휘는 어근 역할을 하는데, 이는 경동사와 결합하여 서술어의 핵심 부분을 이룬다. 사건술어 가운데 BECOME은 주로 사건의 변화를 강조한다. 이 변화된 사건이 궁극적으로 어떤 상태로 나타난다. CAUSE는 사역을 표현한다. BECOME과 CAUSE가 한 조합으로 나타날 때 이들은 같이 사건의 인과 관계를 표현한다. 본고에서는 ‘NP2+무섭다’ 구문을 분석하기 위해 심리형용사를 객관화하는 장치, 보조적 연결어미 ‘-어’와 보조용언 ‘하다’로 구성되는 ‘-어 하다’ 통합구성을 사건의미 BECOME으로 이해하며, 심리술어를 사동화하는 장치, 보조적 연결어미 ‘-게’와 보조용언 ‘-하다’로 구성되는 ‘-게 하다’ 통합구성은 사건의미 CAUSE로 이해한다. 한국어의 특성 상 서술어 사건을 형성하는 경우 본용언은 바로 보조용언과 연결될 수 없고, 보조적 연결어미를 매개로 해서 부가구조(adjunction structure)를 이룬다. 한국어에서 이런 부가구조는 핵 부가 구조<sup>19)</sup>로 기술된다. 이런 사건구조적 관점으로 ‘NP2+무섭다’ 구문의 서술어는 BECOME과 CAUSE까지 포함하는 ‘무섭-+BECOME+CAUSE’로 이해할 수 있다.

(19) 호랑이가 무섭다.



17) 용언 번역은 김윤신(2006:34) 참고함.

18) 한국어에서는 ‘공부를 하다’와 같은 ‘NP2를 하다’ 구문은 경동사 구문으로 다루어지는 경우가 많다. 이 구문에서 경동사 ‘하-’는 의미적으로는 기여하는 바가 거의 없고 다만 시제나 논항을 결정하는 역할을 한다. 그 외에 사건의미를 표현하기 위해 사용하는 용언도 있다. 이들은 혼자 나타날 수 없고, 무조건 동사와 같이 사용된다. 경동사는 음성 형태가 있는 경우 외에 음성 형태가 없는 영 형태도 존재한다. 만약 경동사가 영 형태로 나오면, 동사는 경동사의 자리로 옮겨져야 한다. 이동된 동사가 경동사와 병합하여 부가구조(adjunction structure)를 형성한다. 이러한 이동의 원리는 Chomsky(1995)의 최소주의 이론에 따라 분석될 수 있다. (Chomsky 1995 참고)

19) 본용언은 보조적 연결어미를 매개로 보조용언에 핵 부가되어 있으므로 본용언과 보조용언 사이에는 핵-핵 관계가 성립하게 된다.

20) 송복승(1995)에서 NP2는 ‘NP1+A<sub>심리</sub>’라는 명제의 외부에서, 그 명제에 대한 ‘원인’을 제공하는 부가어와 유사한 기능을 수행하고 있다고 한다.

(19)의 나무그림에는 분해된 부분의 병합을 통해 다층적 통사구조를 반영하였다. 위에 보이듯이 ‘호랑이가 무섭다’ 구문은 두 개의 경동사로 이루어져 있다. ‘호랑이’는 사건 원인의 사동주이고, ‘무섭-+BECOME+CAUSE’는 사건 변화의 결과이다. 경동사 BECOME이 통사부에 도입되고 ‘무섭-’이 음성 형태가 없는 경동사 BECOME 자리로 인상하여 AP ‘pro 호랑이 무섭다’와 병합하고 vP2를 형성한다. 사건의 변화를 강조하는 BECOME의 작용으로 ‘pro 호랑이 무섭-’은 ‘pro 호랑이 무섭-+BECOME’로 전환된다. 이때 ‘pro 호랑이 무섭-+BECOME’의 의미는 ‘pro가 사건의 변화를 겪었고, 변화의 결과가 바로 호랑이에 대한 겁이 난 것이다’라고 해석할 수 있다. 이어 경동사 CAUSE가 통사부에 도입되고 v2 ‘무섭-+BECOME’는 사역을 표현하는 경동사 CAUSE의 작용으로 CAUSE의 자리로 이동하여 v1을 형성한다. 한편 사동의 작용으로 ‘호랑이’는 원래 보충어(complement) 자리에서 명시어(specifier) 자리로 이동하며, 이는 최종적으로 v1’과 결합하여 vP1을 형성한다. 전체 사건구조의 의미 관계는 ‘호랑이가 한 일을 일으키게 했다. 이 일은 pro가 사건의 변화를 겪었고, 변화의 결과는 바로 겁이 난 것이다’라고 해석할 수 있다.

‘NP1+NP2+무섭다’ 구문과 ‘NP2+무섭다’ 구문은 같은 심리형용사 구문인데 이들 구문은 주어와 ‘무섭다’의 관계에 따라 구문의 의미가 달라진다. 전자 같은 심리형용사 구문에서 ‘직접 느끼는 관계’를 맺고 있다면 후자 구문은 사동적인 관계를 맺는다. 따라서 심리형용사 구문은 이런 두 가지 의미적인 특징을 지닌다고 말할 수 있다. 하지만 같은 심리형용사 구문이라도 두 가지 의미 특징을 동시에 다 가질 수 있는 것은 아니다. 경험주가 직접 느낀다는 의미에서 심리형용사 구문이 기본적으로 수용 가능하다면, 사역 의미를 지닌 구문은 언제든 다 수용 가능하지 않다. 이러한 결과는 심리형용사 구문이 외부 대상의 속성을 표현할 수 있는 시기가 아직 성숙하지 않았기 때문일 수 있다. 예를 들어, ‘\*그 개가 부럽다’와 같은 예문의 수용성이 떨어지는 이유로 ‘부럽다’라는 심리형용사가 표현하는 심리 상태가 여러 사람에게 의하여 아직 공감될 수 없기 때문에 수용하기 어렵다고 가정할 수 있다.

그럼 왜 사역 의미를 갖는 특정 구문을 통해 외부 대상 속성을 표현할까 질문할 수 있다. 우리 다음과 같은 해석을 간주할 수 있다. ‘무섭다’는 애초에 심리 상태를 표현하는 의미 특징만 취한 심리형용사이다. 하지만 외부 대상의 속성을 표현할 때, 어휘부에 관련된 형용사가 없으며, 따라서 심리형용사 ‘무섭다’를 차용하여 통사적인 조작을 함으로써 속성의 의미를 표현한다. 심리형용사 구문은 다른 의미의 기능을 할 때 어휘 자체의 형태나 의미가 변화하지 않은 이상, 구조적인 변화가 분명히 있을 것이다. 따라서 심리형용사 구문이 외부 대상 속성을 표현하는 기능을 갖는 것은 심리형용사가 속성을 표현하는 기능을 취하기보다는 심리형용사 구문의 사동화된 결과이다.

여기까지 심리형용사 구문은 두 가지의 의미적 특징을 포함한다고 설명했다. 하나는 기본 의미 유형에 해당하는 심리 상태를 표현하는 것이고 또 하나는 사역 의미를 가지는 의미 기능이다. 심리술어 구문에서 사역 의미를 가지는 것은 한국어에서만 있는 것이 아니라 다른 언어의 심리술어 구문에서도 확인할 수 있다.

(21) 가. 他无聊。(그는 지루하다.)

나. 他的课无聊。(그의 수업은 지루하다.)

(21)은 중국어 심리형용사 ‘无聊(지루하다)’ 구문에서 (21가)의 경우 경험주 ‘他(그)’가 구문의 주어이고 (21나)의 경우 ‘他的课(그의 수업)’이 구문의 주어이다. (21나)와 같은 유형에 속하는 심리형용사는 영어의 V-ing 형용사와 유사하며, (22)에 보이듯이 ‘자극(사동주)+ (使我们) + V’의 형식으로 잠재적인 경험주가 출현할 수 있다.



(22) 他的课 (使 我们) 无聊。

그의 수업은 (우리로 하여금) 지루하다.

이 유형의 어휘 분해 이론에서 밑줄로 친 부분은 실제 표층에는 외면적으로 구현되지 않으나 통사 부에서 잠재적인 의미기능을 담당하고 있음을 의미한다. 즉 '他的课无聊'(그의 수업은 지루하다)라는 문장에서 '无聊(지루하다)'는 '他的课(그의 수업)'자체의 속성이라기보다는 잠재된 경험주인 '我们(우리)'를 지루하게 한다는 뜻으로 해석된다. 따라서 이 유형의 기본적인 구조는 위치럼 경험주가 잠재적으로 억제된 사동 구문을 취하는 것으로 볼 수 있다.

한편 영어 문법에서 형용사는 동사와 대등한 위치에 있지 않다. 영어의 문장에서 형용사는 명사를 수식하거나 'be' 동사의 보어가 되어 'be'동사와 함께 하나의 서술어를 이룬다. 하지만 한국어, 중국어 심리형용사 구문과 달리, 통사구조는 여전히 똑같으나 주어에 따라 심리형용사의 형태가 달라진다.

(23) 가. I was excited (about the holiday). (나는 휴가 생각에 들떴다.)

나. His story was exciting. (그의 이야기는 아주 짜릿했다.)

(23가)는 excited의 주체가 신남을 느끼는 것을 나타낸다. 그 뜻은 '신남을 느끼다'이고 exciting은 남을 신나게 만드는 것을 나타낸다. 따라서 영어의 심리형용사 구문에서도 사역 의미를 확인할 수 있다. 한편 한국어 심리형용사 구문 'NP2+A<sub>심리</sub>'에서 NP2는 [-사람] 외에 [+사람]도 허용한다. 영어에서도 이런 부분을 확인할 수 있다. "I'm exciting" 문법적으로 사용할 수 있는 문장이고 (23가)과 달리 "I'm exciting"은 내가 남을 excited하게 만든다는 의미가 된다. 다만 실제로 이런 의미로 쓰는 경우가 많지 않을 뿐이다.

## 5. 결론

지금까지의 논의 내용을 요약하면 다음과 같다. 2장에서는 심리형용사 구문에서 인칭제약, 주어와 서술어의 관계, '-시-'의 호응 현상 등 특징을 확인함으로써 'NP2+무섭다' 구문의 '무섭다'는 성상형용사와 다양한 차이가 존재하고 NP2와 '무섭다'의 의미적 관계는 성상형용사 구문에서 NP2와 성상형용사의 의미적 관계와도 다르다는 것을 설명하였다. 즉 심리형용사는 속성을 표현하는 기능을 가진다고 보기 어렵다. 선행연구에서처럼 심리형용사를 성상형용사로 보는 것은 심리형용사 구문 자체에 대해 충분히 인식하지 못한 결과이다. 심리형용사 '무섭다'로 구성된 'NP2+무섭다' 구문은 성상형용사 구문과 본질적인 차이가 있다. 전자는 경험주가 존재하고 후자는 존재하지 않고, 또한 전자는 인칭제약을 고려한 전제 하에 형성된 구문이고 후자는 인칭제약을 고려 안 해도 된다.

3장에서 '무섭다'는 여전히 심리형용사로 보는 입장에서 'NP2+무섭다' 구문에 대해 분석하였다. 'NP2+무섭다' 구문의 도출 과정을 설명하기 위해 인간의 심리 상태를 객관화하는 장치인 '-어 하다' 구성과 사동화 장치인 '-게 하다' 구성을 통사부에 도입했고, 사역 의미를 가진 'NP2+pro+무섭다' 구문은 'NP1+NP2를 무서워한다'의 사동문에서 도출되는 것으로 제안하였다. 도출 과정의 분석에 의하여 '호랑이가 무섭다' 구문은 '나는 호랑이가 무섭다' 구문의 NP1 '나'가 삭제된 후 형성하는 것이 아님을 알게 된다. 'NP2+무섭다' 구문의 서술어는 심리형용사 '무섭다'가 아니라 영 형태의 의미 요소를 포함하는 '무섭-+∅(어 하다)+-게 하다'로 이룬다. 심리형용사 '무섭다'는 이 서술어 사건의

구성요소일 뿐이다. ‘∅(-어 하다)+∅(-게 하다)’는 실제 표면 형태에서 외면적으로 구현되지 않은 부분으로 존재하고, 이 부분은 통사부에서 잠재적인 의미기능을 담당하고 있다. NP2와 ‘무섭다’의 의미적인 관계는 결국 ‘원인-결과’의 관계로 귀납하여 설명하였다.

4장에서는 어휘 분해 이론에 따라 ‘NP2+무섭다’ 구문의 사건구조에 대한 논의를 통해 ‘NP2’와 ‘무섭다’의 사역 의미 관계를 설명하였다. ‘NP2+무섭다’ 구문의 서술어는 단순히 심리형용사 ‘무섭다’가 아니라 BECOME과 CAUSE까지 포함하는 ‘무섭+BECOME+CAUSE’를 이룬다. BECOME+CAUSE는 실제 표면 형태에서 외면적으로 구현되지 않은 부분으로 존재하고, 이 부분은 통사부에서 잠재적인 의미 기능을 담당한다. ‘NP2+무섭다’ 심리형용사 구문이 외부 대상의 속성을 표현하는 것처럼 표면에 드러난다 하더라도, 이는 경험주가 잠재적으로 억제되어 사역 의미를 취한 심리형용사 구문이다.

## 참고문헌

- 국립국어원(2020), 국립국어원 문법성 판단 말뭉치(버전1.0), <https://corpus.korean.go.kr>.
- 김세중(1989), “국어 심리형용사문의 몇 가지 문제”, 어학연구 25, 서울대학교 어학연구소, 151-165.
- 김윤신(2006), “한국어 동사의 사건구조와 어휘상”, 한국어학 30, 한국어학회, 31-61.
- 김정남(2005), 국어 형용사의 연구, 역락.
- 남기심 외 (2019), 표준국어문법론, 한국문화사.
- 남승호(2007), 한국어 술어의 사건 구조와 논항 구조, 서울대학교출판부.
- 박재희(2014), “심리형용사 구문의 기본 문형에 관한 연구”, 동남어문논집 1, 동남어 문학회, 5-34.
- 박재희(2017), “경험주 논항과 심리 표현 형용사 구문의 의미·통사구조”, 泮橋語文研究 46, 반교어문학회, 35-64.
- 서정수(1996), 국어문법, 한양대학교출판원. [SEP]
- 송복승(1995), “국어 사동문과 피동문의 논항구조 연구”, 서강대학교 국어국문학과 대학원, 박사학위논문.
- 이선웅(2012), “格的 概念과 分類”, 어문연구 40, 한국어문교육연구회, 7-32.
- 이익환 · 이민행(2005), 심리동사의 의미론, 서울: 역락.
- 임동훈(1997), “이중 주어문의 통사 구조”, 한국문화 19, 서울대학교 규장각한국학연구원, 31-66.
- 유현경(1998), 국어 형용사 연구, 한국문화사.
- 정연주(2011), “형용사의 경험주 논항 처리 방안”, 한국사전학 17, 한국사전학회, 209-238.
- 최정진(2005), “‘X+-어하-’ 構成에 대한 연구”, 어문연구 33, 한국어문교육연구회, 107-134.
- 황화상(2021), “‘하다’없이 나타나는 ‘않다’에 대하여”, 국어국문학 194, 국어국문학회, 125-155.
- Locke, John(2017), *An essay concerning human understanding*, Hastings: Delphi Publishing Ltd.
- Deng, SiYing(2010), *Chinese formal syntax*, Shanghai: Shanghai Education Press.
- Chomsky, Noam(1995), *The minimalist program*, Cambridge, Mass.: MIT Press.
- Vendler, Zeno(1967), *Linguistics in philosophy*, Ithaca: Cornell University Press.
- Huang, C-T. James(2017), “*Argument Structure, Lexical Decomposition, and Light Verb Syntax*”. *Studies in Prosodic Grammar* 2, 1-31.

## 현대중국어 비부정 不, 没(有)의 문법 기능에 관한 연구

박혁재\* · 김광섭\*\*

- |  |
|--|
| 1. 서론<br>2. 선행연구 검토<br>3. 비부정 不, 没(有)의 허가조건<br>4. 부정 함축<br>5. 결론 |
|--|

### 1. 서론

본 논문의 목표는 현대중국어 비부정(non-negation) 不(bù), 没(有)(méi(yǒu))의 분포와 문법 기능을 규명 하는 데 있다. 현대중국어의 不, 没(有)<sup>1)</sup>는 (1a-b)에서 보듯이 전형적인 부정어이다. 부정어로서 不는 동작·상태를 부정하며 没는 사건의 발생을 부정한다.

- (1) a. 近来, 他 不 忙。  
 jìnlái, tā bù máng  
 recently he NOT busy  
 요즘 그는 바쁘지 않다. [고려대 중한사전]
- b. 邮局 还 没有 开门 呢。  
 yóujú hái méiyǒu kāimén ne .  
 post-office yet NOT open-door Prt  
 우체국이 아직도 문을 열지 않았다. [고려대 중한사전]

흥미로운 사실은 不, 没(有)가 비부정어로 사용될 수도 있다는 것이다. 예를 들어 (2a), (2b)에서는 不, 没(有)를 생략하여도 문장의 진리치에 변화가 없으므로, 不, 没(有)가 비부정 용법으로 쓰이고 있다고 볼 수 있다<sup>1)</sup>.

- (2) a. 他 心脏病 突发, 差点儿 没 死了。  
 tā xīnzàngbìng tūfā , chàdiǎnr méi sǐ le 。  
 he heart-disease attack, almost MEI die Prt  
 그는 심장병이 돌발하여 까딱했더라면 죽을 뻔했다.
- b. 他 做事 粗心大意, 难免 不 受批评。  
 tā zuòshì cūxīndàyi , nán miǎn bù shòupīpíng 。  
 he work careless hard-to-avoid BU be-criticized

\* 한국외국어대학교 중국언어문화학부 강사(piaohetzai@kakao.com).

\*\* 한국외국어대학교 영어통번역학부 교수(kwangsup@hufs.ac.kr)

1) 没는 没有와 같으며, 구어에서 没을 많이 쓴다 (呂叔湘 1999:383).

1) 중국어 전통문법에서는 이를 잉여부정 구문(Redundant Negation Construction)으로 보았다. 이러한 관점은 难免, 差点儿 등 맥락에서 나타나는 단어가 이미 부정의미를 포함하고 있으며, 不, 没(有)가 (명제적) 부정기능을 상실하였음을 전제로 한다.

그는 일을 건성건성하여 비판받는 것을 피할 수 없다. [김종호 2007:98]

不, 没(有)의 문법적 기능에 대해 부정어설(王志英 2012, 陈秀清 2018), 주관적 양 표시설(王蕾 2020) 등이 주장되어 왔으나 현재까지 유력한 견해가 없다고 할 수 있다. 본 연구에서는 비부정의 不, 没(有)를 부정극어라고 주장하고자 한다.

본 연구에서는 비부정 不, 没(有)가 쓰이는 전형적 형식에 대해 그 문법기능을 전반적으로 논의 한다. 비부정 不, 没(有)의 전형적 형식으로는 ① 差点儿+没(~할 뻔하다)+VP, ② 难免/不由得+不+VP(~하지 않을 수 없다), ③ 没(有)+VP+之前((아직) ~하기 전에), ④ 회피·방지·억제·금지·거절 동사+不/没(有)+VP등 이 있다<sup>2)</sup>. 본 연구의 주요 주장은 이들 구문은 모두 부정함축을 허용하며, 이러한 부정함축이 비부정 不, 没(有)를 인허한다는 것이다.

본 연구는 다음과 같이 구성되어 있다. 2장에서는 비부정 不, 没(有)의 문법기능에 대한 선행연구를 검토하고, 3장에서는 비부정 不, 没(有)의 허가조건을 논의한다. 4장의 주요 주장은 위 4개 구문은 모두 부정함축의 의미를 허용하며, 비부정 不, 没(有)는 부정함축을 통하여 인허되는 부정극어임을 주장 한다. 마지막으로 5장에서는 본론의 논의를 요약하고 후속 연구의 과제를 제시한다.

## 2. 선행연구 검토

선행연구는 비부정 不, 没(有)가 쓰인 형식을 전반적으로 논의한 경우와 개별 형식을 논의한 경우로 나눌 수 있다. 본절에서는 비부정 不, 没(有) 전반에 대한 선행연구를 다룬다<sup>3)</sup>. 이들 선행연구는 크게 두 부류로 나눌 수 있다. 첫 번째 부류는 이들 비부정어를 잉여부정성분으로 간주하는 것이다. 비부정 不, 没(有)는 부정어 不, 没(有)와 표기와 발음이 동일하며 분포상 겹치는 점이 있다. 따라서 전통문법에서는 잉여 부정성분으로 간주해왔다. 둘째, 비부정어를 주관적 양표지의 하나로 간주하는 것이다. 1990년대 들어 중국어의 양(quantity) 범주에 대한 연구가 활발히 일어났고, 이런 중에 양에 대한 화자의 주관적인 평가를 담은 표지들이 연구되었다(李宇明 2000). 비부정의 不, 没(有) 역시 주관적 양 표지의 하나로 보는 견해가 나타나게 되었다(王蕾 2020).

### 2.1. 陈秀清(2018)의 잉여 부정성분설

부정어설의 대표적인 연구로 陈秀清(2018)를 들 수 있다. 陈은 비부정 不, 没(有)가 쓰이는 형식을 전반적으로 검토한다. 陈은 差点儿 등 상용어가 부정의미를 내재하고 있는 암시 부정어(implied negator)라고 본다. 그리고 비부정 不, 没(有)는 상용어가 의미구조 속에 가지고 있던 부정어미가 통사부로 부상(floating)하여 나타난 부정 잉여성분(the redundant of negative words)이라고 본다.

- (3) a. 老板 鼻子 差点 没 气歪了, 这是哪 跟 哪儿 啊?  
lǎobǎn bízi chàdiǎn méi qìwāile, zhè shì nǎ gēn nǎr ā?  
boss nose almost MEI flat, this is which follow where PRT?

사장님 코가 화나서 거의 삐뚤어질 뻔 했는데, 도대체 무슨 소리를 하는 거예요?

- b. 老板 鼻子 差点 气歪了, 这是哪 跟 哪儿 啊? [이상 陈秀青 2008: 7]  
lǎobǎn bízi chàdiǎn qìwāile, zhè shì nǎ gēn nǎr ā?

2) 전형적 형식은 差点儿과 같은 상용구가 부정의미를 내포하고 있다. 好+不+AP, V不死, 不/没(有)+NumP+절은 상용구에 부정의미가 없는 비전형적 형식인데, 본 연구에서 다루지 않는다. 유사 형식으로 알려진 无时无刻+(不)+VP, 非VP+(不可)는 문장원형에서 부정어구가 생략된 경우로, 别不是+VP, 不要太+AP, 看我+不+VP, 还不得(děi)+VP는 수사의문문(反问句)으로 보고 논의에서 제외한다.

3) 3장에서 不, 没(有)의 문법기능을 개별 형식으로 논의할 때, 함께 선행연구를 다룰 것이다.

boss nose almost flat, this is which follow where PRT?

사장님 코가 화나서 거의 뺨뚫어질 뻔 했는데, 도대체 무슨 소리를 하는 거예요?

(3a)와 (3b)의 차이는 没의 출현여부만 다르다. 陈에 따르면, (3a)와 (3b)의 의미는 본질적으로 같다. (3a)는 의미적으로 감추어진 부정어가 명시된 것이며 사건의 미발생 등 부정 의미를 부각시키려는 언어환경에서 더 잘 쓰일 수 있다. 그러나 3장에서 논의하듯이, 비부정의 不, 没(有)가 쓰인 경우와 쓰이지 않는 경우 의미적 차이가 있다는 관찰들이 제시되어 왔으며, 또한 비부정어 没(有)가 반드시 나타나야 하는 구 문도 존재한다. 잉여 부정성분설은 이를 설명할 수 없다.

## 2.2. 王蕾(2020)의 주관적 양 표시설

不, 没(有)가 양에 대한 화자의 평가를 함축하는 사례가 있다.

- (4) a. 我要-了一杯 纯麦奶茶, 在 屋-里 坐-了 一会儿。  
wǒ yào-le yì bēi chúnmàinǎichá, zài wū-li zuò-le yíhuìr 。

I ask-ASP one cup pure-wheat-milk-tea in room-LOC sit-ASP for-a-while

나는 순수 밀크티 한 잔을 주문하고 잠시 방에 앉아 있었다.

[이상 王蕾 2020:139]

- b. 我药吃-了 不 一会儿, 李蕴昌 由 老淡 开门 引进 屋里 来-了。  
wǒ yào chī-le bù yíhuìr, lǐyùrchāng yóu lǎodàn kāimén yǐn-jìn wū-li lái-le  
I medicin take-ASP BUfor-a-while, LiYC by old-Dan open-door lead-into room ASP-ASP  
내가 약을 먹은 지 얼마 되지 않아 리원창이 단옹을 통해 문을 열고 방으로 들어왔다.

(4a)에서 一会儿(잠시)는 화자의 양적 판단이 없다. 반면에 (4b)에서 화자는 一会儿이라는 시간이 자기 보기 에 짧다는 관점을 드러낸다. 王蕾(2020)는 이를 비부정 不, 没(有) 전체로 확장 시킨다. 예를 들어, 差点儿(没)(~할 뻔하다)는 ‘어림수량’을 나타내고 难免(不)VP(~하지 않을 수 없다)는 ‘회피량’을 나타낸다.

이러한 양의 개념의 확장은 다음과 같은 문제가 있다. 첫째, 양의 범주를 지나치게 넓게 설명하면, 전 형적인 양의 범주와 거리가 먼 경우도 포함하게 된다. 둘째, 양에 대한 평가를 하기에 앞서 기초가 되는 양 자체를 확인하기 어려운 경우도 무리하게 포함하게 된다.

- (5)(=2b) 他做事 粗心大意, 难免 丕 受批评。

tā zuòshì cūxīndàyi, nán miǎn bù shòupíping 。

he work careless hard-to-avoid BU be-criticized

그는 일을 건성건성하여 비판받는 것을 피할 수 없다. [김종호 2007:98]

- (6) 乘客 没 登机 前 要进行 安全检查。  
chéngkè méi dēngjī qián yào jìnxíng ānquánjiǎnchá.

Passengers MEI board before, must proceed safety-check

승객들이 탑승하기 전에 안전검사를 해야 한다. [김종호 2007:105]

(5)의 难免(不)는 사건발생의 필연성을, (6)의 没~前은 사건간 선후관계를 나타낸다. 양은 시간·공간·수량 등 물리속성과 관련을 맺는데, 难免(不), 没~前은 물리속성과 무관하다. 또 이들은 양의 기본 속성에 속하는 물리적 등급 차이를 구분하는 명시적인 표현을 가지고 있지 않다. 예를 들어, 很(hěn, very)과 같은 부사의 수식을 받지 못한다.

### 3. 비부정 不, 没(有)의 허가조건

본 연구에서는 비부정 不, 没(有)가 부정어도, 주관적 양 표지도 아닌 부정극어(Negative Polarity Item)라고 주장한다. 영어나 한국어 등에서 부정극어는 not, 안/못- 등 명시적 부정어 없이도 쓰일 수 있다. 특히 Linebarger (1987)는 부정극어가 부정함축(Negative Implicature, 否定隱含)에 의하여 인허될 수 있다고 주장하고 있다. 본 연구의 주요 주장은 비부정 不, 没(有)가 부정함축에 의하여 인허되는 부정극어라는 것이다. 기존 연구에서는 不, 没(有)를 부정어로만 한정할 뿐, 부정어에서 파생한 부정극어로는 보지 않았다. 하지만 비부정 不, 没(有)와 공기하는 상용구는 모두 부정의미를 내포하고 있다(陈秀清 2018). 이들 상용구는 1) 추측 难免, 保不定 및 회피·방지·억제·금지·거절의 동사 등 주절동사류, 2) 부사 差点儿, 3) 접속사 之前 등인데, 문장의 의미구성에서 핵심적인 역할을 한다. 본 연구에서는 이들이 부정함축의 의미를 가지며, 이러한 부정함축이 不, 没(有)의 출현을 허가한다고 주장한다. 부정함축을 허용하는 상용구가 不, 没(有)가 부정어가 아닌 부정극어로서 쓰일 수 있는 화맥을 형성한다는 것이다. 다음 예문은 음 운이나 표기상 동일한 표현이라도 부정함축 여부에 따라서 비부정 不, 没(有)의 허가여부가 결정된다는 것을 보여준다.

- (7) 今天 到会的 几乎 有 五千人。  
 jīntiān dào huìde jīhū yǒu wǔqiān rén 。  
 today attend-meeting DE almost be thousand people.  
 오늘 회의에 참석한 사람은 거의 5,000명이 된다 [고려대 중한사전]
- (8) 他 脚下 一滑, 几乎 (没) 摔倒。 [陈秀清 2008 : 34]  
 tā jiǎoxiǎ yīhuá, jīhū méi shuāidǎo  
 He underfoot slip almost MEI fall.  
 그는 발이 미끄러져서 거의 넘어질 뻔 했다.

(7-8)의 几乎는 문자표기나 발음이 서로 동일하나 실체는 부정의 함축 여부에서 차이를 보인다. (6)의 几乎<sub>1</sub>(거의)는 사건의 발생을 의미하나, (7)의 几乎<sub>2</sub>(거의 ~할 뻔하다)는 사건의 미발생을 의미한다. 비부정 不, 没(有)는 (7)에만 출현할 수 있다. 이러한 사실은 비부정 不, 没(有)가 부정적 상용구가 허가해줄 수 있는 부정극어(Negative Polarity Item: NPI, 否定极项)라고 주장할 수 있는 근거를 제공해 준다. 4장에서는 비부정의 4개 구문을 각각 검토하면서, 부정적 상용구의 부정 함축이 어떻게 구성되는지를 밝히고자 한다.

## 4. 부정함축

### 4.1. 差点儿+没+VP<sup>4</sup> ‘~할 뻔하다’

差点儿이 구성하는 문장은 부정의 没이 삽입될 수도 있고, 비부정의 没이 삽입될 수도 있다.

- (9) a. 他 差点儿 考上 大学。 <긍정적 예상 (대학에 붙는 것) - 부정적 현실 결과 (떨어짐)>  
 tā chàdiǎnr kǎoshàng dàxué  
 He almost get-into college.  
 그는 대학에 붙을 뻔했다. (안타깝게도 대학에 붙지 못했다.)

4) 差点儿을 差, 差一点儿, 差一点, 差点, 几乎, 险些로 대체하여 쓸 수 있다.

- b. 他差点儿 没 考上 大学。 <긍정적 예상 (대학에 붙는 것) - 긍정적 현실 결과 (붙음)>  
tā chàdiǎnr méi kǎoshàng dàxué  
He almost NOT get-into college.

그는 하마터면 대학에 합격하지 못할 뻔했다.

- (10) a. 我 骑 自行车 不小心, 差点儿 撞着 小孩儿。

wǒ qí zìxíngchē bùxiǎoxīn , chàdiǎnr zhuàngzhe xiǎoháir 。

I take bike careless, almost ran-into child.

<부정적 기대 (아이를 들이 받지 않는 것) - 긍정적 현실결과 (들이 받지 않음)>

나는 자전거를 타다가 조심하지 않아 하마터면 아이를 들이받을 뻔했다. (다행히 들이받지 않았다)

- b. 我 骑 自行车 不小心, 差点儿 没 撞着 小孩儿。

wǒ qí zìxíngchē bùxiǎoxīn , chàdiǎnr méi zhuàngzhe xiǎoháir 。 [김종호 2007:104-105]

I take bike careless, almost MEI ran-into child.

<부정적 기대 (아이를 들이 받지 않는 것) - 긍정적 현실결과 (들이 받지 않음)>

(안도의 감정을 담아) 내가 자전거를 타다가 조심하지 않아 자칫 아이를 들이받을 뻔했지 뭐야.

差点儿구문은 (9a)처럼 긍정적 기대 - 부정적 현실결과로 해석을 유도하기도 하고 (10a)처럼 부정적 기대 - 긍정적 현실결과로 해석을 유도하기도 한다<sup>5)</sup>. (9b)에서 부정어 没는 결과 해석의 방향을 반대로 유도한다. 반면에 (10b)에서는 비부정어 没는 결과 해석의 유도성에 영향을 주지 않는다. (9-10)의 대조를 통해 모든 差点儿이 비부정의 没를 허가하지 않는 것을 알 수 있다. (10)처럼 差点儿이 부정적 기대의 맥락과 만나는 경우에만, 비부정의 没를 허가한다. 다시 말해, VP가 의미하는 사건이 발생하지 않았으면 하는 부정적 기대가 있을 경우에만 没이 비부정어로 사용될 수 있다. 상용구 差点儿은 그 자체로는 <실제 발생하지 않음>이라는 부정적 의미를 가지고 있으나, 이것만으로는 비부정 没 출현의 충분조건을 제공하지 못한다. 맥락이 갖는 <발생하지 않았으면 좋겠다>라는 부정 함축과 결합하여야 비부정 没의 출현을 허가할 환경이 구성된다.

- (11) 비부정 没 허가의 충분조건 : A + B

A. 맥락적 함축 : <발생하지 않았으면 하는 부정적 기대>

B. 어휘적 의미 : 差点儿 <실제 발생하지 않음>

#### 4.2. 难免/不由得+不+VP ‘~하지 않을 수 없다’

이어서 难免/不由得+不+VP에서 비부정 不의 출현을 허용하는 부정함축의 조건을 살펴보자.

- (12) a. 一个人 难免 犯 一些 错误。  
yíge rén nánmiǎn fàn yìxiē cuòwù .

one man hard-to-avoid make some mistakes.

사람이 약간의 실수를 범하지 않을 수 없어.

- b. 一个人难免不犯一些错误。 [이상 정은 2016:293-294]

yíge rén nánmiǎn bù fàn yìxiē cuòwù .

one man hard-to-avoid BU make some mistakes.

사람이 약간의 실수를 범하지 않을 수는 없는데. (그래도 나는 실수를 범하지 않았으면 해)

5) 일상생활에서는 (9)처럼 부정적 예상 - 긍정적 현실결과로 쓰는 비중이 훨씬 많은 것으로 알려져 있다(鲁承发·陈振宇 2020)

정은(2016:291-294)에 따르면, (12a)에서 难免VP의 화용의미 초점은 'VP가 발생하는 것을 피하는 것이 아주 어려운 것이다'에 있다. (12b)에서 难免不VP의 화용의미 초점은 '(발생을 피하는 것이 아주 어렵지만)VP가 발생하지 않기를 바란다'에 있다. 그는 이러한 예를 기초로 이 형식의 不 화자가 VP의 발생을 희망하지 않음을 나타내는 주관화 표지라고 주장하였다. 정은(2016)의 논의는 양에 대한 논의가 없을 뿐,王蕾(2020)의 주관적 양 표지설(2.2)과 유사하다. 不가 我不企望(나는 바라지 않는다)와 같은 화자의 비희망을 낸다는 주장은 많은 경험적 증거의 뒷받침이 필요해 보인다. 또한 부정어 不가 어떻게 화자의 인식양태까지 나타내게 되었는지도 설명하여야 할 것이다.

본 연구에서는 难免+不+VP에서 不 자체가 어휘적으로 화자의 비희망을 나타낸다고 보다, 부정적 초기 예상(희망)의 맥락에서 출현을 허가 받는다고 본다. 먼저, 难免은 <보편적 비희망과 반하는 예상>이라는 부정함축을 가진다.

(13) a. \*小孩儿还不懂事, 难免听话。

b. 小孩儿还不懂事, 难免不听话。

아이가 아직 사리를 알지 못하여 말 안 듣는 것을 피할 수 없다. [이상 김종호 2007:98]

难免이 어휘적으로 부정을 함축하므로 후속하는 VP는 보편적으로 발생하기를 원하지 않는 사건을 가리켜야 한다. (13a)에서 听话(말을 듣다, 순종하는)는 사람들이 보편적으로 발생하기를 원하는 사건이므로 难免에 뒤이어 올 수 없다. 반면에 (13b)의 不听话(말을 안 듣다)는 사람들이 보편적으로 발생하기를 원하지 않는 사건이므로, 难免에 뒤이어 올 수 있다. 즉 (13b)의 不는 부정어로 해석되어야 한다.

(13a-b)가 보여주듯이, 难免은 差点儿的 경우처럼 이것만으로는 비부정 不의 출현을 허가하기에 부족하다. 难免은 差点儿처럼 부정의 맥락적 함축과 결합해야만, 비부정 不의 출현을 허가해줄 수 있다. (13a)는 难免이 화자의 초기 예상 곧 발생의 희망 여부가 배제된 중립적인 맥락과 결합하였으므로 이들이 不의 출현을 허가하지 못한다. 반면에 (12b)는 难免이 <발생하지 않았으면 좋겠다>는 부정함축의 맥락과 결합하였으므로 이들이 不의 출현을 허가한다. 요약하면 다음과 같다.

(14) 비부정 不 허가의 충분조건 : A + B

A. 맥락적 함축 : <발생하지 않았으면 좋겠다>

B. 어휘적 의미 : 难免 <발생하지 않았으면 좋겠다는 보편적 비희망과 반하는 예상>

### 4.3. 没+VP之前<sup>6)</sup> '(아직) ~하기 전에 (~했었다)'

예문 (15a-b)는 VP之前구문도 또한 비부정 没을 인허한다는 것을 보여준다.

(15) a. 乘客 登机 前 要 进行 安全检查。

chéngkè dēngjī qián yào jìnxíng ānquánjiǎnchá.

Passengers board before, must undergo safety-check

승객들이 탑승하기 직전에 안전검사를 해야 한다.

b. 乘客 没 登机 前 要 进行 安全检查。 [이상 김종호 2007:105]

chéngkè méi dēngjī qián, yào jìnxíng ānquánjiǎnchá.

Passengers MEI board before, must undergo safety-check

승객들이 탑승하기 전에 안전검사를 해야 한다.

6) 没(有) 대신에 未/尚未/还没(有)을, 之前 대신에 以前이나 前을 쓸 수 있다.



之前이 어휘적으로 어떤 점에서 부정적 의미를 함축하는지를 따져보기 전에 한국어와 영어의 해당 표현을 생각하여 보기로 하자. 한국어, 영어의 대응표현인 ‘-하기 전에’ ‘before’는 부정극어를 허가하는 문맥을 제공한다.

(16) 수지는 아무도 보기 전에 몰래 그 자리를 빠져 나왔다.

(17) Susie sneaked out of the place before anyone saw her.

(16-17)가 정문인 이유는 수지가 그 자리를 빠져 나온 시점에 그를 본 사람이 없다는 것을 의미하기 때 문이다. 이를 염두에 두고 (15a-b) 예문을 다시 생각하여 보기로 하자. 예문 (15a)는 단순히 비행기 탑승 하는 절차를 기술한다고 말할 수 있다. 즉 안전검사후 탑승이라는 순서로 진행된다는 것이다. 한편 (15b)는 안전검사없이 탑승도 없다는 것을 의미한다. 즉 안전 검사가 일어난 시점 이전에는 승객이 탑 승하는 사건이 발생해서는 안된다는 것이다. 그러한 면에서 之前-구문은 부정적인 함축을 발생시킨다고 말할 수 있으며, 그러한 부정적인 함축이 没을 인허하는 것이다. (18a)와 (18a)의 경우 没가 반드시 나타 나야 하는데 비슷한 방식으로 설명할 수 있다.

(18) a. 还 没 去 学校 之前, 他 就 把 作业 写 完 了。

hái méi qù xuéxiào zhīqián, tā jiù bǎ zuòyè xiěwán le  
yet MEI go-to school before, he just get homework finish PRT

그는 학교 가기 전에, 그는 숙제 쓰기를 마쳤다.

b. \*还 去 学校 之前, 他 就 把 作业 写 完 了。 [이상 李蔷薇 2020:16-17]

yet go-to school before, he just get homework finish PRT

(19) a. 没 做完 作业 之前, 不能 回家。

méi zuòwán zuòyè zhīqián, bùnéng huíjiā  
MEI finish homework before, cannot go-home

숙제를 다 끝내기 전에, 집에 돌아갈 수 없다.

b. \*做完 作业 之前, 不能 回家。

finish homework before, cannot go-home

예문 (18a)에서 没가 반드시 필요한 이유는 숙제쓰기를 마치고 전에는 학교에 가서는 안 된다는 부정적 함축이 있기 때문이다. 따라서 비록 실제로 학교를 가는 행위가 발생하였다 할지라도 이러한 문맥에서는 没가 쓰일 수 있는 것이다. 예문 (19a-b)도 같은 방식으로 설명할 수 있다. 지금까지의 논의를 요약하면 다음과 같다.

(20) 종속절 비부정 没 허가의 충분조건 : A + B

A. 맥락적 함축 : X라는 사건은 Y라는 사건 없이는 발생되어서는 안된다.

B. 어휘적 의미 : X라는 사건이 Y라는 사건을 선행한다.

#### 4.4. 회피·방지·억제·금지·거절 동사+不/没(有)+VP

避免(피하다), 防备(대비하다), 防止(방지하다), 制止(제지하다), 阻止(저지하다), 抵制(배척하다), 拦阻(막다), 忍住(억제하다), 拒绝/拒(거절하다), 抗拒(반대하다) 동사는 내포절에 비부정의 不, 没(有)를 포함할 수 있다. 이들 동사의 공통점은 후속절에 대해 부정적 사역의미를 갖는다는 점이다. 맥락적 함축은 어떠한지 살 펴보자.

- (21) 对 教育督导室 提出的 督导意见, 拒 不 采取 改进措施的 ;  
 duì jiàoyùdūdǎoshì tíchūde dūdǎoyìjiàn jù bù cǎiqǔ gǎijìncuòshīde  
 to Edu.-Supervision-Office put-forward advice, refuse BU take improvement-measures  
 교육감실에서 제기한 교육감의 의견에 대하여 개선 조치를 하는 것을 거부한다. [김용군 2016:115]

김용군(2016:118)은 拒+不+VP 형식의 비부정 不가 거절을 강조하는 차원에서 사용된다고 주장한다. 이를 좀 더 엄밀하게 말하면, 보편적으로 발생을 희망하지 않는 맥락에서 부정적 사역의 동사 拒와 결합 할 때만 비부정 不가 출현할 수 있다.

- (22) a. 见 爸爸 气 … 喘息 , 他 竭力 忍住 不 哭。 [陈秀清 2008 : 53]  
 jiàn bàba qì … chuǎnxī , tā jié lì rěnzhù bù kū  
 See dad breath catch he do-best hold-back BU cry  
 아버지가 … 헐떡이는 것을 보고, 그는 애써 울음을 참았다.  
 b. 我 使尽全力 想要 忍住 泪水。  
 wǒ shǐjìnquánlì xiǎngyào rěnzhù lèishuǐ  
 I exert-strength try hold-back tear  
 나는 울음을 참으려고 안간힘을 썼다. [고려대 한한중사전]

(22a-b) 모두 화자가 발생을 희망하지 않는 맥락에서 쓰인다. 부정적 사역의 동사 忍住(억제하다)와 보편적으로 발생을 희망하지 않는 '(우는) 사건'이 결합할 때, (22a)처럼 이들에 비부정의 不가 출현할 수 있다. 요약하면 다음과 같다.

- (23) 종속절 비부정 不, 没(有) 허가의 충분조건 : A + B  
 A. 맥락적 함축 : <(화자가) 발생을 희망하지 않음>  
 B. 어휘적 의미 : <부정적 사역>

## 5. 결론

지금까지 현대중국어에서 아직 명확하게 규명되지 못한 비부정 不, 没(有)의 분포와 문법기능에 대해 논의하였다. 기존에는 이들을 모종의 기제로 출현한 부정어 혹은 주관감량 표지로 보았으나 이것으로는 비부정 不, 没(有)의 문법기능을 충분히 설명할 수 없다. 본 연구는 비부정 不, 没(有)을 상용어구 및 맥락이 함축하는 부정의미에 의해 출현을 허가 받는 부정어라고 주장하였다. 하지만, 상용어구와 맥락 중 어느 하나라도 부정의미를 함축하지 않는 경우, 不, 没(有)는 본래의 부정어로 해석된다. 그런 점에서 不, 没(有)의 문법기능에 대해 다음과 같이 일반화할 수 있다.

- (24) 不, 没(有)의 문법기능  
 a. 부정의미를 나타내는 상용어와 부정함축의 맥락이 결합한 경우 : 부정극어  
 b. 나머지 문장, 부정의미를 나타내는 상용어와 부정함축의 맥락 중 어느 하나가 결여된 문장에서 :  
 부정어

본 연구에서는 상용구가 부정의미를 함축하는 네 개의 전형적인 비부정 不, 没(有) 구문을 대상으로 논지를 전개하였다. 앞으로의 연구과제는 비부정의 不, 没(有)가 출현한다고 여겨지는 다른 구문에도 부정극어이론에 입각한 설명을 펼칠 수 있느냐 하는 것이다. 부정극어는 일반적으로 강화(strengthening)의 의

미를 준다고 알려져 있다. 부정극어 不, 没(有)가 이러한 강화의 의미이외에 어떠한 의미를 주는지에 대해서도 좀 더 면밀히 검토할 필요가 있다.

## ※ 참고문헌

- 김용균(2016), 「한중 잉여부정의 형태와 의미」, 『중국인문과학』 64.
- 김종호(2007), 「현대중국어 잉여 부정구문 연구 - 호불(好不)VP와 차점아몰(差点儿没)VP 구조를 중심으로」, 『중국연구』 39.
- 짱용웨이(2021), 「중국어 회고유형 잉여부정구문 연구 - '几乎,险些,差点儿+没+VP' 구문을 중심으로」, 한국외국어대학교 박사논문.
- 정은(2013), 「현대 중국어 '差一点(没)VP' 구문 연구」, 부산대학교 박사논문.
- 정은(2016), 「'难免(不)VP' 구문 연구」, 『중국학』 55.
- 陈秀清(2018), 「现代汉语羡余否定研究」, 华东师范大学 博士论文.
- 董为光(2001), 「语言认知心理对“差点儿DJ”结构的影响」, 『语言教学与研究』
- 3.李治平(2010), 「'难免'和'难免不'」, 『长沙理工大学学报(社会科学版)』 25(2).
- 李蔷薇(2020), 「现代汉语“没VP之前”的句法制图研究」, 北京语言大学 硕士论文.
- 李英哲·贾梅露(1982), 「汉语数量词和否定词关系的探讨」, 『语言教学与研究』 1.
- 鲁承发 陈振宇(2020), 「透视与展望 - “差一点没VP”句式研究60年」, 『语言研究集刊』
- 2.王灿龙(2004), 「说“VP之前”与“没(有)VP之前」, 『中国语文』 5.
- 王蕾(2015), 「差点(没)VP」的形成机制」, 『理论界』 3.
- 王蕾(2020). 『现代汉语羡余否定格式研究』, 上海:复旦大学出版社.
- 王志英(2012), 「现代汉语特殊否定现象认知研究」, 上海师范大学 博士论文.
- 裘荣棠(1980), 「“难免要犯错误”和“难免不犯错误”」, 『语文学习』 4.
- 袁毓林(2012), 「动词内隐性否定的语义层次和溢出条件」, 『中国语文』
- 2.张谊生(2004), 『现代汉语副词探索』, 上海:学林出版社.
- Linebarger, Liliane. 1987. Negative Polarity and Grammatical Representation. *Linguistics and Philosophy* 10: 325-387.

## 국어사전 용례의 윤리성 자동 판별과 제안

송영숙(경희대), 정유남(중앙대)

### 1. 문제제기

본 연구는 자동태깅의 품질 향상을 위하여 사전 용례에 나타난 비윤리 표현을 중심으로 검토하고, 윤리적으로 적절한 사전 용례의 제시 방안을 모색하고자 한다. 대상 데이터는 <표준국어대사전>과 <우리말샘>, <한국어 기초사전> 등 3종의 국어사전에 나타난 용례이다. 사전 용례 자료는 2021년 8월 1일 <https://github.com/spellcheck-ko/korean-dict-nikl>를 통해 추출된 <한국어 기초사전> 266,000건, <우리말샘> 265,000건, <표준국어대사전> 191,000건으로 총 722,000건을 대상으로 한다. 이 데이터를 토대로 1차 금칙어를 통한 욕설 검출, 2차 클라우드 워커에 의한 비윤리 문장 추출, 3차 딥러닝 기반 비윤리 자동 판별기를 통한 비윤리 데이터를 추출 및 분석을 거쳤다. 먼저 금칙어 기반의 기계적 필터링은 고어 등으로 표시된 예문은 제외하고 <https://github.com/doublems/korean-bad-words>에 수집되어 공개된 비윤리 어휘와 2022년에 ‘AI HUB’에 공개된 텍스트 윤리검증 데이터<sup>1</sup> 등을 토대로 만든 비윤리 금칙어 사전을 통해 이루어졌다. 그러나 사전 사전을 기반으로 한 비윤리 문장 추출은 ‘2018년에는 ‘영동 난계 국악 축제’에 초청받아 그들만의 음악으로 한국 전통 음악계의 새로운 일원으로 인정받았으며, ...열렬한 환호를 받기도 하였다.’와 같은 문장은 비윤리 표현이 없음에도 불구하고 ‘18’이라는 숫자나 ‘년’이라는 어휘 등이 포함되어 있어 비윤리 표현으로 검출되었다. 다시 말해서, ‘18’이나 ‘년’이 비윤리 단어 사전에 포함되어 있기 때문에 비윤리성 여부에 관계없이 이들을 포함하고 있는 문장이 모두 비윤리 표현으로 검출되었다는 것이다. 또한 ‘씹덕사’, ‘산돼지’와 같이 명백하게 비윤리 표현이라고 하기 어려운 단어들을 포함하고 있는 문장들도 모두 검출되는 현상도 보였다. 이처럼 기계적 필터링은 비윤리 표현뿐만 아니라 비윤리 표현이 아닌 문장들도 함께 추출하므로 객관적 기준을 가지고 모든 문장을 검토하기 어려운 측면이 있었다. 따라서 본 연구에서는 사람에 의한 검수와 기계적 검수를 병행해야 할 것으로 판단하여 국어사전 예문 전체를 대상으로 사람과 기계적 필터링을 수행했다. 비윤리 문장에 대해 교육을 거친 클라우드 워커에 의해 비윤리 문장으로 판단되는 용례를 추출하도록 하고 이 데이터를 딥러닝 기반의 자동화 API를 통해 분석했다. 사람에 의한 검수는 한국지능정보사회진흥원(이하 NIA) 인공지능 학습용 데이터 구축사업 ‘15-1 일반상식 문장 교정 데이터’에 참여한 클라우드 워커들이 작업해서 총 6,740개의 문장을 추출했다. 자동화 도구는 tunibridge.ai에서 자연어 처리 딥러닝 모델인

<sup>1</sup> 텍스트 윤리검증 데이터는 다음 링크에서 다운로드 할 수 있다.

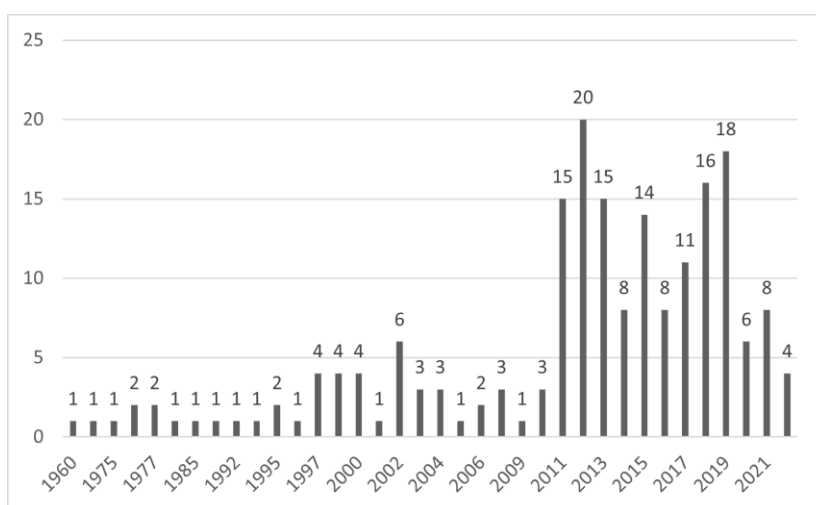
<https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=real m&dataSetSn=558>

Safety Check (Korv0.5)<sup>2</sup> 를 이용하였다. 이를 통해 비윤리 어휘를 추출하고 욕설, 모욕, 폭력/범죄 조장, 외설, 성 혐오, 연령 혐오, 인종/지역 차별, 장애인 혐오, 종교 혐오, 정치성향 혐오, 직업 혐오 등의 11가지로 분류하였다. 그 외에도 0~3까지 4 단계의 강도로 나누어 탐지해 주는 기능이 있었으나 본 연구의 특성에 부합하지 않아 비윤리 강도 자료는 사용하지 않았다. 본 연구에서는 자동으로 분류된 비윤리적 용례를 판별하여 분석하고 자동태깅의 결과에 영향을 미치는 요인을 추가로 검토하고자 한다.

다만, 본 연구에서는 비윤리성 자체가 무엇인지를 판별하는 기준은 논의 대상에서 제외된다. 이는 욕설을 포함한 비윤리 표현의 기능이나 순화를 논의하고자 하는 것이 아니라, 국어사전 용례의 비윤리 자동 판별 결과를 통해 제시된 용례의 적절성과 제시된 용례의 대표성 등을 논의하고, 더 나아가 국어사전의 용례 제시에 대한 바람직한 방향을 제안하고자 하기 때문이다.

## 2. 선행 연구

본 연구는 선행 연구의 검토를 위해 한국 학술 연구 정보 서비스(RISS)에서 학술지 논문을 대상으로 ‘욕설’과 ‘비속어’를 검색하여 관련된 논문 총 193편을 추출할 수 있었다. 또한 ‘비윤리 표현’ 혹은 ‘혐오 표현’으로 진행된 연구도 검토하여 관련 논문을 수집하였다. 이는 그동안 국어학 연구에서 비윤리 표현의 연구 흐름을 파악하기 위한 것이다.



<그림 1> 욕설 및 비속어 연구 증가 추이 그림

辱說考(1960)을 시작으로 욕설이나 비속어의 정의(이선영 2015), 형태나 의미적 특징에 대한 연구(김동연 1998, 장경희 2010, 박미은 외 2022), 청소년의 욕설이나 비속어에 대한 연구(장경희 외 2010, 전은진 외, 2011), 국어 순화의 관점에서 접근한 연구(강희숙 외 2011) 등이 주를 이루었다. 국어 순화의 관점에서 접근한 연구라고 하더라도 강희숙 외(2011)와 김태경 외(2012)에서는 청소년들의 동질감 형성 및 긴장감 해소 등 욕설의 긍정적 기능에도

<sup>2</sup> <https://demo.tunib.ai/safety/>

주목했다는 점이 특징적이다. 2010년에 들어서부터는 인터넷 사용이 본격화되면서 인터넷 댓글과 관련한 논의들이(전병철 2007, 이정복 2018) 증가하는 추세를 보였다. 그간의 연구가 특정 집단에 주목한 것과는 달리 최근에는 모든 집단에 무방비로 노출되어 있는 인터넷 댓글에 대한 다양한 층위의 연구들이 이루어져 왔다. 최근 박진원 외(2021)와 Jeong, Younghoo(2022) 등에서는 비윤리 데이터를 구축한 후, 라벨링을 세분화하여 속성별로 비윤리 표현 또는 혐오 표현 등을 필터링하는 방법을 사용하였다. 박진원 외(2021)의 데이터를 기반으로 만들어진 Safety Check (Korv0.5) 역시 특수 기호를 사용해 비윤리 표현을 마스킹하거나 비속어를 표준어로 교정해 주는 방법을 사용하고 있다.

지금까지의 선행 연구에서 주로 사용된 키워드를 각각의 논문에서 제시하고 있는 주제어를 대상으로 그 빈도를 분석해 보면 다음과 같다. 1960년부터 2022년까지 학술지에 발표된 총 193편의 논문을 ‘파이썬’의 데이터 프레임을 활용하여 키워드의 빈도를 계산한 다음, 워드클라우드를 시각화하였다.



<그림 2> 선행 연구의 주요 키워드

번호	주제어	빈도
1	비속어	37
2	욕설	32
3	slang (teenage slang, slang words )	13(4, 3)
4	은어	12
5	청소년	10
6	abusive language	7
7	swear words (swear word)	7(3)
8	vulgarism	6
9	악성 댓글	6
10	언어폭력	6
:	:	:
1123	힘의 과시	1
합	1464	

<표 1> 선행 연구에서의 주제어와 빈도

지금까지의 국어의 비윤리적 표현의 에 대한 연구는 비윤리적 표현의 긍정적인 측면과 부정적인 측면을 모두 보여 주고 있다. 부정적인 입장에서 비윤리적인 표현을 순화의 대상으로 보고 사전에 무조건 등재하지 말자고 하거나, <표준국어대사전>과 같이 용례를 제시하지 않는 것이 최선의 방법이라고 말하기는 어렵다. 그럼에도 불구하고 용례에서

나타나는 비윤리 표현에 대한 제시 방안을 연구한 논의는 그리 많지 않다. 이정복(2017, 238-239)은 성별, 및 인종, 장애 관련 용례가 부정적 태도를 강화하는 태도가 있음을 밝히고 있고, 조태린(2019) 역시 용례의 중요성을 강조하고 화용 정보나 참고 정보 등의 형식으로 비윤리 용례임을 밝히는 방안으로 모색하고 있다. 봉미경(2020)에서는 성 차별, 세대 및 계급 갈등적 내용을 담은 용례를 지향하고 객관적이고 중립적인 기술 태도를 강조하였다. 따라서 본 연구에서는 사전에 사용된 용례 중 비윤리 표현으로 판단된 용례를 긍정적 용례와 부정적 용례로 나누어 분석하고, 사전에서 용례 제시의 발전적인 방향을 모색해 보고자 한다.

### 3. 비윤리 표현의 빈도와 특징

#### 3.1. 비윤리 표현 추출

<표준국어대사전>, <우리말샘>, <한국어기초사전>의 용례를 대상으로 하여 ‘Safety Check’로 윤리성 자동 판별을 시행한 결과 나타난 비윤리 표현의 빈도는 다음과 같다.

	한국어기초사전	우리말샘	표준국어대사전	총합계
폭력위협/범죄조장	131	192	140	463
외설	72	140	157	369
모욕	85	86	173	344
성 혐오	63	153	110	326
직업 혐오	62	102	57	221
욕설	10	52	75	137
성 혐오, 외설	15	36	67	118
종교 혐오	25	52	28	105
욕설, 모욕	10	33	62	105
장애인 혐오	14	26	36	76
성 혐오, 모욕	7	28	35	70
연령 차별	19	24	21	64
정치 성향 혐오	11	34	13	58
∴	∴	∴	∴	∴
총합계	679	1468	1478	3625

<표 2> ‘Safety Check’로 판별된 비윤리 표현 빈도

클라우드 워커에 의해 비윤리 표현으로 추출된 6,740개 중에서 Safety Check가 비윤리 표현으로 판별해 낸 것은 총 3,625개였다. 각 사전의 표제어와 용례의 수가 서로 다르기 때문에 위의 빈도를 사전별로 비윤리 표현이 나타나는 절대적 기준으로 삼을 수는 없다. 다만 오픈 사전인 <우리말샘>에는 윤리성이 문제가 되는 표현이 월등히 많고, 외국인을 위한

한국어 교육용 사전인 <한국어 기초사전>이나 규범어 사전의 대표적인 <표준국어대사전>에서도 ‘비윤리 표현’의 용례가 나타남을 알 수 있다. 따라서 본 연구는 크라우드 워커와 ‘Safety Check’가 모두 비윤리적 표현으로 판정한 용례를 중심으로 사전 용례의 기술 태도와 방법을 재분석해 보고자 한다. 국어사전의 용례를 수정하거나 보완할 때, 사람이 일일이 선별하기 어렵다. 따라서 딥러닝 기반의 윤리성 자동 판별기를 활용하여 사전 용례의 윤리성을 검토해 보는 것은 아주 유용한 방법이 될 수도 있다. 따라서 본 연구는 국어사전의 용례에서 비윤리적으로 분류된 문장들이 가지고 있는 세부 정보를 살펴봄으로써 국어사전 용례 제시의 바람직한 방안을 모색해 보고자 한다.

### 3.2. 윤리성 자동 판별

윤리성 자동 판별기 ‘Safety Check’는 분석 대상 문장을 ‘비윤리 어휘’와 ‘Toxicity’로 구분한다. 비윤리 표현은 총 325개의 단어가 886번 출현했다.

비윤리 어휘	빈도	비윤리 어휘	빈도	비윤리 어휘	빈도	비윤리 어휘	빈도
계집	76	이년	9	놈이	5	깜둥이	4
놈	39	지랄	9	놈팡이	5	병어리	4
매춘	26	미친	8	미친놈	5	새끼	4
여편네	23	빨갱이	8	발기	5	성기	4
왜놈	23	처먹	8	버지	5	성행위	4
창녀	20	갈보	7	아랫도리	5	시키	4
병신	18	발정	7	육갑	5	아가리	4
년	16	변태	7	자지	5	자살	4
대가리	11	알몸	7	잡놈	5	질름발이	4
딸년	11	강간	6	젓	5	화냥	4
섹스	11	개	6	졸라	5	개돼지	3
저놈	11	쳐	6	처박	5	개망나니	3
젓가슴	10	겁탈	5	튀기	5	꺼져	3
상놈	9	계집년	5	포르노	5	난자	3
성교	9	그년	5	걸레	4	⋮	⋮

<표 3> 비윤리 어휘와 빈도

‘Safety Check’의 윤리성 자동 판별 결과는 딥러닝 모델임에도 불구하고 맥락보다는 특정 어휘에 이끌리는 경우가 있었다. ‘Safety Check’가 일베, 메갈, 워마드, 네이버 지식인, 네이버 뉴스 등의 인터넷 커뮤니티와 Korean HateSpeech Dataset 등의 오픈소스 데이터



세트에서 수집한 50만 개 이상의 혐오 표현을 학습한 것으로 제시(<https://demo.tunib.ai/safety/>참고)되어 있는데 학습의 기반이 된 특정 도메인에서는 빈도 6으로 표시된 '졸라'의 쓰임이 “다른 사람에게 차지고 끈덕지게 무엇을 자꾸 요구하다(조르다)”로 쓰인 경우보다는 비윤리 표현으로 더 자주 쓰였을 것으로 예상된다.

(2) ㄱ. 월급도 일없고 다만 일만 가르쳐 주면 그만이니 어린아이 하나를 써 달라고 **졸라** 대었다.

ㄴ. **미친** 사람에게서 악령을 쫓아내다.

예시 (2ㄱ)에서 '졸라'가 비윤리 어휘라 판단하여 '욕설'로 분류되었고 (2ㄴ)에서는 '미친'이 비윤리 어휘라 '욕설'로 분류되었다. 그러나 이들 어휘는 문장에서 비윤리 표현으로 사용되었다고 보기 어렵다. 기계 처리는 어휘의 형태에 이끌려 탐지하는 오류를 보이기도 한다. 반면에 다음의 (3)과 같이 다소 폭력적인 내용을 포함하고 있음에도 불구하고, 기계처리에서 탐지되지 않은 경우도 있었다.

(3) ㄱ. 그 사체는 날카로운 물체에 마구 **짓찢린** 것이었다.

ㄴ. 다리가 떨어지고 팔이 공중으로 솟구치고 **악살박살**이 되어... 《박종화, 임진왜란》

예시 (3ㄱ)은 '짓찢리다'의 용례이며, (3ㄴ)은 '악살박살'의 용례이다. 문장에서 나타나는 폭력적 표현은 자동 판별에서는 탐지되지 않았다. 이를 통해 명시적으로 비윤리 어휘들이 많이 포함된 도메인뿐만 아니라 비유적 표현이나 암시적 비윤리 표현이 포함된 다양한 도메인에서 비윤리 데이터가 구축되고 학습되어야 할 필요가 있다는 것을 알 수 있다. 본고에서 추출하고 분석에 사용한 사전 용례 데이터를 [https://github.com/songys/ToxicCiD\(v1.0\)](https://github.com/songys/ToxicCiD(v1.0))로 공개하였으나 그 수가 많지 않다. 이후에 영화나 문학 작품 등 비유적 표현 및 다양한 용어가 쓰이는 데이터를 추가로 구축하여 기계학습에서도 균형 잡힌 표상 학습이 이루어져야 할 것이다.

본 연구에서 다루고자 하는 비윤리 표현이 문장에서 명시적으로 드러나는 경우와 명시적으로 드러나지 않는 경우의 예를 제시해 보면 아래와 같다.

(6) ㄱ. 어머니는 늦게 나이가 많은 남자에게 시집 가라고 하였다. [연령 차별]

ㄴ. 귀국선을 탈 때까지 아가리 닥치고 있어! [욕설, 폭력위협/범죄조장, 모욕]

(6ㄱ)은 문장에서 특별하게 비윤리 표현이 명시적으로 드러나지 않은 데에 반해, (6ㄴ)은 '아가리 닥치다'와 같이 비윤리 표현이 명시적으로 드러나 있는 것을 볼 수 있다. 이와 같이, 비윤리 표현이 특정 어휘, 구어절 표현으로 직접적으로 드러나지는 않지만 맥락을 통해 비윤리성이 나타나는 경우가 있다. 윤리성 자동 판별에서는 맥락에서 실현되는 비윤리 표현을 제대로 분류되지 않을 수 있으므로 주의가 요구된다. 따라서 비윤리 표현으로 자동 분류된 결과를 바탕으로 먼저 세부 분류에서 나타나는 예문의 적절성을 살펴보기로 한다. 또한

비윤리 표현으로 자동 분류되지 않은 문장들의 특성을 살펴보고, 사전에서 제시해야 할 맥락 정보에 관하여 제안하고자 한다.

## 4. 윤리성 자동 판별 분석

### 4.1. 사전 용례 분석

비윤리 표현은 세분류가 다중으로 실현되는 특성이 있다. 윤리성 자동 판별의 결과 세부 유형 각각이 비윤리 표현으로 등장하기도 했지만, 다음 표와 같이 두 개 이상의 비윤리 표현이 동시에 나타나는 경우도 많았다.

자동 판별 분류 (Safety Check)	세부 유형 및 특징
욕설과 모욕	욕설과 모욕은 각각의 등장 비율도 높았지만 동시에 나타나는 경우도 많았다.
폭력과 범죄	가장 높은 비율로 등장했고 자주 욕설과 동시에 사용되는 특징이 있었다.
외설	두 번째로 높은 비율로 등장했고 자주 성 혐오와 함께 나타나는 특징이 있었다.
차별	연령과 인종 차별로 구분되어 있으나 전체적으로 차별 표현으로 분류할 수 있다.
혐오	성, 장애인, 종교, 정치성향, 직업 혐오 등으로 구분되어 있으며 차별적 표현과 자주 동시에 나타나는 특징이 있다.

<표 4> 비윤리 표현 세부 유형 및 특징

‘Toxicity’의 범주에는 ‘욕설, 모욕, 폭력/범죄 조장, 외설, 성 혐오, 연령 혐오, 인종/지역 차별, 장애인 혐오, 종교 혐오, 정치성향 혐오, 직업 혐오’가 있다. 본 연구는 추출된 Toxicity의 범주가 몇 개인지를 구별하는 것이 아니라 용례의 윤리성을 검토하고자 하는 것이 목적이기 때문에 유형별로 분류하여 검토하고자 한다. ‘Safety Check’에서 구분하고 있는 ‘혐오’ 범주는 혐오의 대상에 따라 분류한 것이므로 대상의 세분류를 하지 않고 5가지 유형 ‘욕설/모욕, 폭력/범죄, 외설, 차별, 혐오’로 구분해 해당 유형으로 탐지된 사전 용례를 분석하고자 한다.

#### 1) 욕설과 모욕

욕설로 분류된 용례에는 ‘년, 놈, 개자식, 병신, 상놈, 호래자, 새끼’ 등과 같은 표현이 포함된 예들이 나타났다.

(7) ㄱ. 마을 사람들은 그를 미친놈 취급했다. <우리말샘>

ㄴ. 여자로서 자식을 못 낳는다는 것은 병신구실이요 평생을 외롭게 살라는 것 같아서... <표준국어대사전>

(7ㄱ)은 ‘미친놈’이 비유리 표현으로 나타난 것인데 ‘미친놈 취급하다’와 같이 공기하여 쓰이는 용례를 제시하는 것 자체가 문제가 된다고 보기 어렵다. 자동 판별은 어휘 ‘미친놈’에 이끌려 이를 비유리 표현으로 탐지했다. (7ㄴ)의 용례에서는 ‘병신’이 비유리 표현인데 ‘자식을 못 낳는다는 것이 병신 구실’이라는 것은 여성의 역할에 대한 가치관 변화를 고려하지 않은 용례라 할 수 있다.

모욕으로 분류된 용례에는 ‘개뿔, 개돼지, 미개, 잡건, 퇴물’ 등과 같은 표현이 포함된 예들이 나타났다.

(8) ㄱ. 기생퇴물의 흔적은 그새 씻은 듯이 가서 버렸다. <표준국어사전>

ㄴ. 어머니가 모르는 남자와 관계를 가져 사생아를 낳았다. <한국어기초사전>

(8ㄱ)은 ‘퇴물’에 대한 비유리 표현으로 자동 분류되었는데, ‘퇴물의 흔적이 가시다’와 같이 ‘퇴물’의 쓰임을 알아보도록 제시한 문장이라 문장 자체는 독립적인 의미로 해석된다. (8ㄴ)은 <한국어기초사전>의 예시인데 명시적으로 비유리 표현이 드러나지는 않았지만, 편견을 조장할 수 있는 용례이므로 주의가 필요하다.

## 2) 폭력과 범죄

폭력/범죄 조장으로 분류된 용례에는 ‘돌림매’, ‘짓밟다’, ‘죽이다’, ‘인육’, ‘주리팅이’ 등이 포함되어 나타났다.

(9) ㄱ. 돌림매로 실컷 때려 줄 것이니. <표준국어대사전>

ㄴ. 도적이 머리를 베고 배를 따고 그 아이 쫓아 나란히 죽이니라. <우리말샘>

(9ㄱ)은 ‘돌림매’가 비유리 표현으로 나타난 것인데 ‘돌림매’가 ‘여러 사람이 한 사람을 돌아가며 때리는 매’라는 의미를 보이기 위해 ‘돌림매로 실컷 때리다’로 제시하였으므로 특정한 폭력이나 범죄 조장의 의미로 보기 어려울 듯하다. 이에 비해, (9ㄴ)의 용례는 <우리말샘>에 실려 있는 용례로 다소 폭력적인 의미를 담고 있으므로 추가적인 설명이 필요하다.

## 3) 외설

외설로 분류된 용례에는 ‘돌림빵’, ‘젓꼭지’, ‘알몸’, ‘매춘’, ‘발정’, ‘고추 호모’, ‘음란물’ 등이 포함되어 있는데, 이들은 주로 ‘간음하다’와 같은 술어와 같이 나타난다.

(10) ㄱ. 백목련이 이제는 웬일로 푸짐한 알몸을 드러낸 듯이 색정적인 자태로 눈앞에 육박해 오는 것이었다. <표준국어대사전>

ㄴ. 중학교 때 남자아이들하고 한방에서 놀던 여자아이가 ‘돌림빵’이 된 경우가 있다고 한다. <우리말샘>

(10ㄱ)은 문학 작품에서 가져온 용례로 ‘알몸을 드러내다’의 쓰임을 보인 것으로 직접적으로 외설적인 표현이라고 할 수 없다. (10ㄴ)의 용례는 오픈 사전인 <우리말샘>에 실려 있는 용례인데, 이는 누구나 볼 수 있는 사전이므로 중학생 남자아이, 여자아이에 대하여 ‘돌림빵’을 설명하는 것은 문제의 여지가 있다. 사전 용례는 언어 현실을 반영하고 있으면서도 해당 어휘의 쓰임을 대표할 수 있는 대표성을 지녀야 하므로, 윤리적으로 적절한 용례가 제시되어야만 한다.

#### 4) 차별

윤리성 자동 판별 ‘Safety Check’의 분류에서는 ‘인종’과 ‘연령’, ‘차별’이 함께 나타난다. 차별이 되는 대상의 나이와 인종에 따라 비윤리적 표현으로 나타나게 된다. 이 유형의 용례에서는 ‘걸레’, ‘늪은이’, ‘튀기’, ‘깜둥이’, ‘양키’ 등과 같은 인종과 나이와 관련된 차별 표현이 포함된다.

(11) ㄱ. 흑인 감독에게 깜둥이라는 비하 발언을 한 선수가 공개적으로 사과를 했다. <한국어기초사전>

ㄴ. 하여간 요새 젊은이들은 시건방지고 버르장머리가 없어서 큰일이다. <한국어기초사전>

ㄷ. 많은 혼혈인들이 청소년기에 ‘튀기’, ‘잡종’, ‘깜시’, ‘깜둥이’ 등 쏟아지는 언어폭력 속에서 어린 마음을 난도질당하며 자란다. <우리말샘>

(11ㄱ)은 ‘깜둥이’가 비윤리 표현으로 나타났는데, 용례를 제시할 때 ‘깜둥이’라는 표현이 비하 발언임을 설명하고 있기 때문에 문제되지 않는다. (11ㄴ)은 요즘 젊은이들에 대한 편견을 조장하고 언어로 인해 사회적 갈등을 초래할 수도 있기 때문에 사전 용례의 대표성과 적절성을 따져 보아 용례를 제시해야 한다. (11ㄷ)은 ‘튀기’, ‘깜둥이’와 같은 비윤리 표현이 실현되었는데, 해당 용례에서 청소년기 혼혈인에 대한 속성을 부여하거나 편견을 불러일으킬 수 있으므로 다른 용례로 대체되어야 할 것으로 보인다.

#### 5) 혐오

혐오 유형에는 혐오의 대상이 무엇인지에 따라 성 혐오, 장애인 혐오, 종교 혐오, 정치 성향 혐오, 직업 혐오 등으로 구분된다. ‘계집, 군바리, 된장녀, 병어리, 빨갱이, 삐끼, 꼬추, 앓은뱅이’ 등이 포함된 용례가 실현되었다. ‘된장녀’, ‘군바리’, ‘빨갱이’와 같은 표현은

한국사회의 현실을 반영하여 이념화된 표현으로 보인다. 이러한 혐오 표현은 보편적으로 성별, 장애 여부에 관해 나타나기도 하고, 한국사회의 특수성 때문에 ‘군대’, ‘남북 분단’ 등이 연관된 표현으로 나타나기도 한다.

- (12) ㄱ. 앓은뱅이네 식구는 꼬추네 식구들보다 대가 약했다.<표준국어대사전>  
 ㄴ. 태어날 때부터 몸에 이상이 있어 바깥출입을 안 한다.<표준국어대사전>  
 ㄷ. 옛날에는 마을마다 굽사등이가 한 사람씩은 있었어요. <우리말샘>  
 ㄹ. 태어날 때부터 왼쪽 다리가 짧은 김 씨는 절름발이라고 놀림을 받은 적이 많았다.<표준국어대사전>

예시 (12ㄱ) ~ (12ㄹ)은 장애인 혐오에 해당하는 용례이다. ‘앓은뱅이’, ‘굽사등이’, ‘절름발이’처럼 명시적인 표현이 나타나기도 하지만, (12ㄴ)과 같이 ‘몸에 이상이 있다’에서 처럼 비명시적인 경우도 있다. 이러한 용례들이 ‘장애인 혐오’로 자동 분류되었지만, ‘혐오’를 드러내는 서술 표현은 나타나지 않기 때문에 과도하게 걸러진 오류라고도 볼 수 있다. 자동 판별의 오분류 문제는 문장 의미를 전체적으로 파악하지 못하고 특정 표현만으로 비윤리성을 탐지한다는 데서 그 원인을 찾아볼 수 있다.

- (13) ㄱ. 그 여자, 얼굴은 별로이지만 그 예쁜 목소리는 사람을 잡는다. <표준국어대사전>  
 ㄴ. 계집이라고 천생 말상을 해 가지고 소박 안 맞으면 거짓말이지.<표준국어대사전>  
 ㄷ. 그대로 평범히 길러서 시집가기 전까지는 아들 겸 앞에 두고 벌어먹다가 몇 해 후에 시집이나 잘 보내자는 작정이다.<우리말샘>  
 ㄹ. 여자는 시집만 잘 가면 되지.  
 ㅁ. 승규는 호모가 좋아하는 얼굴로 이성보다 동성에게 인기가 많았다.<표준국어대사전>

예시 (13)은 성 혐오로 분류된 용례이다. (13ㄱ) ‘여자’와 ‘얼굴’, (13ㄴ) ‘계집’과 ‘말상’과 같이 외모에 대한 가치판단은 윤리적으로 문제가 될 수 있다. (13ㄷ)과 (13ㄹ)에서처럼 ‘여자’와 ‘시집가다’와 같이 편견을 보이는 용례이므로 적절하지 않다. (13ㅁ)은 <한국어기초사전>의 용례인데, ‘호모가 좋아하는 얼굴’이므로 ‘동성에게 인기가 많다’라는 인과관계를 보이는 문장은 문제가 될 여지가 크다. 특히나 사전 용례를 참조하여 한국어 어휘를 학습하는 사전에서 비윤리적 표현은 더욱 문제가 될 수 있다.

- (14) ㄱ. 코딱지만 한 어떤 공산주의 국가가 그렇다고 한다. <표준국어대사전>  
 ㄴ. 나 같은 다원주의자와 관용주의자가 강경한 빨갱이잡이로 바뀐 것이 한반도의 비극적 현상이다. <우리말샘>  
 ㄷ. 나는 아직도 너를 포함해서 친일했던 사람들이 정치를 요리하고 있는 것을 마땅찮게 보는 입장이다. <우리말샘>  
 ㄹ. 독재 정권 시절 대통령의 행동은 옳지 못했다. <표준국어대사전>

예시 (14)는 ‘정치 성향 혐오’와 관련된 용례이다. ‘공산주의’, ‘빨갱이잡이’, ‘친일’, ‘독재 정권’ 등과 같은 표현이 나타나는데, 가치판단적 서술이 문제가 될 여지가 있다. (14ㄱ)은 ‘코딱지만한 공산주의 국가’, (14ㄴ) ‘빨갱이잡이로 바뀐 것이 비극적 현상이다’ (14ㄷ) ‘친일했던 사람이 정치를 요리하고 있다’, (14ㄹ) ‘독재 정권 대통령의 행동은 옳지 못하다’는 용례의 대표성이나 중립성이 결여된 용례로 볼 수 있다. 사전 용례가 언어의 현실을 반영해서 어휘의 쓰임을 잘 보여야 하는 것은 주지의 사실이다. 다만, 용례가 지니는 편견 및 비윤리성으로 인하여 언어 현실의 대표성을 띠지 못한다면 재고해야 한다는 것이다. 지금까지 본고는 윤리성 자동 판별 방법을 활용하여 분류된 결과를 가지고 국어사전 용례의 적절성을 살펴보았다. 다음 절에서는 비윤리적 표현을 해결할 수 있는 사전 용례의 제시 방안을 보이고자 한다.

## 4.2. 사전 용례 제시 방안

본 연구에서는 딥러닝 기반의 윤리성 자동 분류 ‘Safety Check’로 나타난 결과를 바탕으로 용례의 적절성, 대표성, 윤리성 등을 살펴보았다. 일부 용례에서는 관습적으로 굳어져 자칫 편견으로 여겨질 수 있는 표현들이 나타나고 있음을 알 수 있었다. 이는 사전 용례의 윤리성과도 밀접하게 관련되기 때문에 검토가 필요하며, 용례 제시 방안도 고려되어야 한다. 비윤리성 자동 판별을 통해 비윤리적 표현으로 분류된 용례를 윤리적으로 적절한 용례로 교체하거나, 교체가 어렵다면 관련 정보를 제시함으로써 사전 용례의 대표성, 적절성, 윤리성 등을 확보해야 할 것이다. 본 연구에서는 기존 사전의 형식을 존중하되 용례의 교체가 어려운 경우라면 ‘붙임 정보’를 주어 용례의 윤리성이 널리 고려될 수 있도록 제안하고자 한다. <표준국어대사전>, <우리말샘>, <한국어기초사전>의 내부 구조를 반영하여 붙임 정보를 덧붙인 예시를 보이면 다음과 같다.

### 1) <표준국어대사전> ‘곰보’의 예시

확대하기
  축소하기
  인쇄하기
  내 단어장
  내가 살펴본 단어
  일러두기
  도움말

사전 내려받기    ✓ 목록으로 보기    ✓ 내용으로 보기

## 곰-보

발음 [곰:보 

[편집 이력](#)

---

### 「명사」

얼굴이 엷은 사람을 낮잡아 이르는 말. ≍엷보.

- 주인 여자는...박박 엷은 **곰보였으나** 속살이 몹시 희고 꺾이나 암전한 여자였다. <<이정환, 셋강>>

「비슷한말」 앓둑빼기

<출처 : 표준국어대사전 korean.go.kr>

+ 단어장 저장

## 곰-보

표준국어대사전    고려대한국어대사전    우리말샘    <    >

### 명사

1. 얼굴이 엷은 사람을 낮잡아 이르는 말.

주인 여자는... 박박 엷은 곰보였으나 속살이 몹시 희고 꺾이나 암전한 여자였다.

출처 <<이정환, 셋강>>

 차별 또는 비하의 의미가 포함되어 있을 수 있으므로 이용에 주의가 필요합니다. (차별표현 바로알기 캠페인)

<출처 : 네이버 사전 ko.dict.naver.com>

## 곰보 [곰 : 보] [명]

얼굴이 엷은 사람을 낮잡아 이르는 말. ≍엷보.

- 주인 여자는...박박 엷은 **곰보였으나** 속살이 몹시 희고 꺾이나 암전한 여자였다. <<이정환, 셋강>> 「비슷한말」 앓둑빼기

[붙임] ‘곰보’은 차별 또는 비하의 의미가 포함되어 있음, 용례 이용 시 주의 필요

<표준국어대사전>은 웹과 더불어 네이버 어학사전으로도 서비스되고 있다. ‘네이버 사전’에는 ‘차별, 비하’의 의미라 주의를 요한다는 정보가 노출되어 있다. ‘곰보’의 용례를 살펴보면, 문학 작품에서 주인공 여자의 특성을 묘사한 것으로 윤리성 자동 탐지 결과 ‘외설’로 나타났다. 사전에서 이러한 용례를 이용할 때 주의를 요한다는 붙임 정보를 추가할 수 있다. 종이 사전 형식과 달리 지면의 한계에 크게 구애되지 않는 웹 사전에서는 사용자에게 다양한 정보를 제공하기에 용이할 뿐만 아니라, 해당 용례의 윤리성 문제에 대하여 붙임 정보를 통해 사용자에게 알릴 수 있다는 장점이 있다. 본 연구에서 주장하고자 하는 것은 이와 같이 용례에서 나타나는 맥락 정보 등 윤리성을 판단하는 데에 필요한 정보를 충분히 제시하자는 것이다.

## 2) <우리말샘> ‘방불하다’의 예시

<우리말샘>은 사용자가 집필에 참여하는 오픈 사전이다. 누구나 쉽게 접근 가능하고 집필에도 참여할 수 있다는 점에서는 유용하지만, 다같이 만들어가는 사전이라는 점에서 용례의 윤리성 부분에 대한 취약점이 드러날 수 있다.

### 방불하다

품사 「형용사」 「003」 ((주로 ‘...을 방불케 하다’ 구성으로 쓰여)) 무엇과 같다고 느끼게 하다.

실전을 방불케 하는 연습. 붕괴 현장은 전쟁터를 방불케 했다. 시가지 곳곳에서는 콩 볶는 소리를 방불케 하는 소총 소리가 요란했다. 도떼기시장을 방불케 하는 흥정과 아귀다툼... <<박영한, 머나먼 송바강>>그건 이미 군대의 대열이 아니고 피난민 떼를 방불케 하는 광경이었다. <<이병주, 지리산>>수천 명의 죄수들이 발가벗고 서 있는 모습은 인육 시장을 방불케 하였다. <<최인호, 지구인>>

해당 용례는 ‘(‘주로 ~을 방불케 하다’의 구성으로 쓰여) 무엇과 같다고 느끼게 하다’의 의미에 해당한다. 상황에 대한 생생한 묘사를 위하여 작가는 용례로 ‘인육 시장을 방불케 하였다’를 제시하였다. ‘방불하다’라는 표제어 자체는 비윤리성을 포함하고 있지 않으나, 문학 작품의 맥락 속에서 비윤리 표현이 추가된 경우이다. 이러한 용례는 이정복(2017)과 봉미경(2020)에서 지적하고 있는 바와 같이 비윤리적 표현을 강화하는 측면이 있는 용례에 해당한다고 볼 수 있다. 문학 작품에 사용될 때는 앞뒤 맥락을 충분히 제시되어 그 현실의 참담함을 강조하기 위해 쓰일 수밖에 없었을 것이나, 용례에서는 이와 같은 참담함을 굳이 제시했어야만 했는지 다소 납득하기 어렵다. 또한 문학 작품 내에서 사용되는 표현의 수사적 의미와 일상생활 언어에서 사용되는 의미가 서로 다르게 해석될 수 있기 때문에, 해당 표현에 대한 수사적 표현인지 일상어적 표현인지의 여부에 대한 사용역을 밝혀 붙임 정보로 제시할 필요가 있다.

또한 동일한 표제어에 대해 <한국어기초사전> 용례에서는 “수연은 한적하고 조용한 곳에서 쉬는 순간이 천국을 방불케 할 만큼 좋다고 했다”와 같이 제시하고 있다. ‘방불하다’와 같은



표제어는 표제어 자체에 부정적인 의미를 내포하고 있지 않으므로, 표제어에도 긍정과 부정의 용례를 적절히 제시하여 언어 현실을 반영할 필요가 있다. 이와 같이 표제어에 비윤리성이 없음에도 불구하고, 용례에서는 부정적인 표현만이 노출됨으로써 용례의 비윤리적인 측면이 강화되지 않는지 면밀히 주의를 기울여야 할 것이다.

### 3) <한국어기초사전> ‘흰둥이’와 ‘간음하다’의 예시

흰둥이

2. (낮잡아 이르는 말로) 백인.

흰둥이라고 놀리다.

흰둥이라고 부르다.

태어나서 백인을 처음 본 부족 사람들은 그를 흰둥이라고 부르며 무서워했다.

학교 내의 유일한 백인 학생이었던 소년은 흰둥이 소리를 들으며 따돌림을 당했다.

가: 나는 흰둥이나 검둥이, 다 싫어.

나: 피부색이 다르다고 그렇게 편견을 가지면 안 돼.

<한국어기초사전>은 용례를 통해 학습자가 해당 어휘를 이해하고 관련 용례를 활용할 수 있어야 하므로, 특히 학습자 관점에서의 용례 집필을 요한다. 대화 상황에서도 맥락을 이해할 수 있는 예시로 교체하거나, 붙임 정보를 통해서 특정 어휘의 사용이 비윤리성과 관련이 있다는 정보를 제시할 수 있다. (2나)에서 ‘피부색이 다르다고 그렇게 편견을 가지면 안 돼’와 같이 윤리적 피드백을 주는 상황을 제시함으로써 위와 같은 문장을 사용하는 것이 일반적으로 받아들여지기 어렵다는 것을 학습자들에게 간접적으로 알려줄 수 있을 것이다. ‘간음하다’와 같은 용례도 마찬가지이다.

간음하다 (姦淫하다)

결혼한 사람이 자신의 남편이나 아내가 아닌 사람과 성관계를 맺다.

간음한 죄.

두 남녀가 간음하다.

남의 남자와 간음하다.

남의 여자와 간음하다.

유부남과 간음하다.

박 씨는 친구의 부인을 강제로 간음하는 범죄를 저질렀다.

나는 아내가 다른 남자와 간음했다는 사실을 알고 강한 배신감을 느꼈다.

가: 나는 내 아내가 간음한다면 절대 용서하지 못할 것 같아.

나: 나도 내 남편이 다른 여자와 바람을 피우는 건 상상조차 하기 싫어.

<한국어기초사전>의 대화 용례에서는 아내나 남편이 느낄 수 있는 감정을 연결하여 제시함으로써 한국에서는 간음이 받아들여지기 어려운 개념임을 명시적으로 드러내고 있다.

비윤리적인 표제어라고 해서 용례도 반드시 비윤리적일 필요는 없을 것이다. 본 연구에서는 각 사전의 목적에 부합하도록 용례를 교체하거나 붙임 정보를 통해 윤리성 정보를 제공하는 방안을 모색하고자 하였다. 이미 네이버에서 연동된 웹 사전에서는 차별 표현 정보가 명시되어 있었지만, 일부 표제어에만 해당되며 용례에 대한 윤리성 정보는 비교적 많이 나타나지 않았다. 본 연구에서 제안하고자 한 것은 사전 편찬 과정에서 윤리성 정보가 고려되어야 한다는 점이다. 또한 사전 용례의 적절성, 대표성 또한 윤리성과 관련되어 있다는 사실을 함께 언급하였다. 사전의 목적에 맞게 윤리성 정보를 구조화하는 것은 사전학적 관점에서 보다 실제적인 논의가 이루어져야 할 것이다.

## 5. 마무리

본 연구에서는 크라우드 워커와 딥러닝 기반의 윤리성 자동 판별기를 통해 국어사전 용례에서 비윤리 표현으로 검출된 문장을 중심으로 윤리성 자동판별 문제를 논의하였다. 욕설이나 금칙어와 같이 명백한 비윤리 표현 외에도 구어절이나 비유적 표현과 같이 맥락 정보를 통해 실현되는 비윤리 표현에 대해 논의하고, 이러한 비윤리적 용례에는 붙임 정보를 주어 제시하는 사전 기술 방안을 제안했다는 데에 의의가 있다. 다만, 화자의 비윤리적 의도나 이를 들은 청자의 감정에 따른 분류 등 더 세밀한 분류가 가능하나 이에 대해서는 지면을 달리하여 따로 논의하고자 한다.

사전은 언어 현실을 반영하는 역할도 하지만, 언어 사용의 적절성, 규범성, 대표성, 윤리성 등을 보여야 한다는 점도 부인할 수 없다. 사전이 웹 형식을 취하면서 지면의 제한에서도 자유로워졌다. 따라서 우리는 사용자에게 다양한 정보를 제공할 수 있게 되었다.

본 연구에서는 사전 용례의 윤리성을 고려해야 한다는 취지로 딥러닝 기반 자동 판별 분류 결과를 분석해 보았다. <표준국어대사전>은 규범성을 지니는 사전임에도 불구하고 비윤리적 표현이 용례에서 노출되어 있었고, <한국어기초사전>도 학습자들이 용례를 통해 한국어와 한국문화를 이해할 수 있도록 구성되어 있음에도 불구하고, 일부 비윤리 표현의 용례가 나타나고 있음을 확인할 수 있었다. 또한 본 연구는 사전의 용례에서 관습적으로 굳어진 비윤리성이나 편향성도 검토되어야 함을 제안하였다.

딥러닝 기반의 윤리성 자동 판별이 다양한 텍스트의 윤리성을 검증하기 위하여 활용될 수 있는 만큼, 언어 현실을 반영하는 사전 용례에서도 윤리성 판별이 적용 가능함을 알 수 있었다. 비윤리적인 표제어를 사용한 용례의 경우는 명확히 표시해 줌으로써 사용에 유의해야 함을 표시해 줄 필요가 있다. 또한 비윤리적인 표제어임에도 윤리적으로 건강한 용례를 찾아내거나 언어 사용의 현실에서 부정적인 용례를 사용할 경우에는 맥락 정보를 주는 방안 등이 고려될 필요가 있다.

본 연구에서 검토한 사전의 비윤리적 용례를 공개하여 후속 연구들에 도움이 되고자 한다. 국어 사전 저작권이 ‘크리에이티브 커먼즈 저작자표시-동일조건변경허락2.0 대한민국 라이선스’로 되어 있어 본 연구에서도 이러한 사전의 저작권을 준용하여 데이터를 공개하고 업데이트해 나갈 예정이다. 검토된 비윤리 용례 데이터는 일반적으로 데이터 및 코드는 저장소로 널리 쓰이는 “[https://github.com/songys/ToxicCiD\(v1.0\)](https://github.com/songys/ToxicCiD(v1.0))”에 분석에 사용한 용례 전체를 동일 라이선스로 공개하였다. 인터넷 댓글이나 욕설 표현 등에서는 감정적인 욕설이 많은 반면, 본 연구에서 분석한 사전류에서 드러난 비윤리적 표현은 문학 작품 등에 사용된

비유적 표현을 통하여 많이 나타났는데, 욕설 자체가 비유리 표현으로 나타나지는 않았다. 그러나 문장 혹은 상화 맥락 속에서 비유리적 표현을 드러내는 문장들이 나타나는 특징이 있으므로, 이러한 사실을 고려한다면 앞으로 문학 작품에 쓰이는 비유리 표현의 연구 및 개발에 도움이 될 수 있을 것으로 기대한다. 다만 그 수가 충분히 많지 않아서 테스트 데이터로 쓸 수 있는 정도이므로 영화나 문학 작품 등 비유적 표현 및 다양한 용어가 쓰이는 데이터를 추가로 구축하여 기계학습에서도 균형 잡힌 표상 학습이 이루어져야 할 것으로 판단된다.

#### 참고문헌

- 강희숙, 양명희(2011), 청소년의 욕설 사용의 심리적 기제 및 순화 방안, 『한국언어문학』 제79집, 한국언어학회, 255~282.
- 김동진, 한기영(2010), 청소년 언어 사용 실태 조사, 문화체육관광부.
- 김동언(1998), 국어 비속어의 개념과 특징, 인문과학논집 제5집, 강남대학교 인문과학연구소, 229~244.
- 김태경, 장경희, 김정선, 이삼형, 이필영, 전은진(2012). 청소년의 비속어. 욕설. 은어. 유행어 사용 실태와 언어 의식 연구." 국제어문 54.- 43-93.
- 민현식(2019), 금기어로서의 욕설과 공정언어(PC), 사회언어학 27-1, 한국사회언어학회, 25-65.
- 박미은, 정유남(2022). AI 텍스트 말뭉치 기반 비유리적 어휘 연구. 한국어학, 95, 241-276.
- 박진원, 나영윤, 박규병(2021). 비유리적 한국어 발원 검출을 위한 새 데이터 세트. 한국정보처리학회 학술대회논문집, 28(2), 606-609.
- 봉미경(2020). 국어사전 용례 기술의 비판적 검토 - 용례의 내용을 중심으로 -. 언어사실과 관점, 51, 35-55.
- 徐仁錫(1960), "辱說考", 『國語國文學』 22, 국어국문학회. 10~14. 양명희, 강희숙(2011), 초중고 학생들의 욕설 사용 실태와 태도에 대한 연구, 어문학 제111집, 한국어학회, 57~87.
- 원미진(2011), 한국어 학습자 사전의 용례 기술 방법에 대한 연구, 『한국사전학』 18, 한국사전학회.
- 이선영(2015), 비속어와 욕설의 개념에 대하여. 어문논집 64, 중앙어문학회, 59-80
- 이정복·박은하(2019), 네이버 뉴스 댓글의 욕설에 대한 사회언어학적 연구, 사회언어학 27-1, 한국사회언어학회, 153-178.
- 이정복. (2017). 국어사전의 차별 표현 기술에 대한 비판적 분석. 배달말, 61, 199-245.
- 이정복(2018), 뉴스 댓글에서의 욕설 사용 실태와 한국 사회의 금기 문화, 우리말연구 55, 우리말학회, 61-91.
- 장경희(2010), 국어 욕설의 본질과 유형, 텍스트언어학 29, 텍스트언어학회, 410~427.
- 장경희, 이삼형, 이필영, 김태경, 김정선, 전은진, 권우진, 오현지, 이선숙, 정자영, 오상희, 전병철(2007), "댓글에 나타난 욕설의 형태 연구", 언어학연구 11, 한국중원언어학회, 189-205.
- 전은진, 이삼형, 김정선, 김태경, 이필영, 장경희(2011), 문자 언어에 나타난 청소년 언어 실태 연구, 청람어문교육 43, 371-406.

조태린(2019), 언어 사전의 정보적 기능과 윤리적 문제에 대한 소고, 한국사전학 -34, 105-126.

Jeong, Younghoon & Oh, Juhyun & Ahn, Jaimeen & Lee, Jongwon & Mon, Jihyung & Park, Sungjoon & Oh, Alice. (2022). KOLD: Korean Offensive Language Dataset.. ArXiv, abs/2205.11315.

#### 사전류

국립국어원, 표준국어대사전, 홈페이지 URL : <https://stdict.korean.go.kr/main/main.do>

국립국어원, 우리말샘, 홈페이지 URL : <https://opendict.korean.go.kr/main>

국립국어원, 한국어 기초 사전 홈페이지 URL: <https://krdict.korean.go.kr/mainAction>

# The illocutionary degree modifier *wancen* at the discourse level

Arum Kang

Chungnam National University

arkang@ncu.ac.kr

KGGC & KSLI 2022 Joint Fall Conference

November 12, 2022

1

## Main goal & Data

- The main goal is to identify the novel type of discourse intensifier *wancen* in Korean.
- Three types of *wancen* (Kang & Shin 2012; Yeop & Yang 2016; Ahn 2015; Lee & Whang 2017; Lee 2018, a.o.):

(1) **완전**자동우산 샀어요. [Modifier noun 'complete, perfect']

wancen catong wusan sa-ss-eyo.  
complete automatic umbrella buy-Past-Decl  
'I bought an (complete) automatic umbrella.'

colloquial use (new coinage) of  
*wancen* as an intensifier

(2) **완전** 맛있는 빵을 먹었어요. [Degree modifier 'very/really']

wancen masissnun ppangul mek-ess-eyo  
WANCEN delicious bread eat-Past-Decl  
'I ate **very** delicious bread.'

(3) **완전** 서비스가 대박이에요/서비스가 대박이에요, **완전**. [Discourse modifier 'completely']

sepisu-ka taypak-ieyyo, wancen.  
service-Nom amazing-Decl, WANCEN  
'The service is **completely** amazing.'

# Main goal & Data

## 1. Modifier Noun

★  
**완전**<sup>1</sup> 完全

발음

파생어

표준국어대사전    고려대한국어대사전    우리말샘

명사

1. (주로 일부 명사 앞에 쓰여)

1. 필요한 것이 모두 갖추어져 모자람이나 흠이 없음.

    금융 시장의 **완전** 개방.

    노사분규 **완전** 타결.

3

# Main goal & Data

## 2. Degree modifier

**완전** 完全

발음

---

1. 부사(구어적으로) 보통 정도보다 훨씬 더.

    그는 모델 해도 되겠다는 말에는 고개를 저으며 "모델 사이에 있으면 **완전** 별로."라는 말로 자신을 낮췄다.

    출처 <<오에스이엔 2013년 7월>>

    당시 출연진들은 "시중에 파는 음식 같다.", "**완전** 맛있다." 등의 극찬을 쏟았다.

    출처 <<서울경제 2013년 9월>>

출처 : 우리말샘

# Main goal & Data

## 3. Discourse modifier

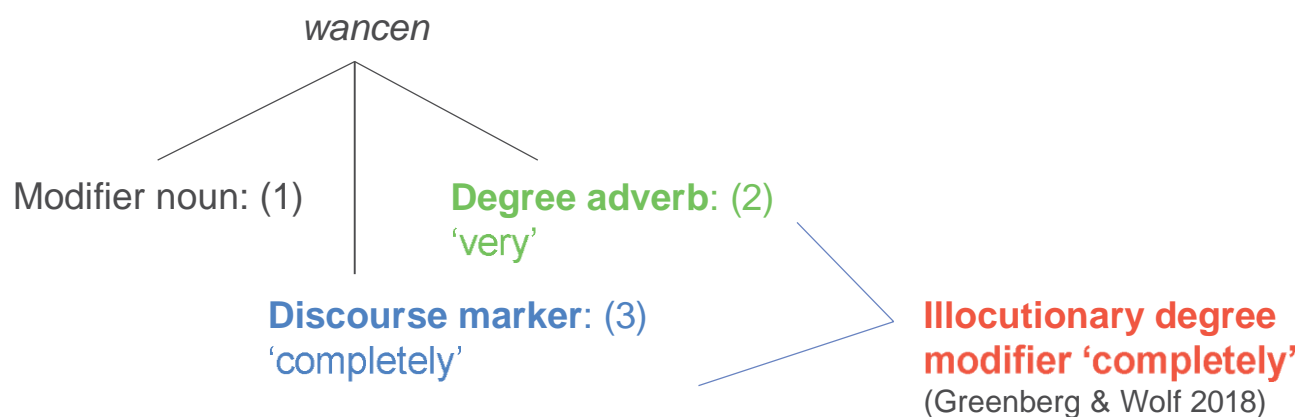
??

NOT included in a dictionary

5

## Proposal: Unified analysis

- What is the exact function of *wancen*?
- How the intensified reading in (2) and (3) arises?



# Proposal

- Unified contributions of *wancen*: **asserting the speaker's subjective maximal degree of probability** towards the utterance.
- It is an **illocutionary degree modifier** modifying a speech act operator *ASSERT* (Greenberg & Wolf 2018).

7

## *wancen* vs. *wancenhi*

(4) 초승달과 상현달은 **완전** 달라요. [Illocutionary degree modifier 'completely']

chosungtal-kwa sanghyental-un **wancen** talla-yo.

the.crescent moon-and the.dark moon-Top WANCEN different-Decl

'The crescent moon and the dark moon are **completely/totally** different.'

(5) 초승달과 상현달은 **완전달**라요. [Lexical degree modifier 'completely']

chosungtal-kwa sanghyental-un **wancenhi** talla-yo.

the crescent moon-and the dark moon-Top completely different-Decl

'The crescent moon and the dark moon are **completely** different.'

🗨️ How are they distinct semantico-pragmatically?



## Subtypes of intensifiers

INTENSIFIERS	AMPLIFIERS
	Maximizers (eg: completely, totally, absolutely) $\approx$ <i>wancen/wancenhi</i> 'completely'
	Boosters (eg: very, so, really)
<hr/>	
DOWNTONERS	
	Approximators (eg: almost, nearly, virtually )
	Compromisers (eg : more or less, kind of, rather)
	Diminishers (eg: partly, merely, slightly)
	Minimizers (eg: hardly, barely, scarcely)

Figure 1. Subtypes of Intensifiers (adapted from Quirk et al. 1985: 590)

9

## Crosslinguistic data on the close relation b/w degree modifier and discourse intensifiers

Beltrama (2015)

- (6) a. The bus is **totally** full. [Degree modifier; lexical *totally*]  
 b. Skiing around Salt Lake is **totally** awesome. [Discourse intensifier; pragmatic *totally*]

(6a): The lexical maximizer in Korean  $\approx$  *wancenhi* 'completely'

(6b): The pragmatic maximizer in Korean  $\approx$  *wancen* 'completely'

👉 Unlike English, Korean employs **distinct words** b/w (6a) and (6b).

# Outlines

## 1. Distributional restrictions

- 1) *wancen* vs. typical degree adverbs *acwu/maywu* ‘very’
- 2) *wancen* vs. *wancenhi* ‘completely’

## 2. Core properties of *wancen*

## 3. Analysis: subjective assertion operator *ASSERT*

## 4. Conclusion and further implications

11

# 1-1. Distributional restrictions: *wancen* vs. *very*

## Grammatical category of predicates co-occurring with degree modifiers (Hong 2002)

### (7) a. Adjectives

#### b. Verbs

- i. 심리동사: ‘존경하다, 숭배하다, 존중하다, 멸시하다, 증오하다, 사랑하다’
- ii. 원망동사: ‘-고 싶다’
- iii. 감정 형용사 파생 동사: ‘그리워하다, 좋아하다, 싫어하다, 미워하다, 슬퍼하다, 기뻐하다, 즐거워하다’
- iv. 형용사 기원 사동사: ‘넓히다, 좁히다, 굽히다, 낮추다, 늦추다’
- v. 형용사 기원 기동 동사: ‘넓어지다, 좁아지다, 커지다, 많아지다, 길어지다, 달라지다, 높아지다, 밝아지다’
- vi. 완성(과정동사): ‘마르다, 시들다, 썩다, 상하다, 앓다, 줄다, 붙다, 익다, 굼다, 붓다’
- vii. 일부동작 동사: ‘먹다’

#### c. Nominal predicates: ‘N+be’

## 1-1. Distributional restrictions: *wancen* vs. *very*

- (8) a. 올해 시카고의 겨울은 **아주/매우**춥다더라. [Adjective]

olhay sikhako-uy kyewul-un **acwu/maywu** chwupta-te-la.

this year Chicago-Gen winter-Top very cold-Evi-Decl

'I heard that this winter in Chicago is **very** cold.'

- b. 올해 시카고의 겨울은 **완전** 춥다더라.

olhay sikhakouy kyewulun **wancen** chwupta-te-la.

this year Chicago-Gen winter-Top WANCEN cold-Evi-Decl

'I heard that this winter in Chicago is **completely** cold.'

13

## 1-1. Distributional restrictions: *wancen* vs. *very*

- (9) a. 나 지금 **아주/매우**집에가고싶다. [원망동사]

na cikum **acwu/maywu** cip-ey ka-kosiph-ta.

I now very home-Loc go-want-Decl

'I **really** want to go home now.'

- b. 나 지금 **완전** 집에 가고 싶다.

na cikum **wancen** cip-ey ka-kosiph-ta.

I now WANCEN home-Loc go-want-Decl

'I **completely** want to go home now.'

## 1-1. Distributional restrictions: *wancen* vs. *very*

- (10) a. 영희가 철수를 **아주/매우** 좋아한대/사랑한대 [심리/감정 형용사 파생 동사]  
 yenghuy-ka chelswu-lul **acwu/maywu** cohahan-tay/salanghan-tay.  
 Y.-Nom C.-Acc very like-Evi/love-Evi  
 'Younghee **really** likes Chulsoo.'
- b. 영희가 철수를 **완전** 좋아한대/사랑한대  
 yenghuy-ka chelswu-lul **wancen** cohahan-tay/salanghan-tay.  
 Y.-Nom C.-Acc WANCEN like-Evi/love-Evi  
 'Younghee **completely** likes Chulsoo.'

15

## 1-1. Distributional restrictions: *wancen* vs. *very*

- (11) a. 저녁 식사 시간이 **아주/매우** 늦추어졌다더라. [형용사 기원 사동사]  
 cenyek siksa sikan-i **acwu/maywu** nucchwuecyessta-te-la.  
 dinner time-Nom very delay-Evi-Decl  
 'I heard the dinner time has been **very** delayed.'
- b. 저녁 식사 시간이 **완전** 늦추어졌다더라.  
 cenyek siksa sikan-i **wancen** nucchwuecyessta-te-la.  
 dinner time-Nom WANCEN delay-Evi-Decl  
 'I heard the dinner time has been **completely** delayed.'

## 1-1. Distributional restrictions: *wancen* vs. *very*

- (12) a. 올해 들어 미세먼지 수치가 **아주/매우** 높아졌다. [형용사 기원 기동동사]

olhay tule miseymenci swuchi-ka **acwu/maywu** nophacy-ess-tay.

this year the fine dust level-Nom very rise-Past-Evi

'I heard that the fine dust level has risen very much this year.'

- b. 올해 들어 미세먼지 수치가 **완전** 높아졌다

olhay tule miseymenci swuchi-ka **wancen** nophacy-ess-tay.

this year the fine dust level-Nom WANCEN rise-Past-Evi

'I heard that the fine dust level has risen completely much this year.'

17

## 1-1. Distributional restrictions: *wancen* vs. *very*

- (13) a. 사랑니를 뽑더니 잇몸이 **아주/매우** 부었어. [완성동사]

salangni-lul ppayssteni ismom-i **acwu/maywu** pwu-ess-e.

wisdom teeth-Acc took.out gum-Nom very swell-Past-Decl

'My gums are very swollen because I took out my wisdom teeth.'

- b. 사랑니를 뽑더니 잇몸이 **완전** 부었어.

salangni-lul ppayssteni ismom-i **wancen** pwu-ess-e.

wisdom teeth-Acc took.out gum-Nom WANCE swell-Past-Decl

'My gums are completely swollen because I took out my wisdom teeth.'

## 1-1. Distributional restrictions: *wancen* vs. *very*

(14) a. 영희 동생이 **아주/매우** 미남이더라. [명사 지정사구]

yenghuy	tongsayng-i	<b>acwu/maywu</b>	minam-i-te-la.
Y.	brother-Nom	very	handsome-Nom-Evi-Decl

'I heard that Younghee's brother is very handsome.'

b. 영희 동생이 **완전**미남이더라.

yenghuy	tongsayng-i	<b>wancen</b>	minam-i-te-la.
Y.	brother-Nom	WANCEN	handsome-Nom-Evi-Decl

'I heard that Younghee's brother is completely handsome.'

19

## 1-1. Distributional restrictions: *wancen* vs. *very*

### • Differences

#### 1. Co-occurring w/ action verbs

(15) a. #나는 이번 BTS 콘서트에 **아주/매우** 갈거야.

na-nun	ipen	BTS	khonsethu-ey	<b>acwu/maywu</b>	ka-lke-ya.
I-Top	this.time	BTS	concert-Loc	very	go-will-Decl

'#(intended) I will very go to BTS concert this time.'

b. 나는 이번 BTS 콘서트에 **완전** 갈거야.

na-nun	ipen	BTS	khonsethu-ey	<b>wancen</b>	ka-lke-ya.
I-Top	this.time	BTS	concert-Loc	WANCEN	go-will-Decl

'I will **completely** go to BTS concert this time.'

## 1-1. Distributional restrictions: *wancen* vs. *very*

- Differences

### 2. Modifying sentences

(16) a. ??/#**아주매우** 서비스가 대박이에요.

<b>acwu/maywu</b>	sepisu-ka	taypakiey-yo.
very	service-Nom	amazing-Decl

b. **완전** 서비스가 대박이에요.

<b>wancen</b>	sepisu-ka	taypakiey-yo.
WANCEN	service-Nom	amazing-Decl

‘The service is **completely** amazing.’

21

## 1-2. Distributional restrictions: *wancen* vs. *wancenhi*

At first glance, *wancen* and *wancenhi* seem to have similar behaviors:

(17) 경기도 광주와 전라도 광주는 **완전히/완전** 다른 곳이에요.

‘Gwangju in Gyeonggi Province and Gwangju in Jeolla Province are **completely** different places.’

(18) 교육부 안에 인문사회계열에 대한 지원 방안이 **완전히/완전** 배제되어 있었다.

‘Within the Ministry of Education, support measures for humanities and society were **completely** excluded.’

## 1-2. Distributional restrictions: *wancen* vs. *wancenhi*

However, they exhibit crucial differences:

① Unlike *wancen*, *wancenhi* does not always co-occur with some adjectives and verbs (Yeop & Yang 2016):

(19) a. #**완전히** 예쁘다 {착하다, 맛있다, 크다, 많다, 좁다...}

wancenhi yeypputa {chakhata, masissta, khuta, manhata, copta...}

completely pretty {good/delicious/big/many/narrow...}

'It's completely pretty {good/delicious/big/many/narrow...}'

b. #나는 그를 **완전히** 사랑한다. {좋아하다, 미워하다, 싫어하다...}

na-nun ku-lul wancenhi salanghanta {cohahata, miwehata, silhehata...}

I-Top him-Acc completely love {like/hate...}

'I completely love {like/hate...} him.'

23

## 1-2. Distributional restrictions: *wancen* vs. *wancenhi*

② Unlike *wancen*, *wancenhi* cannot modify sentences:

(20) a. ??/#**완전히** 서비스가 대박이다.

**wancenhi** sepisuka taypakiey-yo.

completely service amazing-Decl

b. **완전** 서비스가 대박이다.

**wancen** sepisuka taypakiey-yo.

WANCEN service amazing-Decl

'The service is **completely** amazing.'



## 1-2. Distributional restrictions: *wancen* vs. *wancenhi*

③ *wancen* is degraded in information-seeking questions:

(21) a. 물을 물통에 **완전히** 채웠나?

mwul-ul      mwulthong-ey **wancenhi**      chaywess-ni?  
water-Acc    bucket-Loc    completely    fill-Q

'Did you completely fill the bucket with water?'

b. ??/#물을 물통에 **완전** 채웠나?

mwul-ul      mwulthong-ey **wancen**      chaywess-ni?  
water-Acc    bucket-Loc    WANCEN    fill-Q

'#(intended) Did you completely fill the bucket with water?'

25

## 1-2. Distributional restrictions: *wancen* vs. *wancenhi*

④ Only *wancen* can be used as a fragment answer:

(22) A: 너 오늘 도서관에 갈거야?

ne onul      tosekwan-ey      ka-lke-ya?  
you today    library-Loc      go-will-Q  
'Are you going to the library today?'

B: **완전!** / #완전히!

wancen/wancenhi  
'Completely!'

Again, above characteristics reveals the pragmatic level of intensification!

## 2. Core properties of *wancen*

### 1. Pragmatic intensification in subjective assertion

#### ① Scope interaction w/ negation

### 2. Subjective epistemic operator

#### ① Adjective

#### ② Modal

#### ③ Context

27

## Pragmatic intensification in subjective assertion

### ① The scope of *wancen* is necessarily out of negation operator:

(23) a. 교수님의 설명을 **완전히** 이해할 수 없었어

kyoswunim-uy selmyeng-ul **wancenhi** ihayha-l swu.eps-ess-e.

professor-Gen explanation-Acc completely understand-cannot-Past-Decl

'I couldn't completely understand the professor's explanation.' *Neg>완전히, 완전히>Neg*

b. 교수님의 설명을 **완전** 이해할 수 없었어.

kyoswunim-uy selmyeng-ul **wancen** ihayha-l swu.eps-ess-e.

professor-Gen explanation-Acc WANCEN understand-cannot-Past-Decl

'I couldn't completely understand the professor's explanation.' *\*Neg>완전, 완전>Neg*

- *wancenhi* w/ Neg (wide scope): a maximizer 'completely, entirely',
- *wancen* w/ Neg: a **minimizer** 'the least bit, an inch' (≈ 하나도, 조금도)

## Pragmatic intensification in subjective assertion

② When *wancen* modifies a sentence, it cannot occur in negatives:

(24) #완전 서비스가 대박이지 않다.

*wancen*    sepisu-ka    taypak-ici.anh-ta

WANCEN    service-Nom    amazing-Neg-Decl

'#(intended) The service is not completely amazing.'

29

## Pragmatic intensification in subjective assertion

### 2-1. subjectivity: adjective

#### Dimensional vs. Subjective

(25) a. 민수의 새 가방은 **완전 예쁘다**.

minswu-uy    say kapang-un    *wancen*    yeypu-ta.

M.-Gen    new bag-Top    WANCEN    pretty-Decl

'Minsu's new bag is completely pretty.'

b. 민수의 새 가방은 **아주 예쁘다**.

minswu-uy    say kapang-un    acwu    yeypu-ta.

M.-Gen    new bag-Top    very    pretty-Decl

'Minsu's new bag is very pretty.'

(23a) ≥ (23b)

## Pragmatic intensification in subjective assertion

### 2-1. subjectivity: adjective

#### Dimensional vs. Subjective

(26) a. 민수의 새 가방은 **완전** 크다.

minswu-uy	say	kapang-un	wancen	khu-ta.
M.-Gen	new	bag-Top	WANCEN	big-Decl

'Minsu's new bag is completely big.'

b. 민수의 새 가방은 **아주** 크다.

minswu-uy	say	kapang-un	acwu	khu-ta.
M.-Gen	new	bag-Top	very	big-Decl

'Minsu's new bag is very big.'

(24b) > (24a)

31

## Pragmatic intensification in subjective assertion

### 2-2. Subjectivity: modal

#### Epistemic

(27) a. 이번 시즌 LG가 **완전** 우승임

ipen	sicun	LG-ka	wancen	wusung-im	
this.time	championship	this season	LG-Nom	WANCEN	win-Decl

'LG completely wins the championship this season.'

b. **아마** 이번 시즌 LG가 **완전** 우승할 거야.

ama	ipen	sicun	LG-ka	wancen	wusungha-l ke-ya.	
maybe	this.time	championship	this season	LG-Nom	WANCEN	win-will-Decl

'Maybe LG will completely win the championship this season.'

32

## Pragmatic intensification in subjective assertion

### 2-2. Subjectivity: modal

#### Deontic

(28) a. #완전 먹어라.

wancen mek-ela.

WANCEN eat-Imp

'#Eat completely.'

b. #/???이번 시즌 LG가 완전 우승해야 돼.

ipen sicun LG-ka wancen wusunghay-yatway.

this.time championship this season LG-Nom WANCEN win-must-Decl

'#LG must win the championship this season completely.'

33

## Pragmatic intensification in subjective assertion

### 2-3. Subjectivity: Context

Not allowing information-seeking Question:

(29) a. ??완전 대박이냐?

wancen taypaki-ni?

WANCEN amazing-Q

'#Is it completely awesome?'

b. 완전 대박이지?

wancen taypaki-ci?

WANCEN amazing-CI

'It is completely awesome, isn't it?'

By using the suffix *-ci*, the speaker seeks agreement from the listener on what she believes is true (Hyosang Lee 1991).

### 3. Analysis

(30) The contribution of *wancen*:

**ASSERT(*wancen*(p)) = The speaker asserts that the degree of subjective probability (i.e. credence) is maximal**

- *ASSERT*: the speech act operator involving a credence degree (cf. Yalcin 2007, 2010)
- *wancen*: the illocutionary degree modifier manipulating the degree of credence to maximal

35

### 3. Analysis

(31) Basic dynamic entry for ASSERT (Krifka 2014):

$$\llbracket \text{ASSERT} \rrbracket = \lambda p. \lambda c. \lambda c'. c' = \langle c_{\text{sp(eaker)}}, c_{\text{h(earer)}}, c_t, C_w \cap \{w: \text{assert}(p)(c)\} \rangle$$

(32) Revised entry for ASSERT supplementing with credence degree argument (Greenberg & Wolf 2018):

$$\llbracket \text{ASSERT} \rrbracket = \lambda p. \lambda d. \lambda c. \lambda c'. c' = \langle c_{\text{sp(eaker)}}, c_{\text{h(earer)}}, c_t, C_w \cap \{w: d \in S_{\text{cre}} \wedge \text{assert}(p)(d)(c)\} \rangle$$

### 3. Analysis

(33) *wancenhi* as a typical degree modifier (Kennedy & McNally 2005)

$$\llbracket \text{wancenhi} \rrbracket = \llbracket \text{completely} \rrbracket : \lambda G. \lambda x. \exists d = \text{MAX}(S_G) \wedge G(x)(d)$$

(34) *wancen* as an **illocutionary degree modifier** (adapted from Greenberg & Wolf 2018):

$$\llbracket \text{wancen} \rrbracket = \lambda G. \lambda p. \lambda d. \lambda c. \iota c'. c' = \langle C_{\text{sp(eaker)}}, C_{\text{h(earer)}}, C_t, C_w \cap \{w: \text{MAX}(S_{c're}) \wedge G(p)(d)(c)\} \rangle$$

37

### 4. Conclusion and further implications

- Languages parameterize their pragmatic intensifiers (i.e. degree modifiers).
- In Korean, pragmatic intensifier can be encoded by means of *wancen*.
- The historical path further needs to be studied in the connection of modified nouns to a usage of pragmatic intensifier.

**Thank you!**



# Evaluating the World Knowledge of Language Models:

English Backshift Phenomenon

황동진 신운섭 이지은 송상현; 한국생성문법학회 한국언어정보학회 가을 공동학술대회; 2022.11.12

## 1. Introduction

# Motivation

## 1. Introduction

- The ability to understand the distributed/syntactic linguistic feature in a language model has been studied. (Marvin and Linzen 2018, Goldberg 2019)
- Recent trends in AI research seek to find whether language models have 'common sense' beyond these 'characteristics of language.' (Bosselut et al. 2019, Zellers et al. 2018)
  - How does the language model understand 'world knowledge'?

# Claim

## 1. Introduction

- Can current pre-trained LMs capture language phenomenon regarding world knowledge?
- In what form and in what way does the language model represent the phenomenon and world knowledge?

# Significance

## 1. Introduction

- Shows that language models can represent knowledge beyond linguistic features
- A new perspective on the role of linguists in evaluating the competence of language models

## 2. Backgrounds

# English Backshift

## 2. Backgrounds

### Backshift

- When the time reference of the original utterance does not apply when the utterance is reported, it is often required to change the tense forms of the verbs (Quirk et al. 1985, Comrie 1986, Michaelis 2006)
  - 1)
    - a. John said that he was too busy to join the party.
    - b. John said, "I am too busy to join the party."

# English Backshift

## 2. Backgrounds

### Obligatory Backshift

- A. The content in the reported speech is clearly past, and it is invalid at the time of utterance.
  - 1) Socrates said that he {was, \*is} a citizen, not of Athens, but of the world.
- B. The content of the reported speech occurred long before the time of the utterance, or it is not valid at the time of the utterance
  - 2)
    - a. The waiter told me that lunch {was, \*is} then being served.  
(The waiter uttered the reported speech 5 hours ago)
    - b. They thought that prison conditions {had, \*has} improved.  
(The validity of the content in the quoted clause is questionable)

## English Backshift

### 2. Backgrounds

#### Optional Backshift

A. The time reference of events is clear with other expressions in the sentence

3) Ann explained that the exhibition {finished, had finished} the preceding week.

B. The time reference of the original utterance is valid at the time of the reported utterance.

4) a. She said that she {was, is} studying Business Administration.

b. Their teacher told them that the earth {moves, moved} around the sun.

## English Backshift

### 2. Backgrounds

5) a. Card games use a deck of cards as their central tool. (General fact)

b. A cat has nine lives. (Proverb)

c. Narita is the main airport of Tokyo. (World knowledge)

6) a. She said card games use a deck of cards as their central tool.

b. She said a cat has nine lives.

c. She said Narita is the main airport of Tokyo.

# World Knowledge in LM

## 2. Backgrounds

### SuperGLUE Diagnostic Dataset (Wang et al. 2020)

- [...] SuperGLUE, a new benchmark designed to pose a more rigorous test of language understanding.
- GLUE includes an expert-constructed, **diagnostic dataset** that automatically tests models for a broad range of linguistic, commonsense, and world knowledge.



# World Knowledge in LMs

## 2. Backgrounds

Common Sense	World Knowledge
Difficult to express as facts	Can <b>clearly be expressed</b> as facts
Expect to be <b>possessed by most people</b>	Can be <b>broader and less common</b>
<b>Independent</b> of cultural or educational background	<b>Dependent</b> of cultural or educational background
Basic understanding of physical and social dynamics	Geographical, legal, political, technical, or cultural knowledge

# World Knowledge in LMs

## 2. Backgrounds

Commonsense VS. World Knowledge

- World Knowledge

- 7)      a. The reaction was strongly exothermic **entails** The reaction media got very hot.
- b. There are amazing hikes around Mt. Fuji **entails** There are amazing hikes in Japan

- Common Sense

- 8)      Marc Sims has been seeing his barber once a week. **entails** Marc Sims has been getting his hair cut once a week.

## 3. Methodology

# Surprisal

## 3. Methodology

- A value **inversely proportional** to the expectation of a new word in a given chain of tokens (Levy, 2008).

**The chef stirred the savory stew with \_\_\_\_\_.**

<b>a spoon</b>	7.993397	4.100479
<b>a whisk</b>	<u>11.96577</u>	<u>6.620835</u>

# Data

## 3. Methodology

		<b>Source</b>	<b>Matrix</b>	<b>numbers</b>	<b>total</b>
<b>control</b>	<b>past</b>	COCA	He	117	193
			She	76	
	<b>present</b>		He	60	83
			She	23	
<b>experimental</b>		Wiki	He	150	350
		General		92	
		Proverb		73	
		Quotation	Person	35	



# Data

## 3. Methodology

A triangle is a three sided polygon



**He said** a triangle is a three sided polygon

# Data

## 3. Methodology

	<b>Matrix Clause + Embedded Clause</b>	<b>Matrix Clause</b>
<b>Control Group</b>	276	276
<b>Experimental Group</b>	350	350

# Data

## 3. Methodology

- [MASK]

### Wikipedia

- A. He said a triangle **[MASK]** a three sided polygon **is was**
- B. He said a circle **[MASK]** a round two-dimensional shape **is was**
- C. He said a square **[MASK]** a shape with four equal sides **is was**

# Experiment Settings

## 3. Methodology

Model	Parameter Size	Specific Model
BERT	340M	bert-large-uncased
mBERT	110M	bert-base-multilingual-uncased

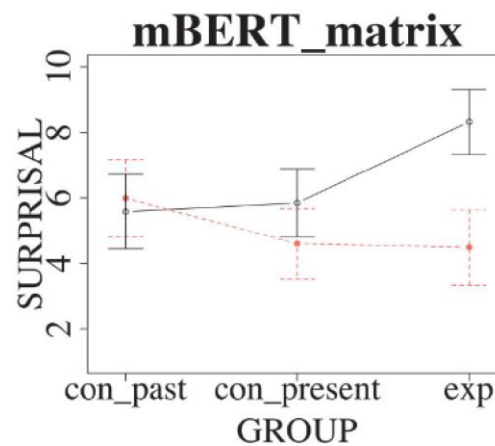
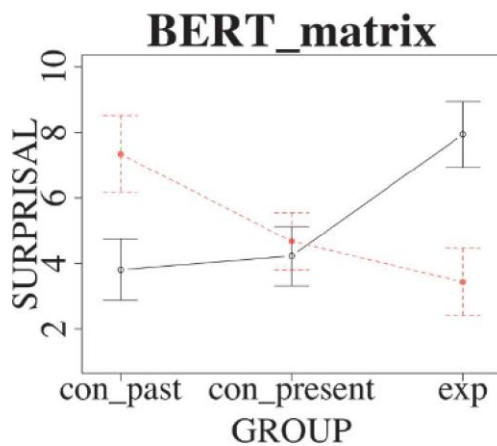
(Devlin et al., 2018)

# 4. Results

## Matrix Clauses

### 4. Results

A triangle [MASK] a three sided polygon. is was



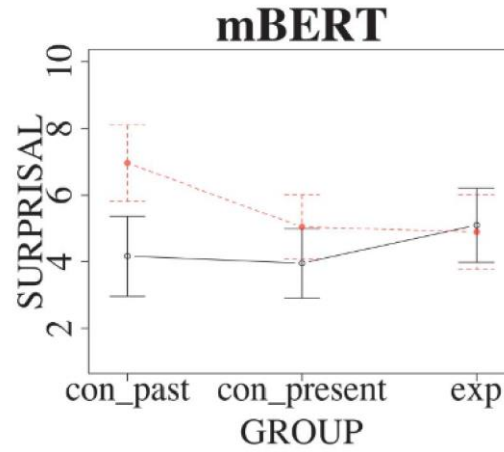
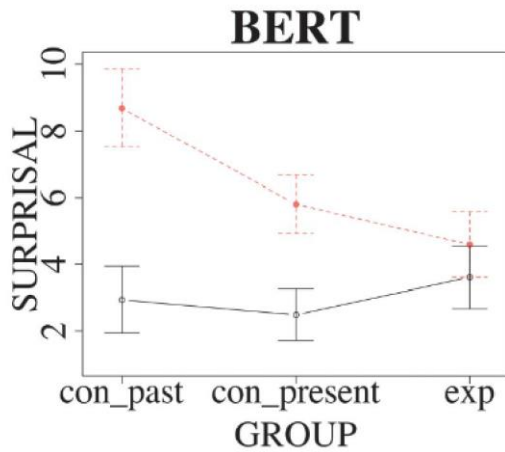
—○— past  $F(2, 1296) = 81.3806$   
 -●- present  $p < .0001$ , '\*\*\*'

$F(2, 1296) = 19.2103$   
 $p < .0001$ , '\*\*\*'

# Embedded Clauses

## 4. Results

He said a triangle [MASK] a three sided polygon. is was



—○— past  
- - - ● - - - present

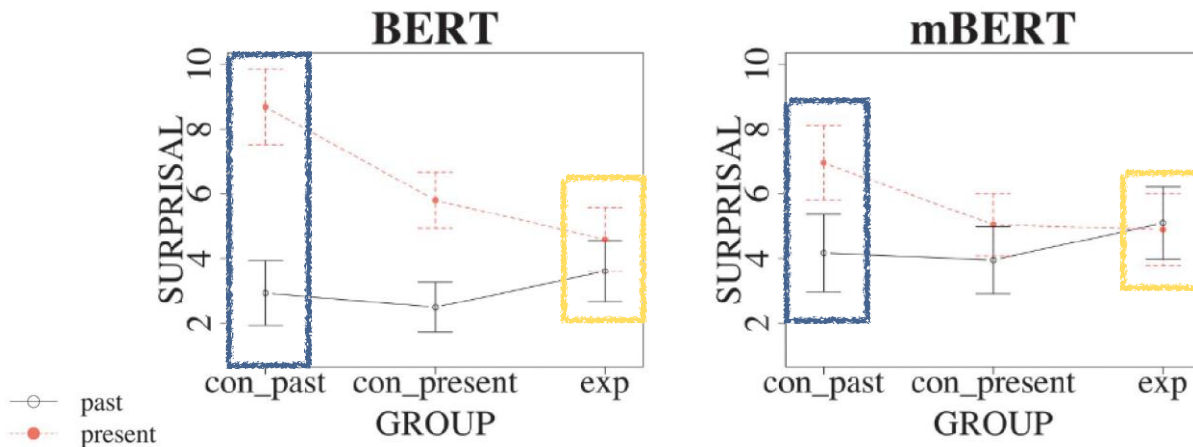
$F(2, 1296) = 29.503$   
 $p < .0001$ , '\*\*\*'

$F(2, 1296) = 9.0048$   
 $p < .0001$ , '\*\*\*'

## 5. Discussion

# Backshift Representation of LMs

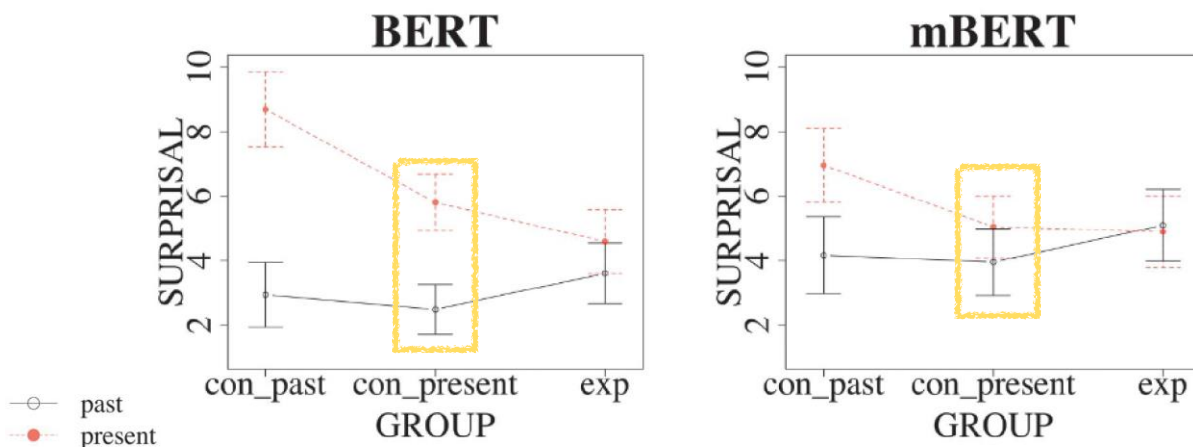
## 5. Discussion



- Results from both models were consistent with the prediction

# Backshift Representation of LMs

## 5. Discussion



- Surprisal of control group sentences with present tense verbs were distributed between the other two groups

## Backshift Representation of LMs

### 5. Discussion

- Sentences with optional (i.e., non-obligatory) backshift according to the applicability at the time of utterance were included in *con\_present*.

9) He said narration [MASK] a way of ordering or reordering events.

is 0.5883667

was 1.8762759

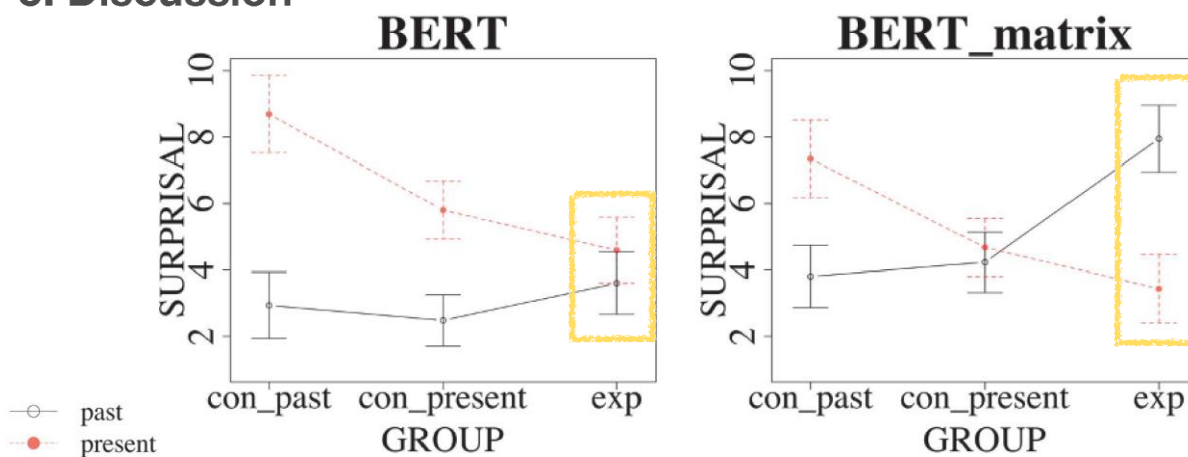
10) She said Barry [MASK] overwhelmed

is 6.8997903

was 0.16579989

## Backshift Representation of LMs

### 5. Discussion



- Both models showed no significant difference in surprisal distribution for the experimental group when the verbs are used in the embedded clause.

# 6. Conclusion

## 6. Conclusion

- PLM such as BERT or mBERT successfully represented the English backshift phenomenon.
- The models also successfully captured the tense usage in matrix clauses.
- Both results revealed LMs representation of the English backshift were more consistent with the body of linguistic research than the actual usage.  
→ Linguists can bring advancement in LM evaluation with linguistic knowledge

# References

- Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for automatic knowledge graph construction. arXiv preprint arXiv:1906.05317.
- Comrie, B. (1986). Tense in indirect speech.
- 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.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- Marvin, R., & Linzen, T. (2018). Targeted syntactic evaluation of language models. arXiv preprint arXiv:1808.09031.
- Michaelis, L. A. (2006). Time and tense. *The handbook of English linguistics*, 220-234.
- Quirk, R. (1985). *A Comprehensive grammar of the English language*. Longman.
- Wang, W., Tian, Y., Wang, H., & Ku, W.-S. (2020). A Natural Language Interface for Database: Achieving Transfer-learnability Using Adversarial Method for Question Understanding. In (pp. 97-108): IEEE.
- Zellers, R., Bisk, Y., Schwartz, R., & Choi, Y. (2018). Swag: A large-scale adversarial dataset for grounded commonsense inference. arXiv preprint arXiv:1808.05326.



Korean Generative Grammar Circle (KGGC) &  
Korean Society for Language and Information (KSLI)  
2022 Joint Fall Conference

November 12, 2022, Korea.

# *Disentangling the underlying structures of Constituent Negation*

Ricardo Etxepare (IKER, UMR5478)  
Uribe-Etxebarria  
r.etxepare@iker.cnrs.fr

&

Myriam

(U. of the Basque Country, EHU/UPV)

myriam.uribe-etxebarria@ehu.eus

## 1. The goal of the talk

To analyze sequences formed by *Negation + QP* (usually characterized as involving Constituent Negation, CN) and investigate the syntactic and semantic structure that underlies these sequences.

**Neg + QP**

- (1) **No** todo el mundo está contento  
Neg all the world is happy  
“Not everyone is happy”

# The

## ■ proposal

- At least **3 different structures** underlying *Negation + QP* sequences. Lges vary regarding which one(s) they choose.

**i. Direct Merge.** Negation merges directly with the QP.

$$[_{QP} \text{ NO } [_{QP} \dots Q \dots ] ]$$

**ii. Association with Focus (*Illusory CN*).** Neg and the QP are born separatedly. They surface together due to the movement of the QP to the Spec of a FocP which is close to a NegP located in a high position of the left periphery.

$$[_{NegP} \text{ NO } [_{FOCP} \text{ QP } \text{ Foc}^{\circ} [_{TP} \dots ] ] ]$$

**iii. Mixed strategy.** The QP has a Foc and a Neg feature and has to move to the specifier of a sentential FocP  $\text{Foc}^{\circ}$

$$[_{FOCP} [_{NEGP/QP} \text{ NO } [_{FOCP} \text{ Foc}^{\circ} [ \text{QP} ] ] ] ] [_{TP} \text{ QP}_i ] ]$$

## ■ The proposal

- At least **3 different structures** underlying *Negation + QP* sequences. Lges vary regarding which one(s) they choose.

**i. Direct Merge.** Negation is from the beginning part of the constituent where the QP/DP is contained. That is, Negation merges directly with the constituent containing the QP.

- Two alternative possibilities in the literature (see Collins 2020).

*Outer Negation:*

$$[_{QP} \text{ NO } [_{QP} \dots Q \dots ] ]$$

*Inner Negation*

$$[_{NP/DP} [_{QP} \text{ NO } Q ] ] [ \text{NP/DP} ] ]$$

We will not focus on the internal structure of these sequences today, and will only focus on whether *Neg+QP* sequences constitute *true* cases of CN, where the negative marker merges directly with the constituent containing the QP and forms a unit with it.

- This strategy is not universally available, and it seems to be unattested in many languages, like Basque.
- Further, we will see that in the case of some lges (a.o. Spanish and Portuguese) this strategy is restricted to a limited set of Qs.

## ■ The proposal

Lges

- At least **3 different structures** underlying *Negation + QP* sequences. Lges vary regarding which one(s) they choose.

**ii. Association with Focus (*Illusory CN*).** Negation does not merge directly with the constituent containing the QP: Neg and the QP are born separately.

They surface together due to the movement of the QP to the Spec of a FocP which is close to a NegP located in a high position of the left periphery.

Association with narrow or  
 $[_{FOCP} \text{NO } [_{\text{QP}} \text{ Foc}^{\circ} ] ] ]$

- focus/Illusory CN is an available strategy to construct subclausal scope for negation in many lges. It depends a.o. on the presence of focus fronting in the relevant languages.
- It is present in the 3 languages under analysis here. It is the only strategy available in Basque.

## ■ The proposal

- At least **3 different structures** underlying *Negation + QP* sequences. Lges vary regarding which one(s) they choose.

**iii. Mixed strategy.** The QP has a Foc and a Neg feature and has to move to the specifier of a sentential FocP .

$[_{FOCP} [_{NEGP/QP} \text{NO } [_{FOCP} \text{ Foc}^{\circ} [ \text{QP} ] ] ] ] ]$

- The QP has a Foc and a Neg feature and has to move to the specifier of a sentential FocP
- Out of the 3 lges analyzed here, it is only possible in Portuguese.

## ■ The proposal

- At least **3 different structures** underlying *Negation + QP* sequences.

Lges vary regarding which one(s) they choose.

**Table I:** Strategies underlying Neg+QP sequences

	Direct Merge	Association with Focus ( <i>Illusory CN</i> ).	Mixed strategy
Spanish	✓	✓	X
Portuguese	✓	✓	✓
Basque	X	✓	X

### ● Crosslinguistic variation in the expression of CN

- The choice of one strategy or the other is dependent on **several factors/properties of the languages**, which contribute to the different behavior exhibited by the so-called Constituent Negation (CN) across different structures and languages.
  - (i) The type of QPs involved.
  - (ii)& (iii) The features and structural position of the type of Negation involved (Zanuttini 1997, Ramchand, 2004; Schapansky, 2002, 2010; De Clerq, 2013, 2016; Poletto, 2017, a.o.).
  - (iv) The status of fronting operations and their articulation with economy considerations (we will not discuss this today).
- In this talk we will discuss these three strategies and provide evidence for them on the basis of Spanish, Portuguese and Basque.

## 2. Disentangling the structures underlying *NEG+QP*

What is the structure underlying *Neg+QP* sequences?

### 2.1. The *Direct Correspondence Approach: Direct Merge [Neg+QP]*

→ In Spanish, a **wide range of quantifiers** can be immediately preceded by *Neg* giving rise to *NEG+QP*, among others:

- a) **evaluative existential quantifiers** (*muchos* “many” / *pocos* “few”).
- b) **degree quantifiers** (*demasiado* “too many/much”) & **comparatives of degree** (*más/menos de* “more than/less than”),
- c) **universal quantifiers** (*todos* “all”, *todo el mundo* “all the world/everyone”).

- |     |                                   |                               |
|-----|-----------------------------------|-------------------------------|
| (2) | a. <i>No muchos</i>               | “not many”                    |
|     | b. <i>No pocos</i>                | “not few”                     |
|     | c. <i>No demasiados</i>           | “not too many”.               |
|     | d. <i>No más/menos de</i>         | “not more than/not less than” |
|     | e. <i>No todos /todo el mundo</i> | “not all”/“not everyone”      |

→ Given that sequences of *NEG+QP* seem to be widely available in Spanish, it would then be **tempting to assume** that a “**constituent negation/CN**” analysis **may be unproblematically extended** to all the attested cases, **as stated by the Direct Correspondence Approach in (3)** below.

### (3) *Direct Correspondence Approach:*

a.  $[_{QP} \text{ NO } [_{QP} \dots \text{ Q} \dots ]]$

b.  $[_{NP/DP} [_{NO} \text{ Q}] [_{NP/DP} ]]$  (Collins’ Inner Negation, 2017)

This is the position defended by many authors, who suggest a unified treatment of such sequences.

(cf. Sánchez López 1999, Kim & Sag 2002; Collins and Postal, 2014; Collins, 2016, 2017).

**However**, as we will show next, if we (minimally) complexify the relevant sequences of negation and QP by adding a Case Marker (CM)/preposition (P), **the *Direct Correspondence Approach* gives rise to a number of structural paradoxes.**

### 2.1.1. A syntactic paradox:

**Word order asymmetries affecting the *Neg+QP* sequence.**

When we add a Case Marker (CM) /Preposition (P) to sequences involving Negation and a QP, there are two possible relative orders between Neg and the CM/P available.

(4) **i.** Negation precedes de Case Marker /Preposition

***NEG* + *P/CM* + *QP***

**ii.** The Case Marker / Preposition precedes Negation

***P/CM* + *NEG* + *QP***

### 2.1.1.1. **NEG + Prep/CM + QP**

The relative order in which negation precedes the CM/P seems to be **available to all the quantifiers that can combine with negation.**

(Most of the examples in the handout are borrowed from CREA (*Corpus de Referencia del Español Actual*) or Google).

(5)

- a.... este tipo de deporte que **no a mucha gente** le gusta jugar  
 this type of sport that neg to many people CL.dat likes play.inf  
 “...this type of sport, which not many people like playing.”
- b. Hecho que molestó **no a pocos de los periodistas** cansados por actos a medias  
 thing that bothered neg to few.pl of the journalists tired by performances halfhearted  
 “ ... a thing that bothered many of the journalists who were fed up with half-hearted performances”.
- c. **No a todo el mundo** le sientan igual las bromas  
 neg to all.sg. the.sg. world.masc.sg. CL.dat affect equally the jokes  
 “ Not everyone is equally affected by jokes.”
- d. **No a todos** les interesa llegar tan lejos  
 neg to all.pl CL.dat interests get.inf that far  
 “Not everybody is interested in getting that far.”

(5)

- g. Hay casos extremos, **no a todo dios** le sienta igual todo  
 there.are cases extreme, neg to every god CL.dat affects equally all  
 “There are extreme cases, not everybody is equally affected by everything.”  
 ([https://www.google.es/?gfe\\_rd=cr&ei=FqARVo\\_TheSs8wf\\_qZzoDQ&qws\\_rd=ssl#q=%22](https://www.google.es/?gfe_rd=cr&ei=FqARVo_TheSs8wf_qZzoDQ&qws_rd=ssl#q=%22))
- e. **No a cualquiera** hubiera enviado el Superior a hablar con el joven  
 neg to anyone would.have sent the superior to speak.inf with the young  
 “The headmaster wouldn’t have sent just anyone to speak to the young man.”
- f. Si bien **no a cada uno** le resulta fácil establecerse...  
 if well neg to each one cl-DAT turns.out easy establish.inf.refl to  
 “Even if not everyone finds it easy to set out...”

**Table 2. Sequences of NEG +CM/Prep +QP**

<b>Neg</b>	<b>CaseMarker/ Prep</b>	<b>QP</b>
<b>NO</b>	<i>a</i> <sub>to</sub>	<i>Muchos/-as</i> NP “Many NP”
		<i>Pocos/-as</i> NP “Few NP”
		<i>Todos/-as</i> DP “All DP”
		<i>Todo el mundo</i> “Everyone”
		<i>Cualquier</i> NP “Any NP”
		<i>Todo dios</i> “Everyone”
		<i>Cada</i> NP “Each NP”

<sup>to</sup>, the word order alternation

Note: Although for easy of exposition our examples involve the preposition *a* we illustrate here is not limited to this preposition, and involves a wider range of prepositions other than *a*. See Etxepare & Uribe-Etxebarria (2018) and (2021) for examples with other prepositions.

- Summarizing: When Negation precedes the Case Marker (CM) /Preposition (P) Negation seems to be able to combine with any type of QPs.

**i.** Negation precedes de Case Marker /Preposition

***NEG + P/CM + QP***

- But remember that there is a second word order possible:

**ii.** The Case Marker / Preposition precedes Negation

***P/CM + NEG + QP***

As we will see now, in this case the situation is more complex.



### 2.1.1.2. Prep/CM + **NEG** + QP

- As we will show next, the **word order** where NEG follows the CM/P does **not** seem to be **available to all quantifiers**.
- **This word order is possible for the following quantifiers.**

#### ■ Muchos (“many”)

- (6) ... recuerdo que **a no muchos** les gustó ese traje  
 ... remember-I that to neg many.pl CL.dat.pl liked that suit  
 “I remember that not many liked that suit.”

(<http://blogdesuperheroes.es/cine-el-misterioso-personaje-de-escuadron-suicida-es-deadshot-ni-nadie-solo-un-espontaneo-que-se-colo-en-el-set>)

#### (7) ■ Pocos (“few”)

- a. En realidad son varios los interrogantes que preocupan **a no pocos trabajadores**  
 in reality are several the questions that concern to neg few.pl workers  
 “To be honest, there are several issues that concern many (lit: not few) workers...”  
(<https://books.google.es/books?id=LtUiDWWyapgC>)

#### ■ Todos/-as (“all”)

The sequence *CM/P+NEG+QP* is also possible for (some but not all) speakers when the quantifier involved is *todos/-as* (“all”).

(8)

Es un evento que **a no todos** les parece claro  
 is an event that to neg all CL.dat.pl seems clear  
 “It is an event that not everyone considers (to be) clear.”

[From CREA, IX/2015]

■ **Todo el mundo** (lit: “all the world” =“everyone”)

→ In the case of ***todo el mundo*** (lit: “all the world” =“everyone”) speakers also seem to split into 2 groups:

some speakers don't license sequences where this QP is immediately preceded by negation while others do.

→ Still, although much less frequent than with other quantifiers, examples of this sort can also be found in oral examples & written texts:

(9) ***Todo el mundo*** (lit: “all the world” =“everyone”)

Sin embargo, ***a no todo el mundo*** le gusta un medio tan abierto.  
 However to neg all the world CL.dat likes a media so open  
 “However, not everyone likes such open(-minded) media.”  
 (es.bab.la › *Diccionario bab.la* › *Español-Alemán* )

■ **Cualquiera** (“anyone”).

→ Finally, although also possible for many speakers, this combination seems to be more restricted for quantifiers like ***cualquiera*** “anyone”.

→ We did not find a single example in monitored corpora, but we did find a few on the internet:

(10) ***Cualquiera*** (“anyone”)

Con el tiempo aprendí que ***a no cualquiera*** se le llama amigo  
 With the time, learnt.I that to neg anyone cl.IMP cl.dat calls friend

“As time went by, I learnt that you cannot call just anybody a friend.”

(<https://twitter.com/.../status/309346974564495360>)

But there are some quantifiers that do not admit such a configuration.

■ *Cada NP* (each NP) & *todo dios* (all god = everyone)

→ We have not found a single instance either in internet or elsewhere of sequences like *a no cada NP* (“to not each one”), or *a no todo dios* (“to not everyone”).

The examples would have looked as follows:

- (11) a. \* **A no TODO DIOS** le gusta eso  
**to not every god** cl.dat likes that  
 “Not everyone likes that.”
- b. \* **A no CADA ESTUDIANTE** le gusta eso  
**to neg each student** cl.dat likes that  
 “Not each student likes that.”

Note: In this talk we will now address the question of why these combinations are not possible. See Etxepare & Uribe-Etxebarria (2018) for an analysis of this problem.

*Table 3. Sequences of CM/Prep + Neg+ QP*

CaseMarker/ Prep	Neg	QP
<i>a</i> <sub>to</sub>	<b>NO</b>	<i>Muchos/-as</i> NP “Many NP”
		<i>Pocos/-as</i> NP “Few NP”
		<i>Todos/-as</i> DP “All DP”
		<i>Todo el mundo</i> “Everyone”
		<i>Cualquier</i> NP “Any NP”
		<i>Todo dios</i> “Everyone”
		<i>Cada</i> NP “Each NP”

with a QP depending

Table 4. Asymmetries in the possibilities of combining Neg on the relative order with respect to CM/Prep.

Neg	Prep	Muchos/-as NP QP "Many NP"	Prep	Neg	Muchos/-as NP QP "Many NP"
NO	<i>a</i> <sub>to</sub>	<i>Pocos/-as</i>	<i>a</i> <sub>to</sub>	NO	<i>Pocos/-as</i> NP "Few NP"
		<i>Todos/-as</i> NP "Few NP"			<i>Todos/-as</i> DP "All DP"
		<i>Todo el mundo</i> "All DP"			<i>Todo el mundo</i>
		<i>Cualquier</i> NP "Any NP"			"Everyone"
		<i>Todo dios</i>			
		<i>Cada</i> NP "Each NP"			"Everyone"
		"Everyone"			
		"Each NP"			"Each NP"

● **Summarizing:**

→ The combination in (12), where Neg first combines with the QP, is not possible for some Qs, like *todo dios* ('every god' = 'everyone') or *cada* ('each').

(12) \* [<sub>CM/P</sub> *a* [<sub>QP</sub> NO | QP TODO DIOS CADA | ] ]

Direct merge of Neg+QP not possible

● **Paradox:**

→ But if this conclusion is correct, it immediately raises a question regarding the syntax of grammatical *Neg+QP* sequences like those in (13a-b), which are well formed and seem to involve the structure in (12), which we have just concluded is an illicit combination.

(13) a. *No todo dios* es amigo del vino  
Neg every god is friend of the wine  
"Not everyone is friends with wine"

[<sub>QP</sub> NO | QP TODO DIOS | ]

→ The paradox extends to similar cases with *cada* 'each'.

(13) b. A casa de tu tía mas *no cada día*  
To house of your aunt, but neg each day.

[<sub>QP</sub> NO | QP CADA | ]

## ▪ Conclusion:

- How do we interpret these results?

Since the structure that corresponds to the **illicit combination** is that in (14), where Neg directly merges with the QP, we could hypothesize that the reason why (14) is illformed is because Neg cannot directly merge with these two QPs (QP<sub>cada+ NP</sub> and QP<sub>todo dios</sub>). We represent the illicit part of the structure **within a dark grey box**.

Direct merge of Neg and these QPs is not possible

(14) \* [PP Prep [ NO [ QP *todo dios / cada + NP* ] ] ]

- If direct merge between Neg and the QP is not possible with certain quantifiers, then

→ We need an alternative underlying structure for some of the sequences involving *Neg + QP*.

This leads us to the **2<sup>nd</sup> strategy** underlying Neg+QP sequences

## 2.2. The *Indirect Correspondence Approach* or *Illusory CN: Association with Focus*

### *Our proposal:*

- i. Some *Neg + QP* sequences are not derived by direct merge; rather, they follow from a focus fronting movement operation of the QP to a position adjacent to a “clause-external” negation, which occupies a high position in the left periphery.

(15) [NEGP **Neg** [FOCP **QP** Foc<sup>o</sup> [IP... QP... ] ] ]

- ii. This operation is available to all QPs that may undergo fronting.

In the following subsections we present evidence in favor of this proposal.

## 2.2.1. Word Order at the sentential level, and within the NEG+QP sequence

- Lasnik (1972) and Postal (1974) observed that so-called **constituent negation (CN) triggers changes in word order**:
  - in English, CN is not available everywhere: negated constituents of the form **Neg+QP must be in preverbal position**.

(16) a. \*The students solved not all of the problems  
 b. Not all of the problems were solved by the students

(17) a. \*Joe kissed not many models  
 b. Not many models were kissed by Joe [Postal 1974]

### 2.2.1.1. Lasnik (1972): Pre-sentential negation

In order to account for this asymmetry, Lasnik proposed that the **negation involved in constituent negation is clause external**.

(18) *Pre-sentential Negation* (Lasnik, 1972):  $S \rightarrow \text{Pre-S NP VP}$

- *not all* and *not many* in (16a) and (17a) are surface constituents, following a rebracketing operation that requires the adjacency of the Pre-sentence and the QP.
- The preverbal position of negated QPs follows from the left peripheral position of negation, and conditions on adjacency that determine the possibility of rebracketing.

### 2.2.1.2. Collins' Negated Quantifier Constraint

- Collins and Postal (2014) and Collins (2017) have argued for an alternative analysis of the preverbal position of CN.
  - An analysis of CN based on a presentential negation cannot account for examples like (19), with 2 instances of CN (it fails to account for the fact that CN can occur following the auxiliary, and misses the generalization that Neg immediately precedes the QP).
 

(19) Not often do not many people want to dance
  - (20) **Alternative analysis: Negated Quantifier Constraint (NQPC)** [adapted from Collins 2017]:  
 An overt occurrence of [NOT QP] c-commands a clause-mate T.
    - The ultimate motivation for this is that at LF, quantifiers must take scope over the clause, but negation does not undergo QR, so fronting is necessary for the QP to take scope over T. Stranding Neg and QR-ing the quantifier would lead us into problems, because traces of movement are of type <e>, and negation must modify entities that include the type <t>.

(21) a. \* I saw not every student at the party  
 b. I saw not quite every student at the party

  - (21b) is good because *not* only has scope over the degree modifier *quite*, so it does not block QR of the QP at LF, as shown in (21b'):
- (21b') I saw [ [not quite] [every student ] ] at the party

- However, as we will see next, **neither of these analyses alone can account for the whole complexity of sequences involving Neg and a QP in Spanish**

### 3. Our Proposal

→ To account for all the Spanish facts under consideration, we need to allow (at least) two different structural options underlying the *Neg+QP* sequences.

- ***Direct Correspondence (True Constituent Negation, TCN)***

- TCN can be structurally described as in (22): direct merge of Neg + QP.

(22) [QP **NO** [...QP...]]

- In Spanish, TCN is available to a restricted set of quantifiers. The availability of (22) may vary crosslinguistically.

- ***Indirect Correspondence (Illusory Constituent Negation, ICN)***

- No direct merge of *Neg* + *QP*.

- The adjacency between Negation and the QP results from focus fronting of the QP to a position adjacent to a “clause-external” negation, as in (23).

(23) [<sub>NEGP</sub> **NO** [<sub>FOCP</sub> **QP** **Foc<sup>o</sup>** [<sub>IP...</sub> QP...]]]



- This option is open to all QPs that may undergo fronting.

● **Evidence in favor of our proposal:**

- The **Spanish sequences involving Neg and a QP show a greater freedom** in their syntactic distribution than the comparable sequences in English, and there are several **asymmetries in their distribution depending on the relative order displayed by Negation and the CM/Prep**.

- **1st asymmetry: preverbal vs. postverbal position**

- **CM/Prep+ Neg + QP** : both, **preverbal and postverbal positions licensed**.

(24) a. Les gusta **a no pocos**      b. **A no pocos** les gusta  
 CL like      CM neg few                      CM neg few      CL.dat.pl like  
 “Not few of them like it”                      “Not few of them like it”

- The grammaticality of (24a) is hard to explain if we only adopt Lasnik’s analysis.
- The existence of both (24a) and (24b) is hard to explain if we only adopt Collins’ approach.

- **Neg + CM/Prep + QP** : **Only preverbal position** is licensed

(24) c. **No a pocos** les gusta      d. \* Les gusta **no a pocos**  
 Neg CM few      CL.dat.pl like                      CL.dat.pl like      neg CM few  
 “Not few of them like it”                      “Not few of them like it”

- **2nd asymmetry: possibility of being licensed as a topicalized constituent**

→ **CM/Prep+ Neg + QP** sequences **can be topics**

(25) a. [ **A no pocos** ] los domingos les gusta ir al futbol  
 to neg few      the Sundays CL-Dat like to.go to football  
 “lit: Not few people, on Sundays, they like going to a football match”

→ **Neg + CM/Prep + QP** sequences **cannot be topics**

b.\* [ **No a pocos** ] los domingos les gusta ir al futbol  
 to neg few      the Sundays CL-Dat like to.go to football  
 “lit: Not few people, on Sundays, they like going to a football match”



- **Conclusion:**

(i) **Neg + Prep/CM + QP** sequences are **not possible in postverbal position** but they are **possible in preverbal position**.

(ii) **Neg + Prep/CM + QP** sequences are **not licit topicalized constituents** (in contrast to Prep + Neg + QP<sub>WEAKQ</sub> sequences)

→ We argue that **these asymmetries follow from the fact that Neg does not merge directly with the CM/P + QP unit**.

- The sequence Neg + CM/P + QP follows from the fronting of [CM/P + QP] to the specifier of a high FocP close to a high clausal negation.

- The sequence **no<sub>NEG</sub> a<sub>TO</sub> pocos<sub>FEW</sub>** in (25b) cannot be topicalized because **no<sub>NEG</sub>** and **a<sub>TO</sub> pocos<sub>FEW</sub>** do not form a constituent.

→ The correct generalization is (26).

(26) **Generalization: Case/Preposition-external negation is related to QP fronting.**

Question: What does this follow from? How can we account for this?

**The word order asymmetry** arising between Case/Prep external negation and Case/Prep internal one (and the subsequent related asymmetries we have discussed) now follow(s) **from two different syntactic structures:**

(27=23) a. **Illusory CN: Sequences with Neg external to CM/Prep**

[<sub>NegP</sub> **No** [<sub>FocP</sub> [**a pocos**] Foc<sup>0</sup> [<sub>IP</sub> ... [<sub>VP</sub> ... [a pocos] ... ]]]]

b. **True CN: Sequences with Neg internal to CM/Prep**

[<sub>IP</sub> Les gusta [<sub>VP</sub> ... [ **a** [<sub>QP</sub> **no** [<sub>QP</sub> **pocos** ] ] ] ] ]

→ The hypothesis entails that the **scope position of negation is external to the DP with certain quantifiers (todo<sub>EVERY</sub> dios<sub>GOD</sub> /cada<sub>EACH</sub> NP)** as well as in the case of sequences with the order Prep+Neg+QP.

## 4. Where is Negation in *Illusory CN*?

- Under our approach, in *illusory CN* the adjacency between negation and the QP is an illusion produced by fronting of the QP into the specifier of a functional head which is in a very local relation with a very high negation.
- We now present **two additional pieces of evidence** in favor of our analysis of some *Neg+QP* sequences as involving *Illusory CN* (indirect approach) (see Etxepare & UE 2018 for additional evidence):
  - i) Cases in which **the sequence *NEG+QP* is interrupted by high clausal adverbs** (Sect. 4.1.)
  - ii) The **existence** of clauses with **double negation** (Sect. 4.2.)

### 4.1. Clausal Adverbs

- In cases of **illusory CN**, where negation is external, ***Neg* and the [ (*Prep*) *QP*] do not need to be immediately adjacent to each other and can be separated by an adverb taking clausal scope.**
- This adverb can be a **modal adverb** (28a,b), or an **evaluative one** (28c), both with clausal scope.
- Direct **morphosyntactic evidence for the clausal scope of the adverb in (28a)** comes from the fact that **it licenses the occurrence of subjunctive inflection**, as one expects from modal adverbs with sentential scope:

(28) a. **No** *tal vez* **a todos** *les* *guste*<sub>subjunct</sub> *eso*  
*neg perhaps to all.pl CL.dat like*<sub>subjunct</sub> *that*  
 “Perhaps not everybody likes that.”

b. *Cosas...que quise compartir de una forma que* **no** *quizás* **a todos** *les gustó*  
 things that wanted(I) share.inf of a way that neg perhaps to all.pl CL.dat liked  
 “Things that I wanted to share in a way that perhaps not everyone liked.”

(forums.lan.leagueoflegends.com/board/showthread.php?t=71433)

c. ...**no** *afortunadamente* **a todos** *los que fueron mis subalternos* *en la Nacional*,  
 neg fortunately to all.pl the that were my subordinates in the National,

*sino a algunos que...*  
 but to some that

“Fortunately not to all those who were my subordinates on the Nacional but to some that...”

→ Notice that in (28a), repeated below, the verb *guste* ‘likes’ is in the subjunctive mood.

(28) a. **No** tal vez **a todos** les *guste*<sub>subjunct</sub> eso  
 neg perhaps to all.pl CL.dat like<sub>subjunct</sub> that  
 “Perhaps not everybody likes that.”

→ That the subjunctive in (28a) is licensed by the modal adverb and not by negation is shown by the fact that, without the adverb, the subjunctive is not licensed:

(29) \* No a todos les *guste* *subjunctive* eso  
 neg to all CL-dat.pl like.3.psg. *subjunct* that  
 “Not everybody likes that.”

→ If in order to license subjunctive mood in the finite form the modal adverb must be able to c-command it, it then follows that the adverb *tal vez* is not embedded in a purported constituent headed by Neg or the Q, as in (28’); rather the structure must be as in (28’).

(28’) \* [No<sub>NEG</sub> tal vez<sub>PERHAPS</sub> a<sub>TO</sub> todos<sub>ALL</sub>] V subjunctive incorrect structure for (28a)  
 (28’’) [No<sub>NEG</sub>] [tal vez]<sub>PERHAPS</sub> [a<sub>TO</sub> todos<sub>ALL</sub>] V subjunctive correct structure for (28a)

→ The existence of sequences like those in (28a-c), where the linear adjacency between negation and the QP is interrupted by sentential adverbs, provides direct evidence against a direct merge analysis of these *NEG-QP* sequences, where Neg would be merged or adjoined in the extended projection of the QP.

→ Furthermore, the fact that in cartographic terms the element that interrupts the sequence is a very high adverb provides strong evidence that negation in those cases belongs in the high CP-domain.

### ● Further Evidence:

→ Further evidence that the position Neg occupies in Illusory CN (Neg<sub>1</sub> in (30)) must be different from the one occupied by regular sentential negation (Neg<sub>2</sub> in (30)), comes from the fact that the latter (Neg<sub>2</sub>) follows (instead of preceding) evaluative and modal adverbs, as shown in (31):

(30) [<sub>NEGP1</sub> **Neg<sub>1</sub>** [<sub>MODALP</sub> **High Adverbs** [<sub>FOCP</sub> **QP** *Foc*<sup>o</sup> [<sub>NEGP2</sub> **Neg<sub>2</sub>** [QP...]]]]]

(31) a. **Afortunadamente / tal vez** **no** han venido  
 fortunately /perhaps, neg have(they) come  
 “Fortunately/perhaps, they didn’t come.”

b. \***No afortunadamente / probablemente** han venido

→ The Neg we find in *Illusory CN* corresponds to **Neg<sub>1</sub>** in (30), while **sentential negation** corresponds to **Neg<sub>2</sub>** in (30)

## 4.2. Sentences with Double Negation

- If the negation involved *in illusory CN* is not standard sentential negation, **we predict the possible existence of cases involving double negation.**
- **This prediction is borne out.** Although it has been claimed that this type of structure does not exist (Rivero 1970), examples of these sort are easily found under the appropriate discourse conditions.
- **These structures typically involve objections to a previously uttered negative proposition**, as in (32), gathered from a discussion website. The heading subject is the proposition “animals that do not smell” (32a).

Following discussion about the heading subject (the idea being that they don't smell), someone has been led to say that “not all animals do not smell” (32b):

- (32) a. ¿Animales que no huelan?  
 animals that neg smell?  
 “Animals that don't smell?”
- b. Pues **no** todos **no** huelen  
 the.truth.is neg all neg smell  
 “The truth is that not all of them don't smell” ([www.faunaexotica.net](http://www.faunaexotica.net))

- The following pair provides **another instance of this type of structure**:

The news heading says that “the Spanish jobless do not wish to work the land” (30a); a reply to this statement in the comment section of the news counterargues saying that “**not** all Spanish jobless don't wish to work the land” (33b):

- (33) a. Los parados españoles no quieren trabajar en el campo  
 the jobless Spanish neg want work.inf in the land  
 “The spanish jobless do not want to work the land.”
- b. Que ponga en el titular “algunos parados”...  
 That says SUBJUNCT. in the headline some jobless  
 que **no** todos **no** quieren trabajar en el campo  
 that neg all neg want work.inf in the land  
 “The headline should say “some jobless”, because not all of them don't want to work the land.”

- Further evidence for our analysis of Illusory CN is the fact that **we can combine double negation and** (subjunctive) licensing **high clausal adverbs** of the type we have just discussed, giving rise to well formed examples like (34):

- (34) Que **no** **probablemente** todos **no** quieran<sub>SUBJUNCT</sub> trabajar en el campo.  
 That neg probably all neg want<sub>SUBJUNCT</sub> work in the field  
 “That not probably all don't want to work in the field”

## ● Conclusion:

- possibility of projecting a **higher Neg** whose function is to object  
 → Over and above standard sentential negation in Spanish, there is the  
**(No<sub>Neg1</sub>)**,  
 to an (implicit or explicit) previous assertion, and which can be combined  
 with **sentential Neg (No<sub>Neg2</sub>)**.

(35) [<sub>NegP1</sub> **No<sub>1</sub>** [<sub>XP\*</sub> (high adverbs) X<sup>0</sup> [<sub>FocusP</sub> **QP** Foc<sup>0</sup> [<sub>NegP2</sub> **No<sub>2</sub>** [<sub>IP</sub> ... **QP** ... ] ] ] ] ]

- Although we will not have time to discuss this in detail here, we defend that **this high Neg occupies a position within the structure of the Speech Act.**

[Under a neoperformative model like the one defended by Wiltschko, this high Neg would occupy a position within the structure of the Speech Act Phrase, probably as a head in the ResponseP (see Wiltschko & Heim 2016, Uribe-Etxebarria 2018, and Etxepare & Uribe-Etxebarria 2021 ms. for related discussion)].

## 5. Mixed Strategies

- Consider what the predictions of our working hypothesis (the existence of two strategies for constituent negation) are for possible lexicalization parameters in the expression of constituent negation.
- We claim that **direct merger of NEG and QP is only possible for some QPs, while as in some other cases the sequences under analysis are derived by movement of the QP to a high Spec/FocP close to a high Negation sitting in the left periphery.**
- Due to the impact that the type of QP has in the possibility of licensing the direct merge strategy, we **predict to find languages with Neg+QP sequences in which the negative particle that directly merges with some QPs differs from the one that combines with other QPs.** European Portuguese is such a language.

### 5.1. European Portuguese

- European Portuguese (EP) is a language that makes a **clear distinction between sequences of Negation and Strong Quantifiers** on the one hand, and **Negation and Weak Quantifiers** on the other. (Peres, 2013):

- (36)
- The negative particle for Neg + **strong QP** sequences is always **nem**
  - The negative particle in Neg + **weak QP** is **não** (identical to the sentential one)

- **Negation + QP<sub>STRONG</sub>**

→ When combining with Negation: **Nem** + QP<sub>STRONG</sub>.

→ Besides being preceded by the negator **nem**, Strong Quantifiers **always occur in preverbal position**.

(37) a. **Nem todos os estudantes** tinham lido o livro  
 Neg all.pl the.pl student.pl have read the book  
 “Not all the students have read the book”

b. **Nem sempre** podemos estar ben dispostos  
 Neg always can.1ppl be well disposed  
 “Not always can we be well disposed”

→ As we will see next, **Neg+ QP<sub>STRONG</sub> sequences differ from** the behaviour exhibited by **Neg+ QP<sub>WEAK</sub> sequences** both **regarding the negator** they can cooccur with **and their position in the sentence** (preverbal/postverbal).

- **Negation + QP<sub>WEAK</sub>**

**Não** + QP<sub>WEAK</sub>

→ Weak Quantifiers are preceded by the negator **não**.

→ **Não** + QP<sub>WEAK</sub> can occur in **postverbal position**

(38) Estavam na festa **não muitos estudantes**  
 there.were PREP party not many students  
 “There were not many students at the party”

(39) **Não muitas pessoas** conheciam o segredo do cofre  
 neg many people knew the secret of.the coffer  
 “Not many people knew the secret of the coffer”

→ Ordinary **association with focus** in Portuguese uses ***não***, not ***nem***:

(40) a. **Não** vineram a reunião **todos os estudantes estrangeiros**  
 Neg they.came to meeting all the student.pl foreign.pl  
 “Not all of the foreign students came to the meeting”

b. \* **Não** **todos os estudantes estrangeiros** vineram a reunião  
 Neg all the student.pl foreign.pl they.came to meeting  
 “Not all of the foreign students came to the meeting”

→ If ***não***, not ***nem***, is the **negator for association with focus** (where negation is external to the focalized constituent), then **(41a=35) does not involve external negation, so its underlying structure cannot be the indirect one in (41b):**

(41=35) a. **Nem** **todos os estudantes** tinham lido o livro  
 Neg all the students have read the book  
 “\*Not possibly all the students have read the book”

b. [<sub>NegP</sub> **Neg****NÃO** [<sub>FocP</sub> **QP** Foc<sup>o</sup> [<sub>IP</sub>... QP...]]]]

→ Peres claims explicitly that **the sequence *Nem* + *QP* cannot be separated by anything** in Portuguese (Peres, 2013: 483):

(42) \* **Nem** possivelmente **todos os estudantes** tinham lido o livro  
 Neg possibly all the students have read the book  
 “\*Not possibly all the students have read the book”

→ Interestingly, **the form of the negative particle for Strong Quantifiers, *nem*, is the same as the one for scalar negation**, the equivalent of ***not even*** in English (Peres, 2013 : 484), or ***ni*** in Spanish:

(43) **Nem** **os melhores alunos** responderam à pergunta  
 Neg the best.pl student.pl answered the question  
 “Not even the best students answered the question”

→ In the case of ***Nem* + QP<sub>TODOS</sub>**, the reading of «***not even***» does not make sense.

(44=37a) **Nem** **todos os estudantes** tinham lido o livro  
 Neg all.pl the.pl student.pl have read the book  
 “Not all the students have read the book”

→ One way of connecting the interpretation of ***nem*** in cases like (43) and ***nem*** in cases involving QP<sub>TODOS</sub> like (44=37a) is by **taking *nem* to be associated to a focus feature, implied e.g. by the scalar English particle *even*.**

→ **The focus feature sets up a set of alternatives, over which negation operates.**

→ ● The **focus feature** would be **involved in the obligatory fronting of *nem todos*** in (44=37a).

● This **feature is internal to the fronted constituent**, in fact it constitutes **part of the lexical description of *Nem*.**

### • 3<sup>rd</sup> For strategy in Neg + QP sequences: Mixed Strategy.

→ languages like **Portuguese**, we are forced to expand the typology of Direct Approaches into Direct Merge (with focus) plus Fronting:

→ The relevant structure is thus one that fronts a phrase headed by negation, and includes a focus feature:

$$\text{Foc}^0 \text{ [TP QP}_i$$

(45) [FOCP [IEGP/QP **NO** [FOCP [ Foc<sup>0</sup> [ QP ] ] ] ] ]

→ The **existence of the mixed strategy** in Portuguese **places the locus of cross-linguistic differences at least partly in the featural constitution of negation and the availability of QP internal focalization.**

## 5.2. Italian varieties

→ Consider in this regard Italian too. The structure in (45) is represented also in other Romance varieties. Consider the following facts about **Standard Italian *non*** (from Manzotti and Rigamonti, 1991) **in comparison with Spanish *no***.

(46) a. **Non** una parola a María !  
not one word to Maria  
“Not even one word/not a single word to María”

b. **N-i** una palabra a María

(47) a. Ha giocatto **non** bene  
has.3psg played non well  
“He didn’t play well”

b. Ha jugado **no** \*(**muy**) bien  
has.3psg played non well

(48) a. María **non** ha voluto parlare **che** con lei  
Maria neg has wanted talk.INF that with them  
“Maria did not want to talk but to them”

b. María **no** ha querido hablar \*(**más**) **que** con ellos

(49) a. **Non** un LIBRO ha comperato  
neg one book has bought  
“Not even one book she bought”

b. **N-i** un libro ha comprado  
n-i one book has bought  
“Not even one book she bought”



- **Non** in Standard Italian seems to possess a **scalar feature** that is lacking in Spanish **no**.
- For constituent negation, it is significant that **in those dialects of Italian that possess a split negation, constituent negation appears to be always headed by the scalar negative particle.**
- Poletto (2017: 95) for instance observes that **in Paduan constituent negation requires the sequence of two negation markers, None of them the minimizer *miga* « crumb »:**

- (50) a. miga      Neg1 Neg2 all  
 “Not everyone”
- b. miga      Neg1 Neg2 few  
 “Not few”

## ■ CONCLUSIONS

We have shown that **sequences of NEG-QP** may correspond to **three different syntactic configurations** (51-53).

- i) • The adjacency between the Neg particle and the quantifier results from **directly merging negation with the QP or with Q** (as in Collins, 2017) (51a/b).
  - In those cases, the **scope of negation is determined by its merge position.**
- ii) • The **adjacency between Neg and the QP results from focus fronting** of the QP to a position immediately dominated by a very high Negation, Neg1, as in (52).
  - In this case, the **scope of negation is determined by association with focus.**
- iii) • In the 3<sup>rd</sup> configuration (53), which can be viewed as a combination of the previous two, **Neg is merged to a focus phrase that takes the QP as its complement.** The scope of negation is determined by association with focus. The presence of a focus feature in the QP may force pied-piping of the whole structure to a preverbal focus position.

- (51) a. [<sub>QP</sub> NO [<sub>QP</sub> Q ] ]
- b. [<sub>NP/DP</sub> [NO Q] [<sub>NP/DP</sub> ] ] (Collins' Inner Negation, 2017)
- (52) [<sub>NegP1</sub> Neg<sub>1</sub> [<sub>FocF</sub> [[QP] ← Foc<sup>o</sup> ... [<sub>TP</sub> ... t ... ]]]]
- (53) [<sub>FocF</sub> [[<sub>NegP/QP</sub> Neg [<sub>FocF</sub> Foc<sup>o</sup> [[QP] ] ] ] ← Foc<sup>o</sup> [<sub>TP</sub> ... t ... ] ]]

## 6. CN IN THE ABSENCE OF CN: THE CASE OF BASQUE

- Payne (1985), many languages cannot directly negate a QP such as *many*, *few* or *all*.
- Basque is a case in point: there is **no equivalent of English *not many* or *not all***.
- Sequences of Neg QP are strongly rejected by all speakers (Basque allows CN only in a very limited set of cases, in which negation always precedes a degree modifier in adjectival or adverbial phrases).  
 (55) \**Ez guztiak pasko etorri dira*  
 not.all det.pl/ many / few come are  
 “Not all of them/many/few came”

In

CN

- order to obtain a meaning similar to the one that corresponds to in other lges, **Basque resorts to association with focus**: more specifically, it resorts to **a structure involving negation and a lower Focus position**, as in (56b) (E&UE 2008/9),

(56) a. *GUZTIAK ez dira agertu*  
 all neg are shown.up  
 “ALL OF THEM didn’t show up”

a’.

”

[<sub>FOCP1</sub> *GUZTIAK* <sub>ALL</sub> [<sub>NEGP</sub> *ez* <sub>Neg</sub> [<sub>IP</sub> t *dira agertu* ]]]

Neg QP

b. *Ez dira GUZTIAK agertu*  
 Neg are all shown.up  
 “Not ALL OF THEM showed up”

b’. [<sub>NEG</sub> *ez* <sub>Neg</sub> [<sub>IP</sub> *dira* [<sub>-FOCP2</sub> *GUZTIAK* <sub>ALL</sub> [<sub>VP</sub> *agertu* ]]]]

- In negative sentences, focus can occur in 2 positions: **i)** in a preverbal focus position external to TP, to the left of sentential Neg (*high* FocusP1, as in (56a)), or **ii)** in a position internal to the IP (*low* FocusP2, as in (56b)), preceding the lexical verb but following the Aux, which marks the edge of the IP (De Rijk, 1996 ; E&UE 2008/9).

- **One intriguing property of Basque**, a language which as we have shown in (55) does not allow direct merge of negation and a QP, **is that sequences of the type *Neg+QP*, impossible as part of a full clause, nevertheless seem to be possible under clausal ellipsis**, as in (57).

(Note that in (57) the remnant subject has ergative case, which is only possible under a Tensed transitive clause; see Rezac et al. 2014).

(57) a. *Gehienek lortu dute, baina ez guztiek*  
 most.ERG achieved have but not all.ERG  
 “Most of them got it, but not all of them”

b. *Batzuek lortu dute, baina ez askok*  
 some.erg achieved have but not many.erg  
 “Some got it, but not many”

- **ANALYSIS PROPOSED**: What is the derivation underlying sequences Neg+QP in (57)?

We analyze these **Neg+QP sequences in (57)** sequences as **resulting from a process of IP-ellipsis** (Merchant 2001, Holmberg 2016). More specifically, we argue that, **just as in the case of some cases of Neg+ (CM/Prep)+QP cases in Spanish, they involve**:

- Focalization of the QP to a high focus position, followed by clausal ellipsis
- A high Neg, **Neg1**, (different from sentential negation, **Neg2**) sitting in a high position of the left periphery.

▪ **ANALYSIS PROPOSED FOR CASES LIKE (57) (CONTINUATION)**

(57) a. Gehienek lortu dute, baina ez guztiek  
 most.ERG achieved have but not all.ERG  
 “Most of them got it, but not all of them”

**a) Focalization of the QP, which moves to the Spec of a high FocP1. This is followed by IP (or NegP) ellipsis** (Merchant 2001, Holmberg 2016).

The QP (in this case the subject) undergoes focus fronting to the left periphery of the clause, an option available for both affirmative (58a) and negative (58b) sentences. (Structure (58b) would correspond to examples like (56a)).

But **sentence initial focalization with affirmative sentences** cases like **sentence initial focalization with negative sentences**  
 (58) a. [<sub>FocP1</sub> guztiek<sub>ALL</sub> Foc<sup>o</sup> [<sub>IP</sub> guztiek lortu dute ]]] b. [<sub>FocP1</sub> guztiek<sub>ALL</sub> Foc<sup>o</sup> [<sub>NegP</sub> ez [<sub>IP</sub> guztiek lortu dute ]]]]

**the negation that precedes the QP in (57) cannot correspond to sentential negation, because IP-external focus obligatorily precedes sentential negation in Basque;** in other words, sentential negation follows sentence initial focus, as shown in (58b). (Laka, 1990; Etxepare & Uribe-Etxebarria 2008)  
**b) We propose that, just as in Spanish, there is another Negation, Neg1, sitting in a high position of the left periphery, which is the one involved in the Neg+QP sequences of the type illustrated in (57).**

(59) [<sub>NEGP</sub> **Neg** [<sub>FOCP</sub> **QP** Foc<sup>o</sup> [<sub>NEGP2</sub> **Neg<sub>2</sub>** [QP...]]]] (Neg2= sentential Neg)

■ **Evidence for our analysis of examples like (57):**

**1. Neg1+QP sequences can be interrupted by clausal adverbs.** Just as we saw in Spanish for illusory CN in (28), modal/evaluative adverbs can break the adjacency between Neg1 and QP in Basque.

(59) A: Jendeek kasu egin dute ?  
 people.ERG attention done Aux  
 “Did people pay attention?”

B: **EZ<sub>NEG1</sub>** **seguru** **aski** **guztiek<sub>i</sub>** [<sub>IP</sub> **t<sub>i</sub>** ]<sub>ellided</sub>  
 Neg sure enough all.ERG  
 “Not probably all of them”

C: **EZ<sub>NEG1</sub>** **zoritxarrez** **guztiek<sub>i</sub>** [<sub>IP</sub> **t<sub>i</sub>** ]<sub>ellided</sub>  
 Not unfortunately all.ERG  
 “Not unfortunately all of them”

**This sequence is impossible with sentential negation** in an ordinary declarative sentence with focus fronting when ellipsis is not involved (the same facts obtain in Spanish, as shown above in (31))

(60) \* **EZ<sub>NEG2</sub>** **zoritxarrez** **GUZTIEK** kasu egin dute  
 neg unfortunately all.erg attention done have

- **Evidence for our analysis of examples like (57):** for the Spanish cases involving illusory CN, if is not standard sentential

2. **Parallel to the prediction** we made the negation involved in Basque in structures like (57) negation, **we predict the possible existence of cases involving double negation.**

→ **This prediction is borne out.**

(61) A: <sup>Zenbatek</sup> <sup>erg</sup> askok <sup>not</sup> ez<sub>NEG2</sub> dute hori egin?  
How many.erg askok have that.abs done  
How many didn't do it?

B1. a. <sup>Ez<sub>NEG1</sub></sup> askok  
Not many.erg  
'Not many (didn't do it)'

b. <sup>Ez<sub>NEG1</sub></sup> askok<sub>MANY</sub> [ <sup>ez<sub>NEG2</sub></sup> askok<sub>MANY</sub> dute<sub>HAVE</sub> hori<sub>THAT</sub> egin<sub>DONE</sub> ]

↑ movement to high Spec/FocP
↑ clausal (NegP) ellipsis

- The answer in (61Ba) has the structure in (61B1b). **Although the sentential negation (ez<sub>NEG2</sub>) present in the question is not pronounced in the answer due to** the process of clausal ellipsis (in this case, NegP ellipsis), the interpretation of this example makes it clear that we are in the presence of two negations: **Neg1**, sitting in a high left periphery position, and a sentential negation, **Neg2**, contained within the elided NegP.

- **Evidence for our analysis of examples like (57):**

3. Furthermore, just like in Spanish (see (34-35)), **we predict that the existence of Basque sentences involving both a high clausal adverb and double negation (Neg1 and Neg2) in the cases involving illusory CN.**

→ **This prediction is borne out.**

(61) A: Zenbatek <sup>erg</sup> askok <sup>not</sup> ez<sub>NEG2</sub> dute hori egin?  
How many.erg not have that.abs done  
How many didn't do it?

B2. <sup>Ez<sub>NEG1</sub></sup> zorionez/seguru aski askok  
Not luckily / probably many.erg  
'Not luckily many (didn't do it)'

- (61B2) has the structure in (62), which is exactly the same as (35), with the only difference that the Basque example in (61) involves clausal (NegP) ellipsis.

(62=35) [ <sub>NegP1</sub> No [ <sub>XP\*</sub> (high adverbs) X<sup>0</sup> [ <sub>FocusP</sub> QP [ <sub>Foc</sub> [ <sub>NegP2</sub> No [ <sub>IP</sub> ... QP ... ] ] ] ] ] ] ]

↑ movement to high Spec/FocP
↑ clausal (NegP) ellipsis

1. Basque does not license true constituent negation with QPs. The only way to get a structure with a meaning similar to CN is by association with focus, in a configuration with a negation that c-commands the QP from a higher position in the clausal skeleton.

■ **Summarizing:**

2. This is possible in the configuration in (63), with sentential negation,  $Ez_{Neg2}$ , and a QP in the Spec of a low FocP, FocusP<sub>2</sub> in (63). (Movement of the QP to a higher FocP, FocP<sub>1</sub>, would get the QP out of the scope of Neg2: QP > Neg2).

(62) [<sub>FocusP1</sub> FOC<sub>1</sub><sup>o</sup> [<sub>NegP2</sub>  $Ez_{Neg2}$  [<sub>IP</sub> ... [<sub>FocusP2</sub> QP FOC<sub>2</sub><sup>o</sup> [...QP... ] ] ] ] ] ]

- Further since in this configuration Neg2 and the QP are not adjacent to each other, due to the existence of intervening material between both elements.
- IP ellipsis implies the PF-elision of inner focus site, constituent negation relations expressed at Spell Out by means of an ordinary sentential negation in Basque, since the QP would be elided together with the rest of the material within IP.

3. also possible in the configuration in (64), with a high Neg periphery,  $Ez_{Neg1}$ , and the QP in the Spec of a high FocP, FocP<sub>2</sub>, with clausal ellipsis.

(63) [<sub>NegP1</sub>  $Ez_{Neg1}$  [<sub>FocusP1</sub> QP FOC<sub>1</sub><sup>o</sup> [<sub>IP</sub> [...QP... ] ] ] ] ]

- This structure can give rise to Neg+QP sequences, where Neg and the QP are adjacent to e.o.
- The Neg1+QP sequence can be interrupted by clausal adverbs, which shows that Neg sits in a high position of the left periphery. That Neg1 is not sentential negation can be further shown by the fact that these structures allow 2 negations, Neg1 and Neg2, although the latter is silent due to clausal ellipsis.

[<sub>FocusP</sub> QP FOC [<sub>NegP2</sub>  $Ez_{Neg2}$  [<sub>IP</sub> ...QP... ] ] ] ]

■ **Summarizing:**

4. Although for reasons of space we have limited ourselves to show that the *illusory* cases of CN involve association with focus with a clausal external Neg, the different meanings that some of this sentence-initial illusory cases of constituent negation can have (and in particular with the possibility of interpreting negation with a refutal or a non refutal interpretation) leads us to suggest that there are more than one position available for this high negation in the structure of the Speech Act Phrase. See Wiltschko & Heim (2016), Wiltschko (2017), Uribe-Etxebarria (2018, 2019, 2021) and Etxepare & Uribe-Etxebarria (2021b) for discussion.
5. The fact that sequences apparently identical to true constituent negation structures should arise in a language that does not have independent constituent negation strengthens the idea defended here that a number of sequences analyzed as cases of constituent negation are a sort of cognitive illusion, and must be reinterpreted in terms of association with focus.

# 대단히 감사합니다

# Thank you

# Mila esker

This research was partially funded by **i)** the Agence Nationale de la Recherche ANR, *Basque in the making (BIM)*, 2019-2022 **ii)** the Spanish Ministry of Science, Innovation and Universities (MiCIU) / Spanish Research Agency (AEI) and the European Regional Development Fund (FEDER, UE), MICIU ( VASTRUD, PGC2018-096870-B-I00), **iii)** The Basque Government (Consolidated Research Groups HITT GIC18/09 & IT1537-22), and **iv)** the University of the Basque Country (UPV/EHU Research Groups HITT-GIU18/221).

## Selected References

- Acquaviva, P. 1997. *The Logical Form Of Negation*. Outstanding Dissertations in Linguistics. Revised Edition, New York: Garland Science.
- Albader, M.H. 2018. *La sintaxis de la coordinación distributiva*. Doctoral dissertation, Universidad Complutense, Madrid.
- Al-Horais, N. 2017. "On Negation and Focus in Standard Arabic" *Journal of Universal Language* 18-1: 1-34.
- Aoun, J. et al. 2010. *The Syntax of Arabic*. Cambridge University Press.
- Baunaz, L. 2011. *The Grammar of French Quantification*. Studies in Natural Language and Linguistic Theory 83, Dordrecht: Springer.
- Beghelli, F. & T. Stowell, 1997. "Distributivity and negation: The syntax of *each* and *every*", in A. Szabolcsi (ed.), *Ways of Scope Taking*. Dordrecht: Kluwer, 71-107.
- Bosque, I. 1984. "Negación y elipsis" *E.L.U.A.* 2: 171-199.
- Cavalcante, R. (2012) *Negação anafórica no Português brasileiro: Negação sentencial, negação enfática e negação do constituinte*. Universidade de São Paulo dissertation.
- Chomsky, N. 1995. *The Minimalist Program*, Cambridge, MA: MIT Press.
- Cinque, G. 1990. *Types of A-bar Dependencies*. Cambridge, MA: MIT Press.
- Collins, C. and P. Postal. 2014. *Classical Neg-Raising*. MIT Press.
- Collins, C. 2017. « The Distribution of Negated Quantifier Phrases in English » Ms., NYU
- Collins, C. 2018. « \*NEG NEG » *Glossa: A Journal of General Linguistics*, 3-1.
- [Cruschina\[1\]](#), S. 2011. *Discourse-Related Features and Functional Projections*. Oxford University Press.
- Corver, N. 1997. "The Internal Syntax of the Dutch Extended Adjectival Projection", *Natural Language and Linguistic Theory* 15, 289-368.
- De Clercq, K. 2013. *A Unified Syntax of Negation*. Doctoral dissertation, Universiteit Gent.
- Depiante, M. 2000. *The syntax of Deep and Surface Anaphora*. Doctoral diss., University of Connecticut.
- [Elordieta\[2\]](#), G. and A. Irurtzun. 2009. « The Prosody and Interpretation of Non-Exhaustive Answers » *ASJU* 43 : 205-230.
- Ettxepare, R. 1997. *On the grammatical representation of speech events*, Ph.D. Dissertation, U. Maryland.
- Ettxepare, R. and M. Uribe-Etxebarria. 2005. "Wh-phrases in-situ in Spanish: scope and locality" *Recherches Linguistiques de Vincennes* 33, 9-34.
- Ettxepare, R. and M. Uribe-Etxebarria. 2008. "On Negation and Focus in Spanish and Basque", in X. Artiagoitia & J. Lakarra (eds.), *Gramatika jaietan: Patxi Goenaga irakaslearen omenaldiz*. Special issue of *International Journal of Basque Linguistics and Philology*, UPV/EHU, Bilbao. 287-309.
- Ettxepare, R. and M. Uribe-Etxebarria. 2011. "Foco y Negación de Constituyentes" In Escandell, M. et al (eds) *60 problemas de gramática*. Madrid: Akal. 235-241.
- Ettxepare, R. and M. Uribe-Etxebarria. 2012. "Las preguntas de Qu-in situ en español: un análisis derivacional", in J.M. Brucart & Á. Gallego (eds.), *El movimiento de constituyentes*, Visor: Madrid, 247-268.
- Ettxepare, R. and M. Uribe-Etxebarria. 2018. «Context Sensitive Aspects of Constituent Negation » In A. Gallego et al (eds). *Language, Syntax and the Natural Sciences*. Cambridge University Press. 77-98.
- Ettxepare, R. and M. Uribe-Etxebarria. 2021. «Crosslinguistic variation in Constituent Negation», ms. CNRS & UPV/EHU. [To be published in M.J. Arche & H. Demirdache (eds.), *Footprints of Phrase Structure*.
- Fox, D. 2000. *Economy and Semantic Interpretation*. MIT Press.
- González, R. 2008. *La polaridad positiva en español*. Doctoral Dissertation Universidad Complutense and Instituto Universitario Ortega y Gasset, Madrid.
- Gutierrez-Bravo, R., C.M. Sobrino and M. Uth. 2019. "Contrastive Focus in Yucatecan Spanish" In A. Gallego (ed) *The Syntactic Variation in Spanish Dialects*. Oxford University Press.
- Herburger, E. 2000. *What Counts: Focus and Quantification*. Cambridge, MA: MIT Press.
- Herrera Morera, G. 2012. « El operador *no* en el español de Centroamérica : análisis tipológico ». *Letras* 51 : 147-162.
- Holmberg, A. 2016. *The Syntax of Yes and No*. Oxford University Press.
- Horn, L. 1989. *A Natural History of Negation*, Chicago: University of Chicago Press.

- Israel, M. 2011. *The Grammar of Polarity: Pragmatics, Sensitivity, and the Logic of Scales*, Cambridge: Cambridge University Press.
- Jackendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Kaiser, E. 2006. « Noncanonical negation and information structure in Finnish » In V. Molnar and S. Winkler (eds) *The Architecture of Focus*. Berlin : Mouton De Gruyter. 265-290.
- Horn, L. 1989. *A Natural History of Negation*, Chicago: University of Chicago Press.
- Israel, M. 2011. *The Grammar of Polarity: Pragmatics, Sensitivity, and the Logic of Scales*, Cambridge: Cambridge University Press.
- Jackendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Kaiser, E. 2006. « Noncanonical negation and information structure in Finnish » In V. Molnar and S. Winkler (eds) *The Architecture of Focus*. Berlin : Mouton De Gruyter. 265-290.
- Kim J.B. & I. Sag, 2002. "Negation without head movement". *Natural Language and Linguistic Theory* 20, 339-412.
- Kiss, K. E. 2002. *A Grammar of Hungarian*. Cambridge: Cambridge University Press.
- Klima, E.S. 1964. "Negation in English", in J. Fodor and J. Katz (eds.), *The Structure of Language*, Englewood Cliffs, N. J.: Prentice-Hall, 246-323.
- Köhler, E. 1985. "Nicht-Não: uma equivalencia problematica" In A.C. Franco (ed) *Duas Linguas in Contraste: Portugues et Alemão. Actas do I Coloquio Internacional de Linguística Contrastiva Portugues-Alemão*. Porto: Universidade do Porto. 29-42.
- Krifka, M. 2001. "Quantifying into question acts" *Natural Language Semantics* 9-1: 1-40.
- Lasnik, H. 1972. *Analyses of Negation in English*. MIT doctoral dissertation.
- Lee, C. 2017. "Contrastive Topic, Contrastive Focus, Alternatives, and Scalar Implicatures" In C. Lee et al (eds) *Contrastiveness in Information Structure, Alternatives and Scalar Implicatures*. Springer. 3-21.
- Neeleman, A. and R. Vermeulen. 2012. "Types of focus and their interaction with negation" In A. Neeleman and R. Vermeulen (eds) *The Syntax of Topic, Focus and Contrast. An Interface Approach*. Berlin: Mouton. 227-264.
- Nicoli, F. (1983) *Gramatica Milanese*. Bramante Editrice.
- Manzotti, E. and A. Rigamonti. 1991. "La negazione" In L.Renzi and G. Salvi (eds) *Grande grammatica italiana di consultazione*. Volume II. Bologna: Il Mulino. 245-320.
- McCawley, J. 1991. "Contrastive Negation and Metalinguistic Negation" *Chicago Linguistic Society* 27: 189-206.
- Merchant, J. 2001. *The Syntax of Silence*. Oxford University Press.
- Neeleman, A. and R. Vermeulen (eds). 2012. *The Syntax of Topic, Focus and Contrast. An Interface Based Approach*. De Gruyter Mouton.
- Parry, M. (1997) "Negation" In M. Maiden and M. Parry (eds) *The Dialects of Italy*. London: Routledge.
- Payne, J.R. 1985. "Negation" In T. Shopen (ed) *Language Typology and Syntactic Description. Vol I*. Cambridge University Press. 197-242.
- Penka, Doris. 2011. *Negative Indefinites*. Oxford University Press.
- Peres, J. A. 2013. "Negação" In E.B.P. Raposo et al (eds) *Gramática do Português. Vol I*. Fundação Kalouste Gulbenkian. 461-498.
- Poletto, C. 2017. "Negative Doubling: in favor of a Big-NegP analysis" In S. Cruschina et al (eds) *Studies on Negation. Syntax Semantics and Variation*. Vienna University Press. 81-104.
- Sánchez-López, C. 1999. "La negación", in I. Bosque y V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española, Volumen 2*. Madrid: Espasa, 2561-2634

- Ramchand, G. 2004. "Two types of negation in Bengali" In V. Dayal et al. (eds) *Clause Structure in South Asian Languages*. Kluwer. 39-66.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual* ("Reference Corpus of Contemporary Spanish"). <http://www.rae.es>
- Remberger E.M. 2014. « A Comparative Look at Focus Fronting in Romance » In A. Duffer and A.S. Octavio de Toledo (eds) *Left Sentence Peripheries in Spanish. Diachronic, Variationist and Comparative Perspectives*. Amsterdam : John Benjamins. 383-418.
- Rivero, M.L. 1970. "A Surface Structure Constraint on Negation in Spanish", *Language* 46 (3), 640-666.
- Rizzi 1997. "The fine structure of the left periphery", in Liliane Haegeman (ed.), *Elements of Grammar. A Handbook of Generative Syntax*, Dordrecht: Kluwer, 281-337.
- Rooth, M. 1992. "A theory of focus interpretation", *Natural Language Semantics* 1, 75-116.
- Schapansky, N. 2010. "Further aspects of negation in French" *Lingua* 120: 103-131.
- Temmerman, T. 2012. *Multidominance, ellipsis, and quantifier scope*, Ph.D. Dissertation, Leiden University.
- Tomioka, S. 2010. Contrastive topics operate on speech acts". In M. Zimmermann a,d C. Féry (eds) *Information Structure: Theoretical, Typological and Experimental Perspectives*. Oxford University Press. 115-138.
- Uriagereka, J. 2008. *Syntactic Anchors*. Cambridge University Press: Cambridge.
- Uribe-Etxebarria, M. 1992. « On the structural position of the subject in Spanish, their nature and their consequences for quantification » In J.A. Lakarra and J. Ortiz de Urbina (eds) *Syntactic Theory and Basque Syntax*. Donostia : Gipuzkoako Foru Aldundia. 447-492.
- Uribe-Etxebarria, M. 2002. "In-situ questions and masked movement", *Linguistic Variation Yearbook* 2, 259-303.
- Uribe-Etxebarria, M. 2018. *The Left Periphery and the Syntax-Pragmatics Interface of (Negative) Polar Questions*, Seoul National University International Conference on Linguistics: Theme Session – Crosslinguistic Variation in the Left Periphery at the Syntax-Discourse Interface, Seoul, June 2018.
- Vicente, L. 2006. « Short negative replies in Spanish » In J. Van der Weijer and L. Bettelou (eds) *Linguistics in the Netherlands*. Amsterdam :John Benjamins. 199-211.
- Wiltschko, M. & J. Heim (2016). The syntax of confirmationals. A neo-performative analysis. In: Gunther Kaltenbock, Evelien Keizer and Arne Lohmann (eds.) *Outside the Clause. Form and function of extra-clausal constituent*. Amsterdam: John Benjamins. 303-340.
- Wiltschko, M. 2017. Response particles beyond answering. In Laura R. Bailey & Michelle Sheehan (eds.), *Order and structure in syntax I: Word order and syntactic structure*, 241–279. Berlin: Language Science Press.
- Zanutini, R. 1997. *Negation and clausal structure : a comparative study of Romance Languages*. Oxford University Press.
- Zeljstra, H. 2013. "Not in the first place". *Natural Language and Linguistic Theory* 31, 865-900.