UNIVERSITY OF CALIFORNIA SAN DIEGO

Human and language model experiments in Korean discourse pragmatics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Linguistics with a Specialization in Computational Social Science

by

Hagyeong Shin

Committee in charge:

Professor Farrell Ackerman, Chair
Professor Leon Bergen
Professor Victor Ferreira
Professor Andrew Kehler

2024

The dissertation of Hagyeong Shin is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

2	2nd person
ACC	accusative
ADVZ	Adverbalizer
CF	contrastive focus
COMP	complementizer
COND	conditional
CONJ	conjunctive
CT	contrastive topic
DAT	dative
DEC	declarative
DEF	definite
EVID	evidential
FOC	focus
FUT	future
HAB	habitual
IND	indicative
INST	instrumental
INT	interrogative
LOC	locative
NEG	negative
NOM	nominative
NR	nominalizer
PASS	passive
PL	plural
POL	polite
PRF	perfect
PROG	progressive
PRT	particle
PST	past tense
PUNCT	punctual
SFP	sentence-final particle
SG	singluar
TOP	topic
QUOT	quotative
VOC	vocative

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Farrell Ackerman, for his support in making the realization of this dissertation possible. His depth and breadth of expertise across various disciplines have been invaluable at every step of the dissertation research, particularly in grounding interdisciplinary topics in linguistics. I am grateful to Leon Bergen for inspiring my interest in language models and for engaging discussions on cutting-edge research in the field of computational linguistics. I would like to give special thanks to Victor Ferreira for sharing his expertise in psycholinguistics and for graciously introducing me to the Language Production Lab. Finally, I thank Andrew Kehler for continuously challenging me with his sharp insights into discourse pragmatics and semantics.

I am fortunate to have met such great people beyond the advisor/committee–advisee relationship. I would like to thank Sean Trott for providing me the opportunity to study language models in the context of the interdisciplinary fields of linguistics and cognitive science. I am also grateful to Theodore Jones, Sarah Oh, Tessa Stanier, and Zhien Zhang for their contributions as research assistants, helping to prepare experimental materials and to organize discourse data. For this, I am thankful to Emily Clem for providing the PURL program, which served as a platform for promoting collaboration between undergraduate and graduate student researchers.

I must acknowledge the wonderful community in the Linguistics Department at UC San Diego. All faculty, staff, and students have provided support and made contributions to this dissertation in one way or another. I thank Eric Baković for providing great compassion as the department chair. I also thank Will Styler, the director of the Computational Social Science program, for encouraging exciting research pursuits. I was able to become a better researcher, thanks to meaningful interactions I had with students and aspiring linguists alike. For this, I thank Will Styler again for providing various opportunities to engage in teaching and the community.

This dissertation reflects the academic journey I have undertaken, starting long before I began the Ph.D. program at UC San Diego. I was introduced to computational linguistics and natural language processing at San Diego State University, particularly by Gabriel Doyle and Robert Malouf. I am grateful to them for opening my eyes to the exciting opportunities ahead. From Young Min Baek at Yonsei

University, I learned a great deal about how to carry out data-driven projects end-to-end and the discipline required to become a researcher.

This dissertation was financially supported by the Graduate Student Research and Travel Fellowships from the International Institute at UC San Diego and the research and travel funding from the Linguistics Department. The language model experiments presented in this dissertation relied on resources provided by the Data Science and Machine Learning Platform and the Social Sciences Computing Facility at UC San Diego. All human experiments in this dissertation were approved by the IRB of UC San Diego (#806449).

The material in Chapter 5 has been published in: Shin, H., & Trott, S. (2024). Do language models capture implied discourse meanings? An investigation with exhaustivity implicatures of Korean morphology. In *Proceedings of the Society for Computation in Linguistics 2024*, 150-161. It was adapted for this dissertation by adding background materials in section 5.2, providing more detailed descriptions of results, and using terminology consistent with that of Chapters 1-4.

VITA

2014	B. A. in Communication, Sookmyung University, Seoul
2017	M. A. in Communication, Yonsei University, Seoul
2019	M. A. in Linguistics, San Diego State University, San Diego
2024	Ph. D. in Linguistics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Human and language model experiments in Korean discourse pragmatics

by

Hagyeong Shin

Doctor of Philosophy in Linguistics with a Specialization in Computational Social Science

University of California San Diego, 2024

Professor Farrell Ackerman, Chair

This dissertation examines the relationship between grammar and discourse in Korean morphology, focusing on the polyfunctionality of the *-lul* and *-nun* markers. The *-lul* marker primarily serves grammatical object functions while also indicating discourse focus or contrastive focus based on context. In contrast, the *-nun* marker primarily indicates discourse topic or contrastive topic status and can denote grammatical subject or object roles in place of dedicated grammatical case markers distinguishing subjects and objects. The interplay between the grammatical and discourse functions of each marker leads to complex usage patterns particularly associated with the notion of exhaustive interpretation.

In Chapter 1, I present the empirical and descriptive grammatical and discourse notions relevant for the analyses in the subsequent four chapters. In Chapters 2-4, I present psycholinguistic experiments

that investigate usage patterns linked to the *-lul* and *-nun* markers' exhaustive interpretation. Chapter 2 examines how these markers interact with differences in shared knowledge between interlocutors. The *-lul* marker tends to convey exhaustivity even without identified alternatives in shared discourse, whereas the *-nun* marker's exhaustivity relies more heavily on the shared discourse. Chapter 3 explores the cancelability of exhaustivity, finding the *-lul* marker's exhaustivity to be cancelable, while the *-nun* marker's exhaustivity is non-cancelable. Chapter 4 confirms the optionality of the *-lul* marker and the exhaustivity it conveys, comparing it to null markings.

Chapter 5 evaluates large language models in terms of psycholinguistic experiments from Chapters 2-4. The outputs of large language models do not fully capture the detailed usage patterns of the *-lul* and *-nun* markers' exhaustivity, highlighting the gap between human speakers and the models. In conclusion, this dissertation demonstrates the complex interaction between the grammatical and discourse functions of morphology, and suggests that systems lacking targeted training in discourse pragmatics do not reflect this complexity in natural language.

1 Morphological and syntactic organization in Korean

1.1 Introduction

Linguistic structures are often viewed through the dichotomy of grammar and discourse. Consequently, it is often assumed that a structure is typically thought to have a primary function in the domain of grammar or discourse. For instance, if a morphological structure seems to indicate a subject argument, we tend to characterize it as a case marker serving a grammatical subject function. In contrast, if a morphological marker designates a function such as a discourse topic, we characterize it as a discourse marker. Thus, we often assume that such morphological markers are specialized for either grammatical or discourse functions, rather than either alternating between functions or simultaneously bearing both functions.

A clear distinction between grammatical and discourse functions of morphology is observed in Imbabura Quechua.[1] In Imbabura Quechua, grammatical subjects are typically unmarked, but when a subject conveys new information, the clitic *=mi* is used to indicate its discourse focus status (I. Kwon, 2011). This is illustrated in (1), where the discourse begins with a question from speaker A. In speaker B's response, the subject 'Laura' does not provide new information that answers the question due to the following negation, and it is not marked with *=mi*. However, the second subject 'John' introduces the new information that responds to the preceding question, and it is followed by *=mi*.

[1] All examples of Imbabura Quechua mentioned here are from I. Kwon (2011).

(1) A: pita miku-rka atalpa-ta xuan o laura
who eat-PST chicken-ACC John or Laura
'Who ate the chicken, John or Laura?'

B: Laura na miku-rka-chu xuan**=mi** miku-rka
Laura NEG eat-PST-NEG John=mi eat-PST
'Laura didn't eat, but John ate.'

Continuing with examples from Imbabura Quechua, grammatical objects are overtly marked with *-ta*. In (2), the object *pirkuti* 'rat' is followed by the *-ta* marker in both A's question and B's response. The subject 'Pepe' in B's response, which introduces new information in response to A's question, is marked with the *=mi* clitic.

(2) A: pita wenyuchi-rka pirkuti-**ta**?
who kill-PST rat-ACC
'Who killed the rat?'

B: Pepe**=mi** wenyuchi-rka pirkuti-**ta**
Pepe=mi kill-PST rat-ACC
'Pepe killed the rat.'

When grammatical objects have topical status, the canonical object marker *-ta* is replaced with *-ka*. In (3), for example, the object phrase 'your things, your thoughts' is followed by the *-ka* marker instead of *-ta*. Additionally, the predicate *apa* is marked with *=mi*, indicating the new information conveyed by the quoted phrase.

(3) kunan-**ga** kamba, kamba ya-shka-ta-**ka** apa-sha=mi ni-shka nin
now-TOP 2SG 2SG think-PRF-ADVZ-TOP take-FUT=mi say-PRF QUOT
"'Now your things, your thoughts, I will take,' he said."

Overall, these examples from Imbabura Quechua illustrate a clear distinction between morphological markers that specialize in either grammatical or discourse functions. Korean morphological markers have often been analyzed under the assumption of a similar division between grammar and discourse. Postpositional markers that follow noun phrases are typically categorized as either grammatical case markers or discourse markers. Grammatical case markers are understood to indicate the syntactic relationship between a noun phrase and a verb, while discourse markers signal the discourse status of a phrase, shaped

by its contextual role within the discourse.

For instance, *-i/ka* and *-(l)ul* markers in Korean are classified as grammatical case markers, as *-i/ka* canonically follows grammatical subjects and *-(l)ul* follows grammatical objects. These grammatical case markers are often contrasted with discourse markers. For instance, the *-(n)un* marker is classified as a discourse marker, as it often indicates discourse topic and prominence (I. Kim, 2016; Choi, 1996). The two markers *-i/ka* and *-(l)ul* serving grammatical functions are illustrated in (4) and (5), and the *-(n)un* marker serving the discourse function is illustrated in (5).[2]

(4) Anna-**i** pizza-**lul** kacwewassta.
Anna-NOM cake-ACC brought
'Anna brought (a/the) cake.'

(5) Anna-**un** pizza-**lul** kacwewassta.
Anna-TOP cake-ACC brought
'Speaking of Anna, she brought (a/the) cake.'

However, a closer look into (4) and (5) show that the *-nun* marker's function cannot be simply limited as a discourse-related, as it replaces the grammatical case marker *-i/ka* in (5). Furthermore, as illustrated in (6), the *-lul* marker and the *-nun* marker can alternate to appear after a grammatical object, or these markers do not have to appear after all. If the *-lul* and *-nun* markers each serve a unique function in the domain of grammar and discourse, the three-way alternation presented in (6) should not be permissible.

(6) Anna-i pizza-**{lul/nun/∅}** kacwewassta.
Anna-NOM cake brought
'Anna brought (a/the) cake.'

Thus, the object marking patterns in Korean, as exemplified in (6), challenge the traditional assumption of a strict dichotomy between grammatical and discourse markers. These patterns offer a unique opportunity to explore the intersection of morphology, syntax, and discourse in Korean. In this dissertation, I will investigate the factors that modulate the alternation between the three marking options, *-lul*,

[2]For simplicity, henceforth, the allomorphic variations of *-(l)ul* and *-(n)un* are referred to as *-lul* and *-nun* respectively.

-nun, and null-markings (∅). This exploration aims to illuminate the relationship between grammatical and discourse functions in morphology.

In the remainder of this chapter, I will first introduce the key organizing notions of syntagmatics, paradigmatics, and polyfunctionality, which are central to this dissertation. Following this, I will clarify the patterns under investigation—specifically the alternating markings of *-lul*, *-nun*, and null markings, as illustrated in (6). Syntagmatic patterns of these markings will be described in section 1.3, and paradigmatic patterns of these markings will be discussed in section 1.4. In section 1.5, I will examine the discourse pragmatic interpretations of the *-lul* and *-nun* markers. The current chapter will conclude with a summary of these discussions and an outline of the patterns to be investigated in the experimental settings.

Chapters 2, 3, and 4 present a series of psycholinguistic studies examining the discourse usage patterns of the *-lul* and *-nun* markers. Each chapter provides an overview of relevant studies and reports on a psycholinguistic experiment. Chapter 2 investigates how the *-lul* and *-nun* markers interact with shared knowledge between interlocutors. Chapter 3 explores how these markers impose different constraints on the discourse that follows. Chapter 4 addresses potential confounds from the second experiment and discusses plausible factors influencing the organization of alternative markings in Korean.

In Chapter 5, I test whether large language models can encode the polyfunctionality of linguistic structures. The chapter begins with a brief introduction to key concepts in modern language models and the approach of evaluating these models as psycholinguistic subjects. I then assess a set of pre-trained large language models to investigate whether systems without targeted training on discourse pragmatics effectively capture discourse patterns observed with human speakers in Chapters 2-4. The dissertation concludes with Chapter 6, where I summarize the main findings and discuss their contributions to a broader understanding of the relationship between linguistic structures and their functions, as well as the interplay between grammar and discourse.

1.2 Notions for organization

In linguistics, a PARADIGM is a set of systematically organized forms that a word can take. An INFLECTIONAL PARADIGM, in particular, is a set of word forms indicating different grammatical features, such as gender, number, case, tense, and mood. For example, the noun *book* in English has the inflected form *books*, which indicates its plural number. The verb *to write* in English can inflect to, for instance, *writes*, *wrote*, or *written* to indicate different tenses and aspects. Although different sets of commonly occurring grammatical features are used across languages, each language exhibits its own way of systematically organizing a paradigm and reflecting certain features of a word within its inflectional morphology.

In explaining the systematic organization of a paradigm, two approaches are traditionally employed. One approach is to explore the internal organization of words or word classes, known as SYNTAGMATIC morphology. The other approach is to examine the external organization of relations between words, known as the PARADIGMATICS of a paradigm. Syntagmatic patterns involve the linear organization of segmental and suprasegmental elements, while paradigmatic patterns reflect the relationships between (classes of) words (see Ackerman, Malouf, & Blevins, 2009; Blevins, Ackerman, & Malouf, 2018; Bonami & Strnadová, 2019).

An inflectional paradigm does not always present a one-to-one relationship between an inflectional marker and a feature expressed by it. The same syntagmatic marker can serve multiple different purposes in a systematic way. This phenomenon is referred to as POLYFUNCTIONALITY (Stump, 2014). An instance of polyfunctionality in inflectional morphology is observed in the Noon language (Niger-Congo; Senegal), particularly with the verbal suffix *-u(s)*. This suffix appears as *-u* in word-final position and as *-us* otherwise. It expresses passive voice, as shown in (7) (a) and (b). As shown in (7) (c) and (d), the same suffix expresses plural number.

(7) An inflectional suffix in Noon (Niger-Congo; Senegal) (Soukka, 2000: 176, cited in Stump, 2014)

a. Mi lím-**u** ga Padee.
I have.child-PASS in Fandène
'I was born in Fandène.'

b. Jën-aa ñamsi (=ñam-**us**-i) na maalu.
fish-DEF eat-PASS-HAB with rice

‘Fish is eaten (habitually) with rice’

c. ɓeti-caa ham-**u** ga feet-aa.
women-DEF dance-PL at feast-DEF
‘The women dance at the feast.’

d. yaal-caa ka’seera (=kaɗ-**us**-ee-ra) Dakka
men-DEF leave.for-PL-PAST-PUNCT Dakar
‘The men had left for Dakar.’

Patterns in (7) above may appear as cases of homonymy, where two distinct suffixes (one *-u(s)* for passive voice and the other *-u(s)* for plural number) have the same morpho-phonological realization. However, sentences in (8) below illustrate that the *-u(s)* suffix cannot be explained as a case of homonymy. The *-u(s)* suffix is supplanted by *-uunun* in the perfect aspect, whether it expresses passive voice in (a) or plural number in (b). That is, the *-u(s)* is accounted for by the same rule of exponence for the perfect aspect, suggesting that the passive *-u(s)* and plural *-u(s)* cannot be considered two distinct suffixes. Instead, the *-u(s)* must be considered a polyfunctional suffix, a single suffix with multiple functions (passive voice and plural number). Due to its polyfunctionality, ambiguity arises in cases exemplified in (c). The *-uunun* suffix may replace the *-u(s)* suffix, indicating either passive voice or plural number.[3] In this dissertation, instances of polyfunctionality will be observed with Korean morphology.

(8) An inflectional suffix in Noon (Niger-Congo; Senegal)

a. Jën-aa tóoh ñam-**uunun**.
fish-DEF all eat-PASS.PRF
‘The whole fish has been eaten.’

b. Oomaa-caa fool-**uunun** bes-ii tóoh.
children-DEF run-PL.PRF day-DEF all
‘The children have run all day.’

c. *Pe’-caa ñam-**uunun**.
goats-DEF eat-P.PRF
Ambiguous: ‘The goats have eaten.’/‘The goals have been eaten.’

With these three analytic constructions—syntagmatics, paradigmatics and polyfunctionality—in

[3]This is only one instance of polyfunctionality. More on different ways that polyfunctionality arises can be found in Stump (2014).

mind, the next subsection presents the syntagmatics of Korean morphology and syntax. Various syntagmatic patterns of morphological markings and word order in Korean are described, and the subset of object-marking patterns are identified as the focus of this dissertation. Following this, paradigmatic contrasts between the alternative object markings are discussed in relation to the functions that each marking option achieves. The object marking patterns examined in the current chapter demonstrate that morphosyntactic functions of alternative markings alone cannot fully explain the inflectional organization in Korean, and discourse functions must be considered. Crucially, paradigmatically contrasting alternative marking patterns reveal that object markers in Korean fulfill functions in both grammar and discourse simultaneously. This phenomenon is explained as an instance of polyfunctionality, in that the markers indicate more than one function across different domains. This chapter closes with highlighting discourse behavioral patterns that require further investigation to explain paradigmatic contrasts between polyfunctional markers.

1.3 Syntagmatic patterns in Korean

1.3.1 Morphology

In Korean, morphological markers typically appear postpositionally after a noun phrase. These markers are often broadly classified into two categories based on the type of relation they indicate: grammatical case markers and discourse markers. Grammatical case markers specify the grammatical relation between constituents. Discourse markers, on the other hand, indicate the relationship between constituents in relation to the discourse context. Table 1.1 (taken from Choi-Jonin, 2008) presents a selection of markers and their functions, classified according to the dichotomy of grammar and discourse.

Table 1.1: Common classification of Korean morphological markers.

Grammatical case markers			Discourse markers		
Label	Form	Function	Label	Form	Meaning
Nominative	*-i/ka*	subject	Topic/Contrastive Topic	*-(n)un*	‘as for’, ‘contrastively’
Accusative	*-(l)ul*	direct object	Inclusion	*-to*	‘also’, ‘too’, ‘indeed’
Dative	*-eykey*	indirect object	Limitation	*-man*	‘only’, ‘solely’
Instrumental	*-(u)lo*	oblique	Addition	*-cocha*	‘even’, ‘as well’

Among the markers classified as grammatical case markers in Table 1.1, the nominative *-i/ka* indicates the grammatical subject relationship with a predicate, while the accusative *-(l)ul* marks the direct object relationship to the predicate.[4] In (9) below, each noun phrase (NP) is accompanied by its grammatical case marker: the subject *nay* ('I') is followed by the nominative marker, the indirect object *Sohee* is followed by the dative case marker *-eykey*, and the direct object *chayk* ('book') is followed by the accusative marker.

(9) Nay-**ka** Sohee-**eykey** chayk-**ul** cwu-ess-ta.
I-NOM Sohee-DAT book-ACC give-PST-DEC
'I gave a book to Sohee.'

The above sentence exemplifies the canonical grammatical case markings in Korean, where each noun phrase is followed by a case marker indicating the grammatical relationship between constituents. The nominative, dative, and accusative markers share the syntagmatic pattern of appearing after a noun phrase to indicate a unique grammatical function of a phrase.

In colloquial Korean, the presence of grammatical case markers after subjects and direct objects is not obligatory. When a noun phrase's grammatical function is clear, the case marker may be omitted. Illustrated with (10), the word *chayk* ('book') is not accompanied by any case markers, yet the sentence remains entirely well-formed in Korean. The grammatical function of the 'book' becomes evident from the clear assignment of grammatical functions to other noun phrases, each marked with the appropriate case markers, and a common sense default interpretation of the clause. Notably, the 'book', as the sole inanimate entity in the sentence, further contributes to its unambiguous interpretation as the direct object.

(10) Nay-ka Sohee-eykey chayk-**(ul)** cwu-ess-e.
I-NOM Sohee-DAT book-(ACC) give-PST-DEC
'I gave a book to Sohee.'

Syntagmatic patterns of discourse markers interact with the grammatical function of a noun phrase. After grammatical subjects and objects, a discourse marker can immediately follow the phrase in place of a grammatical case marker. This is illustrated with (11), where the sentence in (b) continues the

[4]For simplicity, henceforth, the allomorphic variations of *-(u)lo* are referred to as *-ulo*.

discourse started in (a). In (a), the subject *na* ('I') is mentioned initially in the discourse, and it is succeeded by the nominative case marker. In (b), the same subject referent is mentioned for the second time, and it is followed by the *-(n)un* marker, which indicates the topical status of the subject. The direct object *pen* in (b) is followed by the *-to* ('also') marker, which replaces the accusative case marker to indicate that the pen is included in the previous subject-predicate relation (i.e., I-give).

(11) a. Nay-**ka** Sohee-**eykey** chayk-**ul** cwu-ess-e.
I-NOM Sohee-DAT book-ACC give-PST-DEC
'I gave a book to Sohee.'

b. Na-**nun** Sohee-**eykey** pen-**to** cwu-ess-e.
I-TOP Sohee-DAT pen-also give-PST-DEC
'Speaking of myself, I gave also a pen to Sohee.'

While discourse markers appear in place of the nominative and accusative case markers, they appear after an overt grammatical case marker following indirect objects or obliques. In other words, a discourse marker is added agglutinatively to the dative and oblique case markers. This phenomenon is commonly known as the 'double marking,' and is illustrated in (12) and (13). In (12), the indirect object *Sohee* is initially succeeded by the dative *-eykey* and subsequently by the discourse marker *-man*. In (13), the similar pattern is observed with the oblique *pen*. The instrumental *-ulo* appears first, and then the discourse marker *-to* next. Thus, unlike the *-nun* discourse marker immediately following the subject *na* ('I') on its own, the discourse markers appear after the indirect and instrumental case marker, in (12) and (13) respectively.

(12) Na-**nun** Sohee-**eykey**-**man** chayk-**ul** cwu-ess-e.
I-TOP Sohee-DAT-only book-ACC give-PST-DEC
'Speaking of myself, I gave a book only to Sohee.'

(13) Na-**nun** pen-**ulo**-**to** pyeonji-**lul** s-ess-e.
I-TOP pen-INST-also letter-ACC write-PST-DEC
'Speaking of myself, I wrote the letter also with the pen.'

Comparing the syntagmatic patterns of direct objects and other types of objects (e.g., (11) vs. (12) and (13)) reveals an interesting pattern. For indirect objects and obliques, the dichotomy of grammatical

case markers and discourse markers is preserved, as there is a one-to-one relationship between the marker and the grammatical or discourse function it indicates. However, direct objects do not follow this pattern, as the accusative *-lul* can be omitted when its grammatical function is easily inferable and/or when a discourse marker appears to indicate its discourse function. In subsection 1.4.2, I illustrate how alternative marking patterns of direct objects further interact with the syntagmatic positions of different constituents.

1.3.2 Syntax

In Korean, subjects canonically take the sentence-initial position, and verbs canonically take the sentence-final position. When the verb is transitive, the canonical word order of the sentence is SOV. This is illustrated with (14). When there is an indirect object, it is canonically positioned before the direct object, as shown in (15).

(14) Nay-ka chayk-ul ilk-ess-ta.
I-NOM book-ACC read-PST-DEC
'I read a book.'

(15) Nay-ka Sohee-eykey chayk-ul cwu-ess-ta.
I-NOM Sohee-DAT book-ACC give-PST-DEC
'I gave a book to Sohee.'

Fundamental word orders in Korean interact with the discourse status of each phrase and its marking.[5] As an instance, topical phrases tend to be marked with *-nun* and appear in the beginning of the sentence (Choi, 1996). In (16), the *-nun*-marked object 'book' appears in the beginning of the sentence, while deviating from the canonical SOV order exemplified in (14). In (17), the indirect object *Sohee* is marked with *-nun*, and the phrase appears in the beginning of the sentence, deviating from the order illustrated in (15). In these cases, the *nun*-marked phrases are interpreted as topical information.

(16) Chayk-**un** nay-ka ilk-ess-ta.
book-TOP I-NOM read-PST-DEC
'Speaking of the book, I am the person who read it.'

[5]Detailed discussion of the interpretation of discourse notions will be presented in sections 2.3.

(17) Sohee-**eykey**-**nun** nay-ka chayk-ul cwu-ess-ta.
Sohee-DAT-TOP I-NOM book-ACC give-PST-DEC
'Speaking of Sohee (as a recipient), I gave a book (to her).'

Previously, we showed that grammatical case markers may be omitted after subjects and objects. In fact, entire phrases can be omitted when they are already established or contextualized previously in a discourse. In the example provided in (18), a question is directed at the addressee. In the subsequent answer, the addressee has the option to omit the subject phrase, as it is already established that the speakers are having a conversation about the addressee.

(18) Q. Neo-nun nwuku-eykey meo-l cwu-ess-e?
you-TOP who-DAT what-ACC give-PST-INT
'What did you give to whom?'

A. (Na-nun) Sohee-eykey chayk-ul cwu-ess-e.
(I-TOP) Sohee-DAT book-ACC give-PST-DEC
'(I) gave a book to Sohee.'

More than one phrase can be omitted, as long as they refer to contextualized information. In (20), the ensuing response allows for the omission of both the subject phrase and the dative phrase, because they have been contextualized in the question. Unmarked and omitted objects are observed more often in less formal conversational settings, compared to formal and/or non-colloquial Korean (Ko, 2001).

(19) Q. Neo-nun Sohee-eykey meo-l cwu-ess-e?
you-TOP Sohee-DAT what-ACC give-PST-INT
'What did you give to Sohee?'

A. (Na-nun) (Sohee-eykey) Chayk-ul cwu-ess-e.
(I-TOP) (Sohee-DAT) book-ACC give-PST-DEC
'(I, to Sohee) gave a book.'

In sum, these examples illustrate that speakers can selectively omit redundant or contextually evident elements in discourse, showcasing the interaction between morphological markings, sentence structures, and the discourse context in Korean.

1.4 Paradigmatic patterns in Korean

Thus far, three direct object marking options in Korean have been identified. The *-lul* and *-nun* markers can appear alternatively, or there may be no markers at all. In this dissertation, unmarked cases are referred to as 'null-marking' or a 'covert' marker, and are annotated with ∅ in examples. The three marking options share the same syntagmatic pattern, which is to follow a grammatical direct object. Thus, the syntagmatic pattern alone cannot fully explain what one marking option can achieve that other options cannot achieve, or when a particular marking option becomes more appropriate than the others. Given this limitation, this section focuses on paradigmatic contrasts between the three direct object marking options. Subsection 1.4.1 will describe the paradigmatic contrasts between marking options in a canonical word order in Korean, exemplified in (20). Following this, subsection 1.4.2 will explore how these paradigmatic contrasts interact with different word orders, by comparing marking patterns in canonical and non-canonical word orders, exemplified in (20) and (21).

(20) Nay-ka chayk-**{ul/un/∅}** cwu-ess-e.
I-NOM book give-PST-DEC

(21) Chayk-**{ul/un/∅}** nay-ka cwu-ess-e.
book I-NOM give-PST-DEC

Discourse functions of the *-lul*, *-nun*, and null-marking have been explained in terms of INFORMATION STRUCTURE (IS). One way to understand IS is to view it as the informational status of a linguistic element, determined by the knowledge states of both speakers and listeners regarding the discourse context and their awareness of each other's knowledge states (Lambrecht, 1994). Besides this definition of IS, there are a few other approaches to define IS (see Arnold, Kaiser, Kahn, & Kim, 2013). I adopt the conceptualization from Lambrecht (1994), as it effectively captures the nuanced yet distinctive informational statuses associated with Korean markers. These distinctions will be elaborated further in this section. In addition, IS can be assigned to a single constituent, to a phrase, or to an entire sentence. In order to explain alternative object markings in Korean, this dissertation focuses on the IS established for the referent of an object phrase.

In this dissertation, I assume that the knowledge states of interlocutors are twofold. A speaker

has a certain state of knowledge on contents of a discourse, such as entities mentioned in the discourse. This type of knowledge forms the speaker's knowledge about the discourse context. A listener also has specific knowledge about the discourse context which can be distinct from the speaker's knowledge state. For example, the speaker knows (or doesn't know) what the listener knows or doesn't know about the discourse context. Similarly, the listener possesses knowledge about the speaker's knowledge state concerning the discourse context. When both the speaker and the listener are aware of the same piece of information and each recognizes the other's awareness, this information becomes interpretable as SHARED KNOWLEDGE (Clark, 1996) between the interlocutors. In this chapter, I demonstrate that the distinct IS that each of the *-lul* and *-nun* markers indicates stems from the distinct state of shared knowledge between interlocutors.

Ultimately, by describing each marker's unique discourse function, it will be demonstrated that the dichotomy between grammatical and discourse markers is not as straightforward as previously thought. Instead, each marking option can simultaneously fulfill multiple functions in both the grammar and discourse domain. This phenomenon represents polyfunctionality—one syntagmatic pattern serving more than one function. In order to identify each marker's polyfunctionality, in the following, I will refrain from using conventional labels presented in Table 1.1 to refer to the markers. Instead, I refer to each marker with its form in Korean. In the glosses, I adopt the convention of annotating the marker with the function we want to focus on in the given example. For example, if the *-lul* marker is discussed outside of a specific discourse context and only indicates the grammatical object function, I gloss the marker as an accusative (ACC) marker. If the example is given to demonstrate discourse focus status indicated by the marker, I gloss the marker as a focus (FOC) marker.[6]

1.4.1 Morphology

In this subsection, I explore paradigmatic contrasts between markers appearing in the canonical word order (SOV) in Korean. To begin, the overt and covert accusative case marker has been observed as a Differential Object Marking (DOM) phenomenon (see Aissen, 2003; Bossong, 1991). This phenomenon

[6]Remaining agnostic with the marker's label has been adopted in previous literature that pinned down various functions of *-lul* (S.-N. Kwon & Zribi-Hertz, 2008) and *-nun* (I. Kim, 2015, 2016).

is often explained in terms of the semantic features of objects, such as animacy and definiteness. In Korean, however, the presence or absence of the *-lul* marker is explained by the differential discourse status of an object (S.-N. Kwon & Zribi-Hertz, 2008; H. Lee, 2007).

In particular, S.-N. Kwon and Zribi-Hertz (2008) examined a range of object features, including semantic and discourse factors, that could lead to alternative marking patterns in Korean. They demonstrated that semantic features alone cannot fully explain the alternation between null-marking and *-lul*-marking for objects. They argued that designating the discourse status of the object is also essential for a comprehensive explanation of this alternation. Previously, I demonstrated that the *-lul* marker may be omitted when the grammatical object function of a phrase can be easily inferred without the marker. However, when an object is a focused element in the discourse, the *-lul* marker accompanies the object to reflect the focus informational status (S.-N. Kwon & Zribi-Hertz, 2008).

FOCUS information is the part of an utterance that brings new information into the interlocutors' knowledge state. Thus, focus information concerns the changes that the speaker wishes to bring about in the knowledge state of the addressee via the pragmatics of language use. It is also identified as what fills in the gap between what is presupposed between speakers and what is newly asserted by speakers (Lambrecht, 1994).

Because wh-questions require information about a constituent to substitute for a wh-expression, answers to wh-questions are commonly diagnosed to have focus status (Lambrecht, 1994; Zimmermann & Onea, 2011). (22) is presented below as an example. In requesting information about the object, the question presupposes the content 'You gave X to Sohee.' This associates focus status with the object that fills in the gap in the response. Specifically, the 'book' substitutes for the wh-expression from the question, and updates the addressee's knowledge on the object that the speaker gave to Sohee. Thus, it represents the information status of focus.

(22) Q. Neo-nun Sohee-eykey meo-l cwu-ess-e?
you-NUN Sohee-DAT what-ACC give-PST-INT
'What did you give to Sohee?'

A. Na-nun Sohee-eykey chayk-{**ul/#**∅} cwu-ess-e.
I-NUN Sohee-DAT book-FOC give-PST-DEC
'I gave a BOOK to Sohee.'

In the response sentence in (22), omitting the *-lul* marker in the response results in the sentence becoming pragmatically infelicitous.[7] Conversely, when the *-lul* marker is included, the object is interpreted as focal information, ensuring the sentence remains felicitous. Therefore, the association between *-lul* and focus becomes evident when the *-lul*-marked phrase is paradigmatically contrasted with the null-marked phrase.

When one element is singled out from a set of alternatives and is conveyed as new information, it acquires the information status of CONTRASTIVE FOCUS (CF) (Lambrecht, 1994). Various terminologies have been employed to refer to focus information that includes an element of contrastiveness. As an example, S. Dik (1980) subdivided focus information into COMPLETIVE FOCUS, which involves no contrastiveness, and CONTRASTIVE FOCUS, which involves an explicit choice among alternatives. Similarly, Kiss (1998) distinguished INFORMATIONAL FOCUS, which delivers new information without any contrastiveness, and IDENTIFICATIONAL FOCUS, which expresses contrast between alternatives. Despite different labels being used, there appears to be a consensual intuition that there are two types of focus information, one with contrast and one without contrast.

The association between the *-lul* and CF information is demonstrated with (23). In (23), the disjunctive question presents two objects—'book' and 'pencil'—as the set of alternatives. This set of alternatives proposes options that the addressee can choose to form a response. In other words, the question introduces the set of alternatives, which needs to be assumed to convey contrastiveness. When the noun 'book' is picked out from the set and marked with *-lul* in the response, it is presented as new information: the 'book' is identified as the information that satisfies the force of the question. The *-lul*-marked 'book' in the response, therefore, illustrates CF information—focal information with contrastiveness. The null-marking in the response simply presents the 'book' as an expression that refers back to the referent previously mentioned in the question; it does not convey the CF status.

(23) Q. Neo-nun Sohee-eykey chayk-irang yeonpil-cwunge etten ke-l cwu-ess-e?
you-TOP Sohee-DAT book-or pencil-between which thing-ACC give-PST-INT
'Between the book and the pencil, which one did you give to Sohee?'

[7] I use the symbol # to indicate that the structure is syntactically well-formed but infelicitous in its discourse context. Small caps in examples indicate the focus status of a word or a phrase.

A. Na-nun Sohee-eykey chayk-**{ul/#∅}** cwu-ess-e.
I-TOP Sohee-DAT book-CF give-PST-DEC
'I gave the BOOK to Sohee.'

In (22) and (23), null-markings could not replace the *-lul* marker following the focalized object 'book'. Null-markings after the object appearing in the canonical SOV word order indicates topical status of the object. TOPIC information refers to the presupposed information between interlocutors, that is, part of the shared knowledge between interlocutors (Reinhart, 1981; Lambrecht, 1994; Krifka, 2008). Topic information status is often characterized in terms of ABOUTNESS (Kuno, 1972; S. C. Dik, 1978; Reinhart, 1981)—what the speaker intends to communicate about—and GIVENNESS (Prince, 1981; Gundel, 1985; Schwarzschild, 1999; Vallduvi, 1993)—the information available to the interlocutors from the discourse context. Givenness does not necessarily mean that the information must have been explicitly mentioned. Certain semantic relations between entities, such as part-of, subtype-of, member-of relations, or inferences made based on conversational scenarios can establish an entity as given information, even when it is not explicitly mentioned in the discourse (Prince, 1981).

To detect topic information, the interpretation that is semantically equivalent to 'as for' in English is often used (Kuno, 1972; Lambrecht, 1994). To interpret the sentence with the semantics of 'as for', the putative topic is put in the phrase-initial position, and its pronominal form is repeated afterwards. An example is provided in (24), where the topic status and its marking is designated by the 'as for' interpretation of the 'book' in English translation. In (24), the question suggests that the discourse is about 'X who gave a book to Sohee' and makes this information available to the addressee. Therefore, the 'book' is part of the presupposed information between interlocutors, and reflects topical status. As shown below, the response sentence can be paraphrased with the *as-for* construction with the 'book', only when the object is null-marked. When the *-lul* marker is present in (24), the topical status of 'book' is no longer available. This demonstrates that null-markings indicate topical status of the object.

(24) Q: Nu-ka Sohee-eykey chayk-ul cwu-ess-e?
who-NOM Sohee-DAT book-ACC give-PST-INT
'Who gave a book to Sohee?'

A: Nay-ka chayk-**{∅/#ul}** cwu-ess-e.
I-NOM book-TOP give-PST-DEC

'As for the book, I gave it to Sohee.'

Moving onto the *-nun* marker, when it appears after an object within a canonical SOV word order, the object is interpreted with contrastiveness established among other topical elements.[8] I identify the information status of topical elements with additional contrastiveness as CONTRASTIVE TOPIC (CT). Topic information is part of the shared knowledge between interlocutors, and CT status is established when an element is singled out from the set of alternatives that is part of the shared knowledge between interlocutors (Kuno, 1972; Heycock, 2008).

The association between the *-nun* marker and CT is exemplified in (25). In (25), the object 'book' *chayk* appears in its canonical position, following the subject and indirect object phrase. Different interpretations of the null-marked and *-nun*-marked object reflected in (25) (i) and (ii) suggest that we need to distinguish between different types of topics. When the object is null-marked and appears in the canonical SOV order, it is interpreted as topical information. However, the *-nun*-marked *chayk* 'book' conveys the interpretation semantically equivalent to the expression 'as for X between alternatives'. This expression captures the notion that an element is singled out from a presupposed set of alternatives.

(25) Q. Who gave a book and a pencil to Sohee?

A. Nay-ka Sohee-eykey chayk-**{∅/un}** cwu-ess-e.
I-FOC Sohee-DAT book-TOP/CT give-PST-DEC
(i) With ∅: 'As for the book, I gave it to Sohee.'

(ii) With *-nun*: 'As for the book between the book and the pencil, I gave the book to Sohee.'

The association between the *-nun* marker and CT has been observed with constructed examples (Choi, 1996), corpus data (I. Kim, 2015), and intonation patterns (C. Lee, 2006). In the corpus of a Korean radio show, only the 26% of *-nun* marker was used as topic marker, while more than half (53%) instances were associated with CT (I. Kim, 2015). A pitch analysis showed that there is a significant difference in pitch height between the topic-marking *-nun* (150 hz) and the CT-marking *-nun* (over 200 hz), further supporting the existence of multiple functions of the *-nun* marker (C. Lee, 2006).

[8]The *-nun* marker's association with grammatical objects and discourse topic status is observed within a non-canonical OSV word order. This is discussed in detail in the next subsection 1.4.2.

Up to this point, I have identified CF and CT, both of which incorporate the notion of contrastiveness. The two IS statuses are distinguished in terms of the functions that they do in relation to the interlocutors shared knowledge state. CF information, as an instance of focus information, updates interlocutors' shared knowledge state. On the other hand, CT information refers back to the presupposed alternatives. The focal aspect of CF distinguishes CF from topic and CT information. Vice versa, the topical aspect of CT distinguishes CT from focus and CF information.

The difference between CT and focus is well-captured in the account based on QUD (Question Under Discussion) (Büring, 2003). QUD is defined based on the Stalnaker's definition of COMMON GROUND. Stalnaker (2002) defines common ground as the information that interlocutors take for granted, and CONTEXT SET as intersection of the set of propositions that interlocutors take for granted. Using these two notions, Roberts (2006) suggests a QUD model where the question *what is the way things are?* is the ultimate goal that interlocutors strive to answer. Under this model, interlocutors design the discourse to gradually narrow down their common ground to a context set, and ultimately the context set contains relevant propositions as a response to the QUD. Within this framework, contrastive meaning is established through a process wherein interlocutors select alternatives from a set of alternatives, ultimately identifying an alternative that addresses the QUD.

Büring (2003) suggests that CT is established when there exists a set of questions underlying the QUD, and a speaker chooses to respond to a subset of the underlying set of questions. This QUD-based account is used to explain the *-nun* marker's association with CT information (C. Lee, 2003). I illustrated this account in (26). In (26), the context assumes the set of alternatives {book, pencil}, thus the question raises a set of implicit sub-questions listed in parentheses. CT status emerges in the answer of the QUD, when a speaker opts to address only one of the sub-questions under the QUD. In (26), the sub-question regarding only one object (i.e., 'Did you give a book to Sohee?') is answered, out of the two sub-questions under the main question. The 'book', therefore, achieves the CT status.

(26) Context: Sohee could receive a book and/or a pencil.

Q. Neo-nun Sohee-eykey meo-l cwu-ess-e?
you-TOP Sohee-DAT what-ACC give-PST-INT
'What did you give to Sohee?'

(SubQ: Did you give a book to Sohee?)

(SubQ: Did you give a pencil to Sohee?)

A. Na-nun Sohee-eykey chayk-**un** cwu-ess-e.
I-TOP Sohee-DAT book-CT give-PST-DEC
'As for the book between the two, I gave it to Sohee.'

According to the QUD-based account, focus information satisfies the entire information required in the QUD, whereas CT information responds to the part(s) of the QUD, that is, the sub-question(s) under the QUD (Büring, 2003; Roberts, 2012). When responding to the wh-question with a *-lul*-marked object that indicates focus status, the entire information inquired from the wh-question is considered to be fulfilled. However, when the wh-question is responded with a *-nun*-marked object indicating CT status, only the sub-question under the wh-question is considered being responded.

Based on this difference between focus and CT, CF-marking *-lul* and CT-marking *-nun* are contrasted in the example (27) below. In (27), the question introduces the two alternatives {book, pencil} so that this set becomes the shared knowledge between interlocutors. If 'book' is marked with the *-lul* marker, it is interpreted as CF information, which provides information to respond to the wh-question. Accordingly, the *-lul*-marked 'book' is interpreted with focalization, as shown in (i), and no other sub-questions are assumed to exist. On the other hand, if 'book' is marked with the *-nun* marker, it is interpreted as CT information with the interpretation presented in (ii). The *-nun*-marked 'book' is presented as a CT element that refers back to one of the presupposed alternatives, and only the part of the main question ('Did you give the book to Sohee?) is answered.

(27) Q. Neo-nun Sohee-eykey chayk-irang yeonpil-cwunge etten ket-l cwu-ess-e?
you-TOP Sohee-DAT book-and pencil-between which thing-ACC give-PST-INT
'Between the book and the pencil, which one did you give to Sohee?'

A. Na-nun Sohee-eykey chayk-**{ul/un}** cwu-ess-e.
I-TOP Sohee-DAT book-CF/CT give-PST-DEC
(i) With *-lul*: 'As for myself, I gave a BOOK to Sohee.'

(ii) With *-nun*: 'As for myself, as for the book between the two, I gave it to Sohee.'

The paradigmatics introduced thus far showed discourse functions of the *-lul*, *-nun*, and null-

Table 1.2: Paradigmatic contrasts between *-lul*, *-nun*, and null-marking in SOV word order in Korean.

NP-Marker	Grammatical function	NP-Marker	Grammatical function	Discourse function	Predicate
NP-*i/ka*	Subject	NP-{ *lul* / *nun* / ∅ }	Object	Focus/CF CT Topic	VP

marking appearing in the SOV word order. Alternations between the three marking options could not be fully explained in terms of their syntagmatic patterns, because all marking options share the same pattern of following a grammatical object. Paradigmatic contrasts between markers revealed the discourse functions of each marking option. This is summarized in Table 1.2. The *-lul* marker's function is determined as a grammatical case marker or to have additional discourse function, based on the marker's interaction with discourse context. When the marker fulfills additional discourse function, it indicates that a noun phrase is a grammatical object and also a (contrastive) focal element in the discourse. When the *-nun* marker appears after an object, it indicates CT status of the referent as well as the grammatical object function of the referent. The two markers' discourse functions are seen with the contrast with null-markings. When noun phrases are null-marked, it is indicated as topical information, which is part of information presupposed between interlocutors. In summary, the *-lul* and *-nun* markers are polyfunctional in the domain of grammar and discourse, as we observed from the paradigmatic contrasts with the null-marking.

In the SOV word order discussed so far, *-nun* marker is associated with the CT status of object phrases. However, as previously seen with sentence-initial appearances of *-nun* marker in subsection 1.3.2, *-nun* marker is another indication of topical status in Korean. This suggests that the functions of *-nun* marker must be considered in different word orders. Accordingly, in the next subsection, I will discuss how the paradigmatic contrasts of morphology interact with syntax. I will demonstrate that the *-nun* marker's discourse function is determined based not only on the paradigmatic contrasts with other marking options, but also on the interaction with word order.

1.4.2 Interaction between morphology and syntax

The *-nun* marker is associated with different informational statuses when it appears in different word order. Within each sentence in (28) and (29) below, the *-nun* marker is paradigmatically contrasted with the null-marking. The *-nun* marker in (28) is further contrasted with the *-nun* marker in (29), in terms of the syntagmatic position where it appears.

(28) Nay-ka chayk-**{un/∅}** cwu-ess-e.
I-NOM book give-PST-DEC

(29) Chayk-**{un/∅}** nay-ka cwu-ess-e.
book I-NOM give-PST-DEC

This subsection will discuss different discourse functions of the *-nun* marker appearing in the canonical SOV word order (e.g., (28)) and non-canonical OSV word order (e.g., (29)). In the following, I distinguish these two instances of the *-nun* marker, where it is used interchangeably with the null-marking and where it is not, and associate each instance with a distinct type of topic information.

For *-nun* to be used interchangeably with the null-marking and to be interpreted as an indicator of topical status, the *-nun*-marked phrase must appear before phrases with any other IS (Choi, 1996). This constraint on pattern of noun distributions means that subjects, which canonically appear in the phrase-initial position, can be either *-nun*-marked or null-marked and indicate topical status. However, for a *-nun*-marked object to be interpreted as topical, it will need to appear before constituents with other informational statuses.[9] In cases when there is a focal subject, the *-nun*-marked object must appear sentence-initially in a non-canonical OSV word order, in order to be interpreted as topical. I illustrate this in (30), with the subject 'I' in the response conveying the focus information. As previously discussed, the object phrase 'book' in the response can appear without any marker in (i) as topical information. When it appears with the *-nun* marker as topical information, the object phrase must appear in the sentence-initial position, before the focal subject 'I'. This results in the *-nun*-marked object appearing in non-canonical word order illustrated in (ii).

[9]When a subject phrase is elided, an object phrase appears sentence-initially and still considered appearing in the canonical word order.

(30) Q. Who gave a book to Sohee?

A. (i) Nay-ka Sohee-eykey chayk-**{#un/∅}** cwu-ess-e.
I-NOM Sohee-DAT book-TOP give-PST-DEC

(ii) Chayk-**{un/∅}** nay-ka Sohee-eykey cwu-ess-e.
book-TOP I-NOM Sohee-DAT give-PST-DEC
'**As for** the book, I gave it to Sohee.'

From these patterns, we observe that the *-nun* marker's multiple functions are determined in terms of their context of use. When the *-nun* marker appears in place of nominative *-i/ka* or accusative *-lul*, it fulfills the grammatical function of the grammatical case marker it replaces. In addition, the *-nun* marker fulfills discourse functions based on its interaction with syntactic word order. We observed distinct discourse functions of *-nun* by contrasting the marker with the null-marking option. When the *-nun* marker appears after an object in the non-canonical word order, it indicates topical status of the object as well as the grammatical object status. When the *-nun* marker appears after an object within the canonical word order, the marker indicates additional contrastiveness, which could not be indicated by the null-marking option. In summary, from paradigmatic contrasts between the *-nun* and null-marking options, interacting with the canonical and non-canonical word orders, the *-nun* marker exhibits polyfunctionality in the domain of grammar and discourse.

Unlike the *-nun* marker's discourse functions, the *-lul* marker's focus- and CF-marking functions remain consistent in the OSV word order. This is illustrated with (31).[10] In (31), two roommates discuss household chores, and the set of alternatives {cleaning, cooking} is assumed as part of the presupposed and shared knowledge between interlocutors.

[10]The example from Lambrecht (1994) is adapted to Korean. The original example was provided to illustrate the difference between Japanese topic-marking *-wa* (roughly equivalent to Korean topic-marking *-nun*) and CF-marking *-ga* (equivalent to Korean nominative *-i/ka*) following subjects. The original example is provided in (i). Note that the *-wa* marker in (i) is used for topical subjects, not objects. Small caps are from the author as indications of contrasted elements.

(i) Roommates Hanako and Mary discussing household chores:

a. H: Mary-san, anata-wa osoji shite kudasai, watashi-wa oryori shimasu kara.
Mary-VOC you-TOP cleaning do please I-TOP cooking do CONJ
"Mary, YOU do the CLEANING, I'll do the COOKING."

b. M: Ie, watashi-ga oryori shimasu kara; anata-wa hoka-no koto shite kudasai.
no I-NOM cooking do CONJ you-TOP other thing do please
"No, I'll do the cooking; YOU do something ELSE."

(31) Roommates Hana and Mary are discussing household chores, cleaning and cooking.

H: Mary-ya, chengso-**{nun/#lul}** ney-ka hay-cwue, yoli-**{nun/#lul}** nay-ka ha-lkey.
Mary-VOC cleaning-TOP you-NOM do-please, cooking-TOP I-NOM do-FUT
'Mary, as for the cleaning, you could please do it. As for the cooking, I will do it.'

M: Aniya, nay-ka yoli-**{lul/#nun}** ha-lkey; ney-ka chengso-**{lul/#nun}** hay-cwe.
no I-NOM cooking-CF do-FUT; you-NOM cleaning-CF do-please
'No, I'll do the COOKING; you do the CLEANING.'

In the first, Hana's utterance, the *-nun* marker follows each of the two alternatives in the set, and both *-nun*-marked phrases are presented as topical information compatible with 'as for' interpretations. In the following, Mary's utterance, the *-lul* marker follows objects to present each object as newly asserted information that fills in the gap in 'I will do X' and 'You will do Y'. In addition, because each object is singled out from the set, it is contrasted with one another. If Hana's utterance had the *-lul* marker in place of *-nun*, the alternatives could not be associated with topic status. Conversely, if Mary's utterance had the *-nun* marker in place of the *-lul* marker, the objects could not have CF status.

Thus far, the discourse functions associated with the *-lul*, *-nun*, and null-marking in both canonical and non-canonical word orders have been discussed. This is summarized in Table 1.3. Analyzing marking options within the SOV word order alone does not fully account for the different discourse functions of *-nun*. Contrasting *-nun* with null-marking in both canonical and non-canonical word orders reveals that the marker's discourse function is identified not only through contrasts with other marking options but also through contrasts with syntactic word order. In summary, the paradigmatic contrasts of each marking option and word order illustrate each marker's polyfunctionality, fulfilling both grammatical and discourse functions defined within its syntactic environment.

Table 1.3: Paradigmatic contrasts of *-lul*, *-nun*, and null-marking and their interaction with word order.

NP-Marker	Discourse function	NP-Marker	Discourse function	Predicate
SUBJ-*i/ka*	Focus	OBJ-{ *lul* / *nun* / ∅ }	Focus/CF CT Topic	VERB
OBJ-{ *lul* / *nun* / ∅ }	Focus/CF Topic Topic	SUBJ-*i/ka*	Focus/CF	VERB

1.5 Exhaustivity implicature

Considering paradigmatics of the *-lul* and *-nun* in canonical word order, I demonstrated that both markers can indicate contrastiveness. The *-lul* marker indicates contrastiveness for focus information, while the *-nun* marker indicates contrastiveness for topic information. An example illustrating CF and CT is repeated in (32).

(32) Q. Between the book and the pencil, which one did you give to Sohee?

A. Nay-ka chayk-**{ul/un}** cwu-ess-e.
I-TOP book-CF/CT give-PST-DEC
'I gave a book to Sohee.' (→ I didn't give anything else.)

In (32), the *-lul* marker indicates the object 'book' as new information contrasted with alternative 'pencil'. On the other hand, the *-nun* marker indicates 'book' as part of presupposed information contrasted with another presupposed alternative 'pencil'. However, the *-lul* and *-nun* markers in relation to contrastiveness need further discernment. As exemplified in the parentheses in (32), both CF-marking *-lul* and CT-marking *-nun* lead to an exhaustive listing interpretation, where an alternative marked with *-lul* and *-nun* is interpreted as exhaustive information responding to a given question. In this section, I explore differences between the two markers' exhaustive listing interpretation.

In discerning each marker's exhaustive listing interpretation, I will focus on two aspects. First, I will review previous proposals concerning the construction of exhaustive interpretation. One proposal suggests the necessity of contrastiveness for licensing exhausativity, and the other proposal suggests that in the context of wh-questions, contrastiveness is not necessary with focus marked *-lul*. The competing hypotheses will be shown to reflect different perspectives on an empirical issue which motivates the experimental investigation in Chapter 2.

The second aspect is about the nature of the discourse constraints associated with the markers' exhaustivity reading. This concerns whether an exhaustivity reading is cancelable. Exhaustivity is cancelable if the exhaustivity status established in an utterance can be updated or corrected (i.e., canceled) in the following discourse. For instance, if the response sentence in (32) can be felicitously followed by 'I also gave a pencil to Sohee,' the exhaustivity established in the initial response is cancelable. Conversely,

exhaustivity is non-cancelable, if an attempt to cancel the exhaustive status results in a pragmatically infelicitous or logically false utterance. If continuing the response sentence in (32) with 'I also gave a pencil to Sohee' is perceived as infelicitous, the exhaustivity established by the initial response is non-cancelable. Different accounts concerning the cancelability of the *-nun* marker's exhaustivity is another pattern that motivates the experimental investigation in Chapter 3.

1.5.1 Construction of exhaustivity implicature

Exhaustive listing interpretation is commonly explained as a specific type of QUANTITY IMPLICATURE (Gazdar, 1979; Horn, 1984), which is derived from the MAXIM OF QUANTITY—the speakers' and listeners' shared expectation that a sufficient amount of information must be provided in the utterance (Grice, 1989).[11] Based on the Maxim of Quantity, a listener infers that the speaker has said as much as she needs to say with the given utterance and that it provides all the information relevant in the context. This leads the listener to assume that there is no more informative utterance than the current utterance that the speaker conveyed. In other words, exhaustivity refers to a much stronger interpretation effectively inferring that what is not said must not be true or relevant.

For example, in (33), the conversational context states that the set of alternatives {pizza, cake} is part of the shared knowledge between speakers. Based on the Maxim of Quantity, Bob is expected to provide an exhaustive list of items from the set of alternatives that they know that Sue brought. This expectation is also shared between the two speakers. When Bob picks out one alternative from the set to respond in (33), Bob's response is taken as an answer that B provided based on the Quantity maxim (that B must have provided as much information as possible). In other words, because the other alternative from the set of alternatives is not mentioned, it is implied that the unmentioned-alternative must not be the relevant answer to the question. The exhaustive status of the object is thus established, and the denial of the alternatives is implied (noted in the parentheses in (33)).

(33) Alice, Bob and Sue were invited to attend a potluck dinner. They were asked to bring a pizza and a cake. Alice didn't attend the dinner, but Bob did.

[11] More comprehensive overview of the quantity implicatures and Maxim of Quantity will be provided in Chapter 3.

Alice: What did Sue bring?

Bob: Sue brought a pizza. (→ Sue didn't bring anything else.)

It is evident that exhaustivity implicature is constructed when there are more than one alternative to be contrasted. However, there are different hypotheses about how these alternatives become available to create exhaustivity implicature. PRAGMATICS-BASED HYPOTHESES suggest that necessary context to convey exhaustivity is provided from discourse context and interlocutors' interaction with discourse context. On the other hand, SEMANTICS-BASED HYPOTHESES suggest that multidimensional denotations of a linguistic structure express the existence of alternatives, which eventually licenses exhaustivity implicature. In the following, I review the two hypotheses for each of the *-lul* and *-nun* markers.

In relation to the *-lul* marker, there are two hypotheses under the pragmatic-based hypotheses. One account emphasizes the role of discourse context, and suggests that more than one alternative must be contextualized in the discourse to license exhaustivity (Kiss, 1998; C. Lee, 2003). I refer to this hypothesis as CONTEXT-BASED HYPOTHESIS. Context-based hypothesis suggests that the marker's exhaustive interpretation is based on the marker's indication of contrastiveness between alternatives that are explicitly identified in the context. Under this hypothesis, a contextualized set of alternatives must be available for the *-lul* marker to convey exhaustivity. In other words, the *-lul* marker is hypothesized to convey exhaustivity when it indicates CF status in the discourse context where the set of alternatives is established as shared knowledge, exemplified in (33) above.

An alternative account under the pragmatics-based hypothesis views exhaustivity implicature as the result of discourse updating process between interlocutors (Kuno, 1972; Horn, 1981; Zimmermann & Onea, 2011). I refer to this account as EXPECTATION-BASED HYPOTHESIS. Expectation-based hypothesis suggests that having contextualized alternatives is not the necessary condition for conveying exhaustive interpretation, because interlocutors update their expectations based on the discourse context and assume that relevant alternatives must exist.

This expectation-based hypothesis is illustrated with (34), where the wh-question starts the discourse and there are no explicit mentions of alternatives. With the wh-question, the speaker asks about the object that a person can perform the action of giving to Sohee. In this discourse context, alternatives

in the set are not clearly listed.

(34) Q. Neo-nun Sohee-eykey meo-l cwu-ess-e?
you-TOP Sohee-DAT what-ACC give-PST-INT
'What did you give to Sohee?'

A. Na-nun Sohee-eykey chayk-{**ul/#∅**} cwu-ess-e.
I-TOP Sohee-DAT book-FOC give-PST-DEC
'I gave the BOOK to Sohee.' (→ I didn't give anything else to Sohee.)

However, once the question is proposed in (34), the set of alternatives are implicitly delimited to include objects that are reasonable for the addressee to have given to Sohee. Interlocutors update the shared knowledge to include this implicit set of alternatives. As a consequence, the *-lul* marker in the answer sentence is pragmatically preferred in the context of (34) in order to indicate the 'book' as new information that is singled out from some implicit set of alternatives. Thus, under the expectation-based hypothesis, in addition to CF-marking *-lul*, focus-marking *-lul* appearing in the context without contextualized alternatives can convey exhaustivity.

Contrary to the pragmatics-based hypotheses (including both context-based and expectation-based hypotheses), semantics-based hypothesis of focus information argues that a set of alternatives becomes available from multidimensional semantic denotations of focus-marking structures. This has been further elaborated under the label of ALTERNATIVE SEMANTICS (Rooth, 1985).[12] Rooth distinguish between the ordinary semantic value (indicated with $[[\,]]^{O}$) and the focus semantic value (indicated with $[[\,]]^{F}$). These two are distinguished in (35). In (35) (i), the ordinary semantic value of the focus-marked 'book' is simply the referent of the expression. The focus semantic value is the set containing alternatives, that is, all semantic objects of the same semantic type that can presumably replace the 'book.'

(35) A: What did you give to Sohee?

B: I gave a **book** to Sohee.

(i) $[[\text{book}_{F}]]^{O} = \text{book}$

[12]Rooth (1985) and Rooth (1992) are representative literature for alternative semantics, but Rooth (1992) has modified view on the requirement of the contextualized set of alternatives. Rooth (1985) suggests that a set of alternatives does not have to be available from the discourse context, but Rooth (1992) suggest that the discourse context refines a set of alternatives so that the set includes context-relevant alternatives. Semantics-based hypothesis in this chapter is based on the account of Rooth (1985).

(ii) $[[\text{book}_F]]^F = \{\text{book, pencil, flowers...}\}$

Ko (2001) described the *-lul* marker following this account, and defined the *-lul* marker as a focus marker that semantically denotes a set of alternatives. This is illustrated with (36). Following this account, even when the question does not specify alternatives, the *-lul*-marked object 'book' can indicate CF. Not because the listener updates the knowledge state based on the discourse expectations of the given question, but because the *-lul* marker itself informs the listener that some relevant alternatives exist.

(36) A: What did you give to Sohee?

B: Na-nun Sohee-eykey chayk-**ul** cwu-ess-e.
I-TOP Sohee-DAT book-FOC give-PST-DEC
'I gave a book to Sohee' (→ I didn't give anything else.)

(i) $[[chayk_F]]^O$ = book

(ii) $[[chayk_F]]^F = \{\text{book, pencil, paper, pen...}\}$

Moving on to the *-nun* marker's construction of exhaustivity implicature, previously introduced QUD-based account of CT and *-nun* suggests that the marker conveys exhaustivity only when relevant sub-questions exist under the QUD (Büring, 2003; C. Lee, 2003). This is the pragmatics-based hypothesis of *-nun*, suggesting that the marker conveys exhaustivity only when a set of alternatives is identifiable in the discourse context. Unlike the *-lul* marker, whose pragmatics-based hypotheses disagree on whether contextualized alternatives are not required, the *-nun* marker's pragmatics-based hypothesis suggests that a contextualized set of alternatives is required to license exhaustivity. Thus, under the pragmatics-based hypothesis, the *-nun* marker is distinguished from *-lul* in that each and every alternative in the set must be identifiable by the interlocutors to convey exhaustivity.

Previously introduced QUD-based account suggests that CT status is established when there is some unanswered portion of the QUD. In (37), A's question comprises a list of sub-questions, based on the conversational context. In B's response, only one sub-question ('Did Sue bring a pizza?'), rather than the entire QUD, is addressed with the *-nun* marker. The CT status of the object 'pizza' is thus established. Therefore, if the context doesn't specify that there is more than one alternative being discussed, there would be no sub-questions left unanswered. As a result, CT status and exhaustivity implicature cannot be

established.

(37) Alice, Bob and Sue had planned to attend a potluck dinner. Guests were asked to bring a pizza and a cake. Alice didn't attend the dinner, but Bob did.

A: What did Sue bring?

(SubQ: Did Sue bring a pizza?)

(SubQ: Did Sue bring a cake?)

B: As for the pizza between the two, Sue brought it. (→ Sue didn't bring anything else.)

Once CT status is established by the underlying sub-questions and the *-nun* marker, exhaustivity implicature is constructed following the same process where quantity implicatures are inferred. If the speaker wanted to convey that 'Sue brought both the pizza and the cake,' they could have answered to all portions of the QUD. Because the part of QUD ('Did Sue bring a cake?') is not answered by B, B's answer is taken as an exhaustive response to the main question and results in the denial of the unanswered subquestion about the 'cake' (i.e., 'Sue didn't bring a anything else.'). Thus, when only one sub-question is addressed, it implies that the unanswered sub-question(s) must not be relevant under the QUD, and licenses exhauastivity implicature.

Previously mentioned semantics-based hypothesis for the *-lul* marker suggested that the semantic function of the marker denotes a set of alternatives, thus the marker doesn't require contextualized alternatives to convey exhaustivity. Semantics-based hypothesis of *-nun* identifies the marker as a semantic quantifier 'not all' ($\exists x[x \neq \alpha \wedge \neg\beta(x)]$), which semantically presupposes the existence of contextualized alternatives. Unlike the *-lul* marker under semantics-based hypothesis (e.g., (36)), the *-nun* marker does not directly enumerate relevant alternatives but presupposes only the existence of alternatives (J. Kim, 2018). Accordingly, discourse context needs to identify alternatives that can be used to fulfill the *-nun* marker's semantic meaning of 'not all.' This account is illustrated with (38). The *-nun*-marked 'book' is likely to cause the speaker A to follow up for a clarification on the status of the set, because the marker presupposes the existence of relevant alternatives.

(38) A: Neo-nun Sohee-eykey meo-l cwu-ess-e?
you-TOP Sohee-DAT what-ACC give-PST-INT
'What did you give to Sohee?'

B: Na-nun Sohee-eykey chayk-**#un** cwu-ess-e.
I-TOP Sohee-DAT book-CT give-PST-DEC
'As for the book between some unknown things, I gave it to Sohee.'

A: What do you mean by *chayk-un*? Was there anything else that you could have given to Sohee?

Thus far, I introduced different proposals on how the *-lul* and *-nun* markers construct exhaustivity implicatures. This is summarized in Table 1.4. Under the pragmatics-based hypothesis of *-lul*, context-based hypothesis suggests that a set of alternatives must be contextualized in a discourse context (C. Lee, 2003). Under the pragmatics-based hypothesis of *-lul*, expectation-based hypothesis suggests that the *-lul* marker can convey exhaustivity with or without explicitly stated set of alternatives (Kuno, 1972; Horn, 1984). The semantics-based hypothesis of the *-lul* marker suggests that alternatives are not required in the discourse context, because the marker's semantic denotation can contextualize alternatives. For the *-nun* marker, both pragmatics-based and semantics-based hypothesis suggests that a set of alternatives must be provided in the discourse context to license exhaustivity. From this discussion, it is evident that there is a disagreement on how the *-lul* marker licenses exhaustivity implicature. The conflicting hypothesis summarized in Table 1.4 will be experimentally attested in Chapter 2.

Table 1.4: Hypotheses on whether a set of alternatives must be contextualized in the discourse to license exhaustivity.

	Contextualized alternatives	
	-lul	*-nun*
Pragmatics-based	Required (Context-based) Not required (Expectation-based)	Required
Semantics-based	Not required	Required

1.5.2 Discourse constraints of exhaustivity implicature

Now we move on to contrasting how the cancelability of the *-lul* and *-nun* marker's exhaustivity are previously discussed. Under the pragmatics-based hypothesis of *-lul*, exhaustivity implicatures

established from either CF information (context-based hypothesis, C. Lee, 2003) or focus information (expectation-based hypothesis, Kuno, 1972; Horn, 1981) are both cancelable. The *-lul* marker's exhaustivity implicature established with CF information and its cancelability are illustrated in (39). Considering B2 as a follow-up of the preceding response B1, B2 successfully accomplishes the cancellation of the exhaustivity. Thus, B's first and second responses together (i.e., B1 and B2) form a coherent and felicitous response to A's question. Alternatively, B2′ can follow B1, and the two responses together (i.e., B1 and B2′) create another felicitous response to A's question.

(39) Context: Alice, Bob and Sue were invited to attend a potluck dinner. They were asked to bring a pizza and a cake. Alice didn't attend the dinner, but Bob did.

A: Sue-nun mwe-l kacyeka-ss-eo?
Sue-TOP what-ACC bring-PST-INT
'What did Sue bring?'

B1: Pizza-**lul** kacyew-ass-e.
pizza-CF bring-PST-DEC
'Sue brought exactly the pizza.' (→ Sue didn't bring anything else.)

B2: Cake-to kacyew-ass-e.
cake-also bring-PST-DEC
'Sue also brought the cake. (→ ~~Sue didn't bring anything else.~~)'

B2′: Cake-un ahn kacyew-ass-e.
cake-CT NEG bring-PST-DEC
'Sue didn't bring the cake.'

Under the expectation-based account that does not consider contextualized alternatives as necessary to convey exhaustivity (Kuno, 1972; Horn, 1981), the *-lul* marker in (39) B1 indicates the focus status of 'pizza' and an exhaustive interpretation, even when the given context of 'pizza' is not presumed. Still, this account suggests cancelability, and both the B2 and B2′ are valid continuations of B1.

The semantics-based hypothesis of the *-lul* marker's exhaustivity (Ko, 2001) does not make any specific claims about the cancelability of the *-lul* marker's exhaustivity. According to the alternative semantics account, a set of alternatives, rather than the exhaustivity implicature itself, is part of the marker's semantic denotation (Rooth, 1985). After a set of alternatives is introduced as semantic denotation, exhaustivity is conveyed as a pragmatic implicature. Thus, whether the exhaustivity can be updated or not in

the proceeding discourse is still considered an issue of cancelability of pragmatic implicature under this account, although no specific claims are made about it.

In relation to the *-nun* marker, I introduced a pragmatics-based account based on QUD (Büring, 2003; C. Lee, 2003). Both pragmatics-based accounts Büring (2003) and C. Lee (2003) explain the relation between CT and exhaustivity implicature based on the process of updating common ground and context set (Stalnaker, 2002). However, the two accounts suggest different cancelability for the exhaustivity established by CT information. Büring (2003) suggests that CT information licenses cancelable exhaustivity implicature, while C. Lee (2003) suggests that CT information, specifically indicated by Korean *-nun*, conveys non-cancelable exhaustivity implicature.

These two different accounts are illustrated in (40). According to Büring (2003), B can follow up an initial response B1 with an utterance B2 to update the exhaustive status previously established with CT-marking *-nun*. This account is shown with the canceled exhaustivity in (i) under B2. According to (C. Lee, 2003), B's second response in B2 is infelicitous as a follow-up to the first response B1, as B2 states a non-exhaustive status of 'pizza' that is incoherent with what is implicated in B1. This account is reflected in (ii) under B2. A felicitous continuation to B1 is given in B2′. When the *-nun*-marked object is followed by a statement B2′ which maintains the exhaustive status of 'pizza', the two responses (B1 and B2′) make a felicitous response to A's question.

(40) Context: Alice, Bob and Sue had planned to attend a potluck dinner. Guests were asked to bring a pizza and a cake. Alice didn't attend the dinner, but Bob did.

A: What did Sue bring to the party?

B1: Pizza-**nun** kacyew-ass-e.
pizza-CT bring-PST-DEC
'As for the pizza between the two, Sue brought it.' (→ Sue didn't bring anything else.)

B2: Cake-to kacyew-ass-e.
cake-also bring-PST-DEC
(i) 'Sue also brought the cake. (→ ~~Sue didn't bring anything else.~~)' (Büring, 2003)

(ii) '#Sue also brought the cake.' (C. Lee, 2003)

(iii) '*Sue also brought the cake.' (J. Kim, 2018)

B2′: Cake-un ahn kacyew-ass-e.
cake-CT NEG bring-PST-DEC
'Sue didn't bring the cake.'

Semantics-based hypothesis of *-nun* suggests that exhaustivity conveyed by the *-nun* marker cannot be updated or corrected in the proceeding discourse (J. Kim, 2018). According to this account, the *-nun* marker semantically denotes the meaning 'not all,' which establishes exhaustive listing reading as a semantic meaning, not as a pragmatic implicature. Thus, canceling the exhaustivity of *-nun* brings a semantically contradicting discourse. This account is illustrated in (40) (iii).

This subsection discussed the *-lul* and *-nun* markers' exhaustivity implicatures in terms of their cancelability, summarized in Table 1.5. To summarize, different accounts propose different cancelability of the exhaustive interpretation conveyed with the *-nun*. Under the pragmatics-based approach, exhaustivity implicature from CT is observed as cancelable (Büring, 2003). However, the pragmatics-based and semantics-based hypotheses specifically made for the *-nun* marker suggest that the marker conveys non-cancelable exhaustivity implicature (C. Lee, 2003; J. Kim, 2018). The cancelability of these two markers is tested in the experimental setting in Chapter 3.

Table 1.5: Cancelability of *-lul* and *-nun* markers' exhaustivity implicature proposed by different hypothesis.

	Cancelability of exhaustivity implicature	
	-lul	*-nun*
Pragmatics-based	cancelable	non-cancelable (C. Lee, 2003) cancelable (Büring, 2003)
Semantics-based	not discussed	non-cancelable (J. Kim, 2018)

1.6 Interim summary

In this chapter, I introduced the morphological and syntactic organizations in Korean through syntagmatic and paradigmatic patterns. Paradigmatic contrasts between *-lul*, *-nun*, and null-marking options and their interactions with canonical and non-canonical word orders revealed the discourse status that each option indicates. I demonstrated focus- and CF-marking functions of *-lul*, topic- and CT-marking

functions of *-nun*, and topic-marking function of the null-marking. As a stepping stone towards understanding more complex interactions between morphology and syntax, the remainder of this dissertation focuses on investigating the paradigmatic contrasts between *-lul*, *-nun*, and null-marking appearing in the canonical SOV word order.

In addition to each marker's polyfunctionality, I discussed the *-lul* and *-nun* markers' exhaustive listing interpretation. I introduced various accounts that suggested different discourse behavioral patterns associated with each marker's exhaustive interpretation. In particular, I organized the two discourse usage aspects where the *-lul* and *-nun* markers have been previously described with. Each marker has been discussed in relation to whether a set of alternatives must be provided in the discourse context to convey exhaustive listing interpretation and whether the marker conveys cancelable or non-cancelable exhaustivity interpretation.

These two discourse usage aspects will be used to characterize each marker's association with an exhaustive interpretation. In order to provide the discussion, I report the the two psycholinguistic studies in Chapter 2 and 3, each of which aim to attest the debated patterns of *-lul* and *-nun* markers' exhaustivity interpretations. Chapter 2 will examine the requirement of contextualized alternatives in constructing an exhaustivity interpretation for the *-lul* and *-nun* markers. Chapter 3 will test the cancelability of the *-lul* and *-nun* markers' exhaustivity interpretation.

2 Experiment 1: Shared knowledge

2.1 Introduction

In Chapter 1, I discussed the *-lul* marker's association with focus information, and the *-nun* marker's association with topic information. In addition, when the two markers appear after a noun phrase that is singled out from a set of alternatives, the *-lul* marker indicates CONTRASTIVE FOCUS (CF) and the *-nun* marker indicates CONTRASTIVE TOPIC (CT) status. Despite each marker's association with either focus or topic IS, both markers' contrastiveness IS roles—CF and CT—are associated with exhaustive interpretation. Thus, the *-lul* and *-nun* markers' association with the same discourse interpretation suggests that we need more fine-grained comparison of their discourse functions in order to account for their alternative uses.

I introduced several accounts suggesting how the *-lul* and *-nun* marker's exhaustivity is associated with differences concerning the nature of the sharedness relation between a speaker and a listener. In relation to the *-lul* marker's exhaustive interpretation, on the one hand, it was suggested that focus-marking structure can convey exhaustivity in the context where a set of alternatives is not contextualized (Kuno, 1972; Horn, 1981; Rooth, 1985). That is, a set of alternatives does not need to be assumed as part of shared knowledge prior to the *-lul*-marked focal information being uttered. Instead, focal information is predicted to trigger pragmatic enrichment which leads to the inference of exhaustivity (Horn, 1981). Under this account, the *-lul* marker is predicted to convey exhaustivity when it indicates either focus or CF information of an object, and the availability of a set of alternatives is not a crucial factor that predicts exhaustivity of the *-lul* marker. In the current chapter, I refer to this account as a MAIN EFFECT MODEL.

On the other hand, it was suggested that a set of alternatives must exist as shared knowledge for

Table 2.1: Previous discussions on required status of a set.

	MAIN EFFECTS MODEL (A set of alternatives doesn't need to be available)	INTERACTION EFFECTS MODEL (A set of alternatives must be available)
-lul	Kuno (1972); Horn (1981); Rooth (1985)	Kiss (1998); Rooth (1992); C. Lee (2003)
-nun		Büring (2003); C. Lee (2003); J. Kim (2018)

the *-lul* marker to convey exhaustivity (Kiss, 1998; C. Lee, 2003; Rooth, 1992). This account predicts exhaustivity from the *-lul* marker only when alternatives under contrast are explicitly identified or inferable from the context. Under this account, an exhaustivity interpretation is conveyed by the interaction of the *-lul* marker's focus-marking discourse function and the shared knowledge about a set of alternatives. In this sense, I refer this account as an INTERACTION EFFECT MODEL. Under the INTERACTION EFFECT MODEL, the *-lul* marker is predicted to convey exhaustivity when it indicates CF status, but not when it indicates plain focus status without contrastiveness.

For the *-nun* marker, it is generally agreed that a set of alternatives must be part of shared knowledge to convey exhaustivity (Büring, 2003; C. Lee, 2003; J. Kim, 2018). This means that the *-nun* marker's exhaustivity is explained as an interaction of the marker's discourse function and the shared knowledge about a set of alternatives. Thus, the *-nun* marker's exhaustivity is predicted with the INTERACTION EFFECTS MODEL, but not with the MAIN EFFECTS MODEL, where the *-nun* marker is predicted to convey exhaustivity without a set of alternatives. In other words, the *-nun* marker is predicted to convey exhaustivity when it indicates CT status, but not when it indicates topic status without contrastiveness. These accounts are summarized in Table 2.1.[1]

In the current chapter, I aim to test how each of the *-lul* and *-nun* marker's exhaustive interpretation is influenced by the type of sharedness associated with a set of alternatives. I conduct an experiment to test the two models and examine whether each marker requires a set of alternatives as part of prior knowledge in order to convey exhaustivity. In the following, I first review previous studies confirming that shared knowledge about a set of alternatives is crucial in the processing of exhaustive information.

[1] In Chapter 1, I organized accounts listed in Table 2.1 into pragmatics-based and semantics-based accounts, and identified context-based hypothesis and expectation-based hypothesis under the pragmatics-based account (c.f., Table 1.4). Because the current chapter tests different requirements of the set of alternatives proposed by both pragmatics-based and semantics-based account, I use a different set of notions in this chapter. Main effects model and interaction effects model can either be pragmatics-based account or semantics-based account introduced in Chapter 1.

These results support the interaction between marker and shared knowledge. I then review studies that suggest evidence for the main effect model of the *-lul*. Based on the previous results, I design an experiment where alternatives in a set are identified or unidentified, and nominal objects are marked with either *-lul* or *-nun*. Native speakers of Korean reported the degree of exhaustiveness they perceive when each marker is used with different statuses attributed the set of alternatives. I report results in section 3.4 and discuss further implications in section 3.5. The chapter is concluded in section 3.6.

2.2 Previous studies

Previous studies have demonstrated that a set of alternatives is evoked and used to process information indicated with exhaustivity (Gotzner & Spalek, 2017; C. S. Kim, Gunlogson, Tanenhaus, & Runner, 2015; Washburn, Kaiser, & Zubizarreta, 2011). For example, Washburn et al. (2011) tested the significance of an associative relationship between words and modification of a word with English *only* in information processing. They incorporated the associative relationship and *only* modification as different manipulations of the 'prime word' in stimuli, and assessed experiment participants' processing of the 'target word' given different types of prime word. Specifically, prime words and target words were presented in stimuli written in a form of short narratives. An example of the narrative is illustrated in (1).

(1) a. Christina wants to buy a lock, nails, and a bolt.
b. She needs these to fix her front entrance.
c. Two days ago, she went to a store that didn't have a wide selection.
d. At the store, she was able to buy **[Prime Word: only nails]**.
Target Word: **lock**

In the beginning of the narrative, a sentence introduced a set of three common household items with different degrees of association. For instance, in (1) (a), the first item was the target word (e.g., *lock*), the second item (e.g., *nails*) was not associated with the target, and the third item was commonly associated with the target (e.g., *bolt*). The second sentence (b) assigned a common property to the set introduced in the first sentence, reinforcing the set membership of alternatives. The next sentence (c) was

included to "move the narrative along." The last sentence (d) contained the prime word (e.g., *only nails*) as the last word of the sentence. Prime words in the stimuli were manipulated for their modification with *only* so that they were either modified or unmodified, and for their relationship with a target word. This created six conditions listed in (2).

(2) At the store, she was able to buy...

(i) Focused, associated: *only a bolt*

(ii) Unfocused, associated: *a bolt*

(iii) Focused, unassociated: *only nails*

(iv) Unfocused, unassociated: *nails*

(v) Focused, unrelated: *only a lamp*

(vi) Unfocused unrelated: *a lamp*

The prime and target words had (i) an associated relationship, where the prime word was an associate of the target word (e.g., *bolt* if the target is *lock*); (ii) an unassociated relationship, where the prime was not associated with the target word but contextualized as one alternative in the set from the first sentence (e.g., *nails*); (iii) an unrelated relationship, where the prime is not associated with the target word and not contextualized in the set from the first sentence (e.g., *lamp*). In the experiment, a prime word in one of six conditions listed in (2) appeared in a stimulus.

After all components in a stimulus were presented, participants were shown a target word that might or might not have been associated with the prime word, and asked to judge whether the string of letters (in the target word) was a word or not. The time participants took to make the judgment was recorded as their reaction time, which was used to evaluate how the *only* marking on prime word and the association between prime and target words facilitated information processing.

The first key result was that the focused, unassociated condition (e.g., *only nails* in (2) (iii)) was significantly faster than the focused, unrelated condition (e.g., *only a lamp* in (2) (v)). Note that both unassociated and unrelated words did not have strong a priori semantic associations with the target word that would have been learned outside of the experimental setting. However, when unassociated and

unrelated words were marked with *only*, unassociated words—the word that has been contextualized as part of a set of alternatives—presented faster processing time. This indicates that participants actually utilized a contextualized set of alternatives to process *only*-marked information.

The second key result was that the focused, unassociated condition (e.g., *only nails* in (2) (iii)) reported significantly faster reaction times than the unfocused, unassociated condition (e.g., *nails* in (2) (iv)). This finding supports the idea that *only* marking facilitates the process of retrieving an unassociated item that is previously established as part of shared knowledge (e.g., *nails* are provided in (1) (a)). This result also supports the idea that explicitly given shared knowledge about a set of alternatives is being utilized for processing information associated with exhaustive listing. Overall, in Washburn et al. (2011), the role of contextualization is found significant in processing *only*-marked information. As this result demonstrates the interaction between the focus-marking structure and a set of alternatives, it supports the account of the INTERACTION EFFECT MODEL.

Gotzner and Spalek (2017) studied more than one focus-marking structure in German to investigate the connection between activation of a set of alternatives and exhaustive interpretation. They designed experimental stimuli that began with a pair consisting of a context sentence and a critical sentence. As illustrated in (3), the context sentence introduced a set of two alternatives (e.g., 'judge', 'witness').

(3) Context sentence: Der Richter und der Zeuge verfolgten die Beweisführung
'The judge and the witness followed the argument.'

The critical sentence was manipulated to present one alternative from the set with one out of four focus-marking structures described in (4). The structures included (a) H* accent known to signal focus information without contrastiveness, (b) L+H* accent known to trigger contrastive interpretation, (c) L+H* with a focus particle *nur* 'only', and (d) L+H* with a focus particle *auch* 'also'.

(4) Critical sentence:

a. Der [Richter]$_{F}$ glaubte dem Angeklagten (H*)

b. Der [RICHTER]$_{F}$ glaubte dem Angeklagten (L+H*)

c. Nur der [RICHTER]$_{F}$ glaubte dem Angeklagten (L+H*-only)

d. Auch der [RICHTER]$_F$ glaubte dem Angeklagten (L+H*-also)
'(Only/Also) the [JUDGE]$_F$/the [judge]$_F$ believed the defendant.'

After an exposure to a context and a critical sentence, participants were given with a distractor task (a simple math operations that was correct or incorrect, such as 6 + 2 - 7 = 1 or 3 - 1 + 8 = 11). These distractor tasks were implemented to exploit the working memory load. Following the distractor task, either the mentioned alternative (e.g., 'judge' from (4)) or a previously unmentioned word was shown, and participants were asked to answer whether a word had appeared in the previous context story or not. The time that participants took to respond to this question was measured as their reaction time.

The longest mean reaction time was reported from the condition that included L+H* accent and *nur* ('only') particle (i.e., (4) (c)). This suggests that when working memory resources were strained, the processing effort needed to activate alternatives increased with the 'only'. This result is comparable to the result seen in Washburn et al. (2011), where working memory resources were not exploited in the experimental design and the exclusive marking (English *only*) facilitated retrieving a set of alternatives. Taking the results from Washburn et al. (2011) and Gotzner and Spalek (2017), these results suggest that listeners' working memory resources are utilized to identify relevant alternatives when processing exhaustive listing information. That is, processing information that is associated with exhaustivity involves retrieving a set of alternatives from the shared knowledge. This confirms the interaction between exhaustivity-marking structures and shared knowledge about a set of alternatives, which is considered necessary to evoke exhaustivity under the INTERACTION EFFECT MODEL.

E. Wilcox and Spector (2019) demonstrated the effect of shared knowledge in interpreting the exhaustivity status of utterances. In their experiment, participants were shown a visual illustration of a moving scene presented in a form of a comic strip. In the scene, a character committed to moving whatever she can lift down to the curb. The character's utterance to a friend was also shown in a bubble, mentioning what she is capable of lifting. This utterance was presented in one among six format exemplified in (5).

(5) a. I can lift the chair and the footstool.
b. I can lift the footstool and the chair.

c. I can lift the chair but not the footstool.

d. I can lift the footstool but not the chair.

e. I can lift the chair.

f. I can lift the footstool.

Accompanying the utterance, the comic strip showed that a footstool is about half the size of a chair, informing what the utterances in (5) (e) and (f) would entail about the character's capability of lifting other objects. For instance, if the character could lift the chair that appeared bigger than the footstool, it suggested that the character was capable of lifting the smaller footstool as well. When the character tried to move "everything that she can" and moved one object that appeared harder to move than the other (because one appeared bigger than the other), it was plausible to reason that the character also moved the other object that appeared easier to move (i.e., the smaller object). Thus, this experimental setup provided shared knowledge about the relationship between the two objects and the likelihood of moving each object.

If the knowledge about the two objects (e.g., the chair is bigger than the footstool) is incorporated in interpreting an utterance elaborating the character's ability regarding only one object, the utterance stating only one object (5) (e) 'I moved the chair' is likely to be interpreted non-exhaustively, meaning that the character moved the footstool as well. As predicted, the result demonstrated the effect of shared knowledge on interpretations of exhaustive listings, such as (5) (e) and (f). When the shared knowledge was established that one object was easier to move than the other and a character moved everything that she can, utterances stating that the character moved the harder to move object were interpreted non-exhaustively with the interpretation that the character moved both objects.

This result provides evidence that listeners interpret exhaustive listing sentences based on their shared knowledge. This appears as direct support of the INTERACTION EFFECT MODEL, because it confirms that shared knowledge plays an important role in interpreting exhaustive listing sentences. However, generalizing the result also supports the MAIN EFFECT MODEL. When a set of alternatives is not available in discourse context but the utterance conveys an exhaustive listing, pragmatically competent listeners may still interpret the utterance contextually to derive the exhaustive interpretation. Because this study

shows that listeners are adept at pragmatically interpreting exhaustive listings, it supports the pragmatic interpretation process proposed by both the MAIN EFFECT MODEL and the INTERACTION EFFECT MODEL.

The MAIN EFFECT MODEL suggests that focus information can convey exhaustivity when a set of alternatives is not part of shared knowledge. Studies that investigated contrastive interpretation of focus-marking structure provide some suggestive evidence of the MAIN EFFECT MODEL. As Destruel and Velleman (2014) point out, *it*-clefts in English can indicate exhaustive interpretation, as shown in (6) (a), or contrastive interpretation, as shown in (6) (b).

(6) It was John who cooked the beans.

a. → Nobody other than John cooked the beans.

b. → The fact that John cooked the beans contrasts with something in the discourse context.

Although contrastive interpretation does not always entail exhaustivity, experimental results suggest that exhaustivity entails contrastiveness (Destruel & Velleman, 2014; Washburn, Kaiser, & Zubizarreta, 2019). That is, the exhaustive interpretation in (6) (a) always entails the contrastive interpretation in (b), but the contrastive interpretation in (b) does not entail the exhaustive interpretation in (a). Given this close association between exhaustivity and contrastiveness, experimental studies about contrastiveness and contextual alternatives provide some insights on the relationship between exhaustivity and contextual alternatives.

For example, Washburn et al. (2019) investigated whether explicit mention of the set of alternatives affects the interpretation of English *it*-clefts, the syntactic structure of focus information in English. Their experiment included three manipulations in total, all of which appeared before the *it*-cleft was presented to participants. One of the manipulation was about the contrastiveness of exhaustive *it*-clefts. (7) illustrates exhaustive and contrastive clefts, and (8) illustrates exhaustive and non-contrastive clefts.

(7) Exhaustive, Contrastive

Jane and Tom painted furniture. Tom painted a chair. Later, Kevin remarks: "I bet Tom painted only lamps again, didn't he?"

Jane responds: "He doesn't always paint lamps. Yesterday, it was a chair that he painted."

(8) Exhaustive, Non-contrastive

Jane and Tom painted furniture. Tom painted a chair. Later, Kevin remarks: "I bet Tom painted only a chair again, didn't he?"

Jane responds: "Yes. Yesterday, it was a chair that he painted."

For non-exhaustive yet contrastive clefts, they manipulated the explicit mention of a set of alternatives, and whether specific alternatives were required to establish contrastiveness. For instance, (9) and (10) exemplify non-exhaustive and contrastive clefts appearing after an explicitly mentioned set of alternatives (e.g., *a chair, a desk, and a table*). An utterance preceding the clefts was manipulated to contain or not to contain *only*. When *only* was mentioned in the preceding utterance (e.g., (9)), a set of alternatives was not required to interpret the following cleft sentence with contrastiveness. On the other hand, when *only* was not mentioned in the preceding utterance (e.g., (10)), a set of alternatives was required to estabalish contrastive interpretation from the following cleft.

(9) Non-exhaustive, Explicit, Alternatives Not Required

Jane and Tom painted furniture. Tom painted a chair, a desk, and a table. Later, Kevin remarks: "I bet Tom painted only lamps again, didn't he?

Jane responds: "He doesn't always paint lamps. Yesterday, it was a chair that he painted."

(10) Non-exhaustive, Explicit, Alternatives Required

Jane and Tom painted furniture. Tom painted a chair, a desk, and a table. Later, Kevin remarks: "I bet Tom painted lamps again, didn't he?"

Jane responds: "No, he didn't. Yesterday, it was a chair that he painted."

(11) and (12) illustrate non-exhaustive and contrastive clefts appearing after an implicitly noted set of alternatives (e.g., *a variety of furniture*). The utterance preceding the cleft in (11) contains *only*, thus the alternatives were not required to interpret the cleft with contrastiveness. In (12), due to the absence of *only*, the alternatives were required to interpret the cleft with contrastiveness.

(11) Non-exhaustive, Implicit, Alternatives Not Required

Jane and Tom painted furniture. Tom painted a variety of furniture. Later, Kevin remarks: "I bet Tom painted only lamps again, didn't he?"

Jane responds: "He doesn't always paint lamps. Yesterday, it was a chair that he painted."

(12) Non-exhaustive, Implicit, Alternatives Required

Jane and Tom painted furniture. Tom painted a variety of furniture. Later, Kevin remarks: "I bet Tom painted lamps again, didn't he?"

Jane responds: "No, he didn't. Yesterday, it was a chair that he painted."

After presenting experimental stimuli, the experimenters elicited acceptability ratings of the clefts (e.g., "Yesterday, it was a chair that he painted." in (7)-(12)). Given the experimental design, higher acceptability ratings indicate that the status of object defined in each condition in (7)-(12) was an acceptable interpretation of the status of object indicated by the cleft construction. Overall, the exhaustive and non-contrastive condition (e.g., (8)) was the only condition that was rated less natural than any other conditions, and all of the other conditions did not yield different ratings from each other. This result demonstrates that a felicitous interpretation of exhaustive sentences accompanies contrastive interpretation, while, conversely, contrastive sentences do not have to be interpreted with exhaustivity.

The result of Washburn et al. (2019) that is most relevant to our current discussion is that clefts with contrastiveness in (9)-(12) were rated generally natural, regardless of the explicit versus implicit mention of set of alternatives. Specifically, no differences in ratings were found between (9) and (11), and between (10) and (12). This suggests that focus-marking structure can be felicitously interpreted as indicating contrast, even when a set of alternatives is not explicitly mentioned. It is important to observe that their experiment did not directly test how the status of a set of alternatives affects clefts that are exhaustive and contrastive at the same time. Nevertheless, given the findings that the felicitous interpretation of exhaustive sentences entailed contrastiveness and that contrastiveness was conveyed by a focus structure regardless of the presence of the set of alternatives, exhaustivity may still be conveyed without explicit set of alternatives, as long as the sentence indicates contrastiveness. This result appears to support the MAIN EFFECT MODEL.

Destruel and Velleman (2014) provide production data explaining the contrastiveness of *it*-clefts.

Their first experiment measured how frequently speakers produce *it*-clefts given different contexts. In the experimental stimuli, they provided conversational contexts, as illustrated in (13) and (14), and either a subject (e.g., (13)) or an object (e.g., (14)) constituent that participants were instructed to include in their answer sentences.

(13) Your friend says: This bean dip is fantastic. I really want to get the recipe. Who made it?
Answer: Tim.
You say: _____

(14) Your friend says: Ben and Lucy just bought a new house but had some landscaping work to do. There were a few trees way too close to the house. Which one did they cut down?
Answer: the oak.
You say: _____

The manipulation on the grammatical function of an answer (subject or object) was crossed with whether participants composed an answer to a wh-question or a correction to the previous statement. For instance, unlike examples (13) and (14) that included wh-questions, (15) and (16) prompted an information that needed to be corrected. In (15), participants were asked to correct the boldfaced subject *Shannon* with *Tim*, and in (16), the boldfaced object *the big pine* with *the oak*.

(15) Your friend says: This bean dip is fantastic. I really want to get the recipe. I can't believe that **Shannon** made it – she's normally not a very good cook.
Correction: Tim.
You say: _____

(16) Your friend says: Ben and Lucy just bought a new house but had some landscaping work to do. There were a few trees too close to the house. I don't understand why they cut down **the big pine**, though.
Correction: the oak.
You say: _____

After each stimulus in one condition out of (13)-(16) was given, participants were asked to compose a full response sentence including the provided constituent. Experimenters' interest was the rate that participants produce *it*-clefts in their responses. For example, as a response in (13) and (15), participants could provide syntactically unmarked sentences such as *Tim made it* or *it*-cleft sentences such as *It was Tim who made it*. If *it*-clefts are more natural to convey contrastiveness, those would be used more frequently when participants needed to contrast the previous sentence (i.e., (15) and (16)).

Their result showed that participants were more likely to use clefts when they were asked to indicate contrastivenss, responding in correction conditions exemplified in (15) and (16). However, this pattern was observed only with the subject condition (15), and not with the object condition (16). I speculate that the manipulation of subject and object conditions created a confound, arising from differences in the status of sets of alternative. Notice that in object conditions (14) and (16), there was a mention of a set *a few trees too close to the house*. This could be construable as identifying a set of alternatives and lead to an impulse to refine the set to include relevant alternatives selected from the *trees close to the house*. Unlike in the object conditions, the context preamble from the subject conditions in (13) and (15) did not mention any phrases that could identify or implicitly indicate the existence of a set of alternatives.

This difference in the identification of sets of alternatives could explain the observed frequency of clefts in the subject condition. Participants could have produced clefts more frequently in the subject condition because subjects had the focal status, being picked out as a response when a set had not been identified from the previous context. Conversely, the identified status of the set of alternatives in the object condition (e.g., (16)) could have prevented the focus-marking structure from appearing. On such reinterpretation, considering the status of a set of alternatives, at least English *it*-cleft appears to be able to indicate contrastive interpretation when a set of alternatives is not identified.

Taking the findings from Washburn et al. (2019) and Gotzner and Spalek (2017) together, thus far, it appears clear that the felicitous interpretation of focus-marking structures involves contrastiveness. In addition, focus-marking structures do not require a set of alternatives to be explicitly provided to convey contrastive interpreation. This result supports the MAIN EFFECT MODEL, showing that focus-marking constructions convey contrastiveness (which is entailed by exhaustivity) when a set of alternatives is not identified. Instead of the exhaustivity construction based on a contextualized set of alternatives, the MAIN

EFFECT MODEL proposes a mechanism wherein a focus-marked element answering the wh-question is interpreted exhaustively due to the pragmatic interpretation of the question-answer pairs. Several studies reviewed below have shown that wh-questions prompt exhaustivity in their answers, supporting the mechanism proposed by the MAIN EFFECT MODEL.

To begin, Gerőcs, Babarczy, and Surányi (2014) conducted two experiments in Hungarian. The first experiment provided stimuli consisting a wh-question and a response with a focus-marking structure, but the second experiment did not provide the wh-question prior to the focus-marking structure. They compared the result from the two experiments to demonstrate the effect of wh-questions on the interpretation of focus. In the first experiment, the stimuli appeared as illustrated in (17): The context preamble specified a set of alternatives (e.g., *crown*, *fish*, *pyramid*), and a wh-question followed.

(17) Context: This morning the maid found a corpse in one of the apartments of the Hotel Royale. In the pocket of the victim there was a crumpled piece of paper. There were three figures on it: a crown, a fish, and a pyramid. The victim had circled one or more of them.

Question: What had the victim circled?

Following the context and the question in (17), the response sentence was provided in either form in (18) (a) and (b). The response sentences had a preverbal focus construction and a focalized object (e.g., *piramist* 'pyramid' in (18) (a)), or a syntactically unmarked structure and unfocused object ((18) (b)). In the second experiment, participants saw one of the two constructions in (18) (a) and (b) immediately following the context preamble without the wh-question.

(18) a. Az áldozat a piramist karikázta be.
the victim the PYRAMID.ACC circled PRT

b. Az áldozat bekarikázta a piramist.
the victim circled the pyramid.ACC
'The victim circled the pyramid.'

After an exposure to linguistic stimuli, participants were presented with a picture depicting what actually happened in the scene. In critical conditions, these pictures always depicted a non-exhaustive

scenario (i.e., The victim circled more than one figure), thus the response sentence always stating only about one object (e.g., (18)) could not be a complete description of the picture. Participants were asked whether the picture matches the response sentence, given in the form of (18) (a) or (b), and they could choose either 'yes' or 'no' as an answer. The 'yes' answer, which does not deny the previously given response sentence, indicated that the object in the linguistic description (e.g., 'pyramid' in (18)) was interpreted with less exhaustivity. Meanwhile, the 'no' answer, which denies the previous response sentence, indicated that the response sentence was interpreted with more exhaustivity.

When the wh-question was given in the first experiment, 72% listeners responded to the preverbal focus (e.g., (18) (a)) with 'no' response. When the wh-question was not given in the second experiment, the 'no' response was reported with less than half frequency obtained from the first experiment (35%). This result provides experimental evidence that focus-marking structure is more likely to be interpreted exhaustively when it is given as a response to a wh-question. Because a set of alternatives is explicitly provided in the stimuli, this result does not directly show that the focus-marking can convey exhaustivity without a pre-defined set of alternatives. However, the result supports the idea that wh-questions facilitate a focus-marked structure to be interpreted more exhaustively. It will be the task of our current experiment to test whether wh-questions appearing without a set of alternatives provide enough contextual support for the focus-marking *-lul* to convey exhaustive interpretation.

Onea and Beaver (2009) compared a few different focus-marking structures in terms of ways that each structure interacts with the wh-question to convey exhaustivity. In their experiments, stimuli consisted of a visual stimulus describing a scene where two or more people have the same property (e.g., a picture of two people, Marci and Peter, where each has caught a butterfly) and a sentence describing the scene with one among the three forms of spoken sentences in (19)—preverbal focus on the subject in (a), 'only' focus in (b), and default word order intonation in (c).

(19) a. MARCI fogott meg egy lepkét.
Marci caught PRF a butterfly
'MARCI caught a butterfly'

b. Csak MARCI fogott meg egy lepkét.
Only Marci caught PRF a butterfly
'Only Marci caught a butterfly'

c. Marci meg-fogott egy lepkét.
Marci PRF-caught a butterfly
'Marci caught a butterfly'

After being presented with the description, participants were given with options in (20), and were asked to choose a response to the previous sentence presented in one form out of (a)-(c) in (19). If a previous description in (19) is interpreted with more exhaustivity, (20) (c) would more likely to be chosen, because the response directly denies the previous description with 'no'. Conversely, if a previous description is interpreted with less exhaustivity, responses that update the previous description without a denial ((20) (a) and (b)) would be more likely to be selected.

(20) a. Yes, and Peter caught a butterfly too.

b. Yes, but Peter caught a butterfly too.

c. No, Peter caught a butterfly too.

The authors developed a similarly structured experiment for German. Unlike Hungarian, German has no specific focus position, so they compared sentences with focus intonation (A-accent) on the subject and with default intonation in German. The authors did not provide specific examples of materials that they used for the German experiment, but based on their description, the German stimuli would have looked like (21) below. (21) (a) exemplifies a sentence with the focalized subject 'Marci', and (b) exemplifies a sentence with default intonation.

(21) a. MARCI hat einen Schmetterling gefangen.
Marci has a butterfly caught
'MARCI caught a butterfly.'

b. Marci hat einen Schmetterling gefangen.
Marci has a butterfly caught
'Marci caught a butterfly.'

In the second experiment for both Hungarian and German, they included a wh-question to accompany the visual stimuli. For example, participants saw the question (e.g., 'Who caught a butterfly?') alongside the pictorial stimuli, and the descriptions in one form from (19) (in Hungarian experiment) or

(21) (in German experiment) were presented as answers to the question. Thus, comparing the results obtained from the first experiment (where wh-questions were present in stimuli) and the second experiment (where wh-questions were absent in stimuli) demonstrates the significance of wh-questions in constructing exhaustivity in each of Hungarian and German.

Comparing the results obtained with and without the wh-question, Hungarian preverbal focus was interpreted with constant exhaustivity, regardless of the presence of the wh-question. However, for German, its prosodic focus structure was interpreted with more exhaustivity only when the wh-question was provided. The authors interpreted this result in relation to the difference between German and Hungarian: prosodic focus in German has discourse functions other than indicating exhaustivity, while Hungarian preverbal position has a more specialized function indicating exhaustivity. Thus, the result suggests that a structure that has multiple discourse functions and does not always indicate exhaustivity becomes more likely to convey exhaustivity in contexts containing a wh-question. In other words, wh-questions can activate an exhaustive interpretation of a polyfunctional structure whose discourse functions include the possibility of indicating exhaustive listing. Meanwhile, as the result of the Hungarian preverbal suggests, some focus structures may constantly express exhaustivity regardless of the context of their use.

Focusing on the association between wh-questions and exhaustivity, Moyer and Degen (2021) tested whether certain types of wh-questions induce exhaustive interpretations by default. Their goal was to evaluate the semantic account concerning the interpretation of wh-questions, suggesting that wh-questions are interpreted exhaustively by default, and certain linguistic cues (modal auxiliaries and a particular wh-word *who*) can enhance the likelihood of an exhaustive listing interpretation. To investigate this hypothesis, they constructed experimental stimuli using excerpts of the Switchboard corpus (Godfrey, Hilliman, & McDaniel, 1992) and assessed speakers' intuitive interpretations of various types of questions. In each trial, participants were presented with a question (boldfaced in (22)) appearing after the 10 preceding lines of the dialogue, as presented in (22).

(22) Speaker #2: Pretty good.
Speaker #1: I do like to ski.
Speaker #2: Pretty, pretty down there. huh?

Speaker #1: Yeah, I, I said I do like to ski.

Speaker #2: So, **where have you skied?**

Following the dialogue containing a question, participants were asked to rate how likely the question indicated each of the meanings paraphrased in (23).

(23) Based on the sentence in boldface, how likely do you think it is that the speaker wanted to know about each of the following?

a. What is every place...?

b. What is a place...?

c. What is the place...?

d. Something else

The question in (23) (a) requires an exhaustive list of the information asked by the wh-question in (22). Thus, if participants respond that (23) (a) is what the previous wh-question is likely to indicate, it means that the previous wh-question induced exhaustive listing reading. On the other hand, questions in (23) (b) and (c) do not require an exhaustive list as a response. If participants respond that these are likely meanings of the wh-question in (22), it indicates that the question was not interpreted with an exhaustive listing reading.

Moyer and Degen (2021) reported results showing that wh-questions are not interpreted exhaustively by default. Contrary to the semantic account, only *what* questions were interpreted exhaustively, while modal questions (e.g., *Can you...?*) in fact suppressed exhaustive interpretations for certain wh-words. They concluded that exhaustive interpretation is not observed as a default interpretation of wh-questions, but seems to reflect an interaction between specific modals and wh-words. Note that because their items were based on naturally occurring corpus data, explicit or implicit mention of alternatives was not a controlled factor of the experimental design. Thus, their result shows that some wh-words are more naturally associatd with exhaustive listing interpretation, independent of the availability of a set of alternatives. This lends support to the MAIN EFFECT MODEL.

Combining the preceding results concerning wh-questions (Onea & Beaver, 2009, Gerőcs et al.,

2014; Moyer & Degen, 2021), wh-questions are intertpretable as facilitators of exhaustive interpretation. These results suggest that wh-questions may provide enough context on their own to evoke exhaustive interpretations. This supports the idea of the MAIN EFFECT MODEL that a response to wh-questions with focus status is interpreted in terms of exhaustive listing, even when a set of alternatives is not contextualized. Furthermore, it was demonstrated that different IS-marking structures exhibited distinct patterns in their interactions with the wh-questions (Onea & Beaver, 2009). This suggests the plausibility of hypothesizing that the *-lul* and *-nun* markers in Korean may also have its own unique patterns of interaction with wh-questions in conveying exhaustivity.

Studies being reviewed thus far suggest a strong association between exhaustivity and a set of alternatives (Washburn et al., 2011; Gotzner & Spalek, 2017), empirically attested pragmatic reasoning process to derive non-exhaustivity utilizing shared knowledge (E. Wilcox & Spector, 2019), and wh-questions as a context that prompts an exhaustive listing interpretation of focus information (Gerocs et al., 2014; Onea & Beaver, 2009; Moyer & Degen, 2021). Compared to the association with focus/CF information and exhaustivity, the association between CT and exhaustivity is relatively under-investigated. However, a body of work provides neurological evidence that topic information—as its theoretical conceptualization (c.f., subsection 1.3.1) suggests—is grounded with the given information from shared knowledge, while focus information does not have to be. These results support the INTERACTION EFFECT MODEL of the *-nun* marker, suggesting that for the *-nun* marker to indicate CT status and exhaustivity, it must appear with an alternative that is already established as part of the shared knowledge.

As one factor distinguishing topic and focus, I discussed givenness status of information in Chapter 1. Hirotani and Schumacher (2011) studied how morphological markings in Japanese affect discourse processing of given, inferred, and new information. In their experimental stimuli, a context sentence was shown first to set up a referent to have a certain discourse status in the target sentence, which followed the context sentence. For instance, (24) illustrates the three different types of context sentences that established the referent 'conductor' in the target sentence in (25) to have either given, inferred, or new status, when it appears in the target sentence.

(24) a. Given context: Peter-ga sikisya-o syuzaisimasita.
Peter-NOM conductor-ACC interviewed
'Peter interviewed (a) conductor.'

b. Inferred context: Peter-ga ongakkai-o syuzaisimasita.
Peter-NOM concert-DAT reported
'Peter reported on (a) concert.'

c. New context: Peter-ga terebi-o miteimasita.
Peter-NOM television-ACC was watching
'Peter was watching TV.'

(25) Target sentence to follow each of the context:

Peter-niyoruto **sikisya-{ga/wa}** kodomotati-o sidoositeimasita.
Peter-according to conductor-NOM/TOP childern-ACC was teaching
'According to Peter, (a/the) **conductor** was teaching children.'

The given context in (24) (a) was set up so that a referent in a target sentence (e.g., 'conductor' in (25)) was explicitly mentioned. The inferred context in (24) (b) provided an implicit context where a target referent could be identified. The new context in (24) (c) did not provide any context related to the referent appearing in the target sentence. In the target sentence (25), the referent that could have been introduced previously or that was newly introduced, depending on the context sentence previously provided, was presented with either topic-marking *-wa* or focus-marking *-ga*.

Experimenters measured the Event-Related brain Potential (ERP) to capture the processing patterns of the topic-marked and focus-marked phrases that had given, inferred, or new discourse status. The result showed that the new context (e.g., (24) (c) followed by (25)) yielded the highest N400 effect, which is a neurological indicator of an increased processing cost. Specifically, the N400 effect in the new context conditions was not modulated by the markers used in the target sentence. This appears as a natural result, given that the discourse function of the *-wa* marker is identified only with topical information. In the new context conditions (e.g., (24) (c)), the *-wa* marker in the target sentence appeared after an NP with non-topical status. In this case, the *-wa* marker did not provide any linguistic cue to indicate the topic status of an NP. This demonstrates that the topic-marking function of *-wa* is fulfilled by the interaction with the discourse status of an NP.

Further neurological evidence exists to support the association between morphological markings

and exhaustivity. Wang and Schumacher (2013) compared Japanese nominative *-ga*, topic-marking *-wa*, and accusative *-o*. They designed stimuli with a pair consisting of a context sentence and a target sentence, as illustrated in (26)-(28).

(26) Given: Satomi-wa isha-ni kanshashiteimasu-ne.
Satomi-TOP doctor-DAT be grateful-PRT
'Satomi is grateful to the doctor, isn't she?'

Target sentence: Sodesune, isha-**{ga/o/wa}** ...
yeah, doctor-NOM/ACC/TOP ...

(27) Inferred: Satomi-wa shujutsu-o ukemashita-ne.
Satomi-TOP operation-ACC received-PRT
'Satomi underwent an operation, didn't she?'

Target sentence: Sodesune, isha-**{ga/o/wa}** ...
yeah, doctor-NOM/ACC/TOP ...

(28) New: Satomi-wa kanja-o matteimasu-ne.
Satomi-TOP patient-ACC be waiting-PRT
'Satomi is waiting for the patient, isn't she?'

Target sentence: ie, isha-**{ga/o/wa}** ...
no, doctor-NOM/ACC/TOP ...

A target sentence had a target NP marked with one of the three markers *-ga*, *-o*, and *-wa* (e.g., 'doctor' in (26)-(28)). Prior to the target sentence, the context sentence was manipulated to set up a distinct givenness status of the target NP. The context sentence either explicitly mentioned the target NP (e.g., (26)), provided a context where the target NP becomes relevant (e.g., '...operation...' in (27)), or did not provide any relevant context (e.g., (28)). Among the 9 conditions in total (Context: Given, Inferred, New × Marker: *-ga*, *-o*, *-wa*), only the nominative *-ga* appearing in the New context (e.g., (28) with *-ga*-marking following 'doctor') induced exhaustive reading.

In the result, all three markers showed the N400 and Late Positivity effect in the new condition exemplified with (28). This indicates that new information required more processing cost than given or inferred information. Within the new context conditions, the *-ga* marker had the significantly highest N400 effect, reflecting the processing cost when the marker indicated exhaustive contrast. These results provide neurological evidence directly confirming that Japanese nominative case marker *-ga* additionally

serves a discourse purpose of indicating exhaustive listing interpretation.

On the other hand, the *-wa* marker under the new condition did not enhance the N400 nor Late Positivity, indicating that the marker did not indicate exhaustive listing interpretation. I presume that this null-effect of the *-wa* marker in the new condition was due to the discourse context in which it appeared. Their stimuli did not provide a set of alternatives to prompt contrastivenss, thus the *-wa* marker appeared after an NP with non-topical status. Given this context, the *-wa* marker could not indicate the topical nor CT status of the NP, and ultimately could not indicate exhaustivity. This suggests that the discourse function of a topic/CT marker is activated based on shared knowledge, and the marker does not fulfill any discourse function if the informational status of an NP is not aligned with the marker's discourse function.

We can speculate how this result from Japanese may explain patterns of Korean *-lul* and *-nun*. For the focus-marking *-lul*, it appears to support the MAIN EFFECT MODEL, as the grammatical case marker in Japanese indicated its exhaustivity interpretation when it appeared with a new information without an identified set of alternatives. For the *-nun* marker, the result supports the INTERACTION EFFECT MODEL, because the topic/CT marker in Japanese did not show any effect when it appeared for new information without an identified set of alternatives. Overall, this study suggests that markers' discourse functions are modulated by the status of shared knowledge, and each marker's discourse function interacts with shared knowledge in a distinctive way.

In an effort to reconcile the evidently conflicting results relating to the MAIN EFFECT MODEL and the INTERACTION EFFECT MODEL, I conducted an experiment to investigate whether the exhaustive interpretation of Korean *-lul* and *-nun* markers is modulated by the status of shared knowledge. Several studies have shown the close association between a set of alternatives and exhaustive interpretation (Washburn et al., 2011, 2019; Gotzner & Spalek, 2017; Destruel & Velleman, 2014), supporting that manipulating discourse statuses of a set is one way to operationalize shared knowledge associated with an exhaustive interpretation. Thus, the experiment is designed to measure the degree of exhaustivity when an object is singled out from an identified or unidentified set of alternatives, and when the marker varies between *-lul* and *-nun*.

The comparison between Hungarian focus and German focus structures previously reviewed (Onea & Beaver, 2009) provide evidence that linguistic structures with the same discourse function may

interact with shared knowledge in distinctive ways to indicate exhaustivity. In addition, neurological evidence suggests that different morphological markings interact with shared knowledge in distinctive ways (Hirotani & Schumacher, 2011; Wang & Schumacher, 2013). Based on these findings, I predict that the *-lul* and *-nun* markers would exhibit different patterns when a set of alternatives is not identified. The MAIN EFFECT MODEL of the *-lul* marker is supported by previous studies that have shown the possibility that focus information can evoke contrastiveness irrelevant of the shared knowledge. The INTERACTION EFFECT MODEL of the *-nun* marker is supported by the neurological evidence that the topic/CT marker fulfills its discourse function only when a referent is part of shared knowledge.

2.3 Methods

2.3.1 Participants

66 participants were recruited via the online crowdsourcing platform (`https://pickply.com/`). The crowdsourcing platform is operated in South Korea, and anyone with a valid Korean identification can join the website. All participants reported that they were born in South Korea and resided in South Korea at the time of their participation. As compensation, participants were rewarded "points," a currency used on the crowdsourcing platform, which they could later transfer into monetary values. The experiment was approved by the IRB of UC San Diego and conducted according to its guidelines. Participants initially recruited had the age range from 19-57 (mean = 35.7, median = 35). 25 reported their gender as female and 41 reported as male.

2.3.2 Materials

4 experimental conditions were designed in which explicit mention of alternatives and markings are manipulated. Table 2.2 lists the components used to construct one experimental item. The first row in the table indicates where the explicit mention of alternatives is manipulated to yield either IDENTIFIED and UNIDENTIFIED conditions. In IDENTIFIED conditions, a set of alternatives having two alternatives is explicitly stated prior to the wh-question was presented. In UNIDENTIFIED conditions, a set of alternatives is not explicitly mentioned, and the experimental item began with the wh-question instead (presented in

the second row in Table 2.2). Components in the table were presented in the order shown in the table to consist one experimental item. Sentences in italics were presented in Korean, and only one among the two in curly brackets was presented in an actual item.

Table 2.2: Structure of experimental items used in Experiment 1.

{Set of alternatives/∅}	*Ne-lang Yuna-ka sandwich-lang cookie-lul mantul-ki-lo* You-and Yuna-NOM sandwich-and cookie-ACC make-NR-COMP *hay-ss-canha.* do-PST-EVID "You know you and Yuna planned to make sandwiches and cookies."
Wh-question	*Yuna-nun meo-l mantul-ess-eo?* Yuna-TOP what-ACC make-PST-INT "What did Yuna make?"
First response	*Sandwich-**{lul/nun}** mantul-ess-eo.* sandwich-LUL/NUN make-PST-DEC "(Yuna) made sandwiches."
Second response	*Talun ke-n ahn mantul-ess-eo.* other thing-NUN NEG make-PST-DEC "(Yuna) didn't make anything else."
Experimental question	*Do you think Sohee's second message ("Didn't make anything else.") conveys a new meaning that wasn't expressed in the first message?*
7-point scale	*Not at all. It repeats the same message.* ... *Highly so. It delivers a new message.* 1 2 3 4 5 6 7

Experimental items in all four conditions consisted of a pair comprising a wh-question and two response sentences. The wh-question inquired about the object that fills in the gap of *what* in the subject-predicate relationship. The first response sentence fills in the gap raised in the wh-question by stating the subject's relationship with a single object. In IDENTIFIED conditions, this object was singled out from the explicitly stated set of alternatives. In UNIDENTIFIED conditions, participants had not seen the object in the previous discourse. Markings of the object varied between *-lul* and *-nun*. UNIDENTIFIED-LUL and IDENTIFIED-LUL conditions had *-lul* marker appeared after the object, whereas UNIDENTIFIED-NUN and IDENTIFIED-NUN conditions had the *-nun* marker appeared after the object.

The wh-question and responses written in one out of four conditions were presented sequentially in animated video clips (see Appendix A.1 for the illustration). The clip features an interface resem-

bling that of mobile text messages, displaying the question and presenting the two response sentences at 3-second intervals. This design is intended to simulate the online text messaging environment and to encourage evaluations made for colloquial Korean.

The second response sentence in Table 2.2 explicitly stated the exhaustive status of the previously *-lul* or *-nun* marked object. The experimental question asked participants to rate the “newness" of the second response. Participants could rate the newness on a 7-point scale, where higher numbers indicated that the second response was perceived as conveying a newer message. As the exhaustivity interpretation is more strongly conveyed in the first sentence, participants would perceive the second response stating the exhaustivity as a more repetitive (i.e., less new) message. Thus, if a marker conveys more exhaustivity effects, much lower ratings will be observed.

Following the structure presented in Table 2.2, 24 critical items were created, each with different subjects, objects, and predicates. To disguise the manipulations on object markings, 24 fillers were created using the locative marker (*-ey(se)*) and instrumental marker (*-ulo*). As stated in Chapter 1, the *-nun* marker is attached to these markers agglutinatively, instead of replacing them. Thus, fillers had objects marked with one of the four options: *-ey(se)*, *-ey(se)-nun*, *-ulo*, or *-ulo-nun*. Besides using different marking options, all other components were constructed in the same way as critical items. Half of the fillers didn't state the set of alternatives, while the other half did.

One critical item appeared in four different conditions, resulting in a total of 96 stimuli. All participants were presented with the same set of fillers. The critical and filler items were pseudo-randomized so that no more than two critical items or fillers appeared in a row. A Latin Square distribution of the critical items was done over four questionnaires. A participant is randomly assigned to complete only one out of four questionnaires. Therefore, each participant saw one critical item only once, appearing in one condition.

The lists presented UNIDENTIFIED items first, followed by IDENTIFIED items. This ordering was implemented to elicit ratings of UNIDENTIFIED items independently from the effects of IDENTIFIED items. If UNIDENTIFIED and IDENTIFIED items were presented together, participants could have given higher ratings to the UNIDENTIFIED items in comparison to the IDENTIFIED items, perceiving them as newer messages across all UNIDENTIFIED items. Table 2.3 summarizes the structure of a list provided to

Table 2.3: Organization of one counter-balanced list presented to a single participant.

	Type	Object marking	Counts	Order
Part 1: UNIDENTIFIED	Critical items	*-lul*	6	pseudo-randomized
		-nun	6	
	Fillers	*-ey(se)*	3	
		-ey(se)-nun	3	
		-ulo	3	
		-ulo-nun	3	
	Attention check		2	
Part 2: IDENTIFIED	Critical items	*-lul*	6	pseudo-randomized
		-nun	6	
	Fillers	*-ey(se)*	3	
		-ey(se)-nun	3	
		-ulo	3	
		-ulo-nun	3	
	Attention check		2	

one participant. Total of four questionnaires were created in this structure. Within each part, items were pseudo-randomized such that the first two items were fillers. An attention check appeared every 9 or 10 questions, asking participants about the content of the previously presented wh-question.

2.3.3 Data exclusion

Among the initially recruited 66 participants, 3 participants who did not complete the experiment, and another 3 participants who had participated in the pilot study are excluded in the analysis. There were 3 participants who completed the current experiment more than once, and only their first responses are included in the analysis. Among the remaining 60, 23 participants reported that their first language was some other language(s) than Korean.[2] Responses from these participants are also excluded in the analysis. 37 participants whose responses are included in the final analysis had the age ranged from 19-51 (mean = 33.7, median = 33.5) at the time of their participation. 12 reported their gender as female and 25 reported as male.

[2]This number seems pretty high, given that everyone responded that they were born and living in South Korea. The question to survey participants' first language was phrased as 'What is the language that you learned for the first time when you were born and raised? If you learned more than one language at the same time, please select all options.' Response options included 'Korean language' and two of the 'Languages other than Korean' options with a text box next to it to indicate the language.

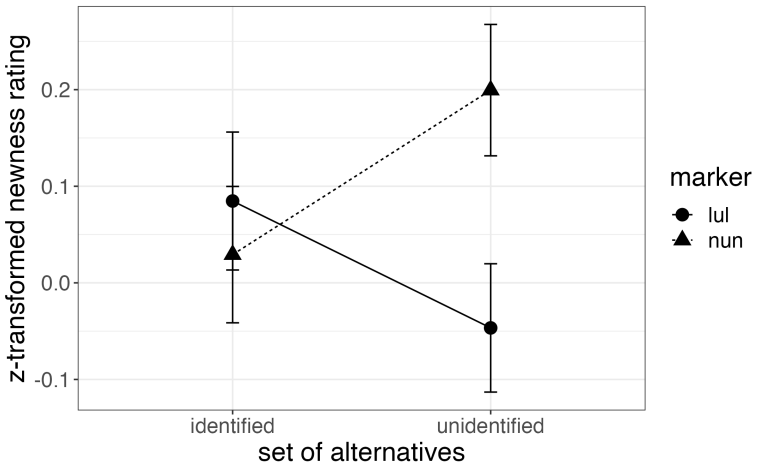


Figure 2.1: Means and 95% confidence intervals (CIs) of z-transformed newness ratings gathered from Experiment 1.

2.4 Results

Results are summarized in the Figure 2.1. In the analysis, ratings responded with the 7-point scale are transformed into z-scores per participant. That is, z-transformation was done for each participant's responses, using the mean and standard deviations of each participant's ratings. This puts the responses gathered from different participants (who may have had different ways of interpreting and utilizing each number on a scale) onto the same scale. Note that the experimental question asked about the newness of the exhaustivity statement (c.f., Table 2.2). Thus, newness ratings reported here can be understood as an inverse representation of the exhaustivity effect. That is, higher newness ratings indicate lower exhaustivity effect, and lower newness ratings indicate higher exhaustivity effect.

Under IDENTIFIED conditions, the *-lul* and *-nun* markers did not differ significantly in terms of the exhaustivity effect ($t(197) = 0.5541$, $p = 0.580$, Cohen's $d = 0.04$). The insignificant difference is visually evident in the left panel of the Figure 2.1, as the two markers' confidence intervals overlap with each another. This is expected based on previous discussions, as both markers have been observed to convey exhaustivity interpretation when a set of alternatives is explicitly provided in the discourse. A difference is observed under the UNIDENTIFIED conditions. When a set of alternatives was not explicitly stated, the exhaustivity effect of the *-lul* marker was observed to be stronger than that of the *-nun* marker ($t(197) = -2.918$, $p = 0.004$, Cohen's $d = 0.21$). This indicates that the *-lul* marker is more likely to convey

exhaustivity than the *-nun* marker when a set of alternatives is unidentified.

To further confirm the differences found under UNIDENTIFIED conditions, a mixed effects model was fit. In the model, z-transformed newness ratings are predicted with fixed factors of marker, set status, and the interaction of the two, with random factors of participants and lexicalization of items. The regression formula is shown in (29). The model fit obtained with the `lmerTest` package in R (Kuznetsova, Brockhoff, & Christensen, 2017). Table 2.4 shows the regression fit of model (29). Higher estimates indicate participants perceived the second response as non-repetitive (i.e. newer). The reference levels of the model were set as Marker: LUL and Set: UNIDENTIFIED.

(29) `z_rating ~ marker*set + (1+marker*set|participant) + (1+marker*set|item)`

Looking into the effect of markers first, change in the marker from LUL to NUN decreases the exhaustivity (β = 0.25, $p = 0.079$). This is visually observable in the figure, as both NUN conditions appear above the reference level condition (LUL-UNIDENTIFIED). Providing the identified set itself does not change the newness ratings, as the IDENTIFIED set is a non-significant predictor of the newness rating (β = 0.13, $p = 0.438$). However, the interaction between the NUN marker and IDENTIFIED set decreases the newness rating (β = -0.30, $p = 0.077$), indicating increased exhaustivity effect. Thus, it suggests that while the *-lul* marker exhibits relatively more constant exhaustivity effect, the *-nun* marker's exhaustivity is modulated by the existence of the set of alternatives.

Post-hoc "simple" pairwise comparisons (Russell, 2024) compared the estimated marginal means

Table 2.4: A mixed effects model fit predicting z-transformed ratings with object marker, status of a set of alternatives, and the interaction of the two, while controlling for the random effects of participants and lexicalized items.

	z-transformed newness rating		
Predictors	Estimates	Standard error	p
(Intercept)	-0.05	0.11	0.669
Marker: NUN	0.25	0.14	0.079
Set: IDENTIFIED	0.13	0.17	0.438
Marker: NUN × Set: IDENTIFIED	-0.30	0.16	0.077
Marginal R^2 / Conditional R^2	0.014 / NA		

obtained from each marker condition separately, holding the set factor fixed at each of the IDENTIFIED and UNIDENTIFIED level. It confirmed that LUL-UNIDENTIFIED and NUN-UNIDENTIFIED are the two conditions where differences in z-ratings are observed (with Tukey correction, $t(34) = -1.78$, $p = 0.084$). Overall, the current data demonstrates that the *-lul* and *-nun* marker exhibit different exhaustivity effect when a set of alternatives is unidentified.

The difference found between the UNIDENTIFIED-LUL and UNIDENTIFIED-NUN conditions, however, is not enough to conclude that the *-nun* marker indeed requires a closed set of alternatives to convey exhaustivity. To make such conclusion, it needs to be further confirmed that the *-lul* marker's exhaustivity is not affected by the status of alternatives, while the *-nun* marker's exhausitivity is modulated by the status of alternatives. In other words, we need to observe insignificant difference between UNIDENTIFIED-LUL and IDENTIFIED-LUL, while observing significant difference between UNIDENTIFIED-NUN and IDENTIFIED-NUN. From the post-hoc pairwise comparisons (Tukey HSD), we observe non-significant difference between *-lul* marker conditions ($t(36) = -0.77$, $p = 0.446$), but also between *-nun* marker conditions ($t(35) = 0.81$, $p = 0.423$). Thus, although the difference between the two markers behavior is found with unidentified set of alternatives, we need more statistical evidence to directly demonstrate that the *-nun* marker's exhaustivity is affected by the set of alternatives, while the *-lul* marker's exhaustivity is not. Further insights found from the result are discussed in the following section.

2.5 Discussion

To directly suggest that the *-nun* marker's exhaustivity effect is modulated by the set of alternatives, having clear baselines of exhaustivity and non-exhaustivity interpretation could be useful. In other words, confirming whether participants actually perceived exhaustivity in any of the conditions can support to make further implications from the result. For this, I considered responses obtained with critical items (having *-lul* and *-nun* markers) with fillers (having locative *-ey(se)* and instrumental *-ulo* with and without *-nun*), shown in Figure 2.2.

On the left panel in Figure 2.2, ratings obtained under IDENTIFIED conditions with critical items and fillers are presented. As we previously discussed, response sentences picking out one alternative from

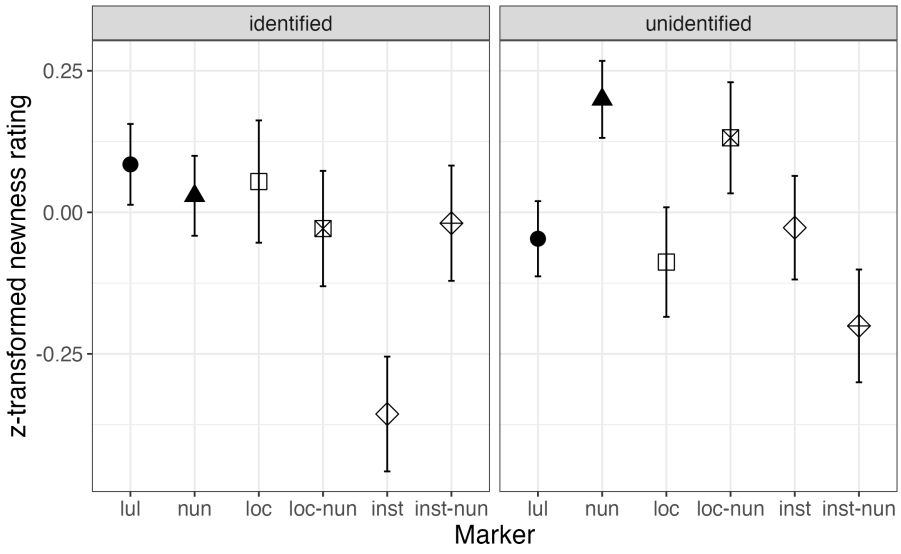


Figure 2.2: Means of z-transformed newness ratings obtained from critical items (*-lul*, *-nun*) and fillers (locative *-ey*, *-ey-nun*, instrumental *-ulo*, *-ulo-nun*) in Experiment 1.

an identified set are naturally likely to be interpreted with exhaustivity, even without any structures explicitly indicating exhaustive interpretation. Ratings are distributed more uniformly without much deviations with an identified set (except the instrumental *-ulo*, which I discuss in a moment) than with unidentified set. This is suggestive of that ratings reported under identified set could be reasonably assumed as an indication of exhaustive interpretation.

Ratings obtained with the instrumental marker with identified set exhibits the strongest exhaustivity effect. One connotation that the instrumental marker may convey is the perfective aspect of the action. For instance, one filler item from the experiment presented the question 'With what did Heesu write down her name on the guestbook?' and the answer 'With a pen', where the 'pen' was marked with the instrumental *-ulo*. In this example, if the action of writing down a name was stated as completed, it must have been odd to imagine that this action could be done with some other different instruments (see Langacker, 1987, for more on perfectives and imperfectives). This can account for the increased newness rating (decreased exhaustivity) observed with having objects marked with *-ulo-nun*. Since the *-nun* marker followed the *-ulo* marker to add discourse interpretations (indicate CT status), the decreased exhaustivity is indicative of the exhaustivity delivered pragmatically.

Thus, I speculate that the instrumental marker conveyed much stronger exhaustivity than other

conditions, due to the grammatical aspect implied by the marker. Considering this interpretation and the instrumental marker fillers showing the score below zero, around zero could be assumed as a range indicating pragmatic exhaustivity, although it is not an empirically confirmed baseline. Another caveat to note in comparing critical items and fillers is that fillers have more observations, because all participants responded to the same set of fillers, whereas critical items were counterbalanced into four different lists.

Under the speculation that the exhaustivity is conveyed pragmatically with *-lul* and *-nun*, now we discuss the exact mechanism that could derive exhaustivity. When a set of alternatives is unidentified, exhaustivity can be conveyed because a set of alternatives is pragmatically assumed, or it can be directly conveyed without any assumptions about the set. In the current experiment, we provided availability of the set of alternatives as an experimental manipulation and did not evaluate whether a set of alternatives actually mediates the participants' exhaustivity interpretation. In other words, the current experiment is not designed to determine the causal relationship between markers, status of a set, and exhaustivity interpretation.

In a future study, one approach that can determine the causality would be to employ an online processing experiment. Studies by Breheny, Katsos, and Williams (2006) and Huang and Snedeker (2009) provide examples of how online processing studies were used to examine the processing patterns of scalar implicatures—whether listeners consider alternative interpretations of scalar terms or they pick out the most relevant interpretation first and consider alternatives later on. Measuring the processing time of focus-marking *-lul* and CF-marking *-lul* (under the assumption that the current result is replicated and both conditions will evoke exhaustivity) would provide more information on whether processing a set of alternatives is a mediating step in calculating exhaustivity.

Going back to the previous accounts listed in Table 2.1, for the *-lul* marker, the current experiment provides evidence to rule out the INTERACTION EFFECT MODEL and to support the MAIN EFFECT MODEL. The INTERACTION EFFECT MODEL suggests that the *-lul* marker is constrained to convey exhaustivity only with an identified set of alternatives. Meanwhile, the MAIN EFFECT MODEL predicts the *-lul* marker to convey exhaustivity irrelevant to the availability of a set of alternatives. In the experiment result, *-lul* marker was associated with consistent exhaustivity, regardless of the existence of a set of alternatives. This pattern is accounted for under the MAIN EFFECT MODEL.

The *-nun* marker's pattern is explained with the INTERACTION EFFECT MODEL, as it showed a decreased exhaustivity effect with unidentified set of alternatives. Although previous accounts resonably proposed an INTERACTION EFFECT MODEL for the *-nun* marker (Büring, 2003; C. Lee, 2003; I. Kim, 2016), each account proposed an exhaustivity of a different nature. Büring (2003) characterizes the exhaustivity from CT-marking structures as conversational implicature, while C. Lee (2003) suggests it is a conventional implicature. Thus, further investigation is needed to tease apart whether each marker's exhaustive interpretation is conveyed as a conversational or conventional implicature. This is further investigated in Chapter 3.

2.6 Conclusion

Overall, this chapter demonstrated that the *-lul* and *-nun* markers vary in terms of how they interact with shared knowledge to convey exhaustivity. Specifically, when a set of alternatives is not part of shared knowledge, the *-lul* marker is more likely to convey exhaustivity than the *-nun* marker. The *-lul* marker is confirmed to exhibit constant exhaustivity regardless of the sharedness of a set of alternatives, while the *-nun* marker's exhaustivity is more likely to be observed when a set of alternatives exists as shared knowledge.

This result is coherent with the associations between focus IS and the *-lul*, and topic IS with the *-nun* marker. When the *-lul* marker marks an object without any previously identified set of alternatives, it indicates a focus status without contrastiveness. When the *-lul* marker singles out an object from a set of alternatives, it indicates a contrastive focus (CF) status. The current experimental result suggests that the *-lul* marker conveys exhaustivity in association with both IS roles, as exhaustivity was found with the *-lul* marker regardless of the status of the set.

The *-nun* marker indicates topic or contrastive topic (CT) status. When the marker appears after an object that has not been previously identified as a member of a set of alternatives, its topic-marking function cannot be fulfilled. In other words, this results in a mismatch between the new informational status of an object and the *-nun* marker indicating the topical status of an object. When the *-nun* marker could not fulfill its topic-marking function, an exhaustivity interpretation was not observed. Thus, the

current experimental result supports the conclusion that each marker conveys exhaustivity in association with its discourse function of indicating the IS of topic or focus.

In sum, the current finding extends the domain where the two markers are paradigmatically contrasted. Each of the *-lul* and *-nun* markers fulfills discourse functions by indicating IS roles, while interacting with shared knowledge in a distinctive manner. The paradigmatic contrast between *-lul* and *-nun* is explained in terms of the different cognitive state of the interlocutor associated with each marker's use.

3 Experiment 2: Cancelability

3.1 Introduction

In Experiment 1, we observed that the *-lul* marker is more likely to induce an exhaustive interpretation than the *-nun* marker when the discourse context does not contain a set of alternatives. This indicates that the two markers differ in conveying exhaustivity in the absence of alternatives within the discourse context. However, when a set of alternatives is identified, both markers are equally likely to induce an exhaustive interpretation. It remains unclear, therefore, how the two markers' discourse interpretation differ when alternatives are present in the discourse context. Accordingly, the current chapter investigates how the two markers differ in conveying exhaustivity when a set of alternatives is present in the discourse context.

In Chapter 1, I described the cancelability of the exhaustive implicature as an aspect where the *-lul* and *-nun* markers may differ. This is summarized in Table 3.1. The CF-marking *-lul* is discussed as conveying a cancelable exhaustivity implicature (C. Lee, 2003). Under the QUD-based account of CT, it is debated whether the *-nun* marker conveys a cancelable exhaustivity implicature (Büring, 2003) or a non-cancelable exhaustivity implicature (C. Lee, 2003). To better understand the associations of the two markers with an exhaustivity implicature, in the current chapter, I report an experiment testing the

Table 3.1: Previously discussed cancelability of exhaustive implicature licensed by the CF-marking *-lul* and CT-marking *-nun*.

	CF-marking *-lul*	CT-marking *-nun*
Cancelable	C. Lee (2003)	Büring (2003)
Non-cancelable		C. Lee (2003); J. Kim (2018)

cancelability of each marker's exhaustive implicature.[1]

In the following section 3.2, I discuss how cancelability can represent the strength of association between a linguistic expression and its meaning. Cancelability has been used as a criterion to classify pragmatic meanings into two types: conversational implicatures and conventional implicatures. After outlining this framework, I introduce previous experimental investigations that utilized cancelability to explore the association between a linguistic expression and its meaning. In section 3.3, I describe the experimental setup that investigates the cancelability of the *-lul* and *-nun* markers' exhaustivity. Experimental results are reported in section 3.4. In section 3.5, I discuss what the observations reveal about the functions of the *-lul* and *-nun* markers in relations between grammar and discourse. The chapter concludes with section 3.6.

3.2 Previous studies

IMPLICATURE is a meaning conveyed beyond the literal content of what is explicitly asserted in utterances: a meaning conveyed beyond the conventional meanings of words (Grice, 1989). Not all pragmatic implicatures are considered to be established based on the same degree of association with a linguistic structure (Ariel, 2008, 2010; Beaver & Clark, 2008; Horn, 2013; Potts, 2007). Among all types of implicatures, CONVERSATIONAL IMPLICATURE is conveyed, relying less on an association with a linguistic structure than on the expectations implicitly shared by interlocutors. This implicit expectation is defined as the COOPERATIVE PRINCIPLE, expressed in (1) (Grice, 1989). Interlocutors share an implicit expectation that each participant will act as a cooperative communicative agent, and this expectation is thought to govern the rational unfolding of conversation.

(1) Cooperative Principle: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

[1]Focus-marking structures and the cancelability of their exhaustivity have been discussed in Kuno (1972); Horn (1981, 2014), as described in Chapter 1. In the current chapter, I focus specifically on the CF-marking *-lul*—the *-lul* marker indicating focus with contextualized alternatives—and its exhaustivity. This is to compare *-lul* and *-nun* when both markers are in a context where they can license exhaustivity.

The Cooperative Principle encompasses a set of CONVERSATIONAL MAXIMS. Four types of maxims—Maxims of Quantity, Quality, Relation, and Manner—were originally proposed in Grice (1989). The Maxims of Quantity, which is most relevant to the discussion of exhaustive interpretation, is presented in (2).

(2) Maxims of Quantity:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

Based on the Maxims of Quantity, a conversational implicature is established in the discourse illustrated in (3). Assuming that the speakers in (3) expect each other to be cooperative, speaker A anticipates that speaker B will provide a response adhering to the Maxims of Quantity. Accordingly, upon receiving B's response, A assumes that B has given the most informative answer possible. The scalar expression *some* has the semantic meaning of 'some and possibly all.' However, if B intended to convey the meaning of 'all of the cake,' she could have used the stronger scalar expression *all* to provide a more informative and accurate response. This leads A to interpret B's utterance pragmatically as meaning 'some but not all of the cake.' This type of interpretations based on the Maxims of Quantity is known as a QUANTITY IMPLICATURE, also known as a SCALAR IMPLICATURE: the denial of a proposition stronger than the one said.

(3) A: Who ate the cake?
B: I ate some of it.
(→ I didn't eat all of the cake.)

Another maxim that is relevant to the exhaustive interpretation is the Maxim of Relation, stated in (4). This maxim states that a speaker must provide relevant information in the context, and leads a listener to presume that the given utterance must be relevant in the context.

(4) Maxim of Relation: Be relevant.

The Maxim of Relation leads to the RELEVANCE IMPLICATURE: an interpretation that speaker's utterance is relevant to the discourse context by implication or explanation. For instance, a conversation in (5) illustrates how the relevance implicature is conveyed. In the context of A's question, B is expected to provide a relevant information that can respond to the question, which is about whether or not she will go to Paul's party. Although B does not state this information explicitly, the given response is interpreted as a relevant answer to the question, and the relevance implicature 'I am not going to Paul's party' is established.

(5) A: Are you going to Paul's party?
B: I have to work.
(→ I am not going to Paul's party.)

EXHAUSTIVITY IMPLICATURE is explained with Maxims of Quantity and Relation. It is an interpretation that denies all other propositions except the one expressed by the given utterance, under the assumption that the other propositions must not be informative or relevant in the given discourse. For instance, in (6), speaker A conveys the question under the expectation that she will hear the informative (given the Maxims of Quantity) and relevant (given the Maxim of Relation) response. Thus, upon receiving B's response, speaker A reasons that any other pieces of information that were not conveyed in the response must not be true or relevant to the preceding question, and ultimately infers the exhaustivity implicature noted in the parentheses.

(6) A: Speaking of John and Mary, who went to the party?
B: John went to the party.
(→ Among the people relevant in the current conversation—John and Mary, no one else but John went to the party.)

As an improvement to the orthodox Gricean framework reviewed above, neo-Gricean proposals

suggest reducing the number of conversational maxims (Horn, 1984, 2013; Levinson, 2000). For instance, Horn (1984) proposed the Q-PRINCIPLE and R-PRINCIPLE listed in (7) as replacements for Grice's four conversational maxims.[2]

(7) Q-Principle: Say as much as you can (given principle R).
R-Principle: Say no more than you must (given principle Q).

Horn's Q- and R-Principles capture how implicatures arise from the alternative uses of linguistic structures with different levels of markedness. Horn (1981) argues that a marked structure reflects a speaker's action of "going out of her way" to use an alternative expression rather than a "simpler" expression. Consider the sentences in (8) and (9):

(8) She got the machine to stop.

a. $\nrightarrow$ She stopped the machine in the usual way.

b. She did not stop the machine in the usual way. (Q-based implicature)

(9) She stopped the machine.

a. She stopped the machine in the usual way. (R-based implicature)

b. $\nrightarrow$ She did not stop the machine in the usual way.

The expression *got to stop* in (8) is considered a marked one, in contrast to the simpler expression *stopped* in (9). Given the marked expression in (8), it is inferred that the speaker must have intended to convey an additional meaning that the simpler expression *stopped* in (9) could not convey. This leads to a Q-based implicature in (8), describing an unusual scene. In contrast, with the unmarked expression *stopped* in (9), it is inferred that the simpler expression conveys everything the speaker needed to convey.

[2]As another neo-Gricean proposal, Levinson (2000) proposed an additional M-PRINCIPLE to Horn's Q- and R-principles and suggested the three principles as replacements for the four Gricean maxims. The M-PRINCIPLE is shown in (i) and is based on Grice's Maxim of Manner.

(i) M-Principle: Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.

Thus, an R-based implicature describing a canonical scene is conveyed.

Using the account of alternative expressions, the scalar implicature illustrated in (3) can be re-examined. The scalar expression *some* in B's response semantically denotes 'some and possibly all,' while an alternative expression *all* semantically denotes 'all' without any other plausible interpretations. Given the alternation between *some* and *all*, when *some* is used in B's response, an R-based implicature is inferred, indicating that the current expression must convey everything the speaker intended. This leads to the interpretation that 'all of the cake' is not the meaning B intended.

These processes of reasoning about an implicature based on conversational maxims or principles are referred to as the CALCULATION of a conversational implicature. Conversational implicatures are established because utterances are interpreted based on the Cooperative Principle and the relevant conversational maxim(s). Thus, pragmatic calculation process is one aspect that identifies conversational implicatures. When exhaustivity implicautre is conveyed as a conversational implicature, we must observe the calculation process where it is established based on the interlocutors' expectations in a given discourse context.

Another feature that identifies conversational implicatures is their CANCELABILITY. Grice originally explained cancelability in the following way:

(10) A putative conversational implicature that *p* is explicitly cancellable if, to the form of words the utterance of which putatively implicates that *p*, it is admissible to add *but not p*, or *I do not mean to imply that p*, and it is contextually cancellable if one can find situations in which the utterance of the form of words would simply not carry the implicature (Grice 1975, [1989]: 44).

Blome-Tillmann (2013) points out two aspects of cancelability described in (10), and refers to the first aspect of cancelability as *explicit cancelability*. As noted, explicit cancelability is observed when utterances that deny a putative implicature *p* (e.g., *but not p*; *but I don't mean to imply that p*) are admissible to add to *p* in the given conversational context. This is the type of cancelability that the current chapter will focus on.[3] Accordingly, to observe the cancelability of an exhaustivity implicature, I will

[3]Blome-Tillmann (2013) refers the second aspect of cancelability as *contextual cancelability*, which is observed when *p* does not hold true in other conversational contexts.

examine whether adding a phrase that denies the exhaustivity is perceived as pragmatically felicitous. For instance, the cancelability of an exhaustivity implicature is observed with the previous example (6) in (11).

(11) A: Speaking of John and Mary, who went to the party?

B: John went to the party. (→ John and Mary did not both attend the party.)

But I don't mean to imply that it was not both who attended the party. Mary also attended the party.

As previously explained, the first sentence in B's response in (11) conveys the exhaustivity implicature noted in parentheses. In order to observe the cancelability of this implicature, as suggested in (10), I added phrases that deny the exhaustivity. In (11), B's following utterances *But I don't mean to imply that...* are admissible to follow the previous utterance in the context. Thus, the implicature conveyed in the previous sentence is considered cancelable.

Both CF and CT information statuses are associated with a cancelable exhaustivity implicature, as presented in Table 3.1. C. Lee (2003) suggested that CF information conveys a cancelable exhaustivity implicature, and Büring (2003) proposed that CT information conveys a cancelable exhaustivity implicature. Following these proposals, when the *-lul* and *-nun* markers convey exhaustivity as a conversational implicature, the exhaustivity interpretation should be felicitously updated by a succeeding phrase, as demonstrated in the response utterance in (12) below.

(12) Context: Interlocutors know that guests were asked to bring a pizza or a cake to the party.

Q. What did Sue bring to the party?

A. Pizza-{lul/nun} kacyewasse. (→ Sue didn't bring anything else.) Cake-to kacyewasse.
pizza-CF/CT brought. cake-also brought
Sue brought a pizza. Sue also brought a cake.

The cancelability of the marker's exhaustivity implicature exemplified in (12) indicates that the exhaustivity implicature is calculated based on conversational maxims (e.g., Maxim of Quantity or R-Principle) and the discourse context of the utterance. The results from Experiment 1 support this charac-

terization of the *-nun* marker's implicature. Because the *-nun* marker licensed an exhaustivity implicature within the relevant discourse context, it is likely that the implicature from the *-nun* marker is calculated as a conversational implicature. On the other hand, the *-lul* marker licensed an exhaustivity implicature even without a relevant discourse context. This suggests that the *-lul* marker may convey exhaustivity implicature through a mechanism different from the one that establishes conversational implicatures.

CONVENTIONAL IMPLICATURE is an implied meaning encoded in a linguistic expression. Conventional implicatures are part of the meaning of an expression, without participating in truth-conditional semantics per se (Horn, 1981). They stem from "idiosyncratic lexical features" (Potts, 2007). Previously, we observed the calculation process of conversational implicatures, where interpretations are established based on interlocutors' expectations about the cooperativeness of others. Unlike conversational implicatures, conventional implicatures do not require a calculation process. Instead, they are triggered by the conventional meaning of a linguistic structure.

An example of conventional implicature is shown in (13) (taken from Horn, 1981). The discourse connective *but* generally indicates that the expectation set by the preceding phrase is not met in the following phrase. In (13), the use of *but* implies that there is an expectation established from the preceding phrase (i.e., *Mr. X is a politician*) and that expectation is not met in the following phrase (i.e., *he is honest*), ultimately implying that the previously established expectation is 'Politicians are usually dishonest.' No interpretive calculation process is involved to reason this implicature, thus it is identified as a conventional implicature triggered by the lexical meaning of *but*.

(13) Mr. X is a politician but he is honest. (→ Politicians are usually dishonest.)

Unlike conversational implicatures, conventional implicatures are not cancelable. This is illustrated in (14). The second sentence in (14) is infelicitous following the first sentence because it attempts to deny the implicature established by the lexical item *but*. The non-cancelability of the implicature is considered evidence of a conventional implicature, indicating that the meaning is more closely encoded by the linguistic expression.

(14) Mr. X is a politician but he is honest. (→ Politicians are usually dishonest.) #But I don't mean to

imply that politicians are usually dishonest. Politicians are usually honest.

In Table 3.1 and the example (12), I reviewed the proposal that associates CT status with a cancelable exhaustivity implicature (Büring, 2003). However, a different proposal suggests that the *-nun* marker, indicating CT status, conveys a non-cancelable exhaustivity implicature (C. Lee, 2003). This proposal is illustrated in (15), where the second sentence in the response ('Sue also brought a cake.') is observed as infelicitous following the previous sentence.

(15) Context: Interlocutors know that guests are asked to bring a pizza or a cake to the party.

Q. What did Sue bring to the party?

A. Pizza-nun kacyewasse. #Cake-to kacyewasse.
pizza-CT brought. cake-also brought
Sue brought a pizza. (→ Sue didn't bring anything else.) #Sue also brought a cake.

Based on the non-cancelability of the implicature, C. Lee (2003) suggests that exhaustivity implicature is a lexically encoded interpretation of the *-nun* marker. However, this proposal does not align with the results observed in Experiment 1. The *-nun* marker was more likely to convey exhaustivity with contextualized alternatives, indicating that the marker's exhaustivity aligns more with the characterization of conversational implicature, which should be cancelable. Additionally, C. Lee (2003) differentiates exhaustivity of CF-marking *-lul* from CT-marking *-nun* and suggests that only the CT-marking *-nun* conveys a non-cancelable exhaustivity. This characterization of the *-lul* marker does not fit with the patterns observed for the *-lul* marker in Experiment 1. The *-lul* marker's implicature showed a pattern closer to conventional implicature in Experiment 1, whereas C. Lee (2003) categorizes it as a conversational implicature. Thus, confirming the cancelability of each marker's exhaustivity will help us better understand how each marker is associated with an exhaustivity implicature—whether it is lexically encoded or licensed through interaction with the discourse context.

Experimental studies have been implementing cancelability as a test to determine whether an exhaustivity implicature is encoded by a linguistic structure (as a conventional implicature) or mainly conveyed by a calculation processes (as a conversational implicature). If an exhaustivity implicature

is less cancelable, it indicates that the implicature is more strongly encoded by the relevant linguistic structure. Conversely, if an exhaustivity implicature is more cancelable, it suggests that the implicature is established in relation to the conversational context, being less associated with the canonical meaning of the linguistic structure. In the remainder of this section, I review experimental studies that utilized cancelability tests in order to gain insights into designing an experiment for Korean markers' exhaustivity implicatures and their cancelability.

Several studies have utilized cancelability to compare various focus-marking structures and their exhaustivity interpretations. For instance, English it-clefts have been debated as to whether their exhaustive interpretation is encoded by the cleft construction itself or conveyed pragmatically. This question is explored in comparison with the lexical exclusive *only*, which conveys exhaustivity as part of its semantic meaning. The cancelability of exhaustive interpretations from cleft constructions and *only* are compared to determine the nature of exhaustive interpretation in cleft constructions. An example of this comparison is illustrated in (16) below (examples from Horn, 1981, modified).

(16) a. It was a pizza that Mary ate. (→ Mary ate nothing but a pizza.) Mary also ate a calzone, and an order of ziti.

b. Mary ate only a pizza. (→ Mary ate nothing but a pizza.) Mary also ate a calzone, and an order of ziti.

The first sentence in (16) (a) illustrates an English it-cleft and (b) illustrates a sentence with the lexical exclusive *only*. Both constructions are considered conveying an exhaustive interpretation, noted in the parentheses. Assume that we want to know whether exhaustivity is a meaning encoded by each construction itself. This is where the cancelability test is used. Comparing the cancelability of the exhaustive interpretation from each construction in (a) and (b) is done by having the second sentence in each of (a) and (b), which denies the exhaustive interpretation of the previous sentence (c.f., explicit cancelability explained in (10); Blome-Tillmann, 2013). We compare the admissibility of the second sentence in the context of each construction, as a diagnosis of whether the exhaustive interpretation is encoded by the structure. This type of cancelability test for clefts and lexical exclusives has been employed in experi-

ments across various languages, including English, Hungarian, German, and French (see Destruel et al., 2015).

To begin, Onea and Beaver (2009), previously reviewed in Chapter 2, devised the so-called 'yes, but' test and compared focus-marking structures in Hungarian in terms of their exhaustivity interpretations. To briefly repeat their experiment design again, experimental stimuli consisted of a visual description of a scene and a spoken sentence describing the scene paired together. The visual stimuli always presented inclusive situations where two characters are associated with the same action (e.g., Marci and Peter each caught a butterfly). Along with the visual stimuli, participants were presented with a verbal description of the scene with one of the structures in (17) (a)-(c). (17) (a) is an example of preverbal focus constructions, which is proposed to convey exhaustivity as a conversational implicature. (17) (b) is an example of sentences using lexical exclusive 'only', which is hypothesized to convey exhaustivity as conventional implicature. (17) (c) is a sentence with canonical word order and prosody, without any particular structures associated with an exhaustivity implicature.

(17) a. MARCI fogott meg egy lepkét.
Marci caught PRF a butterfly
'MARCI caught a butterfly'

b. Csak MARCI fogott meg egy lepkét.
only Marci caught PRF a butterfly
'Only Marci caught a butterfly'

c. Marci meg-fogott egy lepkét.
Marci PRF-caught a butterfly
'Marci caught a butterfly'

After the description of a visual scene was presented using one of the three constructions exemplified in (17), participants were given three response options in (18) and were asked to select one in response to the previously presented sentence.

(18) a. Yes, and Peter caught a butterfly too.

b. Yes, but Peter caught a butterfly too.

c. No, Peter caught a butterfly too.

The three possible responses in (18) are designed to represent different levels of acceptance. The option in (18) (a) continues the conversation by accepting the previous description of the visual scene with the initial 'yes' and an updated description with the phrase following 'and'. Option (b) represents a lower level of acceptance, as it contains a correction of the previous sentence with the phrase following 'but'. Option (c) represents the lowest level of acceptance, as it rejects the exhaustivity status most strongly.

Onea and Beaver (2009) predicted that the 'no' continuation, which overtly disagrees with the previous sentence (e.g., (18) (c)), would be selected more often when stronger exhaustivity is conveyed in the verbal description. This prediction was made based on the consideration that participants would have a greater need to correct a message that is perceived as more erroneous. In other words, their experiment measured the degree of exhaustivity conveyed by different structures exemplified in (17) by examining how likely participants were to deny the description. The 'no' continuation indicates that the previous sentence conveyed a much stronger exhaustivity that requires a stronger denial. Conversely, the 'yes, but' and 'yes, and' continuations reflect a weaker exhaustivity, which could be corrected with less strong denials.

The results showed that each of the three constructions in (17) elicited different continuations from (18) as the most frequently chosen response. The lexical exclusive 'only' (e.g., in (17) (b)) was most frequently followed by the 'no' continuation. The preverbal focus construction in (17) (a) most often elicited the 'yes, but' response. The sentence with default prosody and word order in (17) (c) was most commonly followed by the 'yes, and' continuation. This result is suggested as evidence that the preverbal focus construction in Hungarian expresses weaker exhaustivity compared to the lexical exclusive 'only', and that exhaustivity derived from Hungarian preverbal constructions is less strongly encoded in the structure.

Building on the pattern observed in Onea and Beaver (2009), we can compare the focus-marking *-lul* and the exclusive *-man* in (19). If the exclusive marker *-man* conveys significantly stronger exhaustivity than the *-lul* marker, the response in A2 will likely yield a 'no' continuation, similar to (20) (c). Conversely, if the *-lul* marker conveys relatively weaker exhaustivity, it is more likely to be followed by the 'yes, but' continuation.

(19) Context: Interlocutors know that guests were asked to bring a pizza or a cake to the party.

Q. What did Sue bring to the party?

A1. (Sue-nun) Pizza-**lul** kacyewasse.
(Sue-TOP) pizza-CF brought
Sue brought a PIZZA.

A2. (Sue-nun) Pizza-**man** kacyewasse.
(Sue-TOP) pizza-only brought
Sue brought only a pizza.

(20) a. Ung, kuliko Sue-nun cake-to kacyewasse.
yes and Sue-TOP cake-also brought
'Yes, and Sue brought a cake too.'

b. Ung, haciman Sue-nun cake-to kacyewasse.
yes but Sue-TOP cake-also brought
'Yes, but Sue brought a cake too.'

c. Aniya, Sue-nun cake-to kacyewasse.
no Sue-TOP cake-also brought
'No, Sue brought a cake too.'

However, there are more than one explanation to the result of 'yes, but' test. As Onea and Beaver (2009) suggested, participants might select the 'yes, but' continuation when they perceive the previous sentence as conveying less exhaustivity, which does not require a stronger correction. Contrary to the test intended, however, participants may select 'yes, but' continuation when the previous sentence conveys stronger, thus non-cancelable exhaustivity. If participants perceive the exhaustivity implicature as non-cancelable, responding with 'no'—which cancels the previously conveyed exhaustivity—becomes an invalid option.

To illustrate, participants might be inclined to deny the A2 response in (19) with the stronger rejection in (20) (c), because they perceive that the 'only' marker conveys stronger exhaustivity and requires stronger rejection. Alternatively, participants might be inclined to deny the A2 response in (19) with (20) (a) or (b) instead, because they perceive that the 'only' marker conveys a stronger, thus non-cancelable implicature, which cannot be continued with 'no' responses that deny the previously established implicature. In other words, while Onea and Beaver (2009) considered 'no' responses an indication of stronger exhaustivity previously conveyed, 'no' responses instead might have detected weaker exhaustivity, as it

expresses the cancellation of the exhaustivity. Additionally, the 'yes, but' test has faced other criticisms, which I will review in a moment.

Another experiment compares the exhaustivity of a cleft construction and a lexical exclusive in French (Destruel, 2010). In this experiment, participants were first presented with a wh-question. They then received an answer expressing exhaustivity for the subject (i.e., the response to the wh-word) in one of three constructions illustrated in (21). Specifically, (21) (a) features a *c'est* cleft, (b) uses the lexical exclusive *seulement* ('only'), and (c) presents a sentence without any structures associated with exhaustivity.

(21) Q. Qui est-ce-qui a tapé le voleur?
Who is-that to beat the theif
'Who beat the thief?'

a. C'est le policier qui a tapé le voleur.
it's the policeman who to beat the thief
'It is the policeman who beat the thief.'

b. Seul le policier a tapé le voleur.
only the policeman to beat the thief
'Only the policeman beat the thief.'

c. Le policier a tapé le voleur.
the policeman to beat the thief
'The policeman beat the thief.'

Following the presentation of a question and an answer featuring one of the three constructions, participants were given options to continue the answer with 'yes, and', 'yes, but', or 'no' continuations, similar to the experiment from Onea and Beaver (2009). However, unlike Onea and Beaver (2009), whose experiment focused on focus-markings of subjects, Destruel (2010) designed stimuli for both subjects and objects. Stimuli for objects are illustrated in (22). Specifically, (22) (a) shows a cleft construction for the object, (b) uses the lexical exclusive *seulement* modifying the object, and (c) presents a sentence without any structures associated with exhaustivity.

(22) Q. Qu'est-ce-que le père a brossé?
What the father to brush?
'What did the father brush?'

a. C'est le cheval que le père a brossé.
it's the horse that the father to brushed
'It is the horse that the father brushed.'

b. Le père a brossé seulement le cheval.
the father to brushed only the horse
'The father only brushed the horse.'

c. Le père a brossé le cheval.
the father to brushed the horse
'The father brushed the horse.'

The results demonstrated distinct patterns of cancelability among the different constructions. The 'only' construction yielded the 'no' continuation most frequently, indicating that it conveys the strongest exhaustivity and requires a more definitive rejection. Conversely, the cleft constructions elicited the 'yes, but' response most often, suggesting a weaker exhaustivity that could be updated rather than outright rejected. The default constructions resulted in the 'yes, and' continuation most frequently, reflecting the least strong exhaustivity. These findings align with the results observed in Hungarian by Onea and Beaver (2009), where similar patterns of cancelability were noted.

Interestingly, Destruel (2010) observed different results for subject-focused and object-focused constructions. When the grammatical subject was the focal element modified with 'only', the 'no' continuation was chosen about 94% of the time. On the other hand, when 'only' modified the grammatical object, the 'no' continuation was chosen 50% of the time, significantly less than the 94% observed with subjects. This suggests that if Korean *-man* ('only') marker patterns similar to the French *seulement*, we will observe that the subject-marking *-man* in (23) is interpreted with stronger exhaustivity than the object-marking *-man* in (24).

(23) Q. Who beat the thief?

A. Kyengchal-{i/*un/man} totwuk-ul ttayly-ess-e.
police.officer-FOC/TOP/only thief-ACC beat-PST-DEC
'The police officer beat the thief.'

(24) Q. Whom did the police officer beat?

A. (Kyengchal-i) totwuk-{ul/un/man} ttayly-ess-e.
(police.officer-NOM) thief-FOC/CT/only beat-PST-DEC

'The police officer beat the thief.'

Destruel (2010) does not provide any speculation on why different exhaustivity effects between subject-marking and object-marking *seulement* are observed. However, this result is reminiscent of the interaction between morphology and syntax in Korean, as organized in section 1.4.2. In section 1.4.2, we observed that sentence-initial phrases tend to have topical status in Korean and that the sentence-initial *-nun* marker indicates topical status, not CT status, as illustrated in (23). The result from Destruel (2010) reminds us that an experiment in Korean needs to consider the effect of word order carefully to avoid introducing any unwanted interaction with morphology and syntax. Thus, the experimental stimuli for Korean need to present markers under comparison in the canonical SOV word order, which is the context in which all markers can indicate contrastiveness.

Despite the seemingly clear implications made from the 'yes, but' tests in Onea and Beaver (2009) and Destruel (2010), these tests are criticized for being different from the Gricean cancellation test (c.f., (10) in this chapter). In particular, it has been pointed out that the 'yes, but' test does not diagnose the source of exhaustivity (conventional vs. conversational) but rather the status of exhaustivity as at-issue or non-at-issue (Destruel et al., 2015).

At-issue content is considered a response to the QUD, while non-at-issue content is considered a piece of information that does not directly address the QUD (Tonhauser, 2012). This is exemplified in (25) (taken from Tonhauser, 2012). In (25), Carlos' question *Who stole my money?* establishes the QUD in the discourse.

(25) Context: Carlos' pocket was picked at the party he is attending with Mario.

a. Carlos: Who stole my money?

b. Mario: That man, my mother's friend, stole your money.

Mario's answer to the question in (25) (b) conveys at least the following propositions: (i) Mario is indicating somebody, (ii) the person indicated by Mario is a man, (iii) Carlos had money, (iv) the man Mario is indicating stole Carlos' money, (v) Mario has a mother, and (vi) the man Mario is indicating is Mario's mother's friend. Among these propositions, the fourth proposition (the man Mario is indicating

stole Carlos' money) is the information that responses to the QUD. This type of information that directly addresses the QUD is the at-issue content. The other pieces of information conveyed by the rest of the propositions are non-at-issue contents, as they do not respond to the QUD.

Destruel et al. (2015) suggest that the 'yes, but' continuations reflect that a construction conveys exhaustiveness as non-at-issue meaning and cannot provide evidence for a conversational (in contrast to the conventional) nature of the exhaustive implicature. To support this, Destruel et al. (2015) provide a reanalysis of the results of 'yes, but' tests reported in Xue and Onea (2011). The two types of pragmatic inferences Destruel et al. (2015) re-analyzed were scalar implicatures and relevance implicatures in German, illustrated in (26) and (27).

(26) Die Suppe ist warm.
the soup is warm
'The soup is warm.' (→ Scalar implicature: The soup is not hot.)

(27) Q: Is there any working ATM nearby?

A: Ein paar Schritte weiter ist gleich eine Sparkasse.
a few steps further is just a bank
'There's a bank just a few steps further.' (→ Relevance implicature: The bank will let you withdraw cash.)

Previously, we observed that both scalar implicatures and relevance implicatures as conversational implicatures with examples (3) and (5). These two types of implicatures are established based on pragmatic calculation processes given interlocutors' expectations about rational conversations; There is no particular linguistic structure that triggers these types of implicature. However, Destruel et al. (2015) point out that the two types of implicatures differ in terms of at-issueness: While scalar implicatures are always at-issue, relevance implicatures do not have to be at-issue. In (27), considering the given question as QUD, the relevance implicature does not directly address the QUD, therefore does not convey at-issue content.

Based on these characteristics of scalar implicatures and relevance implicatures, Destruel et al. (2015) compared inferences made in (26) and (27) and their cancelability with the 'yes, but' test. They analyzed the rate of three response options ('yes, and', 'yes, but', and 'no') following scalar implicatures

and relevance implicatures. For instance, following the scalar implicature in (26), they gathered the frequency of each response options listed in (28); following the relevance implicature in (27), they gathered the frequency of each response options listed in (29).

(28) {Ja, und/Ja, aber/Nein,} die Suppe ist heiß.
{yes and/yes but/no} the soup is hot
'{Yes, and/Yes, but/No,} the soup is hot.'

(29) {Ja, und/Ja, aber/Nein,} der Geldautomat ist kaputt.
{yes, and/yes, but/no,} the ATM is broken
'{Yes, and/Yes, but/No,} the ATM is broken.'

If the 'yes, but' test diagnoses at-issueness rather than the conversational vs. conventional nature of the implicature, the 'no' continuation should be observed more frequently with sentences contradicting scalar implicatures (e.g., (29)). However, if the 'yes, but' test diagnoses the conversational vs. conventional nature of the implicature, as intended in Onea and Beaver (2009), both relevance implicatures and scalar implicatures should yield 'yes, but' continuations most often.

As Destruel et al. (2015) speculated, their reanalysis revealed that the 'yes, but' test diagnoses the at-issueness of pragmatic implicatures rather than the conversational vs. conventional nature of the implicature. Relevance implicatures (i.e., non-at-issue implicatures) yielded the 'yes, but' continuations most often, while scalar implicatures (at-issue implicatures) yielded the 'no' continuations most often. Based on the result, Destruel et al. (2015) suggested that the cancelability of exhaustivity implicature observed with Hungarian preverbal focus in Onea and Beaver (2009) cannot lead to the conclusion that preverbal focus conveys exhaustivity as a conversational implicature. An alternative conclusion of Onea and Beaver (2009) suggested by Destruel et al. (2015) is that the cancelability observed with the 'yes, but' test only presents evidence that preverbal focus in Hungarian conveys exhaustivity as a non-at-issue meaning, irrelevant to the conversational vs. conventional nature of it.

Considering the reanalysis and criticism from Destruel et al. (2015), DeVeaugh-Geiss, Zimmermann, Onea, and Boell (2015) tested whether non-at-issue exhaustivity from cleft constructions could still be a conversational implicature. In their tests, DeVeaugh-Geiss et al. (2015) designed experimental stim-

uli utilizing cancelability in a different way from the 'yes, but' test, in order to mitigate the issues of the 'yes, but' test pointed out by Destruel et al. (2015). As illustrated in (30), they included phrases canceling the exhaustivity as part of the stimuli, rather than asking participants to select a continuing phrase as seen in the other 'yes, but' tests.

(30) a. Nur Sabine hat den Tierpark besucht und Anna hat den Tierpark besucht
only Sabine has the zoo visited and Anna has the zoo visited
[Exclusive, At-issue] 'Only Sabine visited the zoo and Anna visited the zoo.'

b. Nur Sabine hat den Tierpark besucht und sie hat den Tierpark nicht besucht.
only Sabine has the zoo visited and she has the zoo not visited
[Exclusive, Not-at-issue] 'Only Sabine visited the zoo and she did not visit the zoo.'

In their first experiment, DeVeaugh-Geiss et al. (2015) compared four types of sentences illustrated in (30) and (31), varying in focus-marking constructions (exclusive 'only' vs. cleft) used in the first phrase and at-issueness (at-issue vs. non-at-issue) of the implicated meanings being canceled in the second phrase. For instance, (30) (a) represents a sentence where an at-issue exhaustivity from 'only' is canceled in the second phrase, and (30) (b) represents a sentence where an non-at-issue meaning of 'only' is canceled. In (31) (a), the at-issue meaning of cleft is canceled, and in (31) (b), the non-at-issue exhaustivity from cleft is canceled.

(31) a. Es ist Sabine, die den Tierpark besucht hat, und sie hat den Tierpark nicht besucht
it is Sabine who the zoo visited has and she has the zoo not visited
[Cleft, At-issue] 'It is Sabine who visited the zoo and she did not visit the zoo.'

b. Es ist Sabine, die den Tierpark besucht hat, und Anna hat den Tierpark besucht.
it is Sabine who the zoo visited has and Anna has the zoo visited
[Cleft, Not-at-issue] 'It is Sabine who visited the zoo and Anna visited the zoo.'

They elicited acceptability of each type listed in (30) and (31) on the scale from 1 to 7. The prediction was that if exhaustivity of the cleft is more cancelable, not because of its non-at-issueness, but because it is a conversational implicature, we should observe a difference between the exclusive's at-issue exhaustivity (e.g., (30) (b)) and cleft's not-at-issue exhaustivity (e.g., (31) (a)): the cleft's exhaustivity should exhibit more cancelability than the exclusive's exhaustivity. However, if exhaustiivty of the

cleft is more cancelable due to its non-at-issueness, its cancellation (e.g., (31) (b)) should have similar acceptability ratings to that of exclusive's non-at-issue meaning being violated (e.g., (30) (b)).

Their results showed that when non-at-issue meanings of clefts are canceled, the mean of rated acceptability improves. Meanwhile, the lexical exclusive 'only' did not show any difference in ratings for at-issue versus non-at-issue meanings. In other words, at-issueness had different interactions with clefts and exclusives. For the lexical exclusive 'only', canceling both its at-issue meaning (i.e., exhaustivity) and non-at-issue meanings resulted in lower acceptability, indicating the non-cancelability of both types of meanings. However, for clefts, canceling the non-at-issue meaning (i.e., exhaustivity) resulted in higher acceptability than canceling the at-issue meaning, indicating the cancelability of exhaustivity. From this result, it was concluded that canceling the exhaustivity of cleft constructions is more acceptable, not because it is non-at-issue, but because it differs in nature from the exhaustivity indicated by 'only'. Given that 'only' indicates exhaustivity as part of its lexical semantics, they conclude that cleft constructions in German must not convey exhaustivity from the structure itself, that is, exhaustivity arises through a conversational implicature of clefts.

Given that DeVeaugh-Geiss et al. (2015) obtained significant result with a different way of utilizing cancelability, it is clear that the cancelability test is a valid test to investigate the association between a linguistic structure and its meaning, as long as the result of cancelability test is correctly interpreted. The discussions on correctly interpreting the cancelability tests proposes implications to be made in relation to the Korean *-nun* marker. In Chapter 1, I introduced the proposal of J. Kim (2018) where the *-nun* marker is identified as a 'not all' quantifier. In particular, it is proposed that the *-nun* marker presupposes the existence of a distinct individual in the alternative set who can be in a potentially contrastive relation as its non-at-issue meaning. With an example in (32), J. Kim (2018) illustrates that the *-nun* marker conveys the non-at-issue meaning that there is another pair that is not Ann and not Dave.[4]

(32) ANN-UN DAVE-HANTEY kisshay-ss-e.
Ann-CT Dave-DAT kiss-PST-DEC
'ANN kissed DAVE.'

a. At-issue meaning: Ann kissed Dave. (i.e., The pair Ann-Dave kissed.)

[4] J. Kim (2018) used capitalization in examples to indicate prosodic emphasis.

b. Non-at-issue meaning: There are at least two elements in a set that are in a contrastive relation with Ann and Dave. (i.e., At least one pair consisted of 'non-Ann' and 'non-Dave' exists.)

J. Kim (2018) characterizes the non-at-issue meaning in (32) (b) as a conventional presupposition of the *-nun* marker. This is illustrated with 'Hey, wait a minute' test in (33). Even when the context does not include any alternatives in (33), the use of the *-nun* marker in A's utterance triggers B to assume that there is another pair who kissed.

(33) Context: Ann and Dave are the only people under discussion.

A: ANN-UN DAVE-HANTEY kisshay-ss-e.
Ann-CT Dave-DAT kiss-PST-DEC
'ANN kissed DAVE.'

B: Hey, wait a minute! What do you mean by 'ANN-UN'? Is there another pair who kissed?

The non-at-issue proposal of the *-nun* marker can be understood in relation to the previous studies addressing non-at-issueness of the exhausitivty implicature (Destruel et al., 2015; DeVeaugh-Geiss et al., 2015). Based on the experimental results that non-at-issue inferences were cancelable, the *-nun* marker's exhaustivity is predicted to be cancelable, if it is associated with the non-at-issue meaning of the marker. For instance, when the *-nun* marker appears after *Ann* in (34) and presupposes the existence of alternatives, the exhaustivity implicature in (c) is established based on the marker's non-at-issue meaning (i.e., non-at-issue conventional presupposition in (a)).

(34) ANN-UN kissed DAVE.

a. Non-at-issue, conventional presupposition: Another pair exists.

b. Non-at-issue, pragmatic inference: Another girl kissed another boy.

c. Exhaustivity implicature: No one else kissed Dave.

(35) Sue also kissed Dave.

The non-at-issue and conventional presupposition proposal by J. Kim (2018) suggests that observ-

ing a cancelable exhaustivity implicature with *-nun* can be interpreted in two ways: either the exhaustivity implicature arises from the marker's non-at-issue meaning, or it is conveyed as a conversational implicature, independent of its at-issue status. For example, if continuing a sentence like in (34) with (35) is perceived as felicitous, it could indicate that the *-nun* marker's exhaustivity is a non-at-issue meaning. Alternatively, it might suggest that the exhaustivity is not lexically encoded by the marker but is instead derived as a conversational implicature.

Psycholinguistic studies reviewed thus far provide behavioral evidence demonstrating that different types of implicatures exhibit distinct cancelability. In addition to the behavioral evidence provided, Drenhaus, Zimmermann, and Vasishth (2011) provide neurological data that validate the role of cancelability in distinguishing different types of implicatures. Specifically, they demonstrated that clefts and lexical exclusive 'only' are associated with distinct neurological patterns reflecting their exhaustive interpretations having different cancelability. Their experimental stimuli in German included sentences with either a cleft construction or a lexical exclusive. To begin, (36) shows sentences having cleft constructions. In (a), the exhaustivity of 'Mary' is canceled in the later part of the sentence, but in (b), the exhaustive status of 'Mary' is maintained.

(36) a. Es ist Maria, die das Klavier spielen kann und auberdem noch Luise und Jana, sagte...
it is Maria the that piano play can and also still Luise and Jana said
'It is Mary that plays the piano and, besides, Luise and Jana, said...'

b. Es ist Maria, die das Klavier spielen kann und auberdem noch die Geige, sagte...
it is Maria the that piano play can and also still the violin said
'It is Mary that plays the piano and, besides, the violin, said...'

Cleft constructions are compared with sentences having lexical exclusives, as illustrated in (37). Parallel to the cleft sentences in (36), sentences with a lexical exclusive in (37) contained the phrase that either cancels (e.g., (a)) or maintains the exhaustive meaning of 'only' (e.g., (b)). If cleft and 'only' constructions are associated with exhaustive interpretations of different nature, pairs of (a) sentences and pairs of (b) sentences in (37) and (38) would result in different ERP patterns.

(37) a. Nur Maria kann das Klavier spielen und auberdem noch Luise und Jana, sagte ...
only Maria can the piano play and also still Luise and Jana said

'Only Mary can play the piano and, besides, Luise and Jana, said...'

b. Nur Maria kann das Klavier spielen und auberdem noch die Geige, sagte...
only Maria can the piano play and also still the violin said
'Only Mary can play the piano and, besides, the violin, said...'

Two distinct ERP patterns were identified in the results. The N400 effect indicates the processing of integrating violated expectations, rather than the violated truth value, while the P600 effect signifies the reanalysis and repair of syntactic interpretations. The N400 effect was observed when the exhaustivity of cleft constructions was canceled (e.g., (36) (b)). This suggests that canceling the exhaustivity of clefts involves integrating an unexpected element rather than reanalyzing syntactic structures. Conversely, the P600 effect was found when the exhaustivity of 'only' constructions was canceled (e.g., (37) (b)), indicating reanalysis and repair of the syntactic interpretation. This contrasting ERP patterns demonstrate that the two constructions convey exhaustivity with different processing patterns. Overall, by demonstrating that the cancellation of conversational vs. conventional exhaustivity is associated with distinct neurological patterns, this study further validates the cancelability test as a tool for diagnosing different types of implicatures.

In summary, studies have proposed cancelability of *-lul* and *-nun* that does not align with the patterns of each marker we observed in Experiment 1. In addition, different proposals are made for the *-nun* marker's cancelability. Thus, an experiment is designed to examine the cancelability of the CF-marking *-lul* and CT-marking *-nun*, with an aim to better understand how each of the *-lul* and *-nun* licenses exhaustive interpretation—whether the marker conveys exhaustivity via pragmatic interpretive process (i.e., as a conversational implicature) or lexically encodes exhaustivity (i.e., as a conventional implicature).

3.3 Methods

3.3.1 Participants

79 participants were recruited via online crowdsourcing platform (`https://gosurveasy.com`). The platform is operated in South Korea, and anyone with a valid Korean identification can join the

website. Among 79, 71 participants completed the experiment, but the remaining 8 participants did not complete the experiment. All participants reported that they were born in South Korea and resided in South Korea at the time of their participation. Upon completion of the experiment, participants were rewarded with "points", a currency used on the crowdsourcing platform, which they could later transfer into monetary values. The experiment was approved by the IRB of UC San Diego and conducted according to its guidelines. 71 participants who completed the experiment reported their ages range from 20-63 (mean = 37, median = 32). 37 reported their genders as female, and 34 reported as male.

3.3.2 Materials

Experimental stimuli were designed to diagnose the cancelability of the exhaustive interpretation of the *-lul* and *-nun* markers. Experimental stimuli varied in terms of the object marking (*-lul* or *-nun*) and the cancellation of the exhaustive interpretations (canceled vs. not canceled). Each experimental question was composed of components illustrated in Table 3.2. The listed components in the table were presented in the order shown to form one experimental question. Italics represent Korean text presented in the actual experiment. Boldfaced components were experimental manipulations, and one option from the curly brackets was presented in the actual experiment.

Each item began with a context preamble. There were always three people in the context preamble (e.g., *Sohee, Mina, and Yuna* in Table 3.2), and these people were described as having planned an activity together (e.g., attending a marathon race). Among the three people, one person did not have the knowledge to respond to the wh-question (e.g., *Sohee didn't keep the promise, so doesn't know who won what in the race*), while the other two people did. The context preamble always listed two alternatives that were relevant to the characters' planned activity. This was to ensure that both *-lul* and *-nun* markers were in the context of licensing exhaustivity. For instance, in Table 3.2, the context contextualizes a set of alternatives, *medal* and *trophy*.

Following the context preamble, a wh-question was presented by the person who did not have knowledge about the event (e.g., *What did Yuna receive?* in Table 3.2). The first response sentence singled out one alternative from the previously presented set (e.g., *medal*), marked with either the *-lul* or *-nun* marker. Based on the context preamble and the appearance of either marker, the first response

Table 3.2: Structure of experimental items used in Experiment 2.

Context	*Sohee, Mina, and Yuna planned to attend the marathon race together. The race had medals and trophies for winners. Mina and Yuna attended the race together. Sohee didn't keep the promise, so she doesn't know who won what in the race. Later, Sohee and Mina exchanged the text messages shown in the next page.*
Wh-question	*Sohee: Yuna-nun meo-l bat-ass-eo?* Yuna-TOP what-ACC receive-PST-INT "What did Yuna receive?"
First response	*Mina: Medal-**{LUL/NUN}** bat-ass-eo.* Medal-CF/CT receive-PST-DEC. "(Yuna) Received a medal."
Second response	*Trophy-**{to/nun mos}** bat-ass-eo.* Trophy-also/CT NEG receive-PST-DEC "(Yuna) received/didn't receive a trophy."
Experimental question	*How appropriate is it that Mina sent the second sentence after the first one as a response to Mina's question?*
7-point scale	*Not at all appropriate* 1 2 3 4 5 6 7 *Highly appropriate*

sentence was designed to convey an exhaustive interpretation.

The second response sentence elaborated on the other alternative from the contextualized set (e.g., *medal*). In the CANCELED condition, the second sentence canceled the exhaustive status established in the first response sentence by marking the object with the additive *-to* and repeating the same predicate. In the MAINTAINED condition, the second response sentence confirmed the exhaustive status of the object mentioned in the first response sentence by elaborating that the other alternative did not hold the same predicate relation stated in the first response sentence. Thus, the experiment used a 2 × 2 design (marker × continuation) with 4 conditions (LUL-CANCELED, LUL-MAINTAINED, NUN-CANCELED, NUN-MAINTAINED) in total.

The context preamble, wh-question, and two response sentences listed in Table 3.2 were presented in an animated video clip (see Appendix A.2). The clip used an interface resembling mobile text messages and displayed the question and the two response sentences at 5-second intervals. This design aimed to mimic a text messaging environment, encouraging participants to respond based on their judgments about

discourse and pragmatics rather than grammar.

Once the stimuli were presented, participants were asked an experimental question: How appropriate is the speaker's second sentence to follow the first sentence? They rated their responses on a 7-point scale (1 = Not at all appropriate, 7 = Highly appropriate). If a marker's exhaustivity is not cancelable, participants would report lower ratings in the CANCELED conditions. Conversely, if a marker's exhaustivity is interpreted as cancelable, the ratings would not differ between the CANCELED and MAINTAINED conditions. A total of 24 critical items were designed for each of the four conditions, resulting in 96 stimuli.

In addition to the 24 critical items, participants were presented with 24 fillers. To disguise the experimental manipulations arising from different permutations of the markers in the first response sentence and the continuation in the second sentence, the fillers included two different markers—-*man* ('only') and *-cocha* ('even')—in the second response sentences, while still having *-lul-* or *-nun*-marked objects in the first response sentences. As mentioned in the previous section, the *-man* marker, being a lexical exclusive equivalent to the English 'only', indicates exhaustivity as its semantic meaning. The *-cocha* ('even') marker signals the inclusion of the noun phrase. These markers appeared in the canceled or maintained continuations to create the fillers, as illustrated in (38).

(38) Q. Yuna-nun meo-l mantul-ss-eo?
Yuna-TOP what-ACC make-PST-INT
'What did Yuna make?'

a. Sandwich-{lul/nun} mantul-ss-eo. Cookie-**man** mantul-ss-eo.
sandwich-CF/CT make-PST-DEC cookie-only make-PST-DEC
[Contradictory] 'Yuna made sandwiches. Yuan only made cookies.'

b. Sandwich-{lul/nun} mantul-ss-eo. Cookie-**cocha** mantul-ss-eo.
sandwich-CF/CT make-PST-DEC cookie-even make-PST-DEC
'Yuna made sandwiches. Yuan even made cookies.'

c. Sandwich-{lul/nun} mantul-ss-eo. Cookie-**man ahn** mantul-ss-eo.
sandwich-CF/CT make-PST-DEC cookie-only NEG make-PST-DEC
'Yuna made sandwiches. Yuan only didn't make cookies.'

d. Sandwich-{lul/nun} mantul-ss-eo. Cookie-**cocha ahn** mantul-ss-eo.
sandwich-CF/CT make-PST-DEC cookie-even NEG make-PST-DEC
[Contradictory] 'Yuna made sandwiches. Yuan didn't even make cookies.'

In the actual experiment, participants saw each lexicalized item appearing in only one type of filler construction among (38) (a)-(d). With the *-man* and *-cocha* markers, the second sentences in (38) (a) and (d) create logically contradictory continuations. Thus, these two types of fillers represent semantic violations (in contrast to the pragmatic violations targeted in the critical items) and provide a baseline to confirm whether ratings of critical items were judgments about pragmatic phenomena rather than grammatical or semantic phenomena. Accordingly, in the following section reporting results, ratings obtained from these two contradictory fillers are compared with the ratings obtained from critical items.

Lastly, the experiment included four attention check questions. These were presented every 9-10 questions throughout the experiment, each positioned after the preamble on the conversational context was presented. The attention check questions asked about the event that the characters from the previously presented preamble planned to do (e.g., *planned to run a marathon* in Table 3.2), and participants were asked to choose the correct summary of the characters' plan from four response options.

Critical items were counterbalanced to create four different versions of the questionnaire, ensuring that each participant saw the same lexicalized critical item only once, appearing in a single condition. Fillers and attention checks were consistent across all participants. Each questionnaire presented questions in a pseudo-randomized order, starting with three fillers at the beginning of the experiment. No more than two critical items or two fillers appeared consecutively, with four attention checks spaced evenly throughout, appearing every 9-10 questions.

3.3.3 Data exclusion

Among the 71 participants who completed the experiment, 11 were excluded because they had participated in the pilot study. This left 60 unique participants who had not participated in the pilot studies. An additional 26 participants were excluded for failing to correctly answer 3 or more attention check questions. The responses from the remaining 34 participants were used in the analysis. All remaining 34 participants reported Korean as their first language, with 1 participant noting they had learned English in addition to Korean from birth. The participants' ages ranged from 20 to 62 (mean = 37.6, median = 32), with a balanced gender distribution (female = 18, male = 16).

3.4 Result

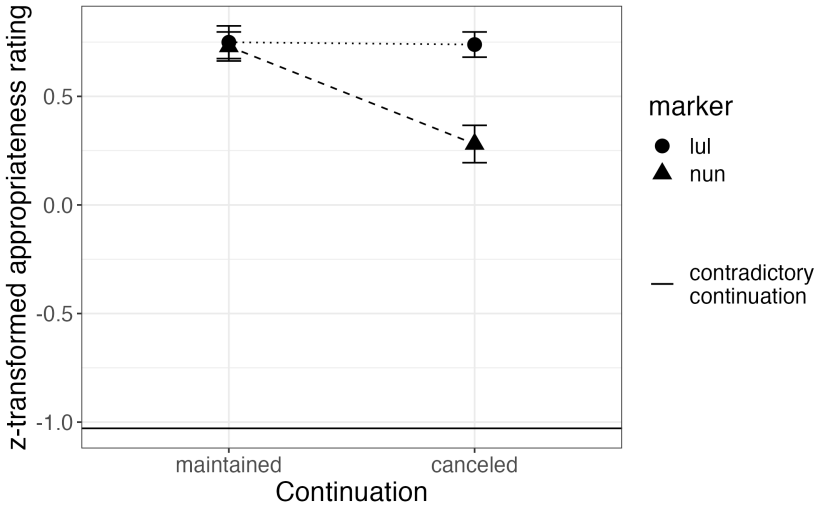


Figure 3.1: Means and 95% CIs of z-transformed appropriateness ratings gathered from Experiment 2.

Ratings of appropriateness reported on the 7-point scale are first transformed into z-scores, standardizing the scores within each participant using the mean and standard deviations of each participant's ratings. This makes the result gathered from different participants more comparable. Figure 3.1 summarizes the ratings. Each point in the figure represents the mean of the z-scores obtained in each of the four critical conditions, and the solid horizontal line in the bottom represents the mean of ratings given to the contradictory fillers (c.f., (38) (a) and (d)). Overall, the NUN-CANCELED condition elicited relatively lower ratings compared to the other three conditions. In addition, four critical conditions, which targeted evaluations on pragmatic uses of the marker, had ratings far above the contradictory fillers. The disparity between critical conditions and contradictory fillers confirms that participants rated critical items for their pragmatic felicity rather than for their semantic well-formedness.

Comparing the *-lul* marker conditions first, the MAINTAINED continuation and CANCELED continuation did not bring significant differences in appropriateness ratings ($t(203) = 0.24, p = 0.810$). This insignificant difference suggests that the *-lul* marker could be followed by either maintained or canceled exhaustivity and a felicitous discourse could be constructed. On the other hand, the *-nun* marker conditions showed significantly lower ratings in the CANCELED continuation ($t(203) = -7.70, p < 0.001$). This demonstrates that the *-nun* marker's canceled exhaustivity brought out infelicitious continuations.

A mixed-effects model was fit to confirm that the changes in object marker modulates the appropriateness of the following sentence. The `lmerTest` package in R (Kuznetsova et al., 2017). The model predicted z-transformed appropriateness ratings with fixed factors of the marker, continuation, and the interaction of the two factors. The model also included random slopes for each participant's interaction with the predictors and random intercepts for each lexicalized item. The regression formula of this 'full model' is shown in (39). Compared with the 'baseline model' in (40), which only included random effects included in the 'full model', the marker and continuation factors and their interaction in the full model significantly improved the prediction of the appropriateness ratings ($\chi^2(3) = 21.38$, $p < 0.001$).

(39) Full model: `z_rating ~ marker*continuation + (1+marker+continuation|participant) + (1|item)`

(40) Baseline model: `z_rating ~ (1+marker+continuation|participant) + (1|item)`

Table 3.3 shows the regression fit of the full model in (39). The reference levels of the model were set as Marker: LUL and Exhaustivity: MAINTAINED. As shown in the table, change in the marker factor to NUN does not bring any changes in appropriateness ratings. CANCELED continuation does not change the appropriateness ratings either. However, the interaction between the NUN marker and CANCELED continuation predicts significantly decreased appropriateness ratings. This demonstrates that canceling the *-nun* marker's exhaustivity implicature is perceived as significantly less appropriate. Post-hoc pairwise comparison confirmed the interaction between the marker and continuation. LUL-MAINTAINED and NUN-MAINTAINED showed significant difference ($\beta = 0.46, se = 0.05, p < 0.0001$), and NUN-MAINTAINED and

Table 3.3: A mixed effects model predicting z-transformed appropriateness ratings with the marker, continuation, and the interaction of the two factors, while controlling for the random effects of participants and lexicalized items.

	z-transformed appropriateness rating		
Predictors	Estimates	Standard error	p
(Intercept)	0.75	0.06	<0.001
Marker: NUN	-0.02	0.04	0.645
Exhaustivity: CANCELED	-0.01	0.07	0.879
Marker: NUN × Exhaustivity: CANCELED	-0.44	0.09	<0.001
Marginal R^2 / Conditional R^2	0.131 / 0.413		

NUN-CANCELED showed significant difference ($\beta = 0.45, se = 0.09, p < 0.0001$).

3.5 Discussion

The *-lul* marker conditions demonstrated positive ratings across scenarios where exhaustivity was either maintained or canceled. This result can be interpreted in two ways due to the marker's polyfunctionality across grammar and discourse domains. Given the *-lul* marker's exhaustive interpretation as discussed in Chapters 2 and 3, one plausible interpretation is that the marker's exhaustivity is cancelable, leading to consistently positive ratings regardless of whether exhaustivity was maintained or canceled. Another plausible interpretation is that these consistently positive ratings might stem from the marker's grammatical function, independent of any discourse-related interpretations. In experimental stimuli, the *-lul* marker always appeared after grammatical objects, yielding grammatically correct constructions. Thus, participants were able to assign positive ratings to the *-lul*-marked sentences for the marker's grammatical appearance, not considering its pragmatic felicity. To verify whether the current results reflect the cancelability of the *-lul* marker's exhaustive interpretation or merely the absence of exhaustivity, exhaustivity of the *-lul* marker needs to be observed in a similar experimental setting. For this, a follow-up experiment is designed in Chapter 4, and direct evidence of the *-lul* marker's association with exhaustivity will be presented. Anticipating results reported in Chapter 4, in the current discussion, I suggest that the observed pattern for the *-lul* marker indicates the cancelability of its exhaustivity implicature, rather than the non-existence of an exhaustivity implicature. Further details on the follow-up experiment are provided in Chapter 4.

The current result demonstrates that the *-nun* marker conveys non-cancelable exhaustivity, given the lower ratings observed with the *-nun* marker followed by the canceled exhaustivity. The claim that CT is associated with cancelable exhaustivity implicature (Büring, 2003) is not borne out, at least with the *-nun* marker indicating CT in Korean. In addition, the proposal that the *-nun* marker's exhaustivity is a non-at-issue meaning (J. Kim, 2018) is not supported by the current result. Considering the cancelability observed for non-at-issue implicatures (Destruel et al., 2015; DeVeaugh-Geiss et al., 2015), if the *-nun* marker were conveying exhaustivity as non-at-issue implicatures, we should have observed cancelability

with the *-nun* marker conditions.

From Experiment 1, it is evident that the *-lul* and *-nun* markers differ in terms of how they are associated with shared knowledge in constructing exhaustivity implicatures. Based on the result that the *-lul* marker is more likely to convey exhaustivity without a set of alternatives, one might be inclined to suggest that exhaustivity implicature is a structurally encoded meaning of the *-lul* marker. Additionally, because the *-nun* marker is likely to convey exhaustivity when its discourse context includes a set of alternatives, one might conclude that the *-nun* marker exhibits exhaustivity as a context-driven implicature, rather than as a structurally encoded meaning. However, the cancelability of the *-lul* and *-nun* markers' exhaustivity observed in the current chapter does not support such clear-cut conclusions.

If the *-lul* marker were to be observed with a stronger association with exhaustivity, as indicated by the result from Experiment 1, the *-lul* marker would have demonstrated non-cancelable exhaustivity in the current experiment. Contrary to this expectation, the *-lul* marker in fact demonstrated cancelable exhaustivity, which is characteristic of a conversational implicature. Additionally, if the *-nun* marker conveys exhaustivity only when contextualized alternatives are available, as observed in Experiment 1, its exhaustivity should have been cancelable, showcasing the feature of a conversational implicature. However, the *-nun* marker exhibited non-cancelable exhaustivity implicature, a feature of a conventional implicature.

Thus, results from Experiment 1 and 2 together do not support the dichotomous classification of the *-lul* and *-nun* as a linguistic structure that conveys either conventional or conversational implicature. In other words, the two markers' patterns do not support the conclusion that one marker conveys exhaustive interpretation more strongly than the other. Instead, each of the two markers has its unique pattern in interacting with discourse context and licensing exhaustive interpretation. The *-lul* marker is likely to convey exhaustivity without the support of discourse context (c.f., Experiment 1), but its exhaustivity can be canceled by the proceeding discourse. The *-nun* marker, on the other hand, requires a support of the discourse context to license exhaustivity implicature, but once the marker conveys the implicature, it acannot be canceled in the proceeding discourse.

These patterns further support the observation made in Chapter 1 about the relation between grammatical and discourse-related functions of Korean morphology. The organization of alternative ob-

ject markings in Korean appears to have a less clear division between grammatical markers and discourse markers. It is not the case that *-lul* marker, which is considered a canonical grammatical case marker, exhibits relatively weaker association with the discourse function of indicating exhaustivity implicature. In addition, the *-nun* marker, which primarily serves as a discourse marker, conveys discourse meanings constrained on the shared knowledge and discourse context. However, when the *-nun* marker conveys exhaustivity implicature within a discourse context, it conveys non-cancelable exhaustivity implicature, showcasing much stronger association to the discourse meaning. Thus, each marker exhibits polyfunctionality, and its grammatical and discourse functions are uniquely blended and weighted in a way that does not classify one marker as a more primarily "grammatical" or "discourse" marker than the other.

The current experiment confirms that the two markers are associated with different discourse patterns even when both markers convey exhaustivity given a set of alternatives. Experiment 1 showed that the *-lul* and *-nun* marker show different patterns when the set of alternatives is not present in the discourse context. Different discourse patterns of each marker's exhaustivity are detected when those patterns are paradigmatically contrasted. Overall, the contrasted patterns of *-lul* and *-nun* marker's exhaustivity demonstrates that linguistic structures are organized such that each form exhibits unique patterns in discourse processing, even if these structures do not reflect clear division of labor between grammar and pragmatics.

3.6 Conclusion

The current chapter provides experimental results further supporting the paradigmatic contrast between *-lul* and *-nun* markers. When both markers convey exhaustivity given a set of alternatives, the *-lul* marker indicates CF information, while the *-nun* marker indicates CT information. Cancelability of each marker's exhaustivity is tested to diagnose the type of association between each marker and its exhaustivity implicature. CF-marking *-lul* showcased cancelability of its exhaustivity implicature, while CT-marking *-nun* exhibited non-cancelability of its exhaustivity implicature. Although the non-cancelability of the *-nun* marker's exhaustivity implicature is confirmed, the *-lul* marker's result requires further confirmation: Because the *-lul* marker can function as a grammatical case marker without a discourse function,

it must be confirmed that the current experimental result is associated with the marker's discourse function, not its grammatical function. In particular, further evidence is required to show that speakers indeed perceive an exhaustivity implicature with the *-lul* marker in the experimental setting similar to the current one. This is directly investigated in the following chapter.

4 Experiment 3: Exhaustivity of *-lul*

4.1 Introduction

In the previous chapter and Experiment 2, we investigated the cancelability of an exhaustive interpretation. The *-lul* marker was associated with consistently positive ratings when exhaustive interpretation was canceled or maintained in the proceeding discourse. This could have resulted if exhaustive implicature from the *-lul* marker were cancelable. However, another possible explanation for this result is that the *-lul* marker only indicated a grammatical object function and did not convey an exhaustive implicature after all, contrary to what the experimental design had intended. That is, positive ratings observed with the *-lul*-marked objects could have been from the marker's grammatical appearances following grammatical objects, not the cancelability of its exhaustivity implicature. Thus, the current chapter aims to confirm the association between the *-lul* marker and exhaustive interpretation. To this end, an experiment is designed to compare the *-lul* marker and null-marking with respect to their association with exhaustive interpretation.

In this chapter, I will first review previously proposed frameworks that explain overt and covert uses of case markings as systematically organized patterns of language. These frameworks mainly explored the association between grammatical features of noun phrases and overt or covert appearances of case markings. However, the discussion about Korean suggests that overt and covert *-lul* markings are determined by the interaction between grammatical and discourse features of noun phrases. With an experiment, this chapter investigates whether exhaustivity is another discourse feature that leads to systematic uses of overt *-lul* markings.

The current chapter continues with section 4.2, where I provide an overview of previous frame-

works explaining overt and covert case markings. I discuss how these frameworks are related to the patterns of *-lul* marking in Korean. I describe the experiment in section 4.3, and report the results in section 4.4. The results are discussed in section 4.5. Given the context of the previous frameworks discussed in section 4.2 and the experimental results of *-lul*, I discuss what generalizations could be made about systematic organizations of marking patterns in language. The chapter is concluded in section 4.6.

4.2 Previous studies

Languages are often hypothesized to be systems of human communication that are systematically structured to facilitate efficient exchange of information (Piantadosi, Tily, & Gibson, 2012; Gibson et al., 2019). Efficiency in communication is achieved when a message is transmitted with minimal effort by both the speaker and listener. Communicative effort is often quantified by the length of messages, because shorter messages are easier for speakers to produce and present fewer signals for listeners to process (c.f., Zipf, 1929; Shannon, 1948). Therefore, as an efficient communication system, language exhibits structures that maximize the chances of successfully conveying the intended message while being as short as possible.

In using a language, speakers often alternate between various linguistic expressions and structures to maximize efficiency (Frank & Jaeger, 2008; Jaeger, 2010). In particular, studies within the Uniform Information Density (UID) framework suggest that speakers make choices that maintain a roughly constant information density per unit of time (Aylett & Turk, 2004; Jaeger, 2006; Levy & Jaeger, 2006; Frank & Jaeger, 2008; Jaeger, 2010). When upcoming information is less predictable, speakers use longer forms to avoid a surge in the flow of information. Conversely, when conveying highly predictable information, speakers may reduce utterance length to conserve communicative effort.

For instance, the complementizer *that* is an optional structure in English. In (1) (a), the pronoun *I* in the complement clause is more easily predicted based on the preceding possessive pronoun *my* in the main clause. In contrast, (b) shows that the pronoun *we* in the complement clause is less predictable. Under the UID account, speakers produce language in a way that avoids a spike in uncertainty and are more likely to use a full form to distribute the uncertainty more evenly across the utterance. Thus, *that*

is expected to be omitted more frequently in predictable contexts like (1) (a) than in (b). A corpus analysis confirmed this tendency, demonstrating the efficiency achieved by the optional use of linguistic expressions (Jaeger, 2010).

(1) a. My boss thinks (that) I am absolutely crazy.

b. My boss thinks (that) we are absolutely crazy.

Similar to how complementizers can be overt or covert in English, the overt and covert uses of case markings are discussed as another example of linguistic efficiency. Overt and covert grammatical case markers in Korean have been labeled with various terminologies—bare nominals (H. S. Lee & Thompson, 1989), case drop (C. Lee, 1994), case ellipsis (H. Lee, 2006a), and zero particle (D.-Y. Lee, 2002). Alternations between overt and covert case markings of direct objects, in particular, have often been discussed under the label of Differential Object Marking (DOM) (c.f., Bossong, 1991). Despite different labels, these terminologies all aim to explain the alternative overt and covert appearances of case markers as systematic patterns.

One framework for overt and covert markings suggests that the prototypicality of objects determines their case markings. The PROMINENCE SCALE of direct objects is proposed as a measure of their prototypicality. Grammatical features of objects, namely the animacy and definiteness features, are often discussed as the aspects that determine the prominence of direct objects. Direct objects whose referents are human are considered the most prominent because direct objects are less likely to be human than animate, and less likely to be animate than inanimate. This forms an animacy hierarchy, as presented in (2) (a). Higher position in the hierarchy noted with the greater-than sign (>) indicates higher prominence. Additionally, definite referents are more prominent as direct objects, because direct objects are less likely to be definite and more likely to be indefinite. These observations are organized into the definiteness hierarchy presented in (2) (b).

(2) a. Animacy hierarchy: Human > Animate > Inanimate

b. Definiteness hierarchy: Personal pronoun > Proper noun > Definite NP > Indefinite specific NP > Non-specific NP

In determining which combinations of animacy and definiteness features are marked and which are not, the principles of ICONICITY and ECONOMY operate in conjunction (see Aissen, 2003). The iconicity principle favors more prominent and less prototypical features being overtly case-marked (Haiman, 1980, 1983). This principle is applied to explain the cross-linguistic tendency where a direct object with higher prominence is more likely to be overtly case-marked (Croft, 1988, 2003; Silverstein, 1976, 1981). Considering that the main function of case markings is to discriminate grammatical relations between phrases (Comrie, 1988), the iconicity principle distinguishes objects with less typical features (e.g., human and definite) from other phrases. Meanwhile, the economy principle operates to avoid case marking. Objects whose referents are lower in the prominence scale (e.g., inanimate and indefinite) have more typical features as grammatical objects and require less disambiguation from other phrases. In this case, the economy principle leads to covert case markings.

The operation of the iconicity and economy principles is illustrated by the *ko* marking in Hindi. In Hindi, a human direct object (i.e., higher on the prominence scale) typically takes the *ko* marking, regardless of whether it is definite or indefinite. This is demonstrated in (3) (taken from Comrie, 1988). In (a), the human object *bacce* 'child' is marked with *ko*, and could be felicitously interpreted as a definite or an indefinite. In contrast, the null-marked 'child' *baccā* in (b) results in an infelicitous sentence (the judgment indicated with '?' comes from the author). This demonstrates the operation of iconicity principle in relation to the prominence scale, enforcing an overt marking on human objects that are high on the prominence scale.

(3) a. Aurat bacce **ko** bulā rahī hai.
woman child ACC calling PROG is
'The woman is calling the/a child.'

b. ?Aurat baccā bulā rahī hai.
women child calling PROG is

When an object is inanimate, the likelihood of *ko*-marking varies according to the referent's definiteness. To begin, inanimate objects that are indefinite do not appear with *ko*. This is illustrated in (4), where the inanimate and indefinite object 'letters' does not appear with *ko*. This suggests that when objects are lower in the prominence scale, the iconicity principle is less prioritized, and the economy

principle is prioritized to avoid case markings.

(4) Patr likhie.
letters write-POL
'Write letters please.'

Inanimate and definite objects may, but do not always, take the *ko* marking. As illustrated in (5) (a), the inanimate object 'letters' appears with *ko* when the referent is definite (as modified by *un* 'those'). However, this is not always the case. The object on the same prominence scale (inanimate and definite) does not appear with *ko* in (5) (b), and the sentence remains felicitous.

(5) a. Un patrõ **ko** paṛhie.
those letters ACC read-POL
'Please read those letters.'

b. Ye patr paṛhie.
these letters read-POL
'Please read these letters.'

These patterns in Hindi demonstrate that both animacy and definiteness hierarchies are considered in determining the *ko*-marking, and that the iconicity and economy principles operate simultaneously to decide where the marking should appear. Besides Hindi, substantial cross-linguistic data show the operation of these principles (Aissen, 2003; Croft, 2003; Comrie, 1988; Haiman, 1983; Silverstein, 1976, 1981). Given this, overt and covert case markings are thought to arise from the tension between the iconicity and economy principles. Overt case markings tend to appear when there is a need to disambiguate the grammatical function of an object due to its less typical grammatical features. Covert markings are more likely to be found when the grammatical function of an object is unambiguous due to its prototypical grammatical features.

Grammatical features in the prominence scale have been considered for overt and covert case markings in Korean. H. Lee (2006b, 2007) reported frequencies of *-lul*-marked and null-marked objects obtained from the CallFriend Korean corpus (Canavan & George, 1996). The *-lul*-marked and null-marked objects were classified based on the referents' animacy, definiteness, and person. Person

features are often considered in addition to animacy and definiteness, as another dimension that reflects the prominence of objects. As illustrated in (6), 1st and 2nd person are considered more prominent and less prototypical features of direct objects, while the 3rd person feature is considered less prominent and more prototypical for direct objects.

(6) Person hierarchy: first, second > third

Following the previous discussion on iconicity and economy principles, if referents are human or animate, 1st or 2nd person, and indefinite, objects are considered having atypical features as a grammatical object. These objects are expected to accompanied by the *-lul* marker. Meanwhile, inanimate, definite, and 3rd person objects are lower in the prominence scale, and they are expected to be null-marked.

In the corpus analysis comparing animate and inanimate objects, *-lul* markings were observed more often with human (45.3%) and animate (49.3%) objects than with inanimate (25.1%) objects. Notably, inanimate objects, which are lower in the prominence scale, were null-marked significantly more often (74.9%) than marked with *-lul* (25.1%). Examples of sentences with null-marked objects found in the corpus are presented in (7) and (8), where the ∅ notation is added by me to indicate where the *-lul* marker would have appeared. Null-markings were found more often with inanimate objects (i.e., lower prominence), such as 'energy' in (7) (b), than with animate objects, such as 'employee' in (7) (b).

(7) Animacy

a. Kun kenmwul-i energy-∅ tel ssu-ya hay.
big building-NOM energy less consume must
'Big buildings should consume less energy.'

b. Wuri hoysa-to naynyun cikwon-∅ camchwuk hay.
our company-also next year employee reduction do
'Our company too will reduce the number of employees next year.'

In relation to definiteness, definite objects were observed more often with *-lul* markings. Pronouns and names, which are high in the prominence scale, were observed with *-lul* markings 67.6% and 57.1% of the time, respectively. All other types of objects besides pronouns and names were observed

more often with null-marking (73.2%) than with *-lul*-marking (26.8%). For instance, in (8) (a), the noun phrase *yeca* 'women' is indefinite, referring to a non-specific group of people. In the corpus, these indefinite phrases were more likely to be null-marked than *-lul*-marked. In (8) (b), on the other hand, the person's name *Hyunswu* is definite, referring to a specific person. These definite phrases were less likely to be *-lul*-marked and more likely to be null-marked.

(8) Definiteness

a. Gongchay-eyse yeca-∅ wuday hapnita.
hiring-in women give priority
'We give priority to women in hiring.'

b. Wuyenhi hakkyo-eyse Hyunswu-∅ mannass-e.
accidentally school-at Hyunsoo met
'(I) came across Hyunsoo at school.'

Lastly, in relation to the person hierarchy, *-lul* markings were observed more often with 1st (45%) and 2nd person (46.7%) than with 3rd person (33.7%). 3rd person objects, which are lower in the prominence scale, were more frequently null-marked (66.3%) than *-lul*-marked (33.7%). To illustrate, 3rd person objects, such as *namca* 'men' in (9) (a), were found more often without the *-lul* marking. Null-marked 1st person objects, such as *wuli* 'we' in (9) (b), were observed less frequently than the null-marked 3rd person objects.

(9) Person

a. kensel hoysa-nun namca-∅ senho hay.
construction company-TOP men prefer do
'Construction companies prefer men.'

b. Youngmi-ka wuli-∅ moim-ey choday hays-e.
Youngmi-NOM us meeting-DAT invitation did
'Youngmi invited us to the meeting.'

Overall, Lee's corpus analysis show that the proportion of *-lul*-marked objects and null-marked objects confirm the general pattern suggested with the iconicity principle and prominence scale, in that objects higher in the prominence scale were more likely to appear with the overt *-lul* marker. Although

the corpus analysis suggests a correlation between marking patterns and features of objects, it does not confirm that certain features of objects cause the overt marking patterns we observe in language. To establish a causal relationship between patterns in language use and the hypothesized principles, previous studies have utilized miniature language learning paradigms in their experiments (Fedzechkina, Jaeger, & Newport, 2012; K. Smith & Culbertson, 2020; Tal, Smith, Culbertson, Grossman, & Arnon, 2022). In these experiments, participants learn an artificially created language, and their use of the language is later assessed to determine whether certain features have caused the observed patterns in their language use.

Fedzechkina et al. (2012) conducted a miniature language learning experiment to test whether the need to avoid ambiguity could be the main motivation shaping the overt and covert marking patterns we observe. Experimenters taught participants an artificial language designed to observe variations in overt and covert marking patterns.[1] The artificial language described events involving an animate subject and an animate or inanimate object (e.g., a chef hugging a referee, a chef lifting a chair). Crucially, input materials presented non-systematic and irregular patterns. The (non-)occurrence of the case marker was not conditioned on the typicality of the argument being marked. In other words, both inanimate (typical feature of objects) and animate (atypical feature of objects) objects occurred with the case marker 60% of the time. Participants encountered SOV 60% of the time and OSV 40% of the time.

After the learning phase, participants produced sentences using the miniature language. Although the input language did not exploit variations in marking patterns as cues to resolve grammatical roles of arguments, participants restructured and reorganized the grammar in their usage of the language. In the sentences they produced, they exhibited a systematic tendency to mark animate objects more often than inanimate objects. Participants' usage showed regularized case markings, which were used as cues to resolve ambiguous argument roles. This supports the operation of iconicity principle in the organization of grammar and language.

Overall, studies reviewed thus far provide evidence of iconicity and economy principles operating in relation to the prominence scale of objects, and overt and covert *-lul*-markings in Korean are also observed under this framework. The relative frequency of null-marked objects compared to *-lul*-marked

[1] Fedzechkina et al. (2012) conducted two experiments. The first experiment investigated object markings, while the second experiment focused on subject markings. I review the results of the first experiment here; the second experiment yielded similar results supporting the iconicity principle.

objects increases as the referent exhibits more typical features as grammatical objects. Considering the causal link found between animacy features and alternative markings (Fedzechkina et al., 2012), the overt and covert *-lul* markings in Korean appears as the result of the operations of iconicity and economy principles.

However, in Chapter 1, I demonstrated that it is not only the grammatical features that determine overt markings in Korean. Discourse features of objects, in particular, Information Structure (IS) types were observed as a factor associated with overt and covert markings in Korean. Then, how would the iconicity principle operate in relation to the discourse features? The account of Horn (1981), reviewed in Chapter 3, demonstrates that the iconicity principle is observed not only with grammatical features but also with discourse and pragmatic meanings. Specifically, Horn contrasts alternative constructions where a speaker has "gone out of her way" with simpler constructions. For instance, the phrase *managed to pass* in (10) (a) is contrasted with simpler construction *passed* in (10) (b). When more complex construction is used in (a), pragmatic implicature is indicated. This is contrasted with (10) (b), where there is no pragmatic implicatures indicated.

(10) a. Bill managed to pass the test. (→ Passing the test was difficult for Bill.)

b. Bill passed the test.

(10) shows the correlation between additional linguistic structure (*managed to*) and the implicature about the unusual description of the scene. Similar to the overt markings associated with less typical grammatical features, in (10), a more complex linguistic expression is associated with a less typical interpretation of a sentence. By showing the close association between alternative linguistic structures and discourse meanings, Horn's account supports the link between the discourse prominence of objects and their markings that I continue to explore in the following.

Previous studies proposed information structures of objects as additional discourse features encoded by case markers in Korean. In particular, studies in the generative grammar framework view morphological markers as realizations of underlying informational statuses, and aim to make generalized predictions on overt and covert appearances of grammatical case markers in Korean based on the infor-

mational statuses of objects (Ko, 2001; S.-N. Kwon & Zribi-Hertz, 2008).

For instance, Ko (2001) suggests that it is the CF status of an object NP that determines the alternation between *-lul*-marking and null-marking. Based on the alternative semantics account (Rooth, 1985), the *-lul* marker is suggested as a semantic operator that additionally denotes a set of alternatives for the focused constituent (c.f., section 1.5.1). This semantic function allows the *-lul* marker to satisfy the CF feature of an NP. In other words, it is the CF feature of an NP that licenses the *-lul* marker, and the marker must attach to the stem NP as an operator of the CF function.

S.-N. Kwon and Zribi-Hertz (2008) adopt Erteschik-Shir's (1997, 2007) framework of f(ocus)-structure to explain the overt and covert alternations of *-lul*-marking. F-structure is structural descriptions of information structure statuses: It is a level of representation distinct from syntactic structure and sensitive to discourse, but still assume to pertain the grammar of a sentence. As illustrated with (11), mapping constituents with their information structure roles results in the f-structure in (11) (b) that is distinct from the syntactic structure in (11) (a) (example taken from Erteschik-Shir, 2006).

(11) Q: What did John eat?

A: He ate the cake.

a. [He ate [the cake]$_{DP}$]$_{VP}$.

b. [He]$_{TOP}$ ate the [cake]$_{FOC}$.

Under this framework, constituents seen in the syntactic structure occupy structural positions in a sentence, and only those syntactic constituents can be assigned with information structure roles in f-structure. In addition to this assumption, S.-N. Kwon and Zribi-Hertz (2008) further presume functional positions in syntactic structures. Functional positions make constituents "visible" in the f-structure, and ultimately allow constituents to be assigned with information structure roles in the f-structure. In other words, syntactic constituents gain discourse functions only when they are in functional positions so that they are visible in the f-structure. They propose that overt *-lul* markings of a constituent makes the constituent visible in the f-structure, allowing it to be assigned with a discourse function. In contrast, null-marked objects fail to be visible in f-structure due to the absence of case markings, resulting in the

absence of discourse functions.

This account from S.-N. Kwon and Zribi-Hertz (2008) is further illustrated with examples in (12) (a) and (b), with infelicity judgments (noted with ??) made by the authors. In (12) (a), the 'book' is marked with *-lul*, making it visible in the f-structure. This visibility allows the argument to have a focus status in the initial phrase and to have a topic status in the second phrase (noted with $\varnothing_z$). In contrast, in (12) (b), the 'book' is null-marked and thus remains invisible in the f-structure. Consequently, the object NP is not assigned with any discourse functions in the initial phrase and cannot be assigned with topic status in the second phrase. Thus, eliding the NP as if it has a topical status in (12) (b) becomes infelicitous.

(12) a. Minswu-nun$_k$ **chayk-ul$_z$** ilk-eya hay $\varnothing_k$ $\varnothing_{\mathbf{z}}$ kot tollyecwu-eya ha-ketun.
Minsu-TOP book-LUL read-must DEC right.away return-must do-because
'(As for) Minsu$_k$, he$_k$ must read the book(s)$_z$ right away because he$_k$ must return {it/them}$_z$.'

b. ??Minswu-nun$_k$ **chayk$_z$** ilk-eya hay $\varnothing_k$ $\varnothing_{\mathbf{z}}$ kot tollyecwu-eya ha-ketun.
Minsu-TOP book read-must DEC right.away return-must do-because
Lit. '(As for) Minsu, he must do his reading right away because he must return {it/them}.'

Thus, Ko (2001) and S.-N. Kwon and Zribi-Hertz (2008) suggest that the informational statuses of focus and CF are another factor associated with overt *-lul* markings. The observation that overt *-lul* markings are seen with additional discourse features but null-markings are not suggest the operation of the iconicity principle. Under the iconicity principle, when objects have discourse features, in addition to grammatical features, additional markings on objects are favored to indicate the existence of discourse features.

H. Lee (2006a) provides experimental data demonstrating that both discourse features and grammatical features jointly determine the prominence of objects and overt *-lul* markings in Korean. In Lee's experiment, objects were manipulated to vary in animacy, definiteness, and CF status. For example, in (13), the target object 'you' in B's utterance has CF status, as it contrasts with 'her mother' in A's utterance and introduces new information. At the same time, its human and definite features make it more prominent as a direct object. In contrast, in (14), the target object 'a nice guy' in B's utterance does not have CF status and is presented merely as new information. At the same time, 'a nice guy' is indefinite,

making it less prominent than the definite 'you' in (13).

(13) [+Contrastive focus, Human, Definite]

A: Jinswu said that Youngmi was looking for her mother all night long.

B: No, she was looking for **{you-*lul*/you-∅}**.

(14) [-Contrastive focus, Human, Indefinite]

A: It's been more than two years since Mijin got a job. How has she been doing lately?

B: She's been doing well. She bought a house last year and is dating **{a nice guy-*lul*/a nice guy-∅}**.

In the experiment, participants were provided with stimuli designed in the structure of (13) and (14) and were asked to choose whether the target object phrase should be *-lul*-marked or null-marked to sound most natural. The result showed that all three factors (CF status, animacy, definiteness) simultaneously and independently influenced object marking. Main effects of these factors and their interactions were significant predictors of null-marking. Objects with CF status and with human and indefinite features yielded more *-lul* marked responses. Among the three factors, CF status was observed as the strongest predictor of null-marking, followed by animacy, and then definiteness. This suggests that discourse features not only contribute to marking patterns in Korean but are also more significant than grammatical features in determining object markings.

This result underscores that both grammatical and discourse factors significantly influence the use of *-lul* markings in Korean. The result reinforces the notion that the principle of iconicity operates across these dimensions, with both CF status and grammatical features contributing to overt case markings. The findings also highlight that CF status can be even more influential than grammatical features in determining *-lul* markings. This allows us to hypothesize that exhaustivity can be another factor that increases *-lul* marking and decreases null-marking.

Despite strong evidence supporting the iconicity principle, it has been suggested that it is not the iconicity principle itself that leads to more markings on more prominent features, but rather the FREQUENCY of features on prominence scale shaping the PREDICTABILITY of the features (Bybee & Hopper,

2001; Haspelmath, 2008, 2021). More frequent features and meanings are naturally more predictable, allowing them to be conveyed with less "coding" (i.e., less markings) without risking to convey ambiguity (Haspelmath, 2021). This relationship between frequency, predictability, and form is proposed as a universal framework for explaining the association between linguistic form and meaning, including the overt and covert forms of case markings and their meanings.

As an instance, Kurumada and Jaeger (2015) investigated the effect of predictability on object markings in Japanese. Similar to Korean, null markings are prevalent in conversational Japanese. In the first experiment of their study, predictability was manipulated through the animacy of objects. For example, participants first listened to sentences structured as (15) and (16). The stimuli sentences contained either animate (e.g., 'student' in (15)) or inanimate (e.g., 'fire engine' in (16)) objects, which appeared with or without the accusative marker *-o*. After being presented with stimuli, participants heard the verb of the sentence (e.g., heard 'saw' in relation to (15) and (16)) and were asked to recall the full sentence that they heard as stimuli and had that verb.

(15) Sensei-ga **seito-{o/∅}** ekimae-de mi-ta-yo.
teacher-NOM student-ACC/∅ station-LOC see-PST-DEC
'The teacher saw a student near the train station.'

(16) Sensei-ga **shobosha-{o/∅}** ekimae-de mi-ta-yo.
teacher-NOM fire engine-ACC/∅ station-LOC see-PST-DEC
'The teacher saw a fire engine near the train station.'

Animacy of objects was predicted to modulate case marking in sentences that participants recalled. Specifically, it was predicted that participants would overtly mark animate objects, even if those animate objects had not appeared with the *-o* marker in the initially presented stimuli. The result confirmed this prediction, and was explained with the predictability-based account. When human subjects and human objects appear together in a sentence without any case markers, the grammatical function of each argument becomes less predictable. This explains that more markings were observed on unexpected (i.e., human) features of objects. However, as pointed out by the authors themselves, this result did not disprove the iconicity principle. That is, the iconicity principle could still explain the result by suggesting that more prominent features of objects (such as humanness) receive more markings.

To provide more direct evidence for the predictability-based account, Kurumada and Jaeger (2015) conducted a second experiment. In the second experiment, participants went through the same protocol as in the first experiment, and received stimuli sentences and recalled those sentences with a verb as a cue. However, the plausibility of stimuli was manipulated in the second experiment, while having the animacy features of objects identical across stimuli. For example, participants were given with sentences illustrated in (17). Both sentences in (17) have human subjects and objects, but due to world knowledge, (a) is more plausible than (b).

(17) a. Hannin-ga keisatsu(-o) yonaka-ni osot-ta-yo.
criminal-NOM police officer-ACC night-at attack-PST-SFP
'The criminal attacked the police officer in the middle of the night.'

b. Keisatsu-ga hannin(-o) yonaka-ni osot-ta-yo.
police officer-NOM criminal-ACC night-at attack-PST-SFP
'The police officer attacked the criminal in the middle of the night.'

If it is not the grammatical feature itself but the predictability shaped by the grammatical feature that influences marking patterns, then sentences with lower predictability should exhibit more overt markings. In the result, it was observed that the overt markings in participants' recalled sentences did not vary in relation to the animacy features, but in relation to the predictability. This result confirms the association between predictability and overt marking patterns, and suggests that grammatical features on the prominence scale may not be directly linked to the marking patterns.

In Kurumada and Grimm (2019), the causal link between predictability and alternative markings is tested. The researchers conducted a miniature language learning experiment where participants learned a language that did not contain the hypothesized correlation between predictability and grammatical number category. Specifically, the miniature language featured two classes of nouns: "individuals," which were presented as single visual referents, and "collectives," which were presented as plural visual referents. In the learning input, "collectives" predominantly occurred with plural meanings (e.g., peas are shown instead of a single pea). Thus, for "individuals" (e.g., father), singular meanings were more predictable, while for "collectives" (e.g., pea), plural meanings were more predictable.

In this language, singular meanings were never marked with additional linguistic encoding. Plural

markings (*-ka*) were not obligatory but appeared in 66% (2/3) of the input given to the learners. According to the predictability account, participants were expected to increase the proportion of plural markings for "individuals" because the plural meanings of "individuals" were perceived as less predictable compared to the more frequently plural-marked "collectives."

In the assessment phase, participants were presented with a set of visual images that they needed to describe using the miniature language. As predicted by the predictability account, participants were more likely to produce the optional plural marker for multiple referents than for a singleton referent. Moreover, participants were significantly less likely to use the plural marker for collectives, even when they appeared as multiples. Since collectives were more frequently shown as plural in the learning input, this result suggests that null-markings were influenced by the predictability of their plural meanings. Participants tended to suppress overt markings when the plural meaning was more predictable.

Kurumada and Grimm (2019) designed a second experiment to determine whether the results of the first experiment could be attributed to the transfer of semantic knowledge from participants' native language (English) to the miniature language. For example, native English speakers might expect a word for 'ants' (i.e., collectives) to appear with a plural marker because 'ants' in English typically involve plural marking. To address this potential confound, the second experiment used visual referents in geometric shapes with no commonly known names, while keeping the grammar of the miniature language identical to that of the first experiment.

Results from the second experiment replicated those of the first experiment. When plural meanings were more predictable, overt markings were observed less frequently. This finding indicates that the results of the first experiment cannot be explained by the transfer of semantic knowledge from participants' native language. To summarize, Kurumada and Grimm (2019) demonstrated that predictability can cause variations in markings by having participants learn an artificially designed language.

As reviewed previously, marking patterns in Korean are influenced by both discourse and grammatical features. How can this be explained based on the predictability account? Previous research has shown that the overt appearance of the *-lul* marker is related to the predictability of discourse interpretation (H. S. Lee & Thompson, 1989; C. Lee, 1994). For instance, H. S. Lee and Thompson (1989) constructed a dataset consisting of informal narratives and dialogues. They analyzed contexts where null-

markings occurred and found that objects are more often null-marked if referents are part of the shared knowledge between interlocutors, established from previous discourse, shared experiences, or cultural background. For example, they compared the frequency of null-markings in single-speaker narratives and dialogues involving more than one speaker. In narratives, shared knowledge between speakers cannot be established due to the nature of the discourse involving only a single speaker. Accordingly, null-markings were observed less frequently in narratives than in dialogues. This observation aligns with the discussion in Chapter 1, where null-marking is associated with topic informational status, defined as presupposed information between interlocutors.

A similar account suggests that null-markings indicate that the information conveyed is rather familiar or expected, based on the interlocutors' shared knowledge (C. Lee, 1994). Examples used to support this account are shown in (18) and (19), with grammaticality judgments (* indicating grammatical/pragmatic ill-formedness) made by the author. In (18) (a), when the object *cha* ('car') is null-marked, the question is interpreted as inquiring about a familiar or expected event. This interpretation contrasts with (18) (b), where the object is *-lul*-marked and the interpretation involves unexpectedness.

(18) a. ney-ka cha-∅ takk-ass-ni?
you-NOM car wash-PST-INT
'Did you wash the car?' (The idea of '[your] washing the car or cars' is familiar or expected.)

b. ney-ka cha-**lul** takk-ass-ni?
you-NOM car-ACC wash-PST-INT
'Did you wash the car?' (I didn't expect it.)

C. Lee (1994) further points out that null-markings cannot appear in rhetorical sentences expressing unexpectedness, as shown in (19). If the *-lul* marker is used overtly, it indicates that "the event involved is rather unexpected and attention is paid." Thus, the generalization from (18) and (19) is that it is not only the informational status of the object referent, but also the expectedness of the event involving the object referent, that influences the overt use of the *-lul* marker.

(19) ney-ka cha-{lul/*∅} takk-a?
you-NOM car-ACC wash-INT
'You say you washed the car—how can I believe it?' (rhetorical question)

The predictability account for Korean markings is experimentally tested in H. Lee (2012). Specifically, the experiment was designed to test whether preferences for subject marking in Korean are modulated by the predictability of the subject argument. In the experiment, the predictability of the subject was manipulated by varying the givenness of the referent. For instance, in the "high predictability" condition illustrated in (20), the first sentence (a) explicitly mentioned the subject referent (e.g., 'Minsu'). The following sentence (b) presented an expression referring to the subject mentioned in the first sentence ('that person'). Half of the stimuli included the nominative *-i/ka* following the referring expression, while the other half did not.

(20) High predictability condition

a. ecey **Minswu-ka** i cip-ul sa-le o-ass-e. haciman na-nun ku
yesterday Minsoo-NOM this house-ACC buy-to come-PST-IND but I-TOP that
salam-hanthey nay cip-ul an phal-keya.
person-DAT my house-ACC NEG sell-will
'Minsoo came (here) yesterday to buy this house. But I won't sell my house to him.'

b. i cip-ul **ku salam-{i/∅}** swipkey phoki an ha-lkeya.
this house-Acc that person-NOM/∅ easily give up NEG do-will
'He (=Minsu) won't give up this house easily.'

In the "low predictability" condition, the first sentence in the stimuli did not explicitly identify a subject referent. This created a new informational status for the subject appearing in the second sentence, as it was mentioned for the first time in the discourse. This setup is illustrated in (21). In this example, the first sentence in (a) does not specify a subject referent, while the second sentence in (b) specifies a subject in response to the information inquired by the wh-question. The second sentence in this low predictability condition could also appear with or without the nominative *-i/ka*.

(21) Low predictability condition

a. i cenhwaki-nun nemwu olay toyse pelilye-ko hayss-nun-tey, acwu
this phone-TOP too old become throw.away-to do-COMP-CONJ, very
melccengha-key pyenha-yss-ney. ike-l **nwu-ka** kochy-ess-e?
fine-COMP change-PST-DEC this-ACC who-NOM repair-PST-INT
'(I) was going to throw away this phone because it is too old, but it became completely fine.

Who fixed (it)?

b. mit-ki elyep-keyss-ciman kuk-el **wuli-{ka/∅}** kochy-ess-e. kochi-l
believe-NR difficult-COND-CONJ that-ACC we-NOM/∅ repair-PST-DEC repair-NR
kes iss-umyen icey wuli-hanthey mathkye.
thing be-COND now we-DAT give
'(It is probably) Hard to believe, but we fixed it. Now if there is anything to fix, give it to us.'

Participants were asked to rate the acceptability of the second sentence on a 5-point scale. Based on the predictability account, null-markings were predicted to be more acceptable when the subject argument in the second sentence was set up with high predictability in the first sentence. The results confirmed this prediction. Null-markings were more acceptable in the high predictability condition than in the low predictability condition, while overt markings were more acceptable in the low predictability condition than in the high predictability condition.

Within the high predictability condition, null-marked subjects received significantly higher mean ratings compared to nominative-marked subjects. Conversely, in the low predictability condition, subjects marked with nominative *-i/ka* were rated as more acceptable than null-marked subjects. Additionally, OSV sentences containing null-marked subjects—typically dispreferred—had increased acceptability in the high predictability condition compared to the low predictability condition. This indicates that the predictability of the subject can mitigate the general dispreference for null-marked subjects. To summarize, H. Lee (2012) provide evidence suggesting that predictability of a phrase established in the discourse context is closely associated with the acceptability of overt and covert markings in Korean.

One thing to note is that the choices between overt and covert markings in Korean did not appear as categorical preferences but as gradient statistical preferences. Experimental results (H. Lee, 2006a, 2012) indicate that preferences for overt and null-marking fluctuated based on grammatical and discourse features of an object, but did not result in extreme preferences (e.g., close to 0% or 100% consensus among participants for the lowest (e.g., 0-point) or highest (e.g., 5-point) ratings) akin to judgments on grammaticality. These gradient preferences for overt and covert markings were also observed in other studies (H. Lee, 2015; Kurumada & Jaeger, 2015; Kurumada & Grimm, 2019). Thus, discourse features and predictability should not be viewed as rigid grammatical principles that produce ill-formed structures

but rather as probabilistic constraints that affect the likelihood of overt markings.

The studies reviewed thus far suggest that overt and covert markings in Korean are influenced by multiple factors, encompassing both grammatical and discourse elements. Various discourse statuses of objects were seen as factors that increase the likelihood of overt *-lul* markings. I reviewed two principles that explain this pattern. Per the iconicity principle, focus discourse status may increase the prominence of objects, and must be indicated with overt markings. Alternatively, under the predictability-based account, discourse status modulates the predictability, and overt markings may appear for less predicted discourse statuses.

Although considering discourse statuses of objects contributed to the understanding of overt and covert *-lul* markings, it has not been tested whether the pragmatic implicature is another factor that modulates overt *-lul* markings. Previous studies (Ko, 2001; H. Lee, 2012) suggest discourse features (focus and CF statuses) that are directly linked with objects can affect their markings. However, it has not been tested whether the exhaustivity implicature that arises from discourse features of objects as well as the interpretation of those features in the discourse context have a direct association with overt *-lul* markings.

Accordingly, I designed an experiment to examine whether exhaustivity implicature is a factor that increases overt *-lul* markings. The experiment is designed to elicit speakers' preferences for *-lul*-marked and null-marked objects when an exhaustivity implicature is either intended or not. Based on previous observations, it is predicted that an exhaustivity implicature will lead to increased use of the overt *-lul* marker. If speakers use the *-lul* marker to convey exhaustivity implicature, this would confirm the association between the *-lul* marker and exhaustivity implicature.

4.3 Methods

4.3.1 Participants

Participants were recruited via the online crowdsourcing platform (`https://gosurveasy.com/`). The crowdsourcing platform is operated in South Korea, and anyone with a valid Korean identification can join the website. 113 participants were initially recruited. All participants reported that they were born in South Korea and resided in South Korea at the time of their participation. As compensation,

participants were rewarded "points", a currency used on the crowdsourcing platform, which they could later transfer into monetary values. The experiment was approved by the IRB of UC San Diego and conducted according to its guidelines. The initially recruited participants had the age ranged from 17 and 63 (mean = 39, median = 39). 55 participants reported their gender as female, and 58 as male.

4.3.2 Materials

In order to observe the *-lul* marker's discourse function under a paradigmatic contrast with null-marking, forced-choice questions were designed to elicit speakers' choices between the overt and covert *-lul* marker. Table 4.1 lists the components used to construct one experimental item. The boldfaced components in the table indicate experimental manipulations, and only the italicized portions were presented in Korean in the actual experiment.

A context preamble illustrated in the first row of Table 4.1 was presented first as a stimulus. It had the same structure as those used in Experiment 2 (c.f., section 3.3.2): The context always identified three characters (e.g., *Sohee, Mina, and Yuna* in Table 4.1), an event they planned to engage in (e.g., *marathon race*), and two alternatives that were relevant to the characters' planned event (e.g., *medal*, *trophy*). In the context preamble, there was one character who did not follow through with the planned event (e.g., *Sohee*), and this character then asks a question presented in the second row of the table.

A question asked by one character was presented following the context preamble. Unlike in Experiment 2, this question appeared in two different forms: a wh-question in the WH-Q conditions, and a disjunction question in the DQ conditions. In the WH-Q conditions, the question contained the wh-word *meo-* ('what'), asking about an object that one of the characters from the context was engaged with (e.g., *What did Yuna receive?*). In the DQ conditions, objects introduced in the context preamble (e.g., *medal*, *trophy*) appeared in a nominal disjunction to form a disjunctive question (e.g., *Did Yuna receive a medal or a trophy?*).

In addition to the presence or absence of the wh-word, the two question formats differ in the appearance of the *-lul* marker. While the wh-word *meo-* ('what') in the WH-Q conditions is followed by the *-lul* marker, nominal disjunctions in DQ conditions do not contain the *-lul* marker. Both question formats are grammatical and pragmatically felicitous in colloquial Korean. Considering that the structure of a

Table 4.1: Structure of experimental items used in Experiment 3.

Context		*Sohee, Mina, and Yuna planned to attend the marathon race together. The race had medals and trophies for winners. Mina and Yuna attended the race together. Sohee didn't keep the promise, so she doesn't know who won what in the race. Later, Sohee sent the text message shown below to Mina.*
Question	**WH-Q**	*Sohee: Yuna-nun meo-l bat-ass-e?* Yuna-TOP what-ACC receive-PST-INT "What did Yuna receive?"
	DQ	*Sohee: Yuna-nun medal-ina trophy bat-ass-e?* Yuna-TOP medal-or trophy receive-PST-INT "Did Yuna receive a medal or a trophy?"
Intended message	**EXHAUSTIVE**	*Mina intends to say that Yuna only received the medal.*
	NON-EXHAUSTIVE	*Mina intends to say that Yuna received both the medal and the trophy.*
Experiment question		*From the sentences below, which one is the best for Mina to send as a response? Sentences below may not include everything that Mina has to say. Regardless, please select the option that you think is the best.*
Response options		• *Medal-ul bat-ass-e.* medal-LUL receive-PST-DEC "(Yuna) received a medal." • *Medal-∅ bat-ass-e.* medal receive-PST-DEC "(Yuna) received a medal."

preceding utterance can affect the structure of an upcoming utterance, even across speakers and dialogue turns (Bock, 1996; Levelt & Kelter, 1982), questions containing the *-lul* marker may lead experimental participants to choose *-lul*-marked responses. Given that the appearance of the *-lul* marker has been discussed in terms of formality and stylistic choices (Ko, 2001), an overt *-lul* marker in the question may lead participants to presume that they must adhere to the overt markings initiated in the question. Thus, the question format was manipulated into two experimental conditions to determine whether *-lul*-marked responses are chosen due to the marker following the wh-word in the preceding question.

Once the question was presented as a wh-question or disjunctive question, participants were prompted with a statement describing an intended message that they needed to convey in their response.

In EXHAUSTIVE conditions, it was stated that the speaker intends to indicate the exhaustive status of one alternative. In NON-EXHAUSTIVE conditions, it was described that the speaker intends to say that both alternatives from the context preamble hold the relation denoted by the predicate in the preceding question, so neither alternative has exhaustive status. These manipulations yielded a 2 × 2 experimental design.

Following the statement about the intended message, participants were asked to pick one option from the two response options. The response options were always a pair of sentences, varying only in the markedness of the object: One response had a *-lul*-marked object, and the other response had an unmarked object. Crucially, the given response options did not elaborate on both objects from the context preamble. Instead, the response options always mentioned only one object from the set. This design was intended to capture the (non-)exhaustivity implied by the *-lul*-marking or null-marking. Because participants had to choose a sentence about only one object from the two, they were prompted to choose the sentence that most closely expressed the intended exhaustivity. In other words, they were not given response options that had a complete description of the (non-)exhaustive status of the object but were instead given options that could imply the (non-)exhaustivity.

The experimental question clarified that the given response options "may not include all of the information" that the speaker has to convey and encouraged participants to pick one option from the two response options that they "think is the best."[2] Experiment questions were given as forced-choice questions; Participants had to choose one option from the two, even if they thought neither was the most optimal response to the question. Therefore, the better option for speakers to choose was the one that would more likely to deliver exhaustivity when exhaustivity was desired, and the one that would not mistakenly evoke exhaustivity when the exhaustive status of objects was not intended. Considering previous discussions (Aissen, 2003; H. Lee, 2006b), subjects from the preceding question and objects in the response options had prototypical features of subjects and objects. Subject referents were always human, and object referents were always non-human and definite. Thus, the null-markings on objects in one of the response options did not create any ambiguity in resolving the grammatical roles.

[2]In the exit survey, participants were asked to freely discuss any thoughts they had about the experiment. Some participants reported that they had to rely on different nuances of the response options, because sentences listing all of the contextualized entities (e.g., "Received both the medal and the trophy.") were not given as a response option. This confirms that the experimental materials and questions were understood as I intended them to be.

In addition, a few other aspects of the experimental stimuli were controlled to ensure that all stimuli could elicit responses based on colloquial and conversational uses of the *-lul* marker. Previous studies have suggested that the overtness of markers in Korean is related to the formality of the conversation and that null markings are rarely observed in formal speech (Ko, 2001). Considering this, all sentences in the experimental items appeared with informal sentence-ending particles. Presenting questions in an informal tone ensured that null-marking was not restricted in the answer sentence due to the formality of the conversation. Next, the response options had elided subjects and ended with a casual sentence-ending particle. This also signaled that participants did not have to choose sentences with all arguments and linguistic structures included, but rather more colloquially natural sentences. Lastly, none of the verbs in the critical items were light verb constructions ('noun(*-lul*) + *ha*(=do)'). Because the *-lul* marker after the complement noun of the light verb *ha* ('do') is considered optional in all genres and styles of Korean speech and writing (H. Lee, 2007), light verbs were not included in the stimuli to avoid introducing any preferences toward null-marking.

To summarize the design of the experimental conditions, the current experiment is designed to observe the association between the overt *-lul* marker and exhaustivity. While the conversational context allowed both overt and null marking options as felicitous, other factors that may promote the overt *-lul* marker were controlled. If exhaustivity is one of the aspects that requires overt markings, we should observe a systematic increase in the use of *-lul* when exhaustivity needs to be conveyed. Conversely, if exhaustivity is not one of the aspects that modulates overt *-lul* markings, then there will be no systematic patterns in the frequent use of *-lul*.

To disguise the fact that the experiment is about the alternative uses of the *-lul* and exhaustiveness, fillers were included in the experiment. After a context preamble with the same structure as the critical items was presented, fillers had a question about a subject. Similar to the critical items, questions in fillers were presented in the form of wh-questions or disjunctive questions. This is illustrated in (22): (a) shows an example of a wh-question asking about a subject with the wh-word *nu(ku)-* 'who', and (b) shows an example of a disjunction question, where two subjects from the context preamble are presented in a disjunction.

(22) Questions about a subject used in fillers

a. Sandwich-nun nu-ka mantul-ess-e?
sandwich-TOP who-NOM make-PST-INT
'Speaking of sandwiches, who made them?'

b. Neo-lang Yuna-chunge nu-ka sandwich-∅ mantul-ess-e?
you-and Yuna-between who-NOM sandwich make-PST-INT
'Between you and Mina, Who made sandwiches?'

Because the context preamble always introduced three characters (e.g., *Sohee, Mina, Yuna planned to make sandwiches together.*) and one of the character who did not participate in the event (*Sohee*) asked the question, contextually plausible answers to the questions in (22) were limited to either 'Mina', 'Yuna', or 'both Mina and Yuna'. Following the question presented in either a form of a wh-question or a disjunctive question, a statement about the intended message is presented, as illustrated in (23). The intended message required to list one character exhaustively, as stated in (23) (a), or to list both characters non-exhaustively, as stated in (23) (b).

(23) a. Exhaustive: Mina intends to say that only Yuna made sandwiches.

b. Non-exhaustive: Mina intends to say that both herself and Yuna made sandwiches.

Following the intended message, two response options about subjects were given in the fillers. In half of the fillers, the two response options had elided object and verb phrases (i.e., S(OV) sentences). These are illustrated in (24). Following the questions presented in (22), the response sentences did not have to include explicit object and verb phrases (e.g., *made sandwiches*).

(24) Mina-{ka/∅} (sandwich-lul) (mantul-ess-e).
Mina-NOM (sandwich-ACC) (make-PST-DEC)
'Mina (made the sandwiches).'

In the other half of the fillers, response sentences had elided object phrases (i.e., S(O)V sentences). These are illustrated in (25).

(25) S(O)V response options:

a. Mina-ka (sandwich-lul) mantul-ess-e.
Mina-NOM (sandwich-ACC) make-PST-DEC
'Mina made it (= the sandwiches).'

b. #Mina-∅ (sandwich-lul) mantul-ess-e.
Mina (sandwich-ACC) make-PST-DEC
Intended: 'Mina made the sandwiches.' (Alternative interpretation: 'X made Mina.')

Similar to the S(OV) responses in (24), *-ka*-marked and null-marked subjects constituted a pair of S(O)V response options in (25). However, unlike in S(OV) sentences where both *-ka*-marked and null-marked subjects created felicitous options, the null-marked subjects in S(O)V sentences are less preferred (e.g., (25) (b)), because they could result in ambiguous interpretations. H. Lee (2006b) also notes that null-marked subjects are generally dispreferred. The *ka*-marked S(O)V responses illustrated in (25) (a) are felicitous because the subject phrase is clearly identified by the overt nominative marker. Thus, by including some response options where the nominative marker is used for disambiguation of grammatical roles, fillers disguised the fact that the experiment is about exhaustivity in the intended message.

To summarize, fillers could consist of either a wh-question or a disjunction question (e.g., (22)), followed by an intended message statement that was either exhaustive or non-exhaustive (e.g., (23)), and then followed by S(O)V responses or S(OV) responses. 24 fillers were created in this structure, each appearing after a distinct context preamble.

There were four attention check questions throughout the experiment. Attention checks had the same components as critical items and fillers, but the response options consisted of a relevant response and an irrelevant response. Relevant responses pertained to one alternative previously mentioned in the context preamble, while irrelevant responses referred to an object not mentioned in the context preamble. For example, after presenting the preamble describing that characters planned to learn Japanese and Chinese, and the intended message stating the alternative(s), the relevant response "Sumi learned Japanese" and the irrelevant option "Sumi learned Taekwondo" were paired and presented as response options. These fillers were intended to check whether participants were paying attention to the experiment, particularly to the intended message statement and the response options.

The experimental materials that each participant saw are summarized in Table 4.2. Each of the 24 critical items appeared in one of four experimental conditions (WH-Q & EXH, WH-Q & NON-EXH, DQ

Table 4.2: Materials that each participant saw in Experiment 3.

Type	Response options		Counts
Critical items	(S)OV	• Obj-*lul* V • Obj-∅ V	24 (6 each for 2 question forms and 2 types of intended message)
Fillers	S(O)V	• Subj-*i/ka* V • Subj-∅ V	12 (3 each for 2 question forms and 2 types of intended message)
	S(OV)	• Subj-*i/ka* • Subj-∅	12 (3 each for 2 question forms and 2 types of intended message)
Attention check		• Relevant response • Irrelevant response	4

& EXH, DQ & NON-EXH), yielding 96 stimuli in total. Critical items were counterbalanced into four lists so that participants saw only one context preamble for each experimental condition. All participants saw the same set of 24 fillers and 4 attention checks. The order of stimuli was pseudo-randomized so that no more than 2 critical items appeared in a row and an attention check appeared every 9-10 questions. The order of response options was also randomized so that marked and unmarked options appeared as the first option by random chance. Similar to Experiments 1 and 2, experimental stimuli were displayed in a mobile text messaging interface, and the response options were presented inside text message bubbles (see Appendix A.3). This design simulated a text messaging environment and provided a colloquial context of the language use.

4.3.3 Data exclusion

Among 113 participants, 21 who had participated pilot studies and Experiment 2 were excluded from the final analysis. From the remaining 92, 2 participants were excluded because they did not indicate that they had learned Korean since birth. In addition, 3 participants who incorrectly responded to 3 or 4 attention check questions were excluded. Among the remaining 87, 8 participants responded to all critical items and fillers with a case-marked option. These participants were identified as speakers who did not consider case marking as an alternative pattern and were excluded from the data analysis.

From the remaining 79, 36 participants selected the null-marked option over the nominative-marked option in S(O)V responses in more than 3 fillers (i.e., more than 25% of this type of filler). Note

that in S(O)V sentences, null-marked subjects are generally dispreferred (H. Lee, 2012) and could technically yield ambiguous interpretations (c.f. (25) (b)). Thus, participants who were not responsive to the null-marked subjects in S(O)V sentences might represent at least three types of speakers. They might lack sensitivity towards infelicitous null-markings or might be hyper-pragmatic speakers who accept ambiguous sentences as long as the relevant discourse context is present to disambiguate the interpretation. The experimental interface simulated a text messaging environment as a conversational setting where case ellipsis is prevalent. Despite the potential ambiguity that null-marked subjects could introduce, participants in this group might have favored null-marked responses due to the specific contextual environment guiding their responses.

Alternatively, these participants might have detected the experimental manipulation regarding exhaustivity and selected null-marked responses based solely on the non-exhaustivity stated in the intended message. Indeed, the 36 participants displayed a systematic behavior of choosing nominative-marked responses when exhaustivity needed to be conveyed, regardless of the structures of the response options (S(OV) and S(O)V). An exclusion criteria were set prior to the experiment to exclude participants who select the null-marked option in 25% or more of S(O)V sentences. Although the number of participants in this group was higher than expected, to adhere to the a priori criteria, these participants were excluded from the main analysis. As a result, 43 remaining participants' responses are included in the final analysis.

4.4 Results

Participants were generally more likely to select the *-lul*-marked responses when the intended message stated that the exhaustive status of an object needed to be conveyed. This pattern was observed in both WH-Q and DQ conditions. The proportion of *-lul*-marked responses compared to null-marked responses is visually evident in Figure 4.1. The exact proportions of *-lul*-marked responses obtained in each of the four conditions are presented in Table 4.3. When the intended message stated the exhaustive status of an object, the proportions of *-lul*-marked responses increased from 57.94% to 65.08% in WH-Q conditions and from 48.81% to 62.70% in DQ conditions.

Fillers were intended to disguise the experimental manipulation and to compare the preferences

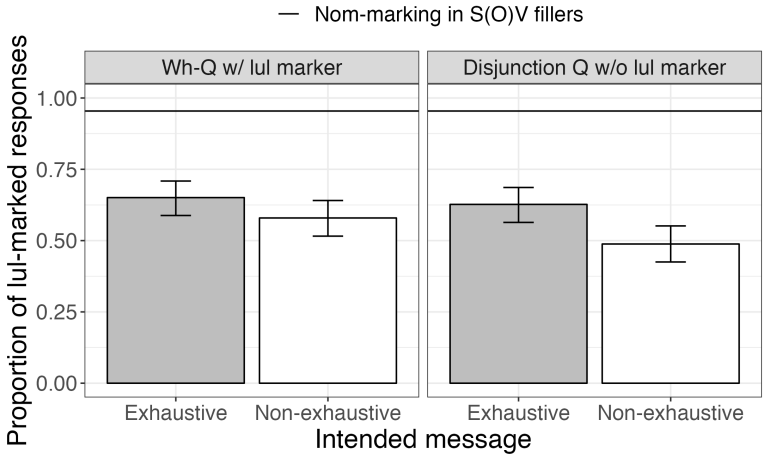


Figure 4.1: Proportion of *-lul*-marked responses and 95% CI.

Table 4.3: Proportion of *-lul*-marked responses in each condition.

	Exhaustive	Non-exhaustive
Wh-Q	65.08%	57.94%
DQ	62.70%	48.81%

for alternating object markings with those for alternative subject markings. Based on previous discussions (H. Lee, 2006b, 2007), subject markings were predicted to show stronger preferences for overt marking. Indeed, as visually observable in Figure 4.1, the proportion of *-lul*-marked responses was less than the proportion of nominative-marked responses in S(O)V sentences, where null-marking could result in ambiguous interpretations.

To confirm that the differences in proportions presented in Table 4.3 are statistically significant in relation to the exhaustivity of the intended message, a series of mixed logit models were fitted (following Jaeger, 2008). All models were fitted with the `lmer()` function from the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2022), with the logistic function set as the link function. The two predictors were sum-coded (WH-Q = 0.5, DQ = -0.5, EXHAUSTIVE = 0.5, NON-EXHAUSTIVE = -0.5), so that each estimated coefficient represents the effect of the predictor.

I started by fitting the maximal random effects and progressively simplified the random effects structure until convergence was reached (following Barr, Levy, Scheepers, & Tily, 2013), eventually obtaining the model specified in (26). The model predicted the log-odds of *-lul*-marked responses using

the formats of the preceding question, the intended message, and the interaction between the two as predictors. Random slopes for participants are included in the model to control for individual differences in responses to the two predicting factors. Random intercepts for items are included to control for effects of the lexicalization of each item.

(26)
```
Response ~ Question*Intended message +
           (1+Question+Intended message|Participant) + (1|Item)
```

Table 4.4 reports the details of the model's prediction. Exhaustivity in the intended message increased the log-odds of *-lul*-marked responses ($\hat{\beta} = 0.68, p < 0.001$). The coefficient of 0.68 indicates that exhaustivity in the intended message is likely to increase the odds of selecting the *-lul*-marked response by a factor of 1.97. This result directly demonstrates the association between the *-lul* marker and exhaustivity. The format of the preceding question did not affect the odds of obtaining *-lul*-marked responses. This is shown by the insignificant interaction effect between the question form and the intended message. Overall, the model confirms that exhaustivity in the intended message was the factor that significantly increased the likelihood of overt *-lul* marker usage.

Table 4.4: A mixed logit model predicting the log-odds of *-lul*-marked responses with question format, exhaustivity in the intended message, and the interaction between the two, while controlling for the random effects of participants and lexicalized items.

Predictor	Estimates	Standard error	z	p-value
(Intercept)	0.52	0.29	1.80	0.071
Question: WH-Q	0.24	0.25	0.93	0.354
Intended message: EXHAUSTIVE	0.68	0.18	3.81	0.000
Question: WH-Q × Intended message: EXHAUSTIVE	-0.52	0.32	-1.63	0.102
Marginal R2 / Conditional R2	0.021 / 0.516			

4.5 Discussion

Current experimental results support the contrasted discourse functions of the *-lul* marker and null-marking. With grammatical features being identical, the *-lul* marker is more likely to appear when the object is associated with an additional discourse interpretation (i.e., exhaustivity implicature). When

objects are not associated with the additional discourse interpretation, the *-lul* marker is less likely to appear. This result confirms the association between the *-lul* marker and exhaustivity. In relation to Experiment 2, the current result confirms that the positive ratings observed with the *-lul* marker indicate its cancelable exhaustivity implicature, rather than a lack of the marker's exhaustive interpretation.

Confirming the exhaustivity implicature of the *-lul* marker provides insights into the organizing principles previously discussed. The current result can be understood through frameworks based on the principles of iconicity and economy as well as predictability. First, the exhaustivity implicature can be seen as an additional discourse feature that elevates the prominence of objects, similar to findings with CF status and *-lul* marking (H. Lee, 2006a). According to the economy principle, null-marking is favored for objects without exhaustivity implicature, while the iconicity principle supports overt markings to encode exhaustivity implicature.

Alternatively, from a predictability-based perspective, exhaustivity tends to be more marked because it represents a less predictable interpretation for objects or for discourse in general. Exhaustive status is typically conveyed in more specific discourse contexts, making it less frequent and thus less predictable (see Haspelmath, 2021). To confirm the significant effect of frequency, accurately estimating the frequency of exhaustivity implicature conveyed in discourse would be ideal, though challenging. Alternatively, an artificial language learning paradigm, similar to those reviewed previously (Kurumada & Grimm, 2019; Fedzechkina et al., 2012), could establish a causal link between frequency, predictability, and *-lul* marking. Language modeling is another viable approach to test the causal relationship between factors influencing grammar organization and language use patterns (see Hamilton, Leskovec, & Jurafsky, 2016; Hawkins et al., 2023).

Although the specific mechanisms proposed by the frameworks of the iconicity and economy principles and predictability differ, these frameworks share a common goal: explaining language as a system designed for efficient communication. Regardless of the principles guiding the overt and covert use of case markings, the discourse functions associated with overt and covert *-lul* markers exhibit efficiency in the morphological system. Instead of requiring overt *-lul* markings for all grammatical objects, Korean grammar allows speakers to use *-lul* selectively to indicate additional discourse interpretations. By omitting *-lul* for more prototypical objects and in more predictable contexts, speakers reduce the production

cost without risking the delivery of inaccurate or ambiguous messages.

In addition to language efficiency, linguistic complexity is another crucial aspect shaped by the organization of a language. Complexity has been measured across various linguistic domains, as well as for languages as a whole, using different metrics (see Juola, 2008; Kortmann & Szmrecsanyi, 2012). One way that morphological complexity is measured is through E(numerative)-complexity and I(ntegrative)-complexity (Ackerman & Malouf, 2013). E-complexity is determined by factors such as the size of the morphological inventory, including the number of exponents and inflectional/declension classes. In contrast, I-complexity is a metric based on the predictability between forms, specifically the entropy of forms. While E-complexity reflects how complex a system appears to both native and non-native speakers, I-complexity reveals how morphological systems are structured to be learnable and usable by native speakers.

It remains a direction of future studies to explore how polyfunctionality of the *-lul* and *-nun* markers would affect the morphological complexity, in particular the E-complexity and I-complexity. Instead of introducing a new marker to indicate exhaustivity, the Korean language implements polyfunctionality and allows the *-lul* and *-nun* markers to indicate exhaustivity. The alternation between overt and covert *-lul* in Korean is an instance where the size of the morphological inventory remains the same, but the inventory of morphosyntactic functions of a form increases. As a result, could this polyfunctionality (one-to-many mappings between a form and its meanings) yield a paradigm that is more or less learnable, and thus more or less complex? Additionally, how would different types of polyfunctionality affect morphological complexity? Would polyfunctionality in one domain (e.g., grammatical functions) be considered more or less complex than polyfunctionality across multiple domains observed in Korean (e.g., grammatical functions and discourse functions)? Investigating these questions, along with different principles that motivate the polyfunctionality of a structure, could further illuminate our understanding of polyfunctionality.

4.6 Conclusion

This chapter examined the exhaustivity implicature of the *-lul* marker by comparing its overt and covert uses. The *-lul* marker tends to appear overtly when objects are exhaustive, while remaining covert

otherwise. This suggests that the organization of alternative markings is shaped not only by grammatical features but also by discourse interpretations. This pattern exemplifies how the Korean object-marking system achieves efficiency: by utilizing a paradigmatic contrast with null-marking, the overt *-lul* marking signals additional discourse functions. In addition, this pattern illustrates how the polyfunctionality of a marker contributes to the shape of the morphological inventory. Instead of introducing new morphological forms, the Korean language leverages an existing case marker to convey additional discourse meanings, thereby avoiding unnecessary expansion of the morphological inventory. In conclusion, the comparison of overt and covert uses of the *-lul* marker highlights the marker's connection to pragmatic implicature and demonstrates how a polyfunctional marker contribute to the organization of a morphological paradigm.

5 Language Models

5.1 Introduction

It has long been believed that language is a unique feature of humans. However, recent advancements of LANGUAGE MODELS (LMs) are challenging this belief. A language model is a system designed to perform language modeling tasks, predicting a word or sequences of words given the input. Specifically, LMs are built to process and produce natural language through probability distributions; Most language models are not built to explicitly encode components essential in discourse pragmatics, such as shared knowledge and communicative intent of human mind. However, despite the lack of explicit training on discourse pragmatics and human mind, some advanced LMs have presented evidence of contextual processing of language. In this chapter, we explore whether the discourse usage patterns of the *-lul* and *-nun* markers can be effectively captured by LMs—systems that lack explicit representations and targeted training of discourse pragmatics.

This chapter is organized as follows. In section 5.2, I overview various language models and how language is represented in the applications of language models. I then review previous studies using the paradigm of psycholinguistic assessments of language models, and discuss language models in handling discourse and pragmatics in natural language. Section 5.3 describes the models I evaluate in the subsequent experiments. In section 5.4, I report Experiment 5, which assesses LLMs' output regarding *-lul* and *-nun* markers' exhaustivity and its cancelability. In section 5.5, I report Experiment 6, assessing how LLMs utilize *-lul* and *-nun* markers in generating exhaustive messages. I discuss the overall result in section 5.6 and conclude the chapter in section 5.7.

5.2 Background

In the current section, I provide an overview of different approaches to language modeling and the representation of linguistic structures and their meanings. I first introduce a purely statistical approach to language modeling and word representations.[1] Next, I discuss neural networks and the representations of words built using these networks.[2] Finally, I introduce representations of language that emerge during the training process of large language models, also referred to as contextualized embeddings. Following this background, I summarize previous evaluations of language models with respect to the discourse and pragmatics of natural language.

A purely statistical approach to language modeling is exemplified by n-gram models. An n-gram model predicts the next word based on the previous $n-1$ words or sequences. For example, a bigram model considers one preceding word, while a trigram model looks at two preceding words. More precisely, n-gram models estimate RELATIVE FREQUENCY, which is the probability of observing a particular word sequence given the frequency of that sequence relative to a preceding sequence (Jurafsky & Martin, 2024, Chapter 2). This is formalized in (5.1), where w_n represents the n^{th} word in a sequence of length N, and $w_{n-N+1:n-1}$ denotes the sequence of words from the start of the sequence up to the $(n-1)^{\text{th}}$ word. The probability of observing a word w as the n^{th} word is thus the ratio of the frequency of the sequence $w_{n-N+1:n-1}$ followed by w to the frequency of observing the sequence $w_{n-N+1:n-1}$.[3]

$$P(w_n|w_{n-N+1:n-1}) = \frac{C(w_{n-N+1:n-1}w_n)}{C(w_{n-N+1:n-1})} \tag{5.1}$$

In addition to language modeling, frequency distributions of words are used to build DISTRIBUTIONAL REPRESENTATIONS of words, where words are represented as vectors (i.e., WORD VECTORS). The idea of representing a word by its frequency distributions comes from what can be called the DISTRIBUTIONAL HYPOTHESIS (DH). DH asserts that the meaning of a word can be defined by its frequent collocations, which are the types of linguistic contexts in which the word appears (Firth, 1968; Harris,

[1] I referred to Manning and Schütze (1999) and Lenci (2008, 2018) for this part.

[2] For this part, I referred to Jurafsky and Martin (2024) and the course slides and notes developed for CS224N: Natural Language Processing with Deep Learning at Stanford, taught by Chris Manning and others.

[3] This approach to probability estimation is an example of maximum likelihood estimation (MLE). See Jurafsky and Martin (2024) Chapter 2 for details on MLE explained in the context of LMs.

1954). When a word *w* appears in a text, its context refers to the surrounding words within a specific range, or "window," that includes a certain number of words before and after *w*. The core idea of the DH is that if we accumulate many contexts of *w*, those contexts will eventually build up a representation of the meaning of *w*. Thus, similarity in meaning between words correlates with similarities in their distributions, specifically, the contexts of the words.

For instance, consider the word *election* appearing in three different contexts in (1). Under the DH, the meaning of *election* is represented by words that co-occur in the context window. If we observe a word that co-occurs more frequently with *election*—such as *vote*—that pair of words is distributed in similar contexts and is considered to have similar meanings. Thus, within distributional representations, the meaning of a word is inferred from patterns of co-occurrence in data.

(1) a. ...as the **election** approaches, local officials are urging citizens to *vote* early...

b. ...*voter* turnout for this year's **election** is projected to be higher than in previous years...

c. ...many *voters* expressed hope that the **election** win would bring about....

In building the distributional representation of a word, we can consider the frequency of a TYPE and a TOKEN. A type refers to a unique instance of a word, while a token refers to each occurrence of that word in a context. For example, in the sentence *The cat sat on the mat*, there are 6 word tokens (including the duplicated *the*), but only 5 word types (unique words: *the*, *cat*, *sat*, *on*, *mat*). To illustrate distributional representations using types, let us assume we are interested in how the three words *election*, *game*, and *win* are represented in the context of news articles. The type of each of these words will consist of a set of target words, $T = \{election, game, win\}$. Additionally, we are interested in how these target words are represented in the context of two dimensions: 'politics' and 'sports,' which are common categories in news articles. The type of these words will consist of a set of context words, $C = \{politics, sports\}$. In this toy example, we will consider news articles published on a given day as our corpus data, and count how often the target words appear within 10 words of the context words. For example, in the sentence *The outcome of this election will significantly impact the future of politics in our country*, one of our target words, *election*, appears within the 10-word window of the context word *politics*.[4]

[4]This is only one way to define the context of a word. See Lenci (2018) section 3.1 for the summary of other approaches.

Co-occurrences of target words and context words obtained from our corpus result in a co-occurrence matrix, as presented in (5.2). In this matrix, each row represents a target word, and each column represents a context word. Each value in the matrix indicates how many times a word type in a row and a word type in a column appear together in the corpus data. For example, the matrix shows that tokens of the word type *election* appear with *politics* 169 times and with *sports* 17 times, in both cases within a 10-word window of each other. Thus, the target word *election* is represented by the vector [169, 17]. This is the distributional representation of the word *election* established through its co-occurrences with context words.

	politics	*sports*
election	169	17
game	8	125
win	76	143

(5.2)

Word vectors of the three target words in (5.2) can be visualized in vector space, as shown in Figure 5.1. The two context words represent the two dimensions: the x-axis represents *politics*, and the y-axis represents *sports*. Each target word type is plotted as a point within these two dimensions. For instance, the target word type *election* is located at (169, 17) in this space, reflecting its 169 occurrences in the context of *politics* and 17 occurrences in the context of *sports*.

Given the word vectors, the similarity between two words is measured by the similarity between their vectors. One of the popular measure of vector similarity is the cosine value of the two vectors, referred to as cosine similarity. Cosine similarity is calculated by dividing the dot product of two vectors ($\mathbf{u} \cdot \mathbf{v}$) by the product of their magnitudes ($||\mathbf{u}||||\mathbf{v}||$):

$$cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||||\mathbf{v}||} \tag{5.3}$$

The cosine ranges from 1 (if the vectors are identical) to -1. The lower the cosine is, the lower the vector similarity will be. For instance, from the matrix in (5.2), we have one two-dimensional vector for each target word: [169, 17] for the word *election*, [8, 125] for the word *game*, and [76, 143] for the word

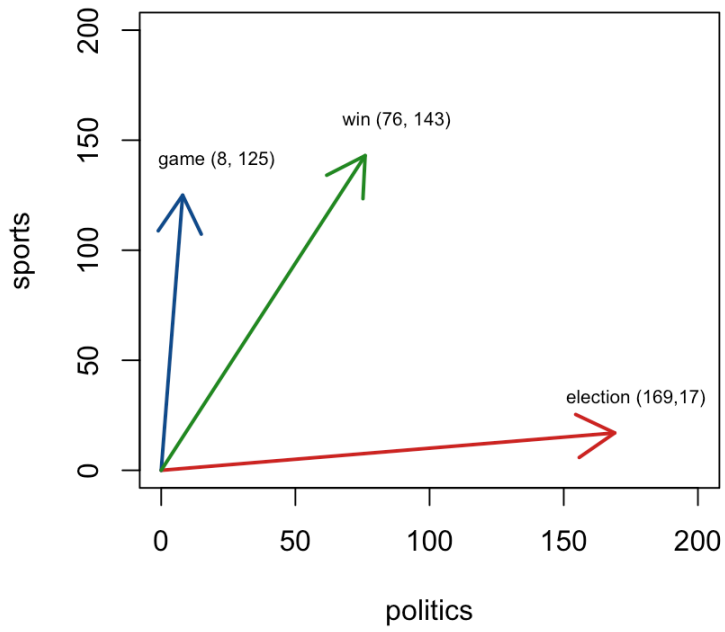


Figure 5.1: Vectors of the target words *game*, *win*, and *election* represented by their co-occurrences with the context words *politic* and *sport*.

win. The cosine similarities of these vectors, calculated using the formula in (5.3), are listed in Table 5.1. For example, word vectors of *win* and *game* result in the highest cosine similarity of 0.91. This indicates that these two words are the most similar in terms of their meanings along the dimensions represented by *politics* and *sports*, compared to other pairs, such as *game-election* or *win-election*.

Table 5.1: Cosine similarities of the word vectors presented in (5.2).

	election	*game*	*win*
election	1		
game	0.16	1	
win	0.56	0.91	1

In the vector space represented in Figure 5.1, the higher cosine similarity is reflected in the smaller angle between the two vectors. Visually, it is clear that the vector for *win* has a smaller angle with the vector for *game*, but a larger angle with the vector for *election*. This is reflected in their cosine similarities

of 0.91 vs. 0.56. Thus, word vectors provide an explicit representation of words, allowing various features of words and the relationships between them to be more transparently reflected in the vector space. These relationships can be measured as similarity between vectors.

The set of words and word vectors considered in (5.2) are one of the simplest cases used for illustration. If we consider how all of the vocabulary in English is represented by more than just two dimensions (i.e., context words), such as *politics* and *sports*, we would have many more vectors with more than 3 rows and 2 columns representing the co-occurrences of words. As the set of target words and the set of context words get larger, the vector becomes more high-dimensional, with more rows and columns. In addition, considering a larger set of target words and context words results in a sparse vector, meaning many entries will be zero. For instance, many words in English may not co-occur with *politics* or *sports* in documented text (e.g., news articles), and we will observe no co-occurrences between certain target words and context words. Therefore, counting co-occurrences for a larger set of target and context words in a corpus dataset will result in a high-dimensional and sparse vector, as exemplified in (5.4).

	election	*politics*	*sports*	*weather*
aardvark	0	0	0	0
aaron	0	0	0	0
⋮	⋮	⋮	⋮	⋮
election	52	169	17	9
⋮	⋮	⋮	⋮	⋮
game	74	8	125	11
⋮	⋮	⋮	⋮	⋮
win	36	76	143	6
⋮	⋮	⋮	⋮	⋮
zyzzyva	0	0	0	0

(5.4)

High-dimensional and sparse vectors turn out to be less useful for making generalizations based on the distributions of words. First, context words may be very similar and/or strongly correlated with one another. For example, in (5.4) above, four context words—*election*, *politics*, *sports*, and *weather*—are

considered distinct. However, *election* and *politics* can easily represent similar contexts. Additionally, due to the nature of corpora, many possible word co-occurrences remain unobserved. For example, the corpus data may not contain the co-occurrence of *aardvark* and *election*, but such a pair could still co-occur in unobserved English conversations.

These disadvantages of count-based vectors can be mitigated by dimensionality reduction techniques such as Singular Value Decomposition (SVD) and Principal Component Analysis (PCA). Here, I will briefly introduce SVD. In practice, SVD is often applied on a co-occurrence matrix containing Positive Pairwise Mutual Information (PPMI) values instead of raw frequency counts. PPMI is a measure that weighs the co-occurrences of words, following (5.4) and (5.5).

$$PMI(w_1, w_2) = log \frac{P(w1w2)}{P(w1)P(w2)} \tag{5.5}$$

$$PPMI(w_1, w_2) = max(0, PMI(w1, w2)) \tag{5.6}$$

For instance, the PPMI values for the frequencies observed in (5.1) are presented in (5.6). We can observe that frequency counts below 100 in (5.1) have a PPMI of 0 in (5.6). Thus, converting to PPMI helps focus on more significant co-occurrences.

	politic	*sport*
elect	0.95	0
game	0	0.83
win	0	0.30

(5.7)

Given the co-occurrence matrix of PPMI as input, SVD decomposes it into three matrices. The first two matrices are the input word matrix (U), which consists of one row vector for each target word, and the transposed context matrix (V^T), which consists of one column vector for each context word. The last matrix we consider is the singular value matrix (Σ), a diagonal matrix with singular values containing singular values that reflect the importance of each dimension.

$$C_{m\times n} = U_{m\times z}\Sigma_{z\times z}(V_{z\times n})^T \tag{5.8}$$

After decomposing a co-occurrence matrix into a product of three matrices, only the first k dimensions are retained, while the others are set to zero, resulting in a "truncated SVD." For instance, by keeping only the $k = 2$ singular values and vectors, we obtain the matrices shown in (5.9).

U	1	2	3	4
election	−0.44	−0.30	0.00	0.00
game	−0.13	−0.33	0.00	0.00
win	−0.48	−0.51	0.00	0.00

Σ	1	2	3	4
1	−0.44	−0.30	0.00	0.00
2	−0.13	−0.33	0.00	0.00
3	−0.48	−0.51	0.00	0.00
4	0.00	0.00	0.00	0.00

(5.9)

This method provides a better approximation of the original matrix C in a reduced-dimensional space while preserving most of the variation in the data. Intuitively, dimensionality reduction techniques allow us to disregard unnecessary details and focus on the essential dimensions of meaning between words. Word vectors that implement dimensionality reduction techniques have proven effective in capturing the meanings of both function words and content words (e.g., Marelli & Baroni, 2015).

As language modeling techniques have advanced, approaches to word representations have also evolved. In particular, rather than building representations that reflect frequency distributions observed in input data, more recent word representations employ neural networks that can optimally predict generalized co-occurrence patterns given in the input data. Inspired by how the human brain processes information, a neural network consists of layers of interconnected nodes, or "neurons," that work together to recognize patterns (Rumelhart & McClelland, 1987a, 1987b). Neural networks have been central to advancements in deep learning and machine learning models, including those developed for language modeling (c.f., Bengio, Ducharme, & Vincent, 2000). Here, I briefly introduce the basic architecture of neural networks in the context of language modeling.

In a typical neural network, data moves through an input layer, hidden layers (including the en-

coder and decoder), and an output layer. The raw data—such as text—enters the neural network through the input layer. The encoder processes the input data and transforms it into its distributional representations. The decoder then takes the encoded representation from the encoder and generates the output, also through the hidden layers. Finally, the output from the decoder is passed to the output layer, which converts the generated information into the desired format (e.g., a word in natural language).

An activation function (e.g., ReLU or sigmoid) in the neurons of a neural network determines whether to dismiss or pass its signal to the next layer. Connections between neurons have weights that adjust as the model learns patterns in the input, while biases help shift the output in conjunction with the weighted sum of inputs. The network is trained using a dataset, adjusting its weights through algorithms like backpropagation to minimize prediction errors. Various architectures of neural networks represent different types of networks, including Recurrent Neural Networks (RNNs) (Elman, 1990, 1991) and Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997; Gers, 2000). In addition to predicting the next word, neural language models have been successfully applied to a variety of natural language processing tasks, such as part-of-speech tagging, named entity recognition, sentiment classification, and machine translation (e.g., Wu et al., 2016).

Instead of counting co-occurrences and applying dimensionality reduction techniques, neural network algorithms can directly create low-dimensional distributional representations by learning to optimally predict the contexts of a target word (Lenci, 2018). The distributional representations obtained from neural networks are often referred to as WORD EMBEDDINGS. Two well-known examples of this framework include Skip-gram with Negative Sampling (SGNS), also known as word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), and GloVe (Pennington, Socher, & Manning, 2014). In this section, I will introduce the SGNS model.

Intuitively, the Skip-gram model scans each position in a given corpus, identifying a target word t and its corresponding context words c. The model then uses the similarity of word vectors for t and c to calculate the probability of a context word given the observed target word (i.e., $P(c|t)$). The probability of a context word given a target word, denoted as $P(c|t)$, is obtained with a prediction function presented in (5.10). Using the notations from Mikolov et al. (2013), we treat a target word as an input word w_I and a context word as an output word w_O. The prediction function involves the dot product of the transposed

vector of the output word ${v'_{w_O}}^\top$ and the vector of the input word v_{w_I}, which compares the similarity between the input and output words. The exponentiation $exp({v'_{w_O}}^\top v_{w_I})$ ensures that the outcome of the dot product is positive. This exponentiated dot product is then normalized across the entire vocabulary, represented by the sum $\sum_{w=1}^{W} exp({v'_w}^\top v_{w_I})$, where W is the number of words in the vocabulary.

$$p(w_O|w_I) = \frac{exp({v'_{w_O}}^\top v_{w_I})}{\sum_{w=1}^{W} exp({v'_w}^\top v_{w_I})} \tag{5.10}$$

In standard Skip-gram model, negative sampling is applied to mitigate the computational cost of this normalization, as it requires summing over a large number of words. The negative sampling method trains a binary logistic regression model to differentiate between a "true pair" (an actual co-occurrence of a target word and a context word observed in corpus) and several "noise pairs" (the target word paired with random non-context words unobserved in corpus). As the result of the Skip-gram with negative sampling, we obtain the probability distribution of the output word (i.e., context word) given the input word (i.e., target word). Given this probability distributions, word2vec is obtained once after a neural network gradually adjusts its parameters to optimize the predictions of these probability predictions.

Word embeddings created with neural networks have been shown to capture the meanings of words effectively, particularly the paradigmatic relationships between them (Mikolov et al., 2013; Pennington et al., 2014; Bonami & Guzmán Naranjo, 2023). For instance, Bonami and Guzmán Naranjo (2023) explored whether distributional representations could elucidate predictive relationships between suffixed words and their stems. Their findings suggest that paradigmatic relations in morphology can be encoded by word embeddings learned by neural networks, highlighting that word meaning helps explain the connections between different words and their paradigmatic relations.[5]

Despite their successes, word vectors and neural word embeddings fall short in capturing the contextual meanings of words. When each word type is represented by a single vector (as shown in (5.2) and Figure 5.1), the various meanings of tokens of that type cannot be effectively captured. For instance, the word *game* can have different meanings in the contexts of (a)-(c) in (2), as well as within an idiomatic

[5]Bonami and Guzmán Naranjo (2023) used another type of word embeddings called GloVe (Pennington et al., 2014). Glove embeddings are created by reducing word-by-word matrix with weighted least-squares regression. This is another instance of utilizing neural network to optimally predict associations between words.

phrase in (d).

(2) a. The children spent the afternoon playing a new board **game**.

b. She often ventured into the forest to capture images of elusive **game** in their natural habitat.

c. Their upcoming match was a crucial **game** for the championship title.

d. The company stayed ahead of the **game** in the competitive market.

If we combine the contexts of all sentences in (2) to represent the word type *game*, the contextual meanings of each token conveyed in (a)-(d) would need to be represented by a single vector. This can result in the mixing of different contextual meanings into one vector. Representations of words built with word types do not reflect the dynamic changes in the contextual meanings of words; hence, this type of representation is often called static word embeddings.

Contextualized embeddings make significant improvements in capturing context-sensitive representations of words (McCann, Bradbury, Xiong, & Socher, 2017; Peters et al., 2018). Rather than assigning one vector per word type, contextualized embeddings assign a separate vector to each token instance. In other words, contextualized embeddings generate a vector for each word that is conditioned on its surrounding context. For example, the word *game* appearing in the four sentences in (2) would be represented by four distinct vectors, one for each sentence. This allows the token's vector to reflect what the word means in that particular context, providing a more nuanced representation of a word's meaning in particular contexts (c.f., N. A. Smith, 2020, for a more detailed summary).

Contextualized embeddings are contrasted with static word embeddings in Table 5.2 (taken from Peters et al., 2018). The table displays a source word and its nearest neighbors, which are words that are similar (or closer in vector space, as shown in Figure 5.1) to the vector of the source word. As indicated in the table, static embeddings using word types reveal that the word *play* is closely related to its token appearances, such as *playing*, *played*, and *players*. It also shows related words like *game*, but other meanings of *play* (e.g., in the context of instruments, performances, etc.) are not captured.

In contrast, the bottom two rows illustrate the difference with contextualized embeddings. As explained earlier, contextualized embeddings represent each instance of *play* based on its specific context.

Table 5.2: Nearest neighbors to *play* obtained with GloVe and biLM, reported in Peters et al., 2018.

Embeddings	Source	Nearest Neighbors
Static (GloVe)	play	playing, game, games, played, players, plays, player, Play, football, multiplayer
Contextualized (biLM)	Chico Ruiz made a spectacular play on Alusik s grounder ...	Kieffer , the only junior in the group , was commended for his ability to hit in the clutch , as well as his all-round excellent play .
	Olivia De Havilland signed to do a Broadway play for Garson ...	... they were actors who had been handed fat roles in a successful play , and had talent enough to fill the roles competently , with nice understatement .

In the first example, *play* refers to its nominal use in the context of a game, and its nearest neighbor reflects another context where *play* is used as a noun. In the second example, *play* refers to a stage performance, and its nearest neighbor correspond to another context of *play* indicating its meaning of a stage performance. This demonstrates that contextualized embeddings can capture context-specific meanings and effectively disambiguate polysemous words (see Ethayarajh, 2019; Liu, Kusner, & Blunsom, 2020, for further demonstrations of the success of contextualized embeddings).

Contextualized embeddings are further advanced with the emergence of LARGE LANGUAGE MODELS (LLMs). Although the definition of LLMs is not fixed, as the technology around these models changes rapidly (Alammar & Grootendorst, 2024), LLMs generally refer to language models that share common features, such as a multi-step training process with large datasets and transformer architecture. To begin, LLMs are trained in a way that is distinctive from traditional neural language models. A conventional method for training neural language models involves starting with pre-trained word embeddings, followed by training the neural network to incorporate the context of inputs while performing language modeling tasks. Thus, this approach uses word embeddings that are trained before the neural network is trained with more data.

In contrast, the modern approach of LLMs relies less on having fixed word embeddings prior to model training. Instead, LLMs utilizes a multi-step approach illustrated in Figure 5.2 (taken from Alammar & Grootendorst, 2024). During the initial pre-training, network parameters are optimized while processing each token in the input data as it appears in its own context. The input data used in LLMs'

pre-training consists of unstructured raw text data (i.e., unsupervised data), and its size is particularly large, sometimes comprising billions or even trillions of words.

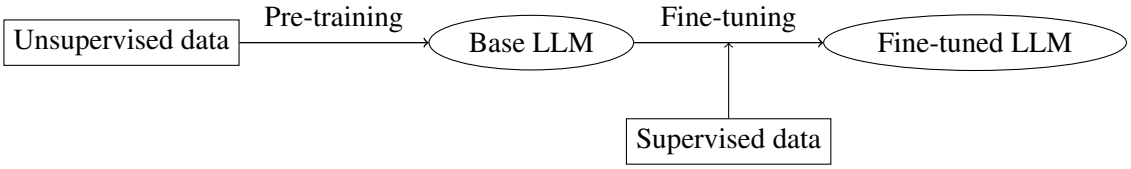


Figure 5.2: A diagram of a multi-step approach of training LLMs, taken from Alammar and Grootendorst (2024).

The model obtained after pre-training is called a base model or foundational model, and the network parameters of the base model construct contextualized embeddings. Thus, the contextualized embeddings of LLMs are representations of words that emerge during the model's pre-training with extensive input data. Due to this approach, the contextualized embeddings of LLMs can represent each token in the vast size of input data as it appears in its own context. The pre-training can be applied to either the encoder, the decoder, or both, and different pre-training methods result in different types of base models.

In the case of pre-training encoders, the models are designed to capture bidirectional context, meaning they are trained to predict a word based on both its preceding and following contexts. This is introduced as the "Masked LM" objective (Devlin, Chang, Lee, & Toutanova, 2019), where input texts are modified to include a masked word, as illustrated in (3). The model is then trained to predict the masked word using its surrounding context. The pre-trained encoder trained with the Masked LM objective is known as the BERT model. Thus, BERT uses word representations that are dynamically informed by the words around them.

(3) I went to the [MASK] last night.

The next approach involves pre-training both the encoder and decoder. In this method, the input is modified by replacing segments of text with unique placeholders of varying lengths. For instance, the original text in (4) is altered to include X and Y, which replace the phrases *went to* and *last*, respectively. The decoder then decodes the spans that were removed from the input. T5 is the model that employs this

pre-training method (Raffel et al., 2020).

(4) Original text: I went to the movie theater last night.

Input: I $< X >$ the movie theater $< Y >$ night.
Target: $< A >$ went to $< B >$ last $< C >$.

The decoder component of neural networks can also be pre-trained, a process often referred to as generative pre-training. In this approach, decoders are trained on the general language modeling task of generating the next words. It has been demonstrated that pre-trained decoders can function effectively as language models, producing convincingly natural sequences of language (Radford, Narasimhan, Salimans, & Sutskever, 2018). Models utilizing generative pre-training are referred to as GPT models.

Within the encoder and decoder components of LLMs, there are transformer blocks, which are neural network layers that effectively capture contextual meanings of language. A transformer block contains two components: a self-attention layer and a feed-forward layer. The self-attention mechanism is particularly relevant to contextual representations of language,[6] as it is designed to process relevant information from other input tokens and positions. It weighs how much "attention" needs to be paid to different parts of the input to process the current token (Vaswani et al., 2017). For instance, when the model is given the sequence of tokens *I went to the movie theater last night, and it was packed...* and is processing the token *it*, which refers to *the movie theater*, the self-attention mechanism weighs how important each context token is for processing the current token *it*. This allows the model to process the token within the context in which it appears. As a result of pre-training these transformer blocks with large amount of data, base LLMs excel at capturing contextual meanings of words. After pre-training, base models go through a further step of fine-tuning, where they are trained to complete specific tasks, such as classifications or following instructions (c.f., Figure 5.2).

Beyond capturing contextual meanings of words, LLMs reflect patterns similar to humans' prediction patterns in language processing. In psycholinguistics, human expectations of words in context are often measured by reaction times, such as reading times recorded by an eye-tracker. Longer reading times for a word indicate greater processing difficulty, suggesting that the word was less expected in the

[6]The feed-forward layer is related to enhancing the model's capacity.

given context. Reading time has been demonstrated as a measurement that is generally proportional to the estimated probability of a word given the context, particularly to the surprisal of a word (Hale, 2001; N. J. Smith & Levy, 2013). Surprisal is defined as the negative logarithm of the probability of an event (e.g., a certain token appearing next). If the next word W is a random variable taking the word w_i with a probability of occurrence $p(W = w_i)$, the surprisal of the word w_i is defined as follows:

$$s(w_i) = -log_2 p(w_i|w_1, w_2, ..., w_{i-1}) \tag{5.11}$$

Surprisals obtained from probabilities assigned by neural LMs have been shown as strong predictors of human reading times (Goodkind & Bicknell, 2018). Additionally, log probabilities assigned by LMs have been suggested as predictors of the N400 effect (E. G. Wilcox, Gauthier, Hu, Qian, & Levy, 2020; Michaelov, Coulson, & Bergen, 2022), a neurological pattern associated with expectedness in humans' language processing. Thus, by investigating LMs' probabilistic representations of linguistic input and output, we can evaluate how well these models had captured patterns in language processing and production.

In particular, LMs have been subject to the question of whether their predictions stem from learning general rules of natural language or from learning specific patterns in the training data. For example, after encountering the pronoun *she* in English, a model might predict the token *goes* to follow, generating a sequence like *she goes to the library*. This prediction could result from the model learning general linguistic rules about subject-verb agreement in English. Alternatively, it might reflect the model's sensitivity towards specific patterns found in the training data, such as a high frequency of the collocation *she* and *goes*. In this case, the model has learned the pattern between these two exact tokens rather than the broader grammatical rule. Consequently, the model might produce incorrect predictions for other pronouns and verbs, generating ungrammatical sequences like *he study linguistics*.

To assess the learning of generalized rules in language, a paradigm has been developed to evaluate LMs with psycholinguistic assessments (e.g., Marvin & Linzen, 2018; Futrell et al., 2019). This research enterprise has explored how LMs represent rules and patterns across various domains, including syntax (Hu, Gauthier, Qian, Wilcox, & Levy, 2020), semantics and pragmatics (Hu, Floyd, Jouravlev, Fedorenko,

& Gibson, 2023), and discourse (Schuster & Linzen, 2022; Pandia, Cong, & Ettinger, 2021; Paperno et al., 2016). In line with these studies, the current chapter evaluates whether LMs can effectively model the discourse functions of Korean morphology discussed in Chapters 2-4.

In Experiment 1 (Chapter 2), we found that exhaustivity implicatures are equally likely to be conveyed by the *-lul* and *-nun* markers when there is a set of alternatives in the discourse context, as seen in examples like (5) A1 and A2. Experiment 2 (Chapter 3) revealed that the *-lul* marker's exhaustivity is cancelable, as shown in (5) A1, while the *-nun* marker's exhaustivity is non-cancelable, illustrated in (5) A2. Lastly, Experiment 3 (Chapter 4) demonstrated that human speakers prefer to use the *-lul* marker over null-marking when intending to convey exhaustivity, as illustrated in (6) A1.

(5) Context: Interlocutors know that guests are asked to bring a **pizza** or a **cake** to the party.

Q. What did Sue bring to the party?

A1. Sue pizza***-lul*** brought. Sue also brought a cake.

A2. Sue pizza***-nun*** brought. #Sue also brought a cake.

(6) Mina intends to say that Sue brought **only** the pizza to the party, among the pizza and the cake.

Q. What did Sue bring to the party?

A1. Sue pizza***-lul*** brought.

A2.#Sue pizza-∅ brought.

Thus, for LMs to detect exhaustivity indicated by the *-lul* and *-nun* markers, they need to recognize multiple discourse entities being discussed as a set of alternatives (e.g., *pizza* and *cake* in (5)) and interpret the presence of the *-lul* or *-nun* marker as a cue to the exhaustivity status of one alternative. When exhaustivity is intended, LMs should produce *-lul*-marked and *-nun*-marked objects more frequently compared to null-marked objects. In relation to these tasks, in the following, we review studies that tested LMs' ability to process discourse context and pragmatic implicatures. I discuss the implications of these studies for evaluating LMs in relation to the discourse functions of Korean morphology.

To begin, LMs are tested for their ability to utilize discourse context (Paperno et al., 2016; Pandia et al., 2021; Schuster & Linzen, 2022). Specifically, Paperno et al. (2016) evaluated LMs on their ability to

track discourse entities using the LAMBADA (LAnguage Modeling Broadened to Account for Discourse Aspects) dataset. An example item from this dataset is presented in (7). Each item in the dataset includes a context text that highlights the target word. For example, the context in (7) describes a dancing scene, making the target word 'dancing' salient.

(7) *Context*: They tuned, discussed for a moment, then struck up a lively jig. Everyone joined in, turning the courtyard into an even more chaotic scene, people now dancing in circles, swinging and spinning in circles, everyone making up their own dance steps. I felt my feet tapping, my body wanting to move.
Target sentence: Aside from writing, I've always loved _____.
Target word: dancing

Following the context text, a target sentence was presented in the input, as illustrated in (7). This target sentence contained an empty slot that needed to be filled with the most natural completion based on the context. In some items, the target word was explicitly mentioned in the context, as in (7) where 'dancing' appears directly. In other cases, the target word was not explicitly stated but was strongly implied by the context. For example, in (8), although the word 'camera' is not mentioned in the context, the context of focusing on the subject of a photograph and clicking the shutter makes 'camera' the salient target word.

(8) *Context*: He shook his head, took a step back and held his hands up as he tried to smile without losing a cigarette. "Yes you can," Julia said in a reassuring voice. "I've already focused on my friend. You just have to click the shutter, on top, here."
Target sentence: He nodded sheepishly, through his cigarette away and took the _____.
Target word: camera

With the LAMBADA dataset designed with the structure exemplified in (7) and (8), Paperno et al. (2016) evaluated LMs using four different architectures: a simple Recurrent Neural Network (RNN) (Elman, 1990), a Long Short-Term Memory network (LSTM) (Hochreiter & Schmidhuber, 1997), a tra-

ditional statistical n-gram language model (Stolcke, 2002), and a Memory Network (Sukhbaatar, Szlam, Weston, & Fergus, 2015). None of these models achieved more than 1% accuracy on this dataset, indicating a lack of capability in predicting the next token based on discourse context. Given these results, tracking a set of alternatives in the discourse context to calculate exhaustivity is likely to be a challenging task for LMs.

Evaluating LMs with the LAMBADA dataset, it is not clear which types of discourse contexts LMs struggled to track. To provide a more granular evaluation of LMs on tasks involving tracking discourse entities, Schuster and Linzen (2022) developed four types of evaluation items using different sentential operators, embedding verbs, and indefinite noun phrases. The first type of item considered the interactions of affirmative sentences and negations, as illustrated in (9) and (10).

(9) Context: Michael baked a cake
Referential continuation: and it was the best thing at the picnic.

(10) Context: Michael didn't bake a cake
Referential continuation: #and it was the best thing at the picnic.

These items included an affirmative sentence as a context sentence and a continuation sentence having a referential expression. The affirmative context sentence included an indefinite NP (e.g., *a cake* in (9)), which introduced a new entity to the discourse, allowing a referential expression (e.g., *it* in (9)) to refer back to the entity. However, when the context sentence contained negation, as shown in (12), the indefinite NP *a cake* did not introduce a new discourse entity. In this case, using a referential expression *it* in the continuation became infelicitous, as illustrated in (10).

As control items, they implemented non-referential continuations, as illustrated in (11). Unlike referential continuations, non-referential continuations were felicitous to appear after both of the context sentences being contrasted. For instance, *Michael baked a cake* and *Michael didn't bake a cake* in (9) and (10) are both felicitous to be followed by the non-referential continuations in (11).

(11) Context: Michael baked/didn't bake a cake

Non-referential continuation: and it's not a big deal.

The second type of item was designed with verbs of different modalities. For instance, the context sentence *Michael baked a cake* was contrasted with *Michael wants to bake a cake*, each followed by the referential continuation *and it was the best thing at the picnic*, creating pairs of felicitous and infelicitous sentences. Additionally, context sentences with the factive verb *know* and the non-factive verb *doubt* were contrasted. The final type of items contrasted implicative verbs like *manage to* and negative implicative verbs like *fail to*. Each type of item had non-referential continuations as control items.

In order to measure how LMs evaluate referential continuations following different context sentences, they compared the relative probabilities: proportion of the probability assigned to the referential continuations out of the sum probability assigned to referential continuations and control continuations. As a result, GPT-2 and GPT-3 overall behaved similarly to humans and generally favored the referential continuation more when the preceding sentence introduced a discourse entity. Schuster and Linzen (2022) saw two plausible interpretations from this result. First, the result could indicate that the models' true capability of making pragmatic judgments based on discourse context. However, it could also indicate that the models learned spurious patterns existed in the training dataset. For example, the model might have learned that clauses with negation are less frequently followed by clauses with *it*, thus assigning lower probabiliites to referential continuations in the context of (10) compared to (9).

To tease apart models' learning of linguistic rules from models' learning of spurious correlations, Schuster and Linzen (2022) conducted the second experiment. As illustrated in (12) and (13), items for the second experiment contained two objects in the context sentences, and the following referential continuation contained the definite NP instead of pronominal NPs *it*. In (12) and (13), the two objects *a shirt* and *a hat* are mentioned as indefinite NPs, and the referential continuation uses the definite NPs *the shirt* and *the hat* to refer to these entities.

(12) Context: Mary found a shirt at the store but she didn't find a hat.
Referential continuation: The shirt/#hat is blue.

(13) Context: Mary didn't find a shirt at the store but she found a hat.

Referential continuation: The hat/#shirt is blue.

Similar to how negation interacted with indefinite NPs to determine felicitous continuations in the first experiment (c.f., (9) and (10)), items for the second experiment contained interactions between indefinite NPs and negations. In (12), for instance, the negation in the second phrase appearing with *a hat* prevents this discourse entity to be introduced. As a result, the definite expression *the hat* becomes infelicitous to follow next. Thus, these items could evaluate LMs' ability to track statuses of two discourse entities, while avoiding evaluating LMs based on spurious patterns (e.g., the pattern of *it* following negations) that they may have learned.

In the second experiment, both GPT-2 and GPT-3 showed the general pattern of choosing the correct continuation, but showed lower accuracy compared to the first experiment. This suggests that LMs become less reliable when interpreting referential continuations involving more than one discourse entity. Authors further analyzed that LMs' size impact their performance on interpreting sentences based on discourse context. The smaller GPT-2 consistently performed the worst, while GPT-3 consistently performed the best. This result suggests that bigger models are more likely to show improvements on contextual understandings of language.

In relation to interpreting Korean object markers as cues to exhaustivity, since GPT-2 and GPT-3 showed above-chance accuracy in both the first and second experiments in Schuster and Linzen (2022), LMs are expected to recognize two objects as contextualized set of alternatives. However, as both GPT-2 and GPT-3 showed lower accuracy in the second experiment when continuation sentences lacked referential expressions, in the context of Korean, LMs may struggle to detect the given statuses of objects if they are mentioned for the second time in a discourse without referential expressions. Incompetency in tracking the given status of objects may lead to LMs failing at identifying the contrastive relation between two previously mentioned objects. Lastly, since GPT-3 consistently outperformed GPT-2 in Schuster and Linzen (2022), also in Korean, larger models are expected to demonstrate greater competency in interpreting discourse statuses of objects.

In addition to tracking discourse entities, demonstrating patterns of the *-lul* and *-nun* markers in natural language involves discerning the two markers as cues to exhaustivity implicatures of distinct can-

celability (c.f., Chapter 3). In other words, LMs need to predict discourse continuations that align with the cancelability of each marker's exhaustivity implicature. Regarding the LMs' ability to predict felicitous discourse continuations, Pandia et al. (2021) explored whether LMs can predict relevant discourse connectives based on the context. They tested three classes of pre-trained LMs—BERT, RoBERTa, and ALBERT—in two stages.

As a preliminary assessment, they compiled 17,476 input items from the Penn Discourse Treebank (Prasad et al., 2008), masked the discourse connectives in each input, and recorded the models' predictions of the missing connectives. An example is illustrated in (14). Sentences from the corpus were modified so that the discourse connective was masked, but the context of the masked word clearly suggested a causal connective must appear in the masked slot in (14).

(14) Mr. Brown was planning to look for new glasses and shoes today. The glasses really are more urgent. **[MASK]**, he now heads towards the optician that a friend recommended. **(correct target: therefore, so)**

In the preliminary assessment using tasks exemplified in (14), most models correctly predicted the discourse connective about half the time, indicating some evidence of using contextual cues to infer the appropriate connective. However, accuracy varied across different types of discourse relations. The models showed higher accuracy for conjunctives (e.g., *and*, *also*, *additionally*), but lower accuracy for causal connectives (e.g., *so*, *thus*, *therefore*). This result was interpreted as the models being quick to predict *and* in contexts where any connective could fit. Researchers speculated whether the models were capturing the fine-grained meanings of *and*, such as its common interpretation implying a temporal sequence, beyond its basic additive meaning.

To test whether LMs detect the multiple meanings of *and*, items like those in (15) were created. Each item contained a cue that either reinforced the temporal implicature of *and* (e.g., *which is to say* in (15) (a)) or canceled it (e.g., *in fact* in (15) (b)). If LMs were sensitive to the temporal implicature of *and*, they should assign a higher probability to the word *before* when the temporal meaning is reinforced in examples like (a). Conversely, LMs should favor *after* when the temporal implicature is canceled, as seen

in examples like (b).

(15) a. Reinforcement test: *Maggie did the paperwork by hand **and** the company bought new computers, **which is to say**, Maggie did the paperwork by hand [MASK] the company bought new computers*. (Ideal target probability: *before* > *after*)

b. Cancellation test: *Maggie did the paperwork by hand **and** the company bought new computers, **in fact**, Maggie did the paperwork by hand [MASK] the company bought new computers.* (Ideal target probability: *after* > *before*)

With these tests about the fine-grained meanings of *and*, Pandia et al. (2021) reported that models did not achieve accuracy above chance. The poor accuracy observed in the reinforcement and cancellation tests (as illustrated in (15)) is contrasted with the higher accuracy seen in the preliminary analysis (shown in (14)). This disparity suggests that while LMs being tested by Pandia et al. (2021) were able to predict discourse connectives to some extent, they were not yet capable of using pragmatic cues (e.g., *which is to say* in (15) (a), *in fact* in (15) (b)) to adjust their predictions of these connectives.

Furthermore, the poor accuracy obtained with the reinforcement and cancellation tests implies that LMs did not effectively discern different continuations based on the fine-grained discourse meanings of *and*. This result suggests that predicting multiple discourse meanings of a function word remains a challenge for LMs. Consequently, predicting the different cancelabilities of the exhaustivity implicated by the *-lul* and *-nun* markers may also pose a significant challenge for LMs, due to the multiple discourse statuses (Focus and CF, Topic and CT) signaled by each marker.

Results on discourse-related tasks so far do not provide strong evidence that LMs will be competent at detecting discourse functions for the *-lul* and *-nun* markers or at predicting pragmatically felicitous continuations to appear after these markers. However, studies that focused on evaluating LMs with making pragmatic reasonings based on linguistic structures showed some evidence demonstrating LMs' capabilities.

To begin, Jeretič et al. (2020) tested whether models trained on the MultiNLI dataset, which focuses on pragmatic inferences (Williams, Nangia, & Bowman, 2018), could learn to infer implicatures

triggered by lexical items such as *some* (implying *not all*) and *or* (implying *not both*). To evaluate this, they developed the IMPPRESS dataset. The dataset included pairs of sentences that shared an implicature relationship, as illustrated in (16), as well as pairs that were logically contradicted, as shown in (17).

(16) Implicature: Jo ate some of the cake. Joe didn't eat all of the cake.

(17) Contradiction: Jo ate some of the cake. Jo ate nothing.

Besides pairs of determiners with implicature relations, such as *some* and *all* in (16), sentence pairs involving connectives (e.g., ⟨or, both⟩), numerals (e.g., ⟨2, 3⟩), modals (e.g., ⟨can, have to⟩), scalar adjectives (e.g., ⟨fine, great⟩), and scalar verbs (e.g., ⟨go towards, get to⟩) were used to form implicature and contradiction relations. In addition, the dataset included sentence pairs with presupposition relations. Examples of these items are presented in (18). To serve as controls for presupposition items, contradictory pairs like those illustrated in (19) were also included in the dataset.

(18) Presupposition

a. It is Sandra who disliked Veronica. Someone disliked Veronica.

b. It is Sandra who disliked Veronica. Exactly one person disliked Veronica.

c. Question: Sue learned why Candice testified. Candice testified.

(19) Pressuposition contradiction: Jo's cat didn't go. Jo's cat went.

The IMPPRESS dataset was used to evaluate a Bag-of-Words (BOW) model, InferSent (Conneau, Kiela, Schwenk, Barrault, & Bordes, 2017), and BERT (Devlin et al., 2019), each trained on the MultiNLI dataset. Results showed that BERT demonstrated an understanding of scalar implicatures related to a subset of determiners including *some* and *all*, but not for other scalar terms. Both BERT and the BOW model also exhibited an understanding of presuppositions triggered by cleft and question constructions. Since presuppositions and scalar implicatures associated with specific lexical items were largely absent from the MultiNLI dataset used for pre-training, any positive performance on IMPPRESS evaluations suggests that the models utilized prior knowledge from their pre-training to make pragmatic inferences.

Thus, Jeretič et al. (2020) interpreted their results as evidence that even if pre-trained LMs may not be able to capture pragmatic inferences, further training with relevant data can train LMs to capture pragmatic inferences.

Schuster, Chen, and Degen (2020) evaluated LMs with a more specific topic in pragmatics focusing on scalar implicatures. They investigated the extent to which neural network sentence encoders can predict the strength of scalar inferences. For example, the lexical scalar term *some* in (20) (a) leads to the pragmatic inference *some, but not all* in (20) (b), whereas its semantic interpretation would be *some, and possibly all*.

(20) a. I like to read *some* of the philosophy stuff.

b. I like to read *some, but not all*, of the philosophy stuff.

They first gathered human speakers' judgments on similarity between pairs of sentences: one containing a scalar term and the other expressing the pragmatic inference of that scalar. Higher similarity ratings between the pairs (e.g., (20) (a) and (b)) was interpreted as an indication of stronger pragmatic inferences. After gathering human similarity ratings, they tested whether neural sentence encoders could predict the human similarity ratings. If neural encoders accurately predicted human ratings, it would indicate their capability to encode pragmatic inferences conveyed by linguistic cues, that is, scalar expressions.

Schuster et al. (2020) reported the result that neural sentence encoders are capable of using linguistic cues to predict human ratings on scalar inferences. However, these encoders did not utilize contextual information as human speakers do when making predictions. Adding preceding conversational context to the model's input did not improve prediction accuracy, and in fact, the model was able to make accurate predictions in most cases without considering the preceding context. Schuster et al. (2020) interpreted this result as evidence of that conversational context was not necessary for the model to draw inferences or that the model did not make adequate use of the context. Given the result that LMs struggled with utilizing discourse context for processing discourse cues, it is suspected that processing exhaustive status based on discourse context and contextualized alternatives could be particularly challenging for LMs.

Using a similar experimental paradigm as Schuster et al. (2020), Li, Schuster, and Degen (2021) investigated neural encoders' capability to predict implicatures associated with *or*, specifically the 'not both' interpretation. They evaluated sentence pairs, such as (21)-(23), where other linguistic cues are included to affect the interpretation of *or*. For instance, in (21), the term *either* could trigger the 'not both' interpretation, while negation in (23) could lead to a 'neither' interpretation.

(21) a. You're **either** liberal *or* conservative.

b. You're **either** liberal *or* conservative, *but not both*.

(22) a. And it naturally, uh, composts *or* stores.

b. And it naturally, uh, composts *or* stores, *but not both*.

(23) a. Well, I **don't** believe in drugs *or* alcohol.

b. Well, I **don't** believe in drugs *or* alcohol, *but not both*.

Unlike the relatively uniform human ratings for implicatures associated with *some* found in Schuster et al. (2020), human similarity ratings for *or* varied significantly across different linguistic contexts. Human ratings revealed that similarity scores for pairs with *either* (e.g., (21)) were higher than those with negations (e.g., (23)), confirming that *either* could be a contextual cue to trigger 'not both' interpretation of *or*. This result is interpreted as evidence presenting the tendency that contextual cues can significantly influence the pragmatic interpretations of *or*. Therefore, predicting human similarity scores for items such as (21)-(23) presents an additional challenge for neural encoders, as they must consider and integrate contextual cues to process the interpretation of *or*.

Li et al. (2021) reported that while the model learned the association between contextual cues and inference strength, it did not perform as well as with *some* implicatures in Schuster et al. (2020). Thus, they concluded that neural networks show the ability to learn scalar inferences encoded linguistically, but struggle with greater variability in interpretations modulated by contextual cues. Regarding the Korean markers *-lul* and *-nun*, this suggests that LMs may face additional challenges in encoding markers' discourse uses. As defined earlier, *-lul* and *-nun* markers are polyfunctional. They appear after object phrases to indicate grammatical object functions, and can signal additional discourse statuses (Focus or CF, Topic

or CT) depending on the context. To accurately identify the relevant discourse function of these markers (or the lack of the discourse function), LMs need to consider contextual cues, similar to how they had to incorporate contextual cues for *or* seen in Li et al. (2021). Thus, the variability in the *-lul* and *-nun* markers' interpretations influenced by contextual cues may further challenge LMs.

Hu et al. (2023) provides a comprehensive evaluation of larger LLMs with various pragmatic phenomena. Their domain of evaluation included seven pragmatic phenomena—deceits, indirect speech, irony, conversational maxims, metaphor, humor, and coherence. For example, (24) illustrates an item providing a task of interpreting indirect speech. After providing the input containing a pragmatic phenomenon (e.g., "Are you going out?"), four response options were presented, consisting of correct interpretation of the indirect speech (1st option), literal interpretation of the sentence (2nd option), and two distractors (3rd and 4th options).

(24) Nate is about to leave the house. His wife points at a full bag of garbage and asks: "Are you going out?" What might she be trying to convey?

1. She wants Nate to take the garbage out.
2. She wants to know Nate's plans.
3. She wants Nate to bring his friends over.
4. She wants Nate to spend more time with the family.

After prompting models with items like (24), they compared the probabilities assigned to each answer token (e.g., the tokens "1", "2", "3", or "4"). The model was considered correct on a given item if it had assigned highest probability to the correct answer token, among all the possible answer tokens for that item. They generated 5 versions of each item by randomizing the order of answer options.

Using evaluation items designed in the structure of (24), Hu et al. (2023) tested GPT-2, T*k*-Instruct (3B), T*k*-Instruct (11B), Flan-T5 (base), Flan-T5 (XL), InstructGPT-3 (ada), and text-davinci-002. These models differed in size, training input, and the incorporation of human feedback during training. Results revealed that larger variants of the same model (e.g., Flan-T5 (XL), compared to Flan-T5 (base); T*k*-Instruct (11B), compared to T*k*-Instruct (3B)) consistently demonstrated better understanding of prag-

matic meanings across the seven phenomena tested. This finding aligns with the previously reviewed findings of Schuster and Linzen (2022), indicating that larger models are more likely to successfully capture pragmatic meanings in natural language.

In addition, it was found that Flan-T5 (XL) and text-davinci-002 not only achieved high accuracy but also made errors similar to those of humans: when selecting incorrect answers, these models tended to choose the literal interpretation of an utterance over distractors. The Flan-T5 model differ from other models being tested in that it is fine-tuned to follow human instructions. Text-davinci-002 is also distinct from others in that it received from human feedback during training. Although the authors avoided making premature conclusions about the importance of instruction tuning and human feedback from these results, their result demonstrated that both of these factors are worth considering when evaluating LMs' capability in pragmatic tasks.

In addition, Hu et al. (2023) reported that LMs demonstrated higher accuracy with conventional implicatures (i.e., implicatures encoded by linguistic structures; see section 3.2) but lower accuracy with conversational implicatures based on social norms and expectations. This finding suggests that LMs may particularly struggle with the exhaustivity implications of *-lul* and *-nun* markers, as both *-lul* and *-nun* markers' exhaustivity exhibit some features of conversational implicatures: The *-lul*'s implicature marker exhibits cancelablity (Experiment 2), and the *-nun* marker's implicature is conveyed based on the discourse context (Experiment 1).

In summary, studies reviewed thus far show that LMs have some capability in encoding pragmatic meanings in natural language (Schuster & Linzen, 2022), especially when trained on dataset targeting pragmatics (Jeretic et al., 2020). However, LMs often fall short in utilizing contextual cues effectively when interpreting scalar expressions (Schuster et al., 2020; Li et al., 2021). In particular, LMs were not competent at predicting discourse entities based on discourse context (Paperno et al., 2016), predicting the interpretations of a discourse connective *and* having multiple meanings (Pandia et al., 2021), and processing conversational implicatures not signaled by linguistic cues (Hu et al., 2023). These results present mixed evidence regarding whether even larger LMs will be able to accurately represent the discourse functions of Korean markers.

In the following, two experiments investigate LLMs' representation of exhaustivity encoded by

the *-lul* and *-nun* markers. Experiment 5 assesses LLMs' sensitivity to exhaustivity implicatures by measuring the probabilities and elicited ratings assigned to sentences that either cancel or maintain these implicatures. This experiment employs a similar paradigm to Experiment 2, allowing us to compare how LLMs' representations of the markers' discourse functions differ from patterns observed in human speakers. Experiment 6 adopts a paradigm similar to Experiment 3 to examine whether LLMs generate *-lul* and *-nun* markers to convey exhaustivity. Details of the experimental setup are provided in the following section.

5.3 Models under evaluation

Previous studies suggest that LMs can capture contextual meanings of words, particularly when both the training data and model size are sufficiently large (Hu et al., 2023; Jeretič et al., 2020; Tenney et al., 2019). Considering this, we evaluated models with varying numbers of parameters to assess whether LMs can effectively capture discourse meanings as their size increases. To mention models with smaller parameters to larger ones, we included KoGPT-2 (125M)[7] and KoGPT-Trinity (1.2B),[8] which are specifically developed and trained for Korean. We also tested Polyglot-Ko models of three different sizes: 3.8B,[9] 5.8B,[10] and 12.8B[11] parameters. These models are part of a multilingual LLM project and were primarily trained with Korean data.

Lastly, we tested text-davinci-003 (henceforth GPT-3)[12] and gpt4-1106-preview (henceforth GPT-4), both accessed with OpenAI API.[13] While the model size of GPT-3 is not publicly available, it is estimated to have 175B parameters (`https://blog.eleuther.ai/gpt3-model-sizes/`). At the time of the study, GPT-4 was considered the largest model, though its exact number of parameters is also not publicly available. These two models are notable for being trained with reinforcement learning from human feedback (RLHF) (c.f., Ouyang et al., 2022) in the fine-tuning process. Despite not being specifically

[7] `https://huggingface.co/skt/kogpt2-base-v2`
[8] `https://huggingface.co/skt/ko-gpt-trinity-1.2B-v0.5`
[9] `https://huggingface.co/EleutherAI/polyglot-ko-3.8b`
[10] `https://huggingface.co/EleutherAI/polyglot-ko-5.8b`
[11] `https://huggingface.co/EleutherAI/polyglot-ko-12.8b`
[12] Accessed before its deprecation on January 4th, 2024.
[13] `https://platform.openai.com/docs/api-reference`

developed for the Korean language, their massive size and extensive training data enable them to exhibit proficiency in Korean. The temperature for each model was set to 0, ensuring that the models assigned probabilities deterministically.

All models being tested share the commonalities of LLMs previously mentioned—pre-training methods, large-scale training data, transformer architecture, and contextualized embeddings. However, the models vary in terms of their access to source code, the data used to train the models, and their embeddings. Ko-GPT models and Polyglot-Ko models are open-source models, meaning that their source code, training data, and embeddings can be accessed freely by the public. Although OpenAI's GPT-3 is not an open-source model, its embeddings can be accessed by the public. Having access to these models' embeddings mean that we could gather and analyze the logits/log probabilities assigned to input and output tokens. Among the set of models tested, GPT-4 was the only model that did not license access to its embeddings. As a result, we used distinct approaches to evaluate these two types of models, as detailed in the following section.

5.4 Experiment 4: Processing patterns

Experiment 4 is designed to compare LLMs' processing of discourse meanings with the patterns observed in human speakers from Experiment 2. The goal is to determine whether LLMs' representations of words reflect the discourse meanings of the *-lul* and *-nun* markers. Specifically, we compare LLMs' and humans' sensitivity to exhaustivity implicatures indicated by these markers and the different cancelability of each marker's implicatures. Details of the experimental items are provided in section 5.4.1, and the results are reported in section 5.4.2.

5.4.1 Materials and measurements

To evaluate LLMs, we created items following the structure of stimuli similar to those used in Experiment 2 (see Table 5.3). 48 items, each with a unique context preamble, were used. Each of these 48 items was presented with 6 different manipulations, as illustrated in Table 5.3, resulting in a total of 288 items. Only the italicized components in the table were presented in Korean, and only one option in

curly brackets was included per item. The example in Table 5.3 written in Korean can be found in the Appendix A.4.

Table 5.3: An illustration of items used in the Experiment 4.

Context	*Sohee knows that Mina and Yuna attended a marathon race which had medals and trophies as prizes.*
Wh-question	*Sohee: Yuna-nun meo-l bat-ass-eo?* Yuna-TOP what-ACC receive-PST-INT "What did Yuna receive?"
First response	*Mina: Medal-**{ul/un}** bat-ass-eo.* Medal-CF/CT receive-PST-DEC "(Yuna) Received a medal."
Second response (Target of the surprisal measurement)	*Trophy-**{nun mos/to/man}** bat-ass-eo.* Trophy-CT NEG/also/only receive-PST-DEC "(Yuna) {Didn't receive/Also received/Received only} a trophy."

To begin, the context sentence shown in the top row of the Table 5.3 was always presented in the beginning of the item. The context preamble from the Experiment 2 is simplified and used in evaluating LMs. This simplified context preamble stated a set of alternatives, followed by a question about one alternative from the set. In preliminary investigations, we used the exact items (without simplifying the context preamble) with LLMs, but this did not yield any differences across conditions. Given LLMs' limited capacity for reasoning Theory of Mind and varying knowledge states of speakers and listeners (Ullman, 2023), the initial context preamble might not have been effective or comprehensible for LLMs in grasping the shared knowledge statuses of characters. In addition, the longer preamble used with human participants might have been too lengthy for the smaller LLMs we tested, considering the size of the context window (i.e., how much prior input LLMs consider when assigning probability to the next word). Thus, it is possible that the longer version of the context preamble used in preliminary investigations could not be factored in when LMs assigned probability scores to the second response, resulting in no systematic differences varying for different markers.

Following the context preamble, a wh-question was presented asking about the object. The response portion of the conversation always consisted of two sentences. The first sentence was manipulated to have different object markings between *-lul* and *-nun*. The second response stated how the other al-

ternative from the set forms the relation with the elided subject. The second sentence either maintained the exhaustivity (with the negation *mos*), canceled the exhaustivity (with the *-to* marker), or contradictied the exhaustivity (with the *-man* marker) established in the first sentence. The contradicted continuation is what we included as fillers in Experiment 2 (c.f., section 3.2.2).

All Ko-GPT series models and Polyglot-Ko series models (i.e., all models except GPT-4) were assessed using the surprisal of the second sentence. To calculate surprisal, logits assigned to each token in the sentence were first converted into log probabilities. We then summed the log probabilities of all tokens in a sentence and normalized by the number of tokens. These were then converted to surprisal. For GPT-3, we skipped the step of converting logits into log probabilities, as the output already provides log probabilities. The last layer of embeddings was used to obtain surprisal, based on the finding that the last layer represents the contextual interpretation of the input (Tenney et al., 2019).

The predictions of the surprisal patterns of our items are as follows. If models can distinguish between logical contradictions and pragmatic violations, they should exhibit notably higher surprisal for contradictory continuations compared to maintained or canceled continuations. If models can perceive the exhaustivity implicature and its cancelability, as observed with human participants in Experiment 2, they should show notably higher surprisal for the canceled continuation following the *-nun*-marked object in the first sentence, compared to the canceled continuation following the *-lul*-marked object in the first sentence.

GPT-4 was evaluated by prompting the model to rate the appropriateness of each type of continuation following the previous sentence. This rating approach was similar to the one used in Experiment 2. However, as previously noted, GPT-4 was given a simplified context preamble. A 7-point Likert scale was provided in the 'system message' in Korean: "Make sure to respond only with a number between 1 and 7. The number 1 indicates inappropriate and 7 indicates appropriate." This system message served as a meta-instruction guiding how the model should respond to each prompt, analogous to the 7-point scale and experimental questions provided after each stimulus in Experiment 2.

To summarize, each of the 7 models was presented with a total of 288 items, consisting of 48 items with 6 different manipulations: LUL-MAINTAINED, LUL-CANCELED, LUL-CONTRADICTED, NUN-MAINTAINED, NUN-CANCELED, and NUN-CONTRADICTED. Each model encountered all items under

each manipulation in a zero-shot manner, meaning each item was presented individually, one at a time. This approach ensured that the models' responses reflected their pre-existing knowledge and patterns learned from training data, rather than any adaptation to the specific evaluation task.

5.4.2 Results

We first report surprisal scores obtained from the Ko-GPT models, Polyglot-Ko models, and GPT-3, summarized in the left six subfigures in Figure 5.3. Higher surprisals are interpreted as indications that the model had lower expectations for encountering the continuation. Conversely, lower surprisals are interpreted as indications that the discourse continuation was evaluated as more expected. Therefore, higher surprisals can be interpreted as equivalent to lower appropriateness ratings from humans, indicating lower pragmatic acceptance. For comparison, the result from Experiment 2 obtained with human speakers are presented in the bottom right subfigure in Figure 5.3.

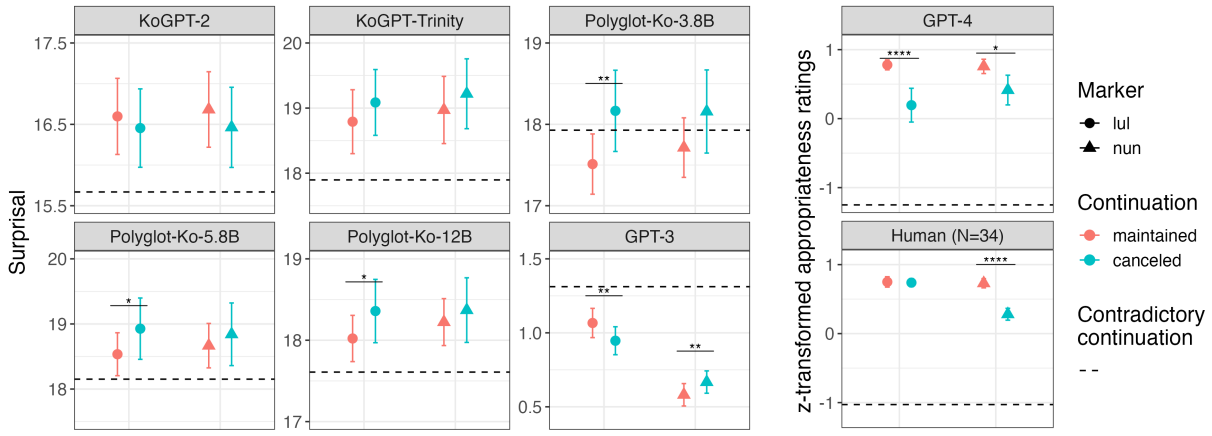


Figure 5.3: Mean surprisals and 95% CIs gathered from KoGPT and Polyglot-Ko models, and mean of z-transformed appropriateness ratings and 95% CIs gathered from GPT-4 and 34 native Korean speakers. Stars indicate adjusted significance levels obtained from paired t-tests with Bonferroni corrections (****: $p < 0.001$, **: $p < 0.005$, *: $p < 0.05$).

Comparing results for logical contradictions (dashed line in Figure 5.3) and pragmatic violations (i.e., *-nun* marker followed by a canceled continuation), it is visually evident that human speakers clearly distinguished between these two types of judgments. From the Ko-GPT models, Polyglot-Ko models, and GPT-3, only GPT-3 successfully differentiated logical contradictions from pragmatic violations, similar

to human speakers. Polyglot-Ko-3.8B reported that canceled continuations following both *-lul* and *-nun* (indicated with blue shapes in the figure) are more surprising (i.e., less expected) compared to contradictory continuations. The remaining four models reported that contradictory continuations as less surprising (i.e., more expected) than any other types of continuations, irrelevant to whether the cancelability of the marker's exhaustivity was violated.

When comparing the interactions of markers and continuations, GPT-3 was the only model to show significantly increased surprisal when a canceled continuation, not the maintained continuation, followed the *-nun* marker ($t(47) = 4.0173$, $p = 0.0002$). KoGPT-Trinity and the three Polyglot-Ko models also exhibited higher surprisals for canceled continuations following *-nun*, but this was not significantly higher than the surprisals for *-nun* followed by maintained continuations. Thus, GPT-3 was the only model sensitive to the non-cancelable exhaustivity implicature associated with the *-nun* marker.

However, GPT-3 also displayed a pattern that diverged from human results. It exhibited higher surprisal when maintained continuations followed the *-lul* marker, compared to when a canceled continuation followed the *-lul* marker ($t(47) = 17.581$, $p < 0.001$). This suggests that GPT-3 predicted the *-lul* marker to be followed by canceled exhaustivity implicature. This contrasts with human results, where the *-lul* marker was perceived as appropriate to be followed by both maintained and canceled continuations. In addition, all Polyglot-Ko models showed increased surprisal for maintained continuations following the *-lul* marker (3.8B: $t(47) = 3.9399$, $p = 0.0003$; 5.8B: $t(47) = 3.0318$, $p = 0.0039$; 12B: $t(47) = 3.1329$, $p = 0.0030$), indicating a divergence from human results.

Now we move on to discuss elicited ratings obtained from GPT-4. Ratings were transformed into z-scores, using the mean and standard deviation obtained within the model. The z-transformed ratings obtained from GPT-4 are presented in the upper right subfigure in Figure 5.3. Comparing the logical contradictions and pragmatic violations first from the figure, it is evident that ratings assigned to contradictory continuations and the other two continuations diverge. GPT-4 assigned generally more positive ratings to pragmatic violations for critical items targeting appropriateness ratings, while giving negative ratings for contradictory continuations targeting truth-conditional judgements. This aligns with human results, confirming that GPT-4 had the sensitivity towards the difference between truth-conditional meanings and pragmatic meanings, and assigned gradient ratings to reflect the model's sensitivity.

GPT-4 assigned lower ratings to the *-nun* marker followed by canceled continuations, instead of maintained continuations ($t(47) = -2.7945$, $p = 0.0075$). This is the pattern coherent with human speakers' patterns, suggesting that the model was sensitive towards the *-nun* marker's non-cancelable exhaustivity implicature. However, the model produced more negative ratings for the *-lul* marker followed by canceled continuations compared to the maintained continuations ($t(47) = -4.7085$, $p < 0.001$), which differed from the pattern observed in human speakers. In other words, GPT-4 interpreted canceled continuations as generally less appropriate than the maintained continuations, regardless of which object marker appeared in the previous sentence.

In order to further investigate each model's sensitivity towards the interaction between markers and continuations, a mixed effects model was fit to predict each model's results. In Experiment 2, z-transformed ratings obtained with human participants were predicted by the mixed effects model, repeated in (25). In mixed effects models predicting LLM results, there was no need to control random effects of participants. Thus, surprisal results were predicted with marker (*-lul* or *-nun*), continuation (repeated or negated; excluding the contradicted continuation to have the same levels as in the humans' model in (25)), and the interaction of the two predictors, while controlling for random effects of each lexicalized item. The regression formula is given in (26). For predicting elicited ratings obtained with GPT-4, the same predictors used in (26) were used to predict z-transformed ratings, as presented in (27).

(25) Human: `z_rating ~ marker * continuation + (1 + marker + continuation|participant) + (1|item)`

(26) Ko-GPT and Polyglot-Ko: `surprisal ~ marker * continuation + (1|item)`

(27) GPT-4: `z-rating ~ marker * continuation + (1|item)`

Predictions made by these three types of mixed effects models are reported in Table 5.4. LLMs are listed from smaller to larger ones in the table. In the first six columns reporting models predicting surprisals, positive coefficients indicate increasing the surprisal, thus decreasing LLM's expectedness. In the last two columns reporting models predicting z-transformed ratings obtained with GPT-4 and human, positive coefficients indicate increase in the appropriateness ratings.

Table 5.4: Surprisal and z-transformed ratings predicted by mixed effects models. ***: $p < 0.001$, *: $p < 0.05$.

Coefficients	β of surprisal (p-value)						β of z-transformed ratings (p-value)	
	KoGPT-2	Ko-Trinity	Poly-3B	Poly-5B	Poly-12B	GPT-3	GPT-4	Human
Intercept	16.60 (***)	18.79 (***)	17.51 (***)	18.53 (***)	18.02 (***)	1.07 (***)	0.78 (***)	0.75 (***)
Marker:NUN	0.08 (0.46)	0.18 (0.30)	0.20 (0.15)	0.13 (0.24)	0.20 (*)	-0.49 (***)	-0.02 (0.85)	-0.02 (0.65)
Continuation:CANCELED	-1.14 (0.21)	0.30 (0.09)	0.65 (***)	0.39 (***)	0.34 (***)	-0.12 (***)	-0.58 (***)	-0.01 (0.88)
Marker:NUN× Continuation:CANCELED	-0.08 (0.64)	-0.05 (0.85)	-0.21 (0.29)	-0.22 (0.17)	-0.19 (0.14)	**0.21 (***)**	0.24 (0.15)	**-0.44 (***)**

Comparing the first two KoGPT series models with Polyglot-Ko models, we observe that the Polyglot-Ko models show sensitivity towards the repeated continuation. As the Polyglot-Ko models reaches 12B parameters, they also show sensitivity towards the markedness of an object. GPT-3 exhibited sensitivity to the interaction between the the marker and continuation factors: Surprisal reported by GPT-3 is predicted to increase when the *-nun* marker is followed by canceled continuation. This suggests that GPT-3 was sensitive towards the non-cancelability of *-nun* marker's exhaustivity, which was also observed with human ratings. Despite the coherent pattern with human ratings, GPT-3 displayed sensitivity to the main effects of the two predictors, which human speakers did not modulate the ratings for. Overall, the first six mixed effects models reported in the table confirm that only GPT-3 was sensitive towards the *-nun* marker's non-cancelable exhaustivity.

Now we discuss the mixed effects model predicting z-transformed ratings reported by GPT-4. GPT-4 demonstrated sensitivity to the main effect of the continuation factor and predicted lower ratings with canceled continuations. This pattern was not observed with human speakers. Although the *t*-test results previously discussed revealed that GPT-4 assigns lower ratings to the *-nun* marker followed by the canceled continuation, mixed effects model shows that it was due to the model's sensitivity towards the repeated continuations, not towards the non-cancelability of the *-nun* marker's implicature.

In summary, only GPT-3 exhibited sensitivity to the cancelability of the *-nun* marker's exhaustivity implicature. We also observed that Polyglot-Ko models and GPT-3 demonstrated modulated surprisal when the *-lul* marker appeared with different continuations (as shown by the round points in Figure 5.3).

This result may be attributed to two potential explanations. First, LLMs may have captured the generalized association between the *-lul* and *-nun* markers' contextual interpretation (i.e., exhaustivity implicature) and cancelability. Under this assumption, GPT-3 appears to have developed sensitivity toward the *-nun* marker's exhaustivity implicature.

However, the results could also stem from models' sensitivity to spurious patterns. For example, GPT-3 might have learned that the *-nun* marker is more frequently followed by negations than non-negated sentences, which could explain the increased surprisal for canceled continuations that do not contain negations. Similarly, GPT-3 may have associated the *-lul* marker with fewer negations, leading to increased surprisal for maintained continuations that do contain negations. Polyglot-Ko models might have learned to expect the *-lul* marker to be followed by maintained continuations with negations, as such patterns could have been more frequent in their training data. In these cases, models have not necessarily made associations between the marker and its discourse interpretation. In order to determine whether the results from Experiment 5 are due to the models' representation of the discourse functions of the *-lul* and *-nun* markers or simply a reflection of spurious patterns, we conducted Experiment 6.

5.5 Experiment 5: Production patterns

5.5.1 Materials and measurements

The evaluation items were designed to assess whether the modulated surprisals and ratings observed in Experiment 4 stem from the LLMs' representation of the discourse meanings linked with the *-lul* and *-nun* markers. We designed 480 stimuli (48 items, each appearing with 10 different manipulations), as illustrated in Table 5.5. In the actual experiment, only the italicized components in the right column are presented in Korean, and only one of the options within the curly brackets is shown in each item (refer to the Appendix A.5 for the Korean version of the item).

Each item began with a wh-question asking about an object, followed by an intended message statement in which the exhaustivity status of the objects varied. When the intended message was EXHAUSTIVE, it explicitly stated that the speaker intended to indicate the exhaustivity of the object in the response (e.g., "only the medal" in Table 5.5). Conversely, when the intended message was NON-EXHAUSTIVE, it

Table 5.5: An example of items used in the Experiment 5.

Wh-question	*Sohee: Yuna-nun meo-l bat-ass-e?* Yuna-TOP what-ACC receive-PST-INT "What did Yuna receive?"	
Intended message	*Mina intends to respond that Yuna received* ***{only the medal/both the medal and the trophy}***.	
Speaker	*Mina:*	
Response sentence for Ko-GPT and Polyglot-Ko models	*Medal-****{lul/nun/∅/man/irang trophy tul ta}*** medal-{LUL/NUN/∅/only/and trophy two both} [With *-lul*, *-nun*, ∅] "(Yuna) Received a medal." [With *-man*] "(Yuna) Received only a medal." [With *irang trophy tul ta*] "(Yuna) Received both a medal and a trophy."	*bat-ass-e.* receive-PST-DEC
OR Response options for GPT-4 and humans	• *Medal-****{lul/nun}*** *bat-ass-e.* • *Medal-∅ bat-ass-e.*	

introduced two objects (e.g., "both the medal and the trophy" in Table 5.5), and indicated that the speaker intended to express non-exhaustivity of those objects.

After the speaker's name was introduced, a response sentence followed for the Ko-GPT series models and Polyglot-Ko series models. This response was structured based on one of five options listed in Table 5.5 and was provided as the speaker's reply to the wh-question. The first three marking options (*-lul*, *-nun*, and null-marking) resulted in three different sentences that varied only in their object markings. Although these sentences shared the same truth-conditional meaning, they differed in their pragmatic implicatures. Specifically, Chapters 2-4 previously demonstrated that while the *-lul* and *-nun* markers could imply the exhaustive status of the objects, the null-marking option lacked this implicature.

The final two response options—using the *-man* marker or a conjunction phrase—were implemented to serve as baseline results. When the intended message indicated the exhaustive status of the object and the *-man* marker was used, the response sentence with the *-man* marker simply verbatim repeated the intended message from the relative clause. Similarly, when the intended message stated non-exhaustive status, the response sentence with a conjunction verbatim repeated the intended message from the relative clause. These are termed as VERBATIM responses.

In addition, another baseline was created by having responses that contradicted the intended mes-

sage at the truth-conditional level. When the intended message was exhaustive, including a conjunction phrase (e.g., "medal and trophy both" in Table 5.5) in the response resulted in a false response. Conversely, when the intended message was non-exhaustive, using the *-man* marker in the response sentence semantically contradicted the intended message. These responses are referred to as CONTRADICTORY responses.

For the Ko-GPT and Polyglot-Ko series models, we evaluated the log probabilities assigned to response sentences with varying object markings. The log probability for each sentence was calculated in the same way as in Experiment 4: logits for each token in the sentence were converted into log probabilities, which were then summed and normalized by the number of tokens in the sentence. For GPT-3, log probabilities were directly accessed without needing conversion. In this experiment, we measured log probabilities instead of surprisal to directly reflect how likely each model was to generate different object markings, offering insights about the models' likelihoods in producing marking variations.

For the Ko-GPT and Polyglot-Ko series models, two distinct types of prompts were generated for each item: one included the intended message statement, while the other did not. To assess how the exhaustivity status in the intended message influenced the log probabilities assigned to each response type, we subtracted the log probabilities of the response sentence when the intended message was included from the log probabilities of the response when the intended message was excluded. This subtraction was intended to reveal the extent to which the exhaustivity status affected the models' interpretation of each response. As a result, each Ko-GPT and Polyglot-Ko series model received 960 prompts in total, with 480 prompts including the intended message statement and 480 excluding it. In the results section, we report the 480 log probability measurements obtained from each pair of inputs (with and without the intended message statement) across all models.

The predicted patterns of log probabilities are as follows. If the LLMs successfully associate the *-lul* and *-nun* markers with the exhaustivity status of the object, we expect the following pattern. When the exhaustivity of the object is intended, the models should assign higher probabilities to responses where the object is marked with either the *-lul* or *-nun* markers. When exhaustivity is not intended, the models should assign higher probabilities to null-marked responses over the *-lul-* and *-nun*-marked responses. These results will indicate whether the models have learned to connect the specific object markings with

the intended exhaustivity status within discourse contexts.

Instead of presenting a response sentence, items for GPT-4 and human participants included a list of response options, as illustrated at the bottom of Table 5.5. These response options always consisted of a pair of sentences that differed only in the markedness of objects. The marked response option contained either the *-lul* or *-nun* marker. Each marked option was paired with a null-marked sentence, generating a list of two response options. If both the *-lul* and *-nun* markers were presented simultaneously as response options, participants or the model might hesitate to choose the *-lul*-marked option due to its canonical grammatical function. By presenting pairs of marked and unmarked sentences, the experiment ensured that neither GPT-4 not the human participants would default to considering the *-lul* marked option solely based on its grammatical role. An example of these forced-choice tasks written in Korean is available in the Appendix A.5.

After viewing the response options, both GPT-4 and human participants were asked to choose the most appropriate response to the preceding wh-question. For GPT-4, we provided the following system message via the OpenAI API: *Please choose the response as you would speak in everyday conversations. Provided options may not express everything that you need to say. Nevertheless, please choose the best option from the two.*. GPT-4 was presented with each prompt twice—once with the response options in one order and then with the order switched. In total, GPT-4 saw 48 items, each followed by two different lists of response options (one with the *-lul* marker and null-marking, and the other with the *-nun* marker and null-marking). Since each item was presented twice with switched response order, GPT-4 evaluated a total of 192 prompts.

Human participants were instructed after each question with the same prompt used in Experiment 3: *From the sentences below, which one is the best for Mina to send as a response? Sentences below may not include everything that Mina has to say. Regardless, please select the option that you think is the best*. In the human experiment, exhaustivity of the intended message (EXHAUSTIVE, NON-EXHAUSTIVITY) and two different lists of response options (one list having the *-lul*-marked and null-marked object, the other list having the *-nun*-marked and null-marked object) were treated as critical experimental manipulations. 24 items appeared with these 4 manipulations, yielding 96 critical items. 96 critical items were counterbalanced into four lists so that each participant saw 24 of them in the experiment, each item ap-

pearing only once in one condition. In addition to 24 critical items, each human participant was presented with 24 fillers on subject markings, and 4 attention check items. Fillers and attention checks were identical to those used in Experiment 3 (c.f., section 4.3.2). The order of the marked and unmarked options in the response lists was randomly decided in every question.

Human participants were recruited and compensated through an online crowdsourcing platform based in South Korea (`https://pickply.com/`). All participants were required to be born in South Korea and to currently reside there. They were compensated with "points," a currency used on the platform, which could later be converted into monetary value. The experiment was approved by the IRB of UC San Diego and conducted according to its guidelines. Originally, 71 participants were recruited. However, following the same data exclusion criteria as in Experiment 3 (see section 4.3.3), responses from 35 participants were included in the final analysis.

Given the results from Experiment 2 and 3, predicted ratings from humans and GPT-4 are as follows. When exhaustivity of the object is intended, human participants are expected to prefer responses with a marked object (either with *-lul* or *-nun*) over responses with a null-marked object. If GPT-4 can utilize the *-lul* and *-nun* markers as cues for exhaustivity, its response patterns should align with those of human participants, favoring responses with marked objects in indicating exhaustive status of objects. Conversely, when the exhaustivity of the object is not intended, human participants are expected to prefer responses with null-marked objects over those with marked objects. If GPT-4 recognizes the *-lul* and *-nun* markers' role in conveying exhaustivity, its choices should similarly favor null-marked responses when non-exhaustivity is intended.

5.5.2 Results

Log probabilities assigned by Ko-GPT series models and Polyglot-Ko series models are reported in Figure 5.4. Solid horizontal lines indicate the mean log probabilities assigned to verbatim responses, such as "Received both the trophy and the medal." when a non-exhaustive message (e.g., both the medal and the trophy) was intended, and "Received only the trophy." when an exhaustive message (e.g., only the medal) was intended. Dashed horizontal lines indicate the mean log probabilities assigned to contradictory responses, such as "Received only the medal" when a non-exhaustive message (e.g., both the medal and

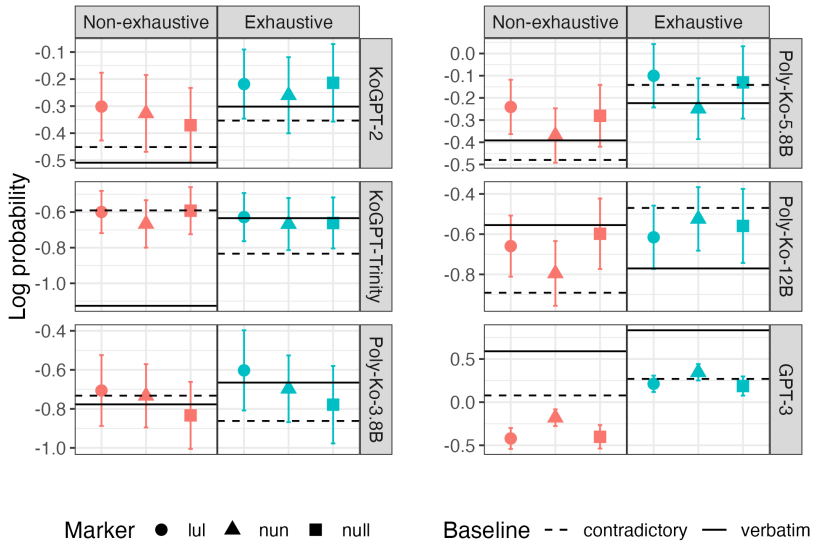


Figure 5.4: The mean log probabilities assigned to *-lul*, *-nun*, and null-marked responses for non-exhaustive and exhaustive messages. Error bars represent 95% confidence intervals (CIs).

the trophy) was intended, and "Received both the trophy and the medal" when an exhaustive message (e.g., only the medal) was intended.

Except GPT-3, all other models did not consistently assign higher probabilities to verbatim responses than to contradictory responses. As presented in Figure 5.4, the three subfigures on the left side show that KoGPT-2, KoGPT-Trinity, and Polyglot-Ko-3.8B assigned higher probability to the contradictory continuations than to verbatim continuations when non-exhaustive message was intended in the input prompt. The other two Polyglot models—Polyglot-Ko-5.8B and Polyglot-Ko-12B—assigned higher probabilities to contradictory continuations when the exhaustive message was intended. These results suggest that models besides GPT-3 were not competent at assigning probabilities based on the intended message presented in the beginning of the input.

Now we move on to the main results obtained with different object markings, indicated with colored points and error bars in Figure 5.4. Results obtained with Polyglot-Ko-12B stand out, as the lowest probabilities were assigned to *-nun*-marked responses when exhaustivity was not intended (indicated with a red triangle under the Polyglot-Ko-12B subfigure). When exhaustivity was intended, the model assigned higher probabilities to the *-nun*-marked responses (indicated with a blue triangle). However, this

result does not provide strong evidence that Polyglot-Ko-12B associated the *-nun* marker with exhaustivity status. The null-marked responses were assigned a similar level of probabilities to *-nun*-marked responses when exhaustivity was intended, indicating that the model did not distinguish between the different exhaustivity implicatures of the two marking options. In addition, because the baseline results did not confirm that the model assigned probabilities to response sentences in the context of the intended message, the modulated probabilities we observe may not reflect the association between markers appearing in the response and exhaustivity status.

In Experiment 4, we observed that GPT-3 exhibited surprisal patterns that were partially comparable to humans' appropriateness ratings. Specifically, it assigned higher surprisal (equivalent to lower appropriateness) when the *-nun* marker was followed by a canceled continuation. However, the results obtained in the current experiment demonstrate that GPT-3 did not associate the intended exhaustivity provided in the prompt with the different marking options. As visually apparent in the bottom right subfigure in Figure 5.4, when exhaustivity was intended in the message, the model did not assign significantly different probabilities to *-lul*-marked or *-nun*-marked responses compared to null-marked responses. Moreover, when exhaustivity was intended in the message, GPT-3 assigned a similar level of probabilities to responses with *-lul*-marked, *-nun*-marked, and null-marked objects, as well as to contradictory responses. This result suggests that GPT-3 did not consider different object markings as cues indicating the exhaustivity status of objects intended to be conveyed.

Log probabilities obtained from the six models, presented in Figure 5.4, were further analyzed using a set of mixed-effects models. We fitted a mixed-effects model to predict the log probability of a response sentence based on the intended exhaustivity status of the object and different markings of the object, while controlling for random effects of each lexicalization of an item (`log probability ~ exhaustivity * marker + (1|item)`). Predictions from the mixed-effects models are reported in Table 5.6. The intercept (baseline) represents the predicted log probability for a non-exhaustive intended message with null-markings. Positive coefficients in the table indicate an increase in the probability assigned by the model.

Seen with the Polyglot-Ko-3.8B, Polyglot-Ko-5.8B, and GPT-3, larger models are more likely to assign distinct probabilities when intended message contained exhaustivity. Although these patterns

Table 5.6: Mixed effects models' prediction of log probabilities reported by Ko-GPT models and Polyglot-Ko models (***: $p < 0.001$. *: $p < 0.05$).

	β of log probability (*p*-value)					
	KoGPT-2	Ko-Trinity	Poly-3.8B	Poly-5.8B	Poly-12B	GPT-3
Intercept	-0.59	-0.83	-0.37	-0.28	-0.60	-0.40
	(***)	(***)	(***)	(***)	(***)	(***)
Intended:EXHAUSTIVE	-0.07	0.05	0.16	0.15	0.04	0.59
	(0.14)	(0.30)	(***)	(***)	(0.28)	(***)
Marker:LUL	-0.01	0.13	0.07	0.04	-0.06	-0.02
	(0.88)	(*)	(0.08)	(0.34)	(0.09)	(0.54)
Marker:NUN	-0.07	0.10	0.04	-0.09	-0.20	0.22
	(0.11)	(0.06)	(0.26)	(*)	(***)	(***)
Intended:EXHAUSTIVE	0.04	0.05	-0.07	-0.01	0.00	0.04
× Marker:LUL	(0.55)	(0.51)	(0.18)	(0.87)	(0.92)	(0.32)
Intended:EXHAUSTIVE	0.07	-0.02	-0.09	-0.03	**0.23**	-0.06
× Marker:NUN	(0.31)	(0.80)	(0.10)	(0.62)	**(***)**	(0.17)

do not align with those observed with human participants, larger models (Polyglot-Ko-5.8B, Polyglot-Ko-12B, GPT-3) demonstrate some sensitivity to paradigmatic selections of marking options, as they are likely to assign significantly different probabilities to *-nun*-marked responses. The result with Polyglot-Ko-12B, illustrated in Figure 5.4, is confirmed as statistically significant (boldfaced in the table): the model assigned a significantly higher probability to *-nun*-marked responses when exhaustivity needed to be conveyed.

Mixed-effects models revealed that no models associated exhaustivity with the *-lul* marker. Since the *-lul* marker functions as a grammatical case marker without considering any discourse context, the current result indicates that encoding dual meanings for a marker that primarily serves as a grammatical case marker must have been more challenging for LLMs. In Experiment 4 (c.f., Figure 5.3), Polyglot-Ko models and GPT-3 exhibited modulated surprisals for the *-lul* marker followed by different continuations. This result is unlikely to stem from associating the *-lul* marker with exhaustivity implicature, given that Polyglot-Ko models and GPT-3 did not associate the *-lul* marker as a cue for conveying exhaustivity in the current experiment. Rather, it is likely that the models' sensitivity towards verb continuations was heightened in Experiment 4 when the more canonical grammatical object marker (*-lul*) appeared.

Now we move on to report the results obtained from GPT-4 and human participants. We elicited

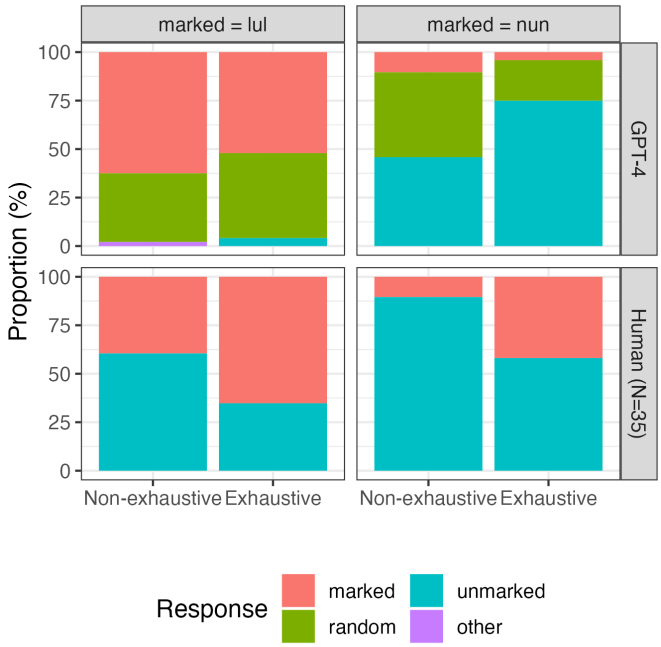


Figure 5.5: Proportions of responses elicited from GPT-4 and from 35 human participants.

responses from GPT-4 and human speakers using forced-choice tasks (c.f., Table 5.5). The results of the forced-choice responses are summarized on the right side of Figure 5.4. The left panels (marked = *-lul*) summarize choices when response options included a pair of sentences with *-lul*-marked and null-marked objects. The 'marked' proportion, colored in red, indicates the proportion of *-lul*-marked responses. The 'unmarked' proportion, colored in blue, indicates the proportion of null-marked responses. The right panels (marked = *-nun*) summarize choices when response options included a pair of sentences with *-nun*-marked and null-marked objects. On the right side of the figure, the 'marked' proportion, colored in red, indicates the proportion of *-nun*-marked responses, while the 'unmarked' proportion, in blue, indicates the proportion of null-marked responses.

GPT-4 frequently made 'random' choices, meaning that the model did not select the same option when the order of the two responses was flipped. Even when disregarding the random choices, the model exhibited a pattern contrary to that found in human responses. When exhaustivity needed to be expressed, GPT-4 more often selected null-marked responses, whereas humans were more likely to mark the object with either the *-lul* or *-nun* marker. Overall, GPT-4 appears incompetency at generating the *-lul* or *-nun* markers to evoke the implicature in our tasks. This result suggests that the decreased ratings observed

with canceled continuations in Experiment 4 (c.f., Figure 5.3) were unlikely to stem from the association of the markers with the exhaustivity implicature.

5.6 Discussion

Overall, we observed that some of the larger models produced results on the *-nun* marker that were coherent with human speakers' patterns. GPT-4—the largest model we assessed—reported lower appropriateness ratings when the exhaustivity implicature of the *-nun* marker was canceled (Experiment 4). GPT-3—the largest model among those we assessed with log probabilities—showed sensitivity to the non-cancelable implicature of the *-nun* marker (Experiment 4). Polyglot-Ko-12B—the second-largest model we assessed with log probabilities—demonstrated competence in utilizing the *-nun* marker to evoke the exhaustivity implicature (Experiment 5). This trend, where patterns closer to human behavior are observed as models scale up, is consistent with previous findings that increases in model size are correlated with performance improvements (Kaplan et al., 2020), as well as earlier evaluations of language models reviewed in section 5.2 (Schuster & Linzen, 2022; Hu et al., 2023).

In addition to model size, GPT-3 and GPT-4 stand out from the other models in that they are trained using human feedback through a technique known as Reinforcement Learning with Human Feedback (RLHF). In Experiment 4, both GPT-3 and GPT-4 exhibited patterns for the *-nun* marker that resembled humans' appropriateness ratings. This result aligns with findings from Hu et al. (2023), which showed that GPT-2 trained with human feedback displayed greater pragmatic competence. Thus, the results for the *-nun* marker suggest that scaling up the model and incorporating RLHF can enhance language models' ability to encode the discourse meanings associated with the *-nun* marker.

Unlike the *-nun* marker, no models exhibited human-like patterns for the polyfunctionality of *-lul* marker. GPT-4, which provided lower ratings when the *-nun* marker's exhaustivity was canceled, also provided lower ratings when the *-lul* marker's exhaustivity was canceled. Although GPT-3 showed sensitivity to the non-cancelable implicature of the *-nun* marker, it did not demonstrate an understanding of the cancelable exhaustivity implicature associated with the *-lul* marker (Experiments 4 and 5). It is important to note that the *-nun* marker primarily functions as a discourse marker and acquires an additional

grammatical function when it replaces the *-lul* marker in a grammatical object position. Conversely, the *-lul* marker denotes a grammatical object without any discourse context and gains an additional discourse function when interpreted within the discourse context. The differing results for the *-lul* and *-nun* markers suggest that encoding the discourse function of a polyfunctional marker that primarily serves as a grammatical marker (e.g., *-lul*) may be more challenging for LLMs compared to encoding the discourse function of a polyfunctional marker that primarily serves as a discourse marker (e.g., *-nun*).

This result is reminiscent of the limited competency that LMs have demonstrated with the interpretation of the discourse connective *and* (Pandia et al., 2021). Processing the temporal sequence implied by *and* (e.g., *I brushed my teeth and went to bed*) requires LMs to incorporate more contextualized interpretations of input compared to processing the meaning of addition commonly conveyed by *and* (e.g., *I baked cakes and cookies*). In Pandia et al. (2021), LMs were less competent at predicting the temporal sequence conveyed by *and*. Our experiments revealed a similar result: LMs were less competent at interpreting the discourse meaning of *-lul*, likely due to the marker's canonical grammatical function requiring more contextual interpretation. Overall, when a linguistic structure conveys multiple meanings, LMs appear to be more competent in processing the meaning that requires less contextual interpretation of the input.

Given these findings, there are some ways to improve the experimental setup employed in the current experiments. It has been pointed out that having a sufficient number of observations is crucial to obtain adequate statistical power in the analysis, not only in the context of psycholinguistic experiments but also in NLP studies (Card et al., 2020). Although our experiments used a number of evaluation items consistent with other psycholinguistic evaluations of LMs, expanding the evaluation dataset could increase the statistical power. Additionally, due to limited computational resources, models with parameters between 12B and 175B could not be tested. It is possible that models within this size range may exhibit different behaviors from the models tested in the current experiments.

What do the current results tell us about the general capabilities of LLMs in capturing discourse functions of morphology? The particular challenge in the current experiments was that models needed to capture the polyfunctionality of morphological markers—both grammatical object functions and exhaustive interpretations established in context. LLMs trained without human feedback do not seem to capture

how humans understand and utilize the polyfunctionality of these markers. However, providing human feedback and scaling up the embedding space resulted in patterns closer to those found with human speakers. Thus, encoding polyfunctionality across the domains of grammar and discourse in natural language does not appear to be an entirely impossible task for models to handle.

Besides the size of the model and the use of RLHF, there are many other factors to consider: the amount of training data, the model architecture, whether the model was trained on multiple languages, and more. A finer-grained analysis of the corpus data will allow us to assess the impact of the frequency of a given morphological marker on the models' ability to generate predictions sensitive to that marker's implicatures. In addition, fine-tuning a model with a dataset annotated for the discourse meanings of markers will allow us to observe whether pre-trained LLMs have the capacity to learn the discourse usage patterns associated with the markers.

5.7 Conclusion

In this chapter, we investigated whether LLMs can encode the meanings of morphological markers that encompass both grammatical and discourse meanings—in other words, the polyfunctionality of the two markers *-lul* and *-nun*. To account for object-marking patterns in Korean, LLMs must be able to capture multiple meanings of the markers: the discourse functions grounded in discourse context, as well as the grammatical functions that are orthogonal to discourse context. Despite LLMs' proven competency in handling general language generation tasks, most of the models we tested failed to exhibit a human-like ability to process and produce different object markings as cues for the discourse statuses of objects. In summary, our results highlight that LLMs without targeted training on discourse pragmatics have limitations in capturing the complex interplay between grammar and discourse in Korean morphology.

The material in Chapter 5 has been published in: Shin, H., & Trott, S. (2024). Do language models capture implied discourse meanings? An investigation with exhaustivity implicatures of Korean morphology. In *Proceedings of the Society for Computation in Linguistics 2024*, 150-161. It was adapted for this dissertation by adding background materials in section 5.2, providing more detailed descriptions of results, and using terminology consistent with that of Chapters 1-4.

6 Conclusion

6.1 Summary of findings

This dissertation aimed to investigate the relationship between grammar and discourse by examining alternating object markings in Korean. The alternating *-lul*, *-nun*, and null-markings discussed in this dissertation are summarized in Table 6.1. The investigation began with an exploration of the syntagmatics and paradigmatics of these three marking options (Chapter 1). Although *-lul*, *-nun*, and null-markings all follow a grammatical object syntagmatically, their paradigmatic contrasts reveal their different associations with discourse statuses of objects. This leads to the first finding on the relation between grammar and discourse: a single structure can serve functions in both domains of grammar and discourse, thus, its function may not be classified dichotomously as either a grammatical one or discourse-related one.

In Chapter 1, I highlighted that the information structure roles associated with each of the *-lul* and *-nun* markers could not fully account for the alternations between the two markers, as both markers' contrastive information structure roles—CF and CT—evoke an exhaustive listing interpretation. This limitation led to the three psycholinguistic studies (Chapters 2-4). The three psycholinguistic studies demon-

Table 6.1: Summary of paradigmatic contrasts *-lul-*, *-nun-*, and null-markings of objects appearing in SOV word order.

	Grammatical function	Discourse function	Without contextualized alternatives	With contextualized alternatives
-lul	Direct object	Focus, CF	Exhaustive interpretation	Cancelable exhaustivity
-nun	Direct object	Topic, CT	No exhaustive interpretation	Non-cancelable exhaustivity
null	Direct object	Topic	No exhaustive interpretation	No exhaustivity

strated that while both markers convey exhaustive interpretations, they are associated with contrasting discourse behavioral patterns.

In Chapter 2, I found that the *-lul* marker is more closely linked to exhaustive interpretation, especially in contexts where alternatives are not explicitly established or shared among interlocutors. The *-nun* marker, however, tends to signal exhaustivity more effectively when the set of alternatives is part of the shared knowledge between the interlocutors. This highlights a difference in how these markers interact with shared knowledge statuses of interlocutors. The *-lul* marker, which can be used as an accusative case marker without any discourse functions, might be expected to have a less prominent discourse function compared to the *-nun* marker, which replaces grammatical case markers to indicate discourse functions. This may lead to the presumption that the *-lul* marker's discourse function is less predominant function of the marker. In addition, comparing the *-lul* and *-nun* markers, one may be inclined to presume that the *-lul* marker's association with discourse functions is weaker than that of the *-nun* marker. However, the results from the first experimental study challenged these assumptions, demonstrating that the *-lul* marker can indeed exhibit significant discourse functions.

This leads to the next finding regarding the relationship between grammar and discourse: one cannot readily assume the predominance of grammatical or discourse functions of a polyfunctional linguistic structure. The polyfunctionality of a linguistic structure across grammar and discourse domains does not imply that it is predominantly associated with one domain. Structures typically classified with grammatical functions should not be assumed to exhibit a weaker discourse function, nor should structures primarily associated with discourse functions be presumed to have a stronger discourse function compared to grammatical structures. Thus, the primary function of a linguistic structure cannot be determined solely based on its more consistent role in one domain (e.g., the grammatical object function of *-lul* which can occur without discourse functions). Even a structure regarded as having a grammatical function as its primary role can display a discourse function that is more significant in certain contexts.

In addition to how each marker interacts with shared knowledge to convey exhaustive interpretation, Chapter 3 revealed that the *-lul* and *-nun* markers show distinct patterns of implicature cancelability. Specifically, the *-lul* marker's exhaustivity implicature is cancelable, while the *-nun* marker's exhaustivity implicature is non-cancelable. This finding challenges the previously discussed distinction between con-

versational and conventional implicatures. The *-lul* marker, which establishes exhaustive interpretation more independently from shared knowledge, exhibits cancelable exhaustivity, indicating a weaker structural association with the implicature. In contrast, the *-nun* marker, which conveys exhaustivity based on shared knowledge, demonstrates non-cancelable exhaustivity, suggesting a stronger structural encoding of the implicature. Thus, different discourse behavioral aspects highlight varying degrees of association between a structure and its discourse interpretation.

Overall, the results demonstrate that the object marking system in Korean is organized systematically, with structures exhibiting the same syntagmatic pattern but showing fine-grained paradigmatic contrasts in discourse usage patterns. The observed contrasts between the *-lul* and *-nun* markers reveal distinct discourse strategies that speakers can use. The *-lul* and *-nun* markers require different levels of shared knowledge to convey exhaustivity, and each marker imposes distinct discourse constraints on the preceding discourse. These differences enable speakers to express (non-)exhaustivity effectively across various discourse contexts. In essence, speakers can choose a marker that best aligns with their intention to communicate exhaustivity given the discourse context. Additionally, speakers may use different markers, either consciously or subconsciously, to convey exhaustivity depending on the shared knowledge context and to impose different cancelability constraints on subsequent discourse.

Psycholinguistic assessments of LLMs showed that these models, which are trained to predict the next tokens based on the distributions from large language inputs, provided only partial evidence that they can capture exhaustivity as a discourse function of markers (Chapter 5). This result aligns with the previous reports that LLMs fall short in capturing discourse meanings, even those models succeed in capturing grammatical patterns of language (section 5.2; Bender & Koller, 2020). These two kinds of linguistic competence are systematically distinguished as formal linguistic competence and functional linguistic competence (Mahowald et al., 2024). Formal linguistic competence encompasses knowledge of the rules and statistical regularities of language, including skills in phonology, morphology, lexical semantics, and syntax. Functional linguistic competence, on the other hand, involves the ability to use language in real-world contexts, relying on world knowledge, discourse context, and social reasoning. The ability to process and produce language based on communicative intentions is part of functional linguistic competence, an area where LLMs have not yet demonstrated much success.

Processing and producing morphological markings in Korean, which may be easily mistaken for tasks related to formal linguistic competence, are demonstrated as tasks related to both formal and functional linguistic competence in this dissertation. The *-lul* and *-nun* markers ultimately serve as discourse strategies that speakers use to communicate intended (non-)exhaustivity given the considerations of shared knowledge and cancelability constraints. Most pre-trained LLMs are trained to complete language modeling tasks based on the forms of language provided as input, without considering the communicative intentions associated with those forms. Consequently, LLMs are not trained to act as communicative agents that make associations between linguistic forms and intended communicative acts (see Andreas, 2022). The LLM outputs reported in Chapter 5 reveals the limitation of LLMs' functional linguistic competence, particularly concerning the shared knowledge, cancelability, and intended exhaustive interpretations. This result suggests that systems that lack targeted training in discourse pragmatics struggle to capture the discourse usage patterns of alternative markings in Korean.

6.2 Future directions

Although the hierarchy between grammatical and discourse functions of polyfunctional markers is not supported in this dissertation, it was not directly investigated whether one function is actually more predominant than the others. This question could be explored in at least three directions. First, studies on online processing patterns could help determine whether functions from multiple domains are assessed simultaneously upon encountering a polyfunctional structure, or whether each function is processed sequentially. Similar to online experiments investigating whether the semantic or pragmatic meaning of some (i.e., 'some and possibly all' versus 'some but not all') is assessed first (Bott & Noveck, 2004; Breheny et al., 2006), one could devise an experiment to investigate the processing of grammatical and pragmatic interpretations of a sentence containing a polyfunctional marker. Understanding the processing order of multiple functions across grammar and discourse domains would provide evidence on whether one function is treated as default, which could suggest whether a hierarchical relationship exists between grammatical and discourse functions in the processing of polyfunctional structures.

In addition to further illuminating the relationship between grammatical and discourse functions,

investigating the hierarchy between these functions through online experiments can yield other valuable contributions. In evaluating LLMs' competency with polyfunctional structures, I used discourse behavioral measurements obtained from human speakers as standards (Chapter 5). However, given that human online processing patterns, such as reading times, have shown close associations with LLMs' surprisal (E. G. Wilcox, Vani, & Levy, 2021; E. G. Wilcox et al., 2020), acquiring online processing patterns for polyfunctional structures could enable more detailed comparisons between human and LLM language processing patterns.

The second direction involves investigating the learning process of polyfunctionality. Native speakers are exposed to structures and their functions within discourse contexts. If a structure is polyfunctional, is there a function that learners acquire first? Do learners acquire one function more easily than the other? An experiment could be designed using a miniature language learning paradigm reviewed in Chapter 4 (e.g., Fedzechkina et al., 2012; K. Smith & Culbertson, 2020). This miniature language could include polyfunctional markers with both grammatical and discourse functions for learners. If learners acquire the grammatical function of the marker first and then move on to associate the marker with discourse functions, it could suggest that the grammatical function of a structure is prioritized in the learning process. Conversely, if grammatical and discourse functions are acquired at similar speeds and levels of mastery, the results can suggest that these domains of function are not distinguished during the learning process.

In Chapter 4, I proposed that the polyfunctionality of the *-lul* marker contributes to language efficiency by avoiding the need for a more complex morphological inventory. This was suggested as a way in which morphology remains a learnable system. Further investigation into the learning process of polyfunctional structures could shed light on how they contribute to a more learnable morphological paradigm. If human learners do not experience significant difficulty in acquiring polyfunctionality compared to learning multiple structures with a single function, it would provide stronger evidence that polyfunctionality indeed supports a more learnable morphological system.

Given the findings from such learning experiments, LLMs' competency with polyfunctional structures can be re-evaluated. If humans acquire a grammatical function first and LLMs demonstrate competency solely with the grammatical function of a marker, it would suggest that the frequency distribution

of language input and the architecture of current LLMs can train those at least for the function that humans learn first. However, if evidence shows that humans learn both grammatical and discourse functions simultaneously, yet LLMs exhibit competency only with the grammatical function, it would indicate that the training process for LLMs is not analogous to human language learning processes.

Lastly, while the current dissertation focused on the synchronic patterns of alternating markers in Korean, future research on diachronic patterns could further illuminate the relationship between grammar and discourse. For instance, if evidence shows that the *-lul* marker was initially used solely for its grammatical function and later acquired a discourse function, this would suggest a link between its grammatical and discourse uses. Conversely, if no evidence indicates which function emerged first, it would support the argument that a structure acquires its functions through evolving usage patterns in discourse.

Various factors have been discussed as sources motivating a linguistic structure to be used as a conventionalized discourse strategy. Cognitive linguists emphasize the role of human cognition in shaping grammar and language, arguing that language reflects how people perceive the world and make pragmatic inferences from linguistic forms (e.g., Lakoff, 1987; Sweetser, 1990; Talmy, 1985). This view is emphasized in the principle of iconicity discussed in Chapter 4. For example, the argument that simpler expressions convey stereotypical interpretations (Levinson, 2000) is based on the iconic relationship between the simplicity of an expression and its discourse and pragmatic meanings.

Sociocultural norms also influence how a linguistic structure evolves into a discourse strategy (e.g., Traugott, 2003). Associating a particular usage pattern with higher social prestige can motivate that pattern to acquire a conventionalized discourse meaning (Labov, 1986, 2001). In Korean, informal and colloquial modes of conversation are sociocultural factors that could increase the overt use of grammatical case markers (Chapter 1 and 3; Ko, 2001).

Discourse linguists emphasize the importance of the usage pattern of a structure within a discourse context (Givón, 2018; Du Bois, 2003). Acknowledging that human cognition, sociocultural norms, and our ability to make pragmatic inferences are all plausible sources that can conventionalize a structure as a discourse strategy, Ariel (2008) aligns with the discourse linguists' perspective. It is suggested that these factors influence the conventionalization of a structure by making its discourse usage pattern (i.e., how the structure is used in a discourse context) more salient. In other words, it is proposed that discourse

usage patterns serve as gatekeepers that regulate other influencing factors and facilitate the development of a structure into a conventionalized discourse strategy.

Although this dissertation did not explicitly weigh the different factors that could motivate the conventionalization of a discourse strategy, the studies of LLMs presented in Chapter 5 provide some indicative results. Conventionalized functions of a linguistic structure can be understood as functions that are more strongly encoded by the structure and are conveyed in instances where the structure is used. The fact that language models, trained without knowledge of communicative intents, failed to utilize polyfunctional markers effectively as discourse strategies suggests that the discourse usage patterns involving communicative intents must not be overlooked in investigations of conventionalized functions.

Future research could explore the significance of discourse usage patterns in the conventionalization of discourse strategies by training language models with different types of data. For instance, one could train a set of models with language data annotated with communicative intentions and another set without such annotations. If the models trained with the annotated data outperform those trained with the non-annotated data, it would suggest that usage patterns grounded in discourse context and communicative intents are crucial factors in how a linguistic structure conventionally encodes a discourse strategy.

6.3 Concluding remarks

In conclusion, this dissertation demonstrated that grammar and discourse do not constitute a dichotomous division but rather a continuous cline. A linguistic structure can serve functions in multiple domains of language. The seemingly complex organization of Korean morphology revealed polyfunctionality that is intricately tied to fine-grained regularities in discourse behavior and cognitive statuses of interlocutors. This polyfunctionality cannot be fully captured by large language models, systems without a targeted training on discourse pragmatics. These results suggest that polyfunctionality should be viewed as an amalgamation of various factors involved in human communication, including cognitive statuses of speakers and communicative intents. To provide a comprehensive understanding of a linguistic structure and its functions along the cline of grammar and discourse, it is crucial to consider not only how natural language is used in real-world contexts but also how conversational agents utilize linguistic structures to

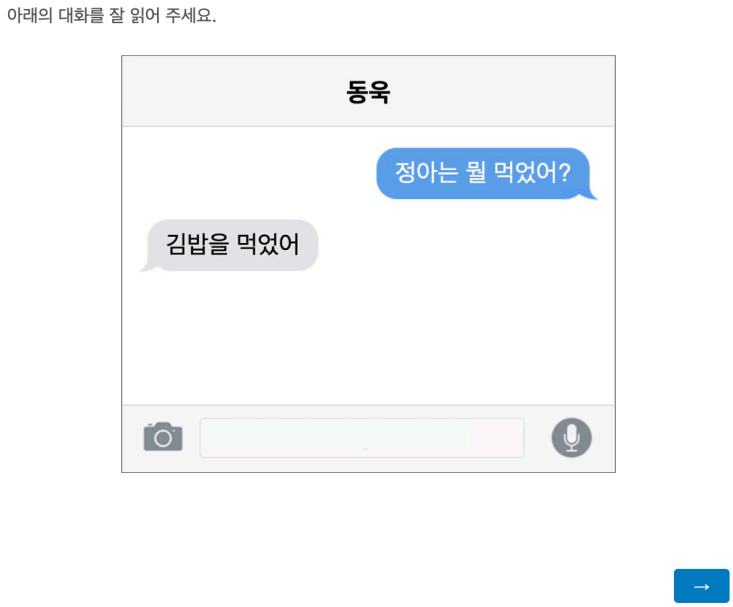
convey their intended message in a discourse.

A Experiment materials

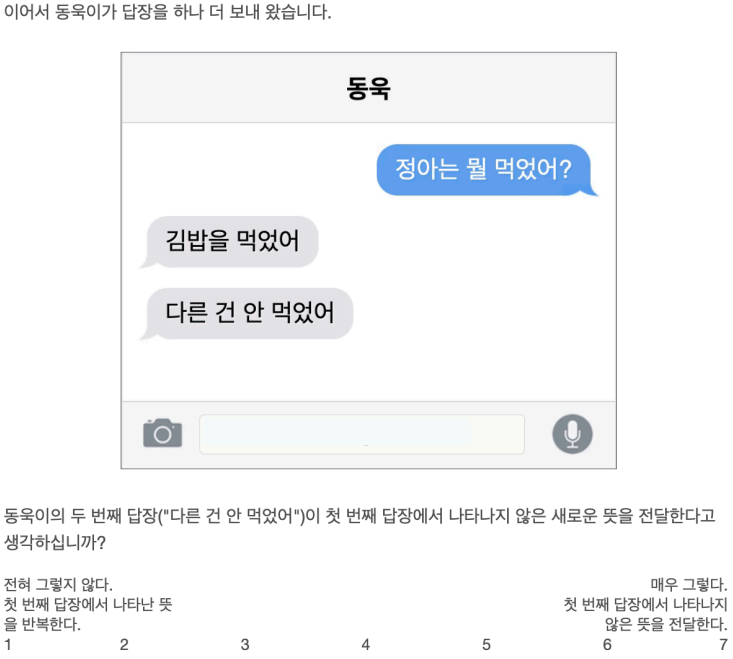
A.1 Experiment 1

Table A.1: List of words used to construct stimuli in Experiment 1.

	Romanized Korean		Translated in English	
	Alternatives	Verb	Alternatives	Verb
1	kimpap, lamyen	mekta	kimbap, ramen	eat
2	cenkong swuep, silsup swuep	tutta	major course, lab course	take
3	antha, homlen	chita	hit, home run	hit
4	sinpal, kapang	sata	shoes, bag	buy
5	lollekhosuthe, pemphekha	thata	roller coaster, bumper cars	ride
6	inthepyu, calyo, selmwun, calyo	mouta	interview material, survey material	collect
7	thakca, sopha	olmkita	table, couch	move
8	yeysan kwanli, hongpo,	mathta	budget management, advertisement	take
9	panci, phalcci	phalta	ring, bracelet	sell
10	khalkwukswu, swuceypi	sikhita	kalguksu, sujebi	order
11	talliki, mellittwiki	ikita	running, long jump	win
12	maykcwu, socwu	masita	beer, soju	drink
13	wuntongcang, kangtang	ccikta	field, auditorium	photograph
14	meytal, sangkum	ttata	medal, prize money	win
15	siksapi, hwaltongpi	nayta	food bill, activity bill	pay
16	sayntuwichi, khwukhi	mantulta	sandwich, cookie	make
17	kkoch, yongton	patta	flower, money	receive
18	selkeci, chengso	topta	dish, cleaning	help
19	theylleypicen, senphwungki	naynohta	TV, fan	put out
20	ungwenpok, ungwen, tokwu	chayngkita	cheer uniform, cheer equipment	pack
21	yeksa mwuncey, kyengcey mwuncey	macta	history question, economy question	correct
22	sangphwumkwen, halinkwen	ppopta	gift card, discount	pick
23	silyek kemsa, hyelayk kemsa	machita	eye exam, blood test	finish
24	tanse, cungke	chacta	evidence, proof	find

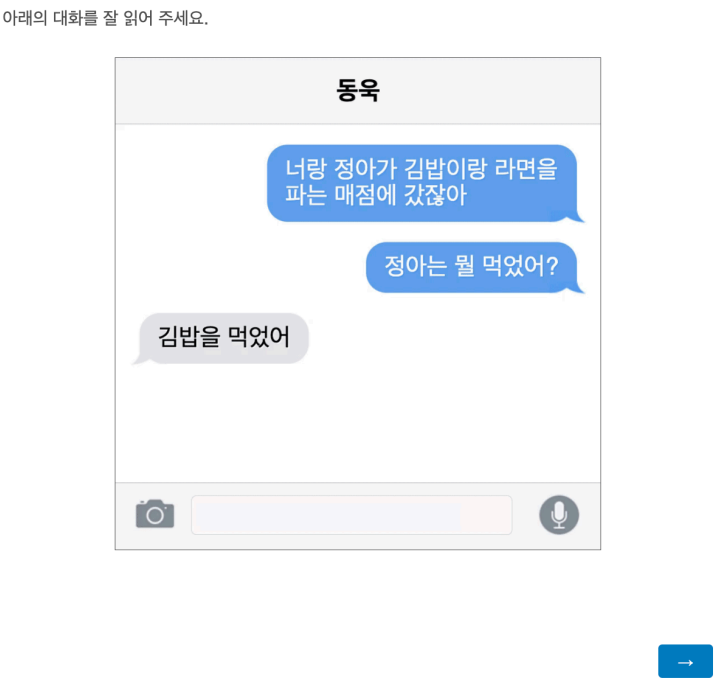


(a) The first page of the item.

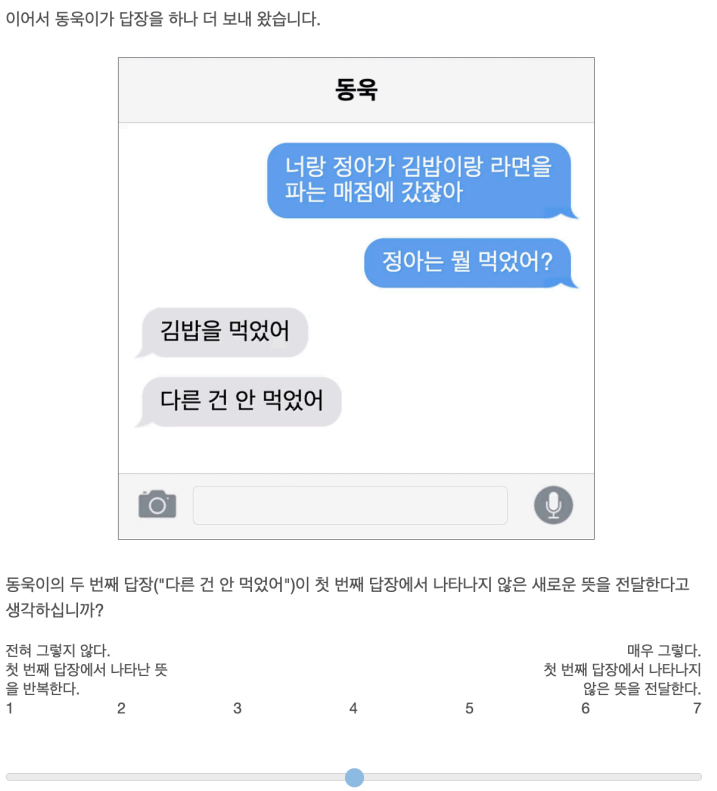


(b) The second page of the item.

Figure A.1: A sample of experiment pages displaying the LUL-UNIDENTIFIED item as presented in Experiment 1.

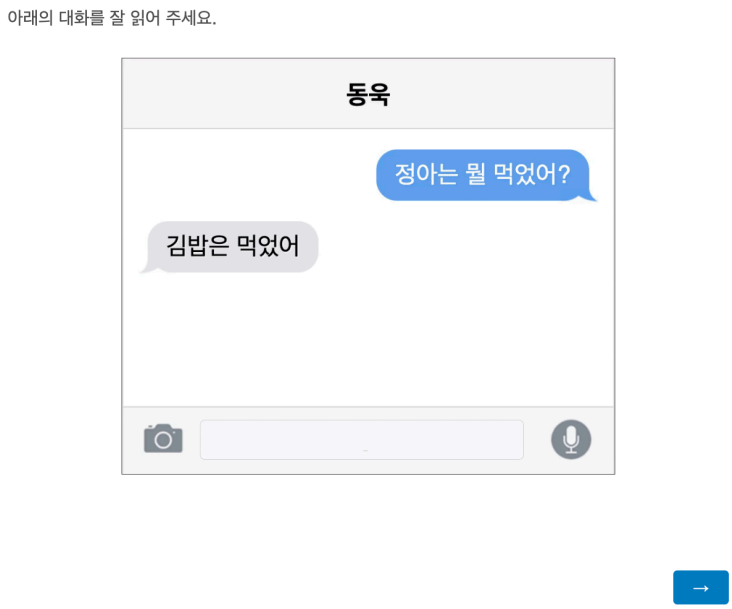


(a) The first page of the item.

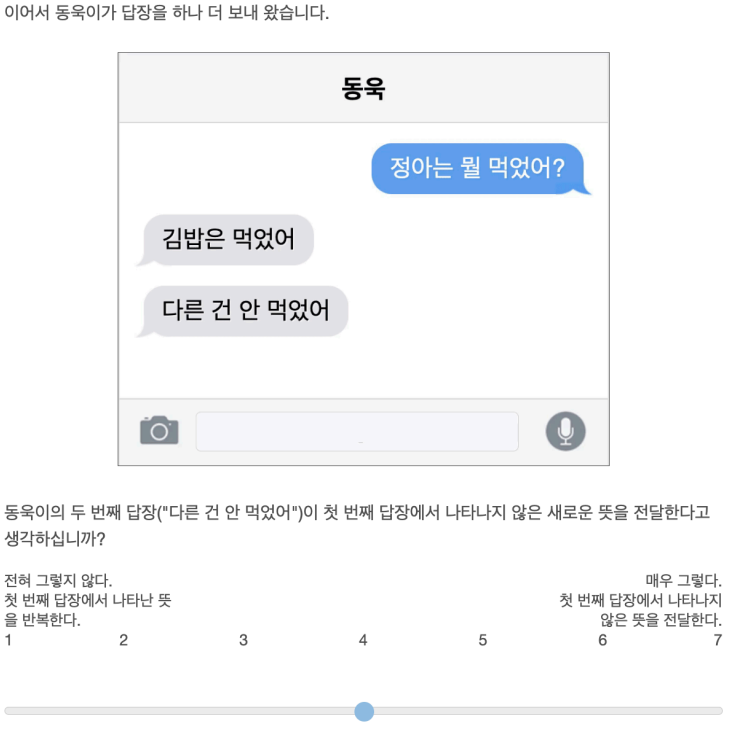


(b) The second page of the item.

Figure A.2: A sample of experiment pages displaying the LUL-IDENTIFIED item as presented in Experiment 1.

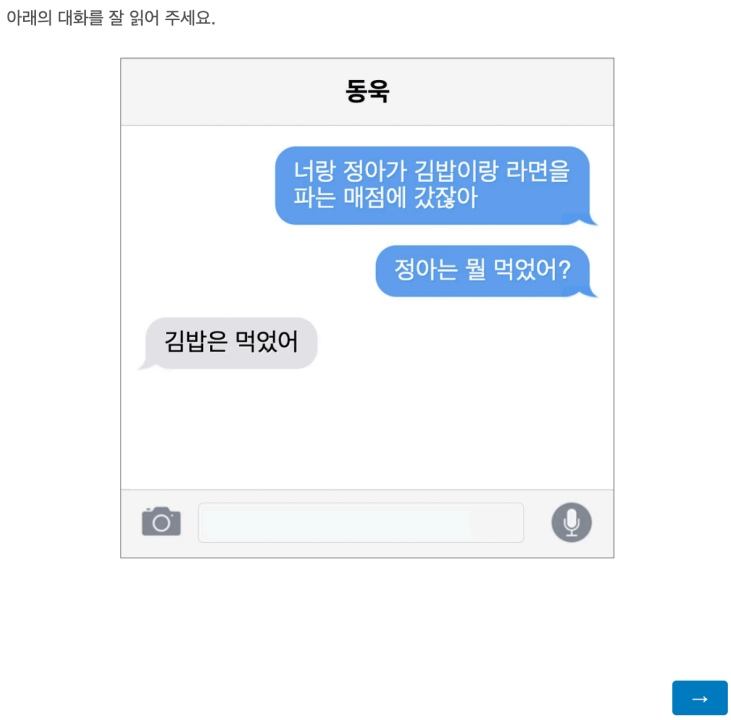


(a) The first page of the item.

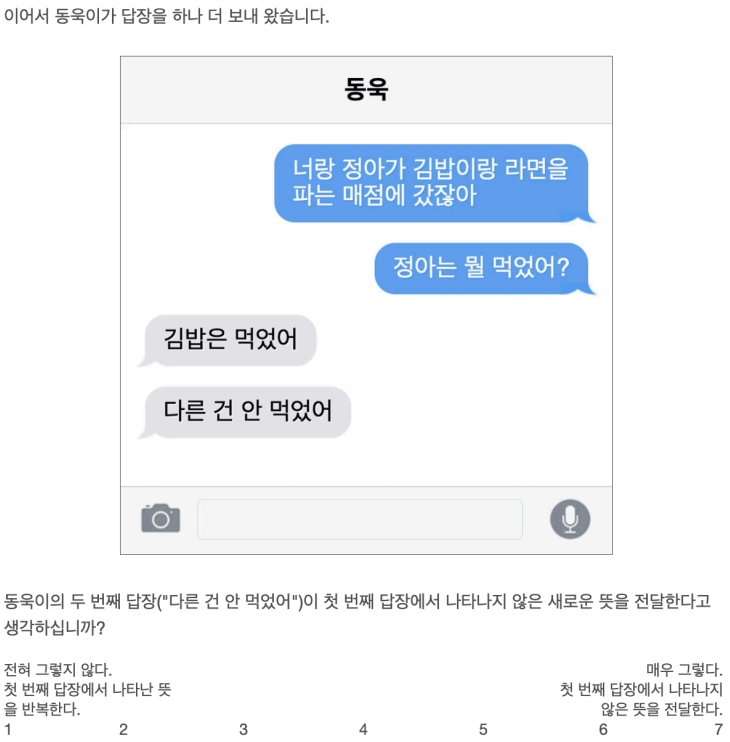


(b) The second page of the item.

Figure A.3: A sample of experiment pages displaying the NUN-UNIDENTIFIED item as presented in Experiment 1.



(a) The first page of the item.



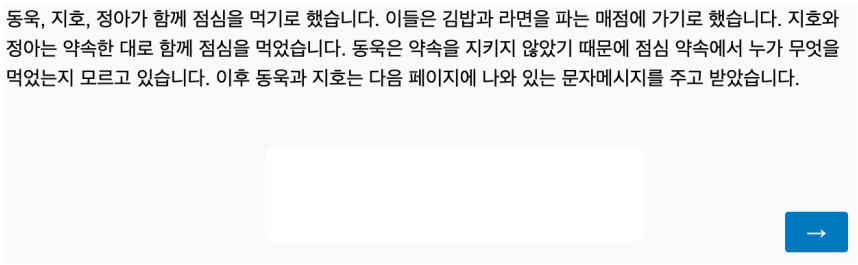
(b) The second page of the item.

Figure A.4: A sample of experiment pages displaying the NUN-IDENTIFIED item as presented in Experiment 1.

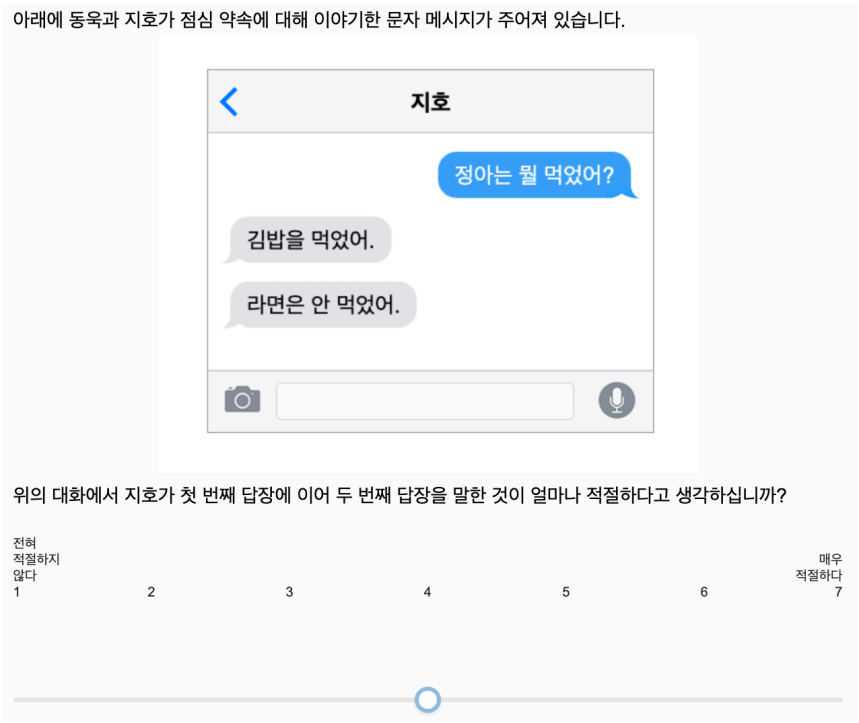
A.2 Experiment 2

Table A.2: List of words used to construct stimuli in Experiment 2.

	Romanized Korean		Translated in English	
	Alternatives	Verb	Alternatives	Verb
1	kimpap, lamyen	mekta	kimbap, ramen	eat
2	cenkong swuep, silsup swuep	tutta	major course, lab course	take
3	inge, meyki	capta	carp, catfish	catch
4	sinpal, kapang	sata	shoes, bag	buy
5	lollekhosuthe, pemphekha	thata	roller coaster, bumper cars	ride
6	inthepyu, calyo, selmwun, calyo	mouta	interview material, survey material	collect
7	thakca, sopha	olmkita	table, couch	move
8	yeysan kwanli, hongpo,	mathta	budget management, advertisement	take
9	panci, phalcci	phalta	ring, bracelet	sell
10	khalkwukswu, swuceypi	sikhita	kalguksu, sujebi	order
11	talliki, mellittwiki	ikita	running, long jump	win
12	maykcwu, socwu	masita	beer, soju	drink
13	wuntongcang, kangtang	ccikta	field, auditorium	photograph
14	meytal, sangkum	ttata	medal, prize money	win
15	siksapi, hwaltongpi	nayta	food bill, activity bill	pay
16	sayntuwichi, khwukhi	mantulta	sandwich, cookie	make
17	kkoch, yongton	patta	flower, money	receive
18	selkeci, chengso	topta	dish, cleaning	help
19	theylleypicen, senphwungki	naynohta	TV, fan	put out
20	ungwenpok, ungwen, tokwu	chayngkita	cheer uniform, cheer equipment	pack
21	theynthu, chimnang	icta	tent, sleeping bag	forget
22	kominhyeng, thokkiinhyeng	ppopta	teddy bear, stuffed bunny	pick
23	leningsyus, cemphusyus	tencita	running shot, dunk	throw
24	tanse, cungke	chacta	evidence, proof	find

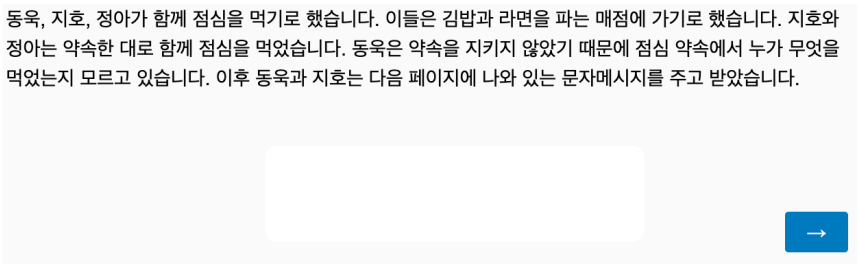


(a) The first page of the item displaying the context preamble.

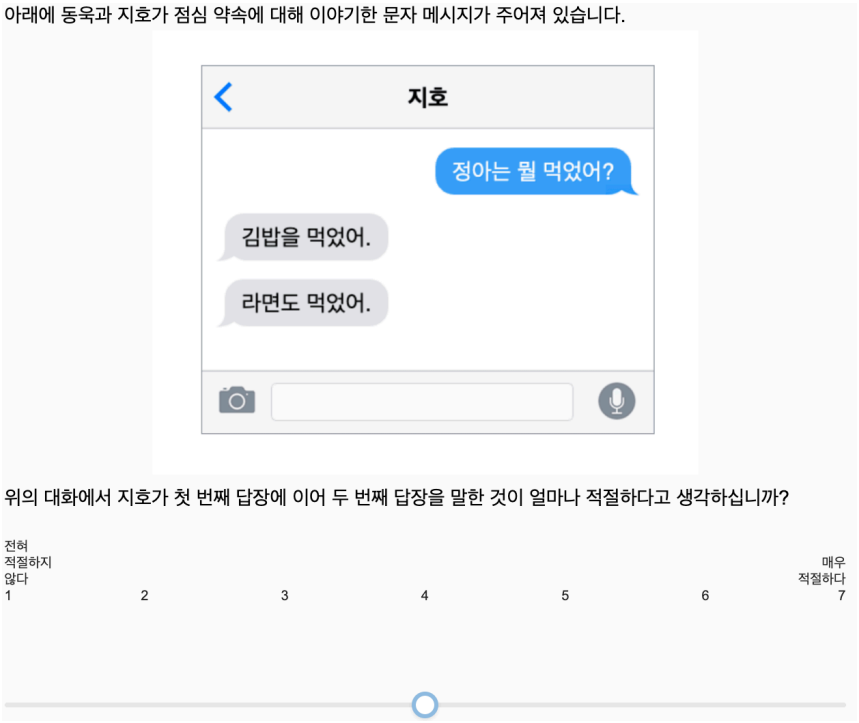


(b) The second page of the item displaying speakers' utterances in text message bubbles, followed by the experiment question and a 7-point scale.

Figure A.5: A sample of experiment pages displaying the LUL-MAINTAINED item as presented in Experiment 2.

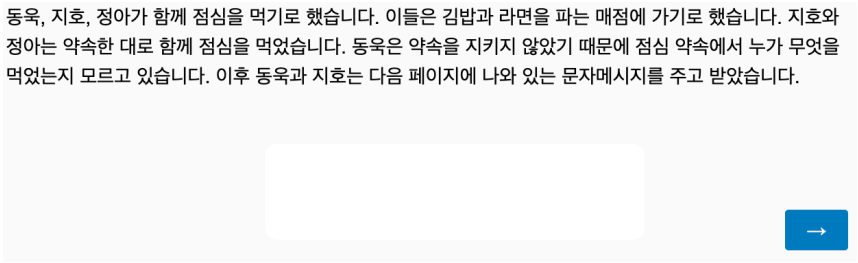


(a) The first page of the item displaying the context preamble.

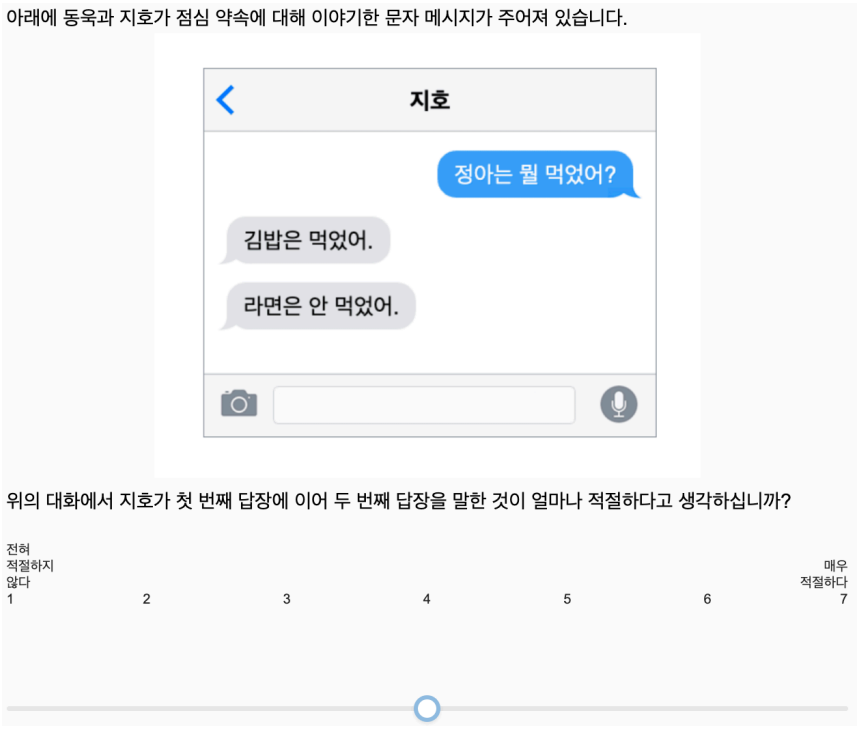


(b) The second page of the item displaying speakers' utterances in text message bubbles, followed by the experiment question and a 7-point scale.

Figure A.6: A sample of experiment pages displaying the LUL-CANCELED item as presented in Experiment 2.

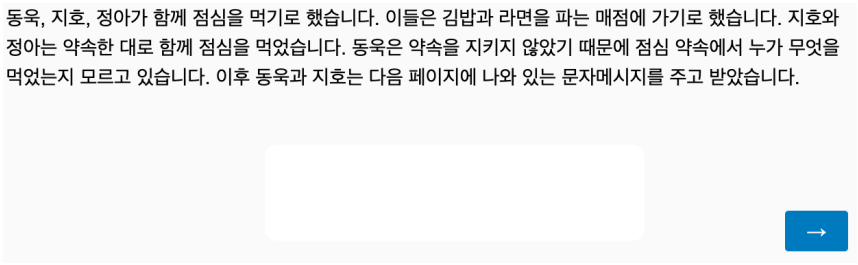


(a) The first page of the item displaying the context preamble.

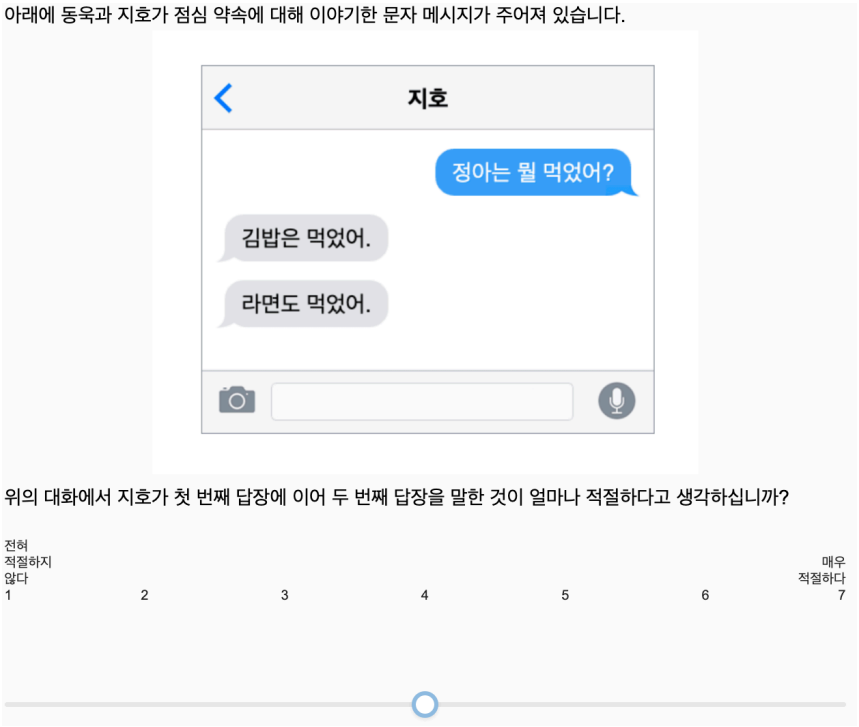


(b) The second page of the item displaying speakers' utterances in text message bubbles, followed by the experiment question and a 7-point scale.

Figure A.7: A sample of experiment pages displaying the NUN-MAINTAINED item as presented in Experiment 2.



(a) The first page of the item displaying the context preamble.



(b) The second page of the item displaying speakers' utterances in text message bubbles, followed by the experiment question and a 7-point scale.

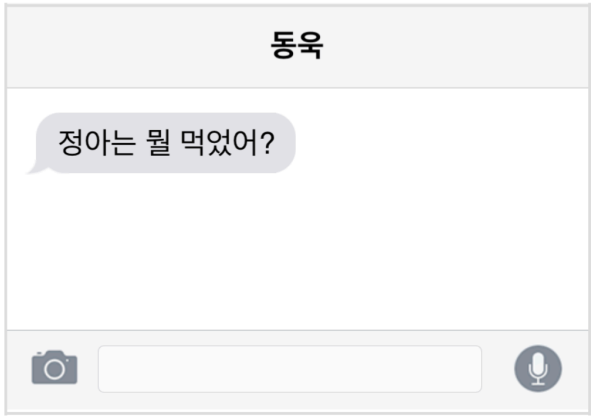
Figure A.8: A sample of experiment pages displaying the NUN-CANCELED item as presented in Experiment 2.

A.3 Experiment 3

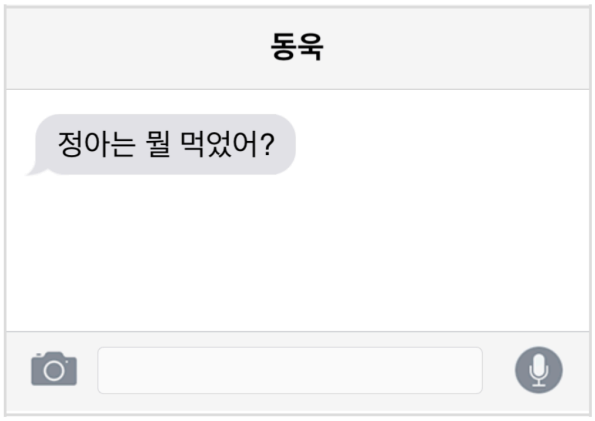
Table A.3: List of words used to construct stimuli in Experiment 3.

	Romanized Korean		Translated in English	
	Alternatives	Verb	Alternatives	Verb
1	kimpap, lamyen	mekta	kimbap, ramen	eat
2	cenkong swuep, silsup swuep	tutta	major course, lab course	take
3	inge, meyki	capta	carp, catfish	catch
4	sinpal, kapang	sata	shoes, bag	buy
5	lollekhosuthe, cailotulop	thata	roller coaster, drop tower	ride
6	inthepyu, calyo, selmwun, calyo	mouta	interview material, survey material	collect
7	thakca, sopha	olmkita	table, couch	move
8	yeysan kwanli, hongpo,	mathta	budget management, advertisement	take
9	panci, phalcci	phalta	ring, bracelet	sell
10	khalkwukswu, swuceypi	sikhita	kalguksu, sujebi	order
11	talliki, mellittwiki	ikita	running, long jump	win
12	maykcwu, socwu	masita	beer, soju	drink
13	wuntongcang, kangtang	ccikta	field, auditorium	photograph
14	meytal, sangkum	ttata	medal, prize money	win
15	siksapi, hwaltongpi	nayta	food bill, activity bill	pay
16	sayntuwichi, khwukhi	mantulta	sandwich, cookie	make
17	kkoch, senmwul	patta	flower, gift	receive
18	selkeci, chengso	topta	dish, cleaning	help
19	theylleypicen, senphwungki	naynohta	TV, fan	put out
20	ungwenpok, ungwen, tokwu	chayngkita	cheer uniform, cheer equipment	pack
21	theynthu, chimnang	icta	tent, sleeping bag	forget
22	kominhyeng, thokkiinhyeng	ppopta	teddy bear, stuffed bunny	pick
23	leningsyus, cemphusyus	tencita	running shot, dunk	throw
24	tanse, cungke	chacta	evidence, proof	find

동욱이, 지호, 정아가 함께 점심을 먹기로 했습니다. 이들은 김밥과 라면을 파는 매점에 가기로 했습니다. 지호와 정아는 약속한 대로 함께 점심을 먹었습니다. 지호는 정아가 점심 약속에서 김밥과 라면을 둘 다 먹었다는 것을 알고 있습니다. 동욱이는 약속을 지키지 않았기 때문에 정아가 점심 약속에서 무엇을 먹었는지 모르고 있습니다. 이후 동욱이는 지호에게 아래에 나와 있는 문자메시지를 보내왔습니다.



(a) The first page of the item displaying the context preamble and the wh-question.



지호는 정아가 김밥만 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 가장 적절한 것은 무엇입니까? 아래에 주어진 문장들은 지호가 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 지호가 보내기에 최선이라고 생각하는 답장을 선택해 주세요.

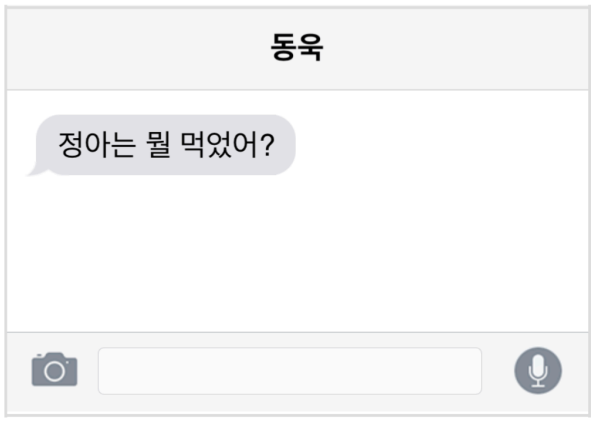
○ 김밥을 먹었어

○ 김밥 먹었어

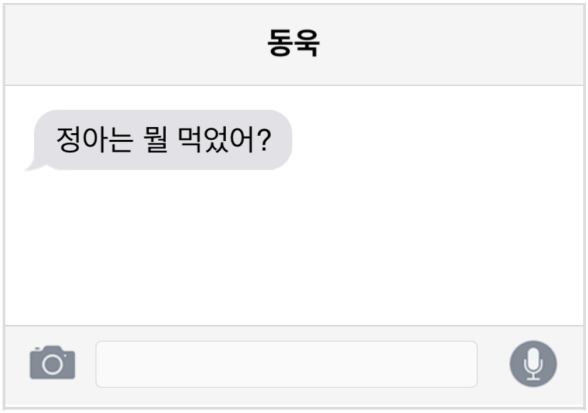
(b) The second page of the item displaying the intended message and response options.

Figure A.9: A sample of experiment pages displaying the WH-Q & EXHAUSTIVE item as presented in Experiment 3.

동욱이, 지호, 정아가 함께 점심을 먹기로 했습니다. 이들은 김밥과 라면을 파는 매점에 가기로 했습니다. 지호와 정아는 약속한 대로 함께 점심을 먹었습니다. 지호는 정아가 점심 약속에서 김밥과 라면을 둘 다 먹었다는 것을 알고 있습니다. 동욱이는 약속을 지키지 않았기 때문에 정아가 점심 약속에서 무엇을 먹었는지 모르고 있습니다. 이후 동욱이는 지호에게 아래에 나와 있는 문자메시지를 보내왔습니다.



(a) The first page of the item displaying the context preamble and the wh-question.



지호는 정아가 김밥과 라면을 둘 다 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 가장 적절한 것은 무엇입니까? 아래에 주어진 문장들은 지호가 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 지호가 보내기에 최선이라고 생각하는 답장을 선택해 주세요.

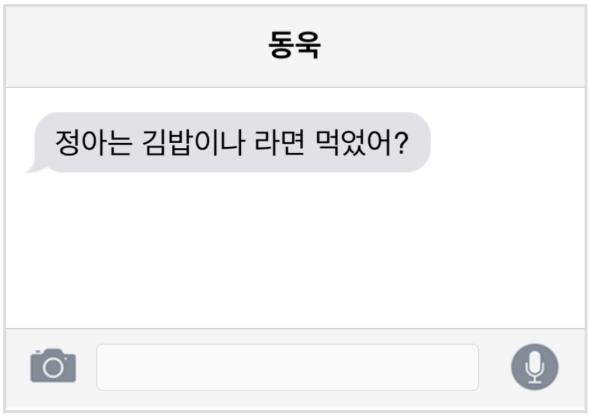
○ 김밥을 먹었어

○ 김밥 먹었어

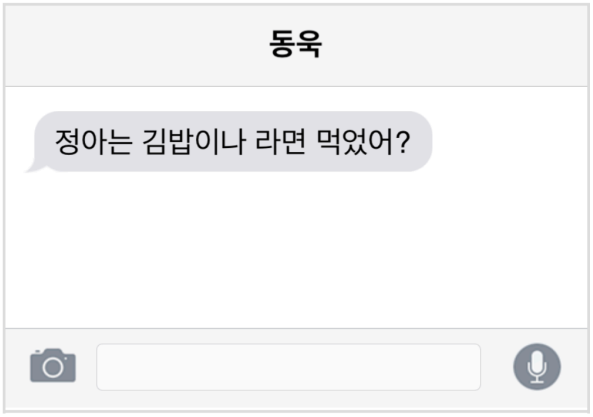
(b) The second page of the item displaying the intended message and response options.

Figure A.10: A sample of experiment pages displaying the WH-Q & NON-EXHAUSTIVE item as presented in Experiment 3.

동욱이, 지호, 정아가 함께 점심을 먹기로 했습니다. 이들은 김밥과 라면을 파는 매점에 가기로 했습니다. 지호와 정아는 약속한 대로 함께 점심을 먹었습니다. 지호는 정아가 점심 약속에서 김밥을 먹었고 라면을 먹지 않았다는 것을 알고 있습니다. 동욱이는 약속을 지키지 않았기 때문에 정아가 점심 약속에서 무엇을 먹었는지 모르고 있습니다. 이후 동욱이는 지호에게 아래에 나와 있는 문자메시지를 보내왔습니다.



(a) The first page of the item displaying the context preamble and the disjunctive question.



지호는 정아가 김밥만 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 가장 적절한 것은 무엇입니까? 아래에 주어진 문장들은 지호가 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 지호가 보내기에 최선이라고 생각하는 답장을 선택해 주세요.

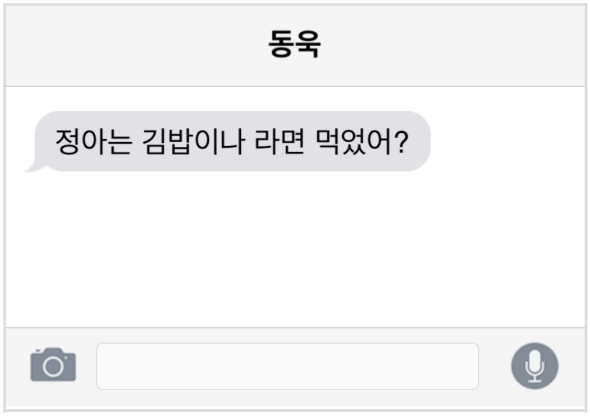
○ 김밥을 먹었어

○ 김밥 먹었어

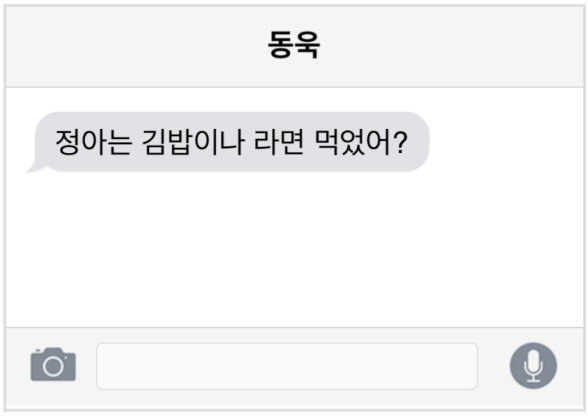
(b) The second page of the item displaying the intended message and response options.

Figure A.11: A sample of experiment pages displaying the DQ & EXHAUSTIVE item as presented in Experiment 3.

동욱이, 지호, 정아가 함께 점심을 먹기로 했습니다. 이들은 김밥과 라면을 파는 매점에 가기로 했습니다. 지호와 정아는 약속한 대로 함께 점심을 먹었습니다. 지호는 정아가 점심 약속에서 김밥을 먹었고 라면을 먹지 않았다는 것을 알고 있습니다. 동욱이는 약속을 지키지 않았기 때문에 정아가 점심 약속에서 무엇을 먹었는지 모르고 있습니다. 이후 동욱이는 지호에게 아래에 나와 있는 문자메시지를 보내왔습니다.



(a) The first page of the item displaying the context preamble and the disjunctive question.



지호는 정아가 김밥과 라면을 둘 다 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 가장 적절한 것은 무엇입니까? 아래에 주어진 문장들은 지호가 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 지호가 보내기에 최선이라고 생각하는 답장을 선택해 주세요.

○ 김밥을 먹었어

○ 김밥 먹었어

(b) The second page of the item displaying the intended message and response options.

Figure A.12: A sample of experiment pages displaying the DQ & NON-EXHAUSTIVE item as presented in Experiment 3.

A.4 Experiment 4

Same list of words used in Experiment 2 (c.f., Table A.2) were used to construct items in Experiment 4.

A.4.1 Materials used to evaluate Ko-GPT, Polyglot-Ko, and GPT-3 models

An example of evaluation items in the LUL-MAINTAINED condition:

> 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
> 동욱: 정아는 뭘 먹었어?
> 지호: 김밥을 먹었어. 라면은 안 먹었어.

An example of evaluation items in the LUL-CANCELED condition:

> 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
> 동욱: 정아는 뭘 먹었어?
> 지호: 김밥을 먹었어. 라면도 먹었어.

An example of evaluation items in the LUL-CONTRADICTED condition:

> 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
> 동욱: 정아는 뭘 먹었어?
> 지호: 김밥을 먹었어. 라면만 먹었어.

An example of evaluation items in the NUN-MAINTAINED condition:

> 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
> 동욱: 정아는 뭘 먹었어?
> 지호: 김밥은 먹었어. 라면은 안 먹었어.

An example of evaluation items in the NUN-CANCELED condition:

> 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
> 동욱: 정아는 뭘 먹었어?
> 지호: 김밥은 먹었어. 라면도 먹었어.

An example of evaluation items in the NUN-CONTRADICTED condition:

> 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
> 동욱: 정아는 뭘 먹었어?
> 지호: 김밥은 먹었어. 라면만 먹었어.

A.4.2 Materials used to evaluate GPT-4

An example of evaluation items in the NUN-CONTRADICTED condition:

System message: 반드시 1과 7사이의 숫자 중 하나로만 답해 주세요.
Input: 동욱이는 정아와 지호가 김밥과 라면을 먹으러 갔다는 것을 알고 있습니다.
동욱: 정아는 뭘 먹었어?
지호: 김밥은 먹었어. 라면만 먹었어.
위의 대화에서 미나가 첫 번째 답장에 이어 두 번째 답장을 말한 것이 얼마나 적절한지 1과 7 사이의 숫자로 답해 주세요. 1은 '전혀 적절하지 않다'는 것을 뜻하고 7은 '매우 적절하다'는 것을 뜻합니다.

A.5 Experiment 5

Same list of words used in Experiment 3 (c.f., Table A.3) were used to construct items in Experiment 5.

A.5.1 Materials used to evaluate Ko-GPT, Polyglot-Ko, and GPT-3 models

An example with intended exhaustivity and *-lul*-marked response:

정아가 김밥과 라면 중에 김밥만 먹었습니다.
동욱: 정아는 뭘 먹었어?
지호: 김밥을 먹었어.

An example with intended exhaustivity and *-nun*-marked response

정아가 김밥과 라면 중에 김밥만 먹었습니다.
동욱: 정아는 뭘 먹었어?
지호: 김밥은 먹었어.

An example with intended exhaustivity and null-marked response:

정아가 김밥과 라면 중에 김밥만 먹었습니다.
동욱: 정아는 뭘 먹었어?
지호: 김밥 먹었어.

An example with intended exhaustivity and verbatim repeated response:

정아가 김밥과 라면 중에 김밥만 먹었습니다.
동욱: 정아는 뭘 먹었어?
지호: 김밥만 먹었어.

An example with intended exhaustivity and contradictory response:

정아가 김밥과 라면 중에 김밥만 먹었습니다.
동욱: 정아는 뭘 먹었어?
지호: 김밥이랑 라면을 둘 다 먹었어.

A.5.2 Materials used to evaluate GPT-4

An item with intended exhaustivity and *-lul*-marked vs. null-marked response options:

System message: 일상적인 대화 상황을 생각하고 답장을 골라 주세요. 주어지는 답장들은 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 두 개의 답장 중에 더 적절하다고 생각하는 답장을 골라 주세요.
Input: 동욱: 정아는 뭘 먹었어?
지호는 정아가 김밥과 라면 중에 김밥만 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 더 적절한 것은 무엇입니까?
김밥을 먹었어. 김밥 먹었어.

The second version of the input above where the order of responses is switched:

System message: 일상적인 대화 상황을 생각하고 답장을 골라 주세요. 주어지는 답장들은 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 두 개의 답장 중에 더 적절하다고 생각하는 답장을 골라 주세요.
Input: 동욱: 정아는 뭘 먹었어?
지호는 정아가 김밥과 라면 중에 김밥만 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 더 적절한 것은 무엇입니까?
김밥 먹었어. 김밥을 먹었어.

An item with intended exhaustivity and *-nun*-marked vs. null-marked response options:

System message: 일상적인 대화 상황을 생각하고 답장을 골라 주세요. 주어지는 답장들은 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 두 개의 답장 중에 더 적절하다고 생각하는 답장을 골라 주세요.
Input: 동욱: 정아는 뭘 먹었어?
지호는 정아가 김밥과 라면을 둘 다 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 더 적절한 것은 무엇입니까?
김밥은 먹었어. 김밥 먹었어.

The second version of the input above where the order of responses is switched:

System message: 일상적인 대화 상황을 생각하고 답장을 골라 주세요. 주어지는 답장들은 말해야 하는 모든 것을 나타내지 않을 수 있습니다. 이와 상관 없이 두 개의 답장 중에 더 적절하다고 생각하는 답장을 골라 주세요.
Input: 동욱: 정아는 뭘 먹었어?
지호는 정아가 김밥과 라면을 둘 다 먹었다고 답하려고 합니다. 아래에 주어진 문장들 중 지호가 답장으로 보내기에 더 적절한 것은 무엇입니까?
김밥 먹었어. 김밥은 먹었어.

References

Ackerman, F., & Malouf, R. (2013). Morphological Organization: The Low Conditional Entropy Conjecture. *Language*, *89*(3), 429–464. doi: 10.1353/lan.2013.0054

Ackerman, F., Malouf, R., & Blevins, J. P. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In *Analogy in grammar* (pp. 54–82). Oxford University Press.

Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory*, *21*, 435–483.

Alammar, J., & Grootendorst, M. (2024). *Hands-On Large Language Models*. O'Reilly.

Andreas, J. (2022). Language Models as Agent Models. In *Findings of the Association for Computational Linguistics* (pp. 5769–5779). Association for Computational Linguistics.

Ariel, M. (2008). *Pragmatics and Grammar*. Cambridge: Cambridge University Press.

Ariel, M. (2010). *Defining pragmatics*. Cambridge: Cambridge University Press.

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: linguistic, cognitive, and processing approaches. *WIREs Cognitive Science*, *4*(4), 403–413. doi: 10.1002/wcs.1234

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, *47*(1), 31–56. doi: 10.1177/00238309040470010201

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beaver, D. I., & Clark, B. Z. (2008). *Sense and sensitivity: how focus determines meaning*. Malden, MA: John Wiley & Sons.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational*

Linguistics (pp. 5185–5198). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.463

Bengio, Y., Ducharme, R., & Vincent, P. (2000). A Neural Probabilistic Language Model. *Advances in neural information processing systems*, *13*.

Blevins, J. P., Ackerman, F., & Malouf, R. (2018). Word and Paradigm Morphology. In J. Audring & F. Masini (Eds.), *Oxford Handbook of Morphological Theory.* Oxford University Press.

Blome-Tillmann, M. (2013). Conversational Implicatures (and How to Spot Them). *Philosophy Compass*, *8*(2), 170–185. doi: 10.1111/phc3.12003

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, *3*(4), 395–421. doi: 10.3758/BF03214545

Bonami, O., & Guzmán Naranjo, M. (2023). Distributional evidence for derivational paradigms. In S. Kotowski & I. Plag (Eds.), *The Semantics of Derivational Morphology* (pp. 219–258). De Gruyter.

Bonami, O., & Strnadová, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, *29*(2), 167–197. doi: 10.1007/s11525-018-9322-6

Bossong, G. (1991). Differential Object Marking in Romance and Beyond. In D. Wanner & D. A. Kibbee (Eds.), *New analyses in Romance linguistics* (pp. 143–170). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/cilt.69.14bos

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. doi: 10.1016/j.jml.2004.05.006

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. doi: 10.1016/j.cognition.2005.07.003

Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Büring, D. (2003). On D-Trees, Beans, And B-Accents. *Linguistics and Philosophy*, *26*, 511–545.

Canavan, A., & George, Z. (1996). *CALLFRIEND Korean.* Philadelphia: Linguistic Data Consortium. doi: https://doi.org/10.35111/71ph-4112

Card, D., Henderson, P., Khandelwal, U., Jia, R., Mahowald, K., & Jurafsky, D. (2020). *With Little Power Comes Great Responsibility.* arXiv.

Choi, H.-W. (1996). *Optimizing structure in context: Scrambling and information structure* (Unpublished doctoral dissertation). Stanford University.

Choi-Jonin, I. (2008). Particles and Propositions in Korean. In *Adpositions: Pragmatics, Semantic*

and Syntactic Perspectives (pp. 133–170). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Comrie, B. (1988). *Language Universals and Linguistic Typology* (2nd ed.). Chicago: University of Chicago Press.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 670–680). Copenhagen, Denmark: Association for Computational Linguistics. doi: 10.18653/v1/D17-1070

Croft, W. (1988). Agreement vs. Case Marking and Direct Objects. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in Natural Language: Approaches, Theories, Descriptions* (pp. 159–179). Stanford, CA: CSLI.

Croft, W. (2003). *Typology and Universals* (2nd ed.). Cambridge University Press.

Destruel, E. (2010). An empirical approach to the pragmatics of the French cest-cleft. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 46, pp. 43–58). Chicago Linguistic Society.

Destruel, E., Velleman, D., Onea, E., Bumford, D., Xue, J., & Beaver, D. (2015). A Cross-Linguistic Study of the Non-at-issueness of Exhaustive Inferences. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions* (Vol. 45, pp. 135–156). Cham: Springer International Publishing. (Series Title: Studies in Theoretical Psycholinguistics) doi: 10.1007/978-3-319-07980-6_6

Destruel, E., & Velleman, L. (2014). Refining Contrast: Empirical Evidence from the English it-Cleft. *Empirical Issues in Syntax and Semantics*.

DeVeaugh-Geiss, J. P., Zimmermann, M., Onea, E., & Boell, A.-C. (2015). Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. *Semantics and Linguistic Theory*, *25*, 373–393. doi: 10.3765/salt.v25i0.3054

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics.

Dik, S. (1980). On the typology of Focus Phenomena. In *Perspectives on Functional Grammar* (pp. 41–74). De Gruyter.

Dik, S. C. (1978). *Functional Grammar*. Amsterdam: North-Holland.

Drenhaus, H., Zimmermann, M., & Vasishth, S. (2011). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics*, *24*(3), 320–337. doi: 10.1016/j.jneuroling.2010.10.004

Du Bois, J. W. (2003). Discourse and Grammar. In *The new psychology of language* (pp. 47–87).

Lawrence Erlbaum Associates.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–225.

Erteschik-Shir, N. (1997). *The Dynamics of Focus Structure*. Cambridge University Press.

Erteschik-Shir, N. (2006). On the architecture of topic and focus. In V. Molnár & S. Winkler (Eds.), *The Architecture of Focus* (Vol. 82). De Gruyter Mouton.

Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface*. Oxford University Press.

Ethayarajh, K. (2019). How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 55–65). Hong Kong, China: Association for Computational Linguistics. doi: 10.18653/v1/D19-1006

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902. doi: 10.1073/pnas.1215776109

Firth, J. R. (1968). *Selected papers of J. R. Firth, 1952-59* (F. R. Palmer, Ed.). Indiana University Press.

Frank, A. F., & Jaeger, T. F. (2008). Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30).

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 32–42). Minneapolis, Minnesota: Association for Computational Linguistics.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.

Gerocs, M., Babarczy, A., & Surányi, A. B. (2014). Exhaustivity in Focus: Experimental Evidence from Hungarian. In *Language Use and Linguistic Structure* (1st ed., pp. 181–194). Olomouc: Palacký University Press. doi: 10.5507/ff.14.24440606.13

Gers, F. A. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, *12*, 2451–2471.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. doi: 10.1016/

j.tics.2019.02.003

Givón, T. (2018). *On understanding grammar: Revised Edition* (1st ed.). John Benjamins Publishing Company.

Godfrey, J. J., Hilliman, E., & McDaniel, J. (1992). SWITCHBOARD: A telephone speech corpus for research and development. In (pp. 517–520).

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)* (pp. 10–18). Salt Lake City, Utah: Association for Computational Linguistics. doi: 10.18653/v1/W18-0102

Gotzner, N., & Spalek, K. (2017). The Connection Between Focus and Implicatures: Investigating Alternative Activation Under Working Memory Load. In S. Pistoia-Reda & F. Domaneschi (Eds.), *Linguistic and Psycholinguistic Approaches on Implicatures and Presuppositions.* Cham: Springer International Publishing. doi: 10.1007/978-3-319-50696-8

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Gundel, J. K. (1985). Shared knowledge and topicality. *Journal of Pragmatics*, *9*(1), 83–107. doi: 10.1016/0378-2166(85)90049-9

Haiman, J. (1980). The Iconicity of Grammar: Isomorphism and Motivation. *Language*, *56*(3), 515. doi: 10.2307/414448

Haiman, J. (1983). Iconic and Economic Motivation. *Language*, *59*(4), 781. doi: 10.2307/413373

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *In Proceedings of North American Chapter of the Association for Computational Linguistics* (Vol. 2, pp. 159–166). Pittsburgh, Pennsylvania: Association for Computational Linguistics. doi: 10.3115/1073336.1073357

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1489–1501). Berlin, Germany: Association for Computational Linguistics. doi: 10.18653/v1/P16-1141

Harris, Z. S. (1954). Distributional Structure. *WORD*, *10*(2-3), 146–162. doi: 10.1080/00437956.1954.11659520

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics*, *19*(1), 1–33. doi: 10.1515/COG.2008.001

Haspelmath, M. (2021). Explaining grammatical coding asymmetries: Formfrequency correspondences and predictability. *Journal of Linguistics*, *57*(3), 605–633. doi: 10.1017/S0022226720000535

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2023). From partners to populations: A hierarchical Bayesian account of coordination and

convention. *Psychological Review*, *130*(4), 977–1016. doi: 10.1037/rev0000348

Heycock, C. (2008). Japanese - WA, - GA, and Information Structure. In S. Miyagawa (Ed.), *The Oxford Handbook of Japanese Linguistics.* Oxford University Press.

Hirotani, M., & Schumacher, P. B. (2011, May). Context and topic marking affect distinct processes during discourse comprehension in Japanese. *Journal of Neurolinguistics*, *24*(3), 276–292. doi: 10.1016/j.jneuroling.2010.09.007

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Horn, L. R. (1981). Exhaustiveness and the Semantics of Clefts. In *Proceedings of the Eleventh Annual Meeting of the North Eastern Linguistic Society* (Vol. 11, pp. 125–142).

Horn, L. R. (1984). Towards a new taxonomy for pragmatic inference : Q- and R-based implicature. In *Meaning, form and use in context.* Georgetown University Press.

Horn, L. R. (2013). Implicature. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 3–28). Malden, MA: Blackwell Publishing.

Horn, L. R. (2014). Information Structure and the Landscape of (Non-)at-issue Meaning. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (1st ed., pp. 108–127). Oxford University Press.

Hu, J., Floyd, S., Jouravlev, O., Fedorenko, E., & Gibson, E. (2023). A fine-grained comparison of pragmatic language understanding in humans and language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 4194–4213). Toronto, Canada: Association for Computational Linguistics.

Hu, J., Gauthier, J., Qian, P., Wilcox, E., & Levy, R. (2020). A Systematic Assessment of Syntactic Generalization in Neural Language Models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1725–1744). Online: Association for Computational Linguistics.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantic-spragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. doi: 10.1016/j.cogpsych.2008.09.001

Jaeger, T. F. (2006). Phonological Optimization and Syntactic Variation: The Case of Optional That. *Annual Meeting of the Berkeley Linguistics Society*, *32*(1), 175. doi: 10.3765/bls.v32i1.3453

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, *59*(4), 434–446. doi: https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. doi: 10.1016/j.cogpsych.2010.02.002

Jeretic, P., Warstadt, A., Bhooshan, S., & Williams, A. (2020). Are Natural Language Inference Mod-

els IMPPRESsive? Learning IMPlicature and PRESupposition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 8690–8705). Online: Association for Computational Linguistics.

Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, F. Karlsson, & K. Sinnemäki (Eds.), *Language complexity: Typology, contact, change* (pp. 89–108). Amsterdam: John Benjamins.

Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). Retrieved from `https://web.stanford.edu/~jurafsky/slp3/` (Online manuscript released August 20, 2024)

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., … Amodei, D. (2020). *Scaling Laws for Neural Language Models.* arXiv. (arXiv:2001.08361 [cs, stat])

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015, June). Context-driven expectations about focus alternatives. *Cognition*, *139*, 28–49. doi: 10.1016/j.cognition.2015.02.009

Kim, I. (2015). Is Korean -(n)un a topic marker? On the nature of -(n)un and its relation to information structure. *Lingua*, *154*, 87–109. doi: 10.1016/j.lingua.2014.11.010

Kim, I. (2016). Can Korean -(n)un mark (contrastive) focus? *Language Sciences*, *56*, 105–117. doi: 10.1016/j.langsci.2016.04.002

Kim, J. (2018). Deriving the contrastiveness of contrastive -nun in Korean. *Linguistics and Philosophy*, *41*(4), 457–482. doi: 10.1007/s10988-017-9227-6

Kiss, K. (1998). Identificational focus versus information focus. *Language*, *74*(2), 245–273. doi: 10.1353/lan.1998.0211

Ko, E.-S. (2001). A Discourse Analysis of the Realization of Object NP Forms in Korean. *Japanese/Korean Linguistics*, *9*.

Kortmann, B., & Szmrecsanyi, B. (Eds.). (2012). *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin, Boston: De Gruyter.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3-4), 243–276. doi: 10.1556/ALing.55.2008.3-4.2

Kuno, S. (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry*, *3*(3), 269–320.

Kurumada, C., & Grimm, S. (2019). Predictability of meaning in grammatical encoding: Optional plural marking. *Cognition*, *191*, 103953. doi: 10.1016/j.cognition.2019.04.022

Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178. doi: 10.1016/j.jml.2015.03.003

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). doi: 10.18637/jss.v082.i13

Kwon, I. (2011). One =Mi: An Evidential, Epistemic Modal, and Focus Marker. In *LSA Annual Meeting Extended Abstracts 2011* (Vol. 2). Linguistic Society of America. doi: https://doi.org/10.3765/exabs.v0i0.554

Kwon, S.-N., & Zribi-Hertz, A. (2008). Differential Function Marking, Case, and Information Structure: Evidence from Korean. *Language*, *84*(2), 258–299.

Labov, W. (1986). The social origins of sound change. *Dialect and Language Variation*, 524–541.

Labov, W. (2001). *Principles of linguistic change* (Vol. Social factors). Oxford: Blackwell.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.

Langacker, R. W. (1987). Nouns and Verbs. *Language*, *63*(1), 53–94.

Lee, C. (1994). Definite/Specific and Case Marking in Korean. In *Theoretical issues in Korean linguistics* (pp. 325–341).

Lee, C. (2003). Contrastive topic and/or contrastive focus. In *Japanese/Korean Linguistics* (Vol. 12, pp. 352–364).

Lee, C. (2006). Contrastive (Predicate) Topic, Intonation, and Scalar Meanings. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus* (Vol. 82, pp. 151–175). Dordrecht: Springer Netherlands.

Lee, D.-Y. (2002). The function of the zero particle with special reference to spoken Japanese. *Journal of Pragmatics*, *34*(6), 645–682. doi: 10.1016/S0378-2166(02)00042-5

Lee, H. (2006a). Iconicity and Variation in the Choice of Object Forms in Korean. *Language Research*, *42*(2), 323–355.

Lee, H. (2006b). Parallel Optimization in Case Systems: Evidence from Case Ellipsis in Korean. *Journal of East Asian Linguistics*, *15*(1), 69–96. doi: 10.1007/s10831-005-3004-1

Lee, H. (2007). Case ellipsis at the grammar/pragmatics interface: A formal analysis from a typological perspective. *Journal of Pragmatics*, *39*(9), 1465–1481. doi: 10.1016/j.pragma.2007.04.012

Lee, H. (2012). Context, Non-Canonical Word Order and Subject-Object Asymmetry in Korean Case Ellipsis. *Discourse and Cognition*, *19*(2), 57–80. doi: 10.15718/DISCOG.2012.19.2.57

Lee, H. (2015). Information structure, topic predictability and gradients in Korean case ellipsis: A probabilistic account. *Linguistic Research*, *32*(3), 749–771. doi: 10.17250/KHISLI.32.3.201512.009

Lee, H. S., & Thompson, S. A. (1989). A Discourse Account of the Korean Accusative Marker. *Studies in Language*, *13*(1), 105–128. doi: 10.1075/sl.13.1.04tho

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, *4*(1), 151–171. doi: 10.1146/annurev-linguistics-030514-125254

Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78–106. doi: 10.1016/0010-0285(82)90005-6

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Levy, R., & Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems* (Vol. 19). MIT Press.

Li, E., Schuster, S., & Degen, J. (2021). Predicting Scalar Inferences From "Or" to "Not Both" Using Neural Sentence Encoders. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2021* (pp. 446–450). doi: 10.7275/XR01-A852

Liu, Q., Kusner, M. J., & Blunsom, P. (2020). *A Survey on Contextual Embeddings.* arXiv. (arXiv:2003.07278 [cs])

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, *28*(6), 517–540. doi: 10.1016/j.tics.2024.01.011

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515. doi: 10.1037/a0039267

Marvin, R., & Linzen, T. (2018). Targeted Syntactic Evaluation of Language Models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 1192–1202). Brussels, Belgium.

McCann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in Translation: Contextualized Word Vectors. In *31st Conference on Neural Information Processing Systems (NIPS 2017).* Long Beach, CA, USA.

Michaelov, J. A., Coulson, S., & Bergen, B. K. (2022). So Cloze yet so Far: N400 Amplitude is Better Predicted by Distributional Information than Human Predictability Judgements. In *IEEE Transactions on Cognitive and Developmental Systems.* doi: 10.1109/TCDS.2022.3176783

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 26th Conference on Advances in Neural Information Processing Systems.*

Moyer, M., & Degen, J. (2021). Who thinks wh-questions are exhaustive? In *Proceedings of the Annual Meeting of the Cognitive Science Society.* doi: 10.31234/osf.io/ky5mn

Onea, E., & Beaver, D. (2009). Hungarian focus is not exhausted. *Semantics and Linguistic Theory*, 342–359. doi: 10.3765/salt.v0i0.2524

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... Lowe, R. (2022). *Training language models to follow instructions with human feedback.* arXiv.

Pandia, L., Cong, Y., & Ettinger, A. (2021). Pragmatic competence of pre-trained language models through the lens of discourse connectives. In *Proceedings of the 25th Conference on Computational Natural Language Learning* (pp. 367–379). Online: Association for Computational Linguistics.

Paperno, D., Kruszewski, G., Lazaridou, A., Pham, N. Q., Bernardi, R., Pezzelle, S., ... Fernandez, R. (2016). The LAMBADA dataset: Word prediction requiring a broad discourse context. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1525–1534). Berlin, Germany: Association for Computational Linguistics.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Doha, Qatar: Association for Computational Linguistics. doi: 10.3115/v1/D14-1162

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 2227–2237). New Orleans, Louisiana: Association for Computational Linguistics. doi: 10.18653/v1/N18-1202

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. doi: 10.1016/j.cognition.2011.10.004

Potts, C. (2007). Into the ConventionalImplicature Dimension. *Philosophy Compass*, *2*(4), 665–679. doi: 10.1111/j.1747-9991.2007.00089.x

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC08).* Marrakech, Morocco: European Language Resources Association (ELRA).

Prince, E. F. (1981). Toward a Taxonomy of Given-New formation. In *Radical pragmatics* (pp. 223–255). New York: Academic Press.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI*.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unied Text-to-Text Transformer. *Journal of Machine Learning Research*, *21*, 1–67.

Reinhart, T. (1981). Pragmatics and Linguistics: an analysis of Sentence Topics. *Philosophica*, *27*(0). doi: 10.21825/philosophica.82606

Roberts, C. (2006). Context in Dynamic Interpretation. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (1st ed., pp. 197–220). Wiley. doi: 10.1002/9780470756959.ch9

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*. doi: 10.3765/sp.5.6

Rooth, M. (1985). *Association with Focus* (Unpublished doctoral dissertation). University of Massachusetts at Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116. doi: 10.1007/BF02342617

Rumelhart, D. E., & McClelland, J. L. (1987a). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. Foundations). The MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1987b). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. Psychological and Biological Models). The MIT Press.

Russell, V. L. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means.* Retrieved from `https://CRAN.R-project.org/package=emmeans`

Schuster, S., Chen, Y., & Degen, J. (2020). Harnessing the linguistic signal to predict scalar inferences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5387–5403). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.479

Schuster, S., & Linzen, T. (2022). When a sentence does not introduce a discourse entity, Transformer-based models still sometimes refer to it. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 969–982). Seattle, United States: Association for Computational Linguistics.

Schwarzschild, R. (1999). Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, *7*, 141–177. doi: 10.1023/A:1008370902407

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, *27*, 379–423, 623–656.

Silverstein, M. (1976). Hierarchy of Features and Ergativity. In R. M. W. Dixon (Ed.), *Grammatical Categories in Australian Languages* (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.

Silverstein, M. (1981). Case marking and the nature of language. *Australian Journal of Linguistics*, *1*(2), 227–244. doi: 10.1080/07268608108599275

Smith, K., & Culbertson, J. (2020). *Communicative pressures shape language during communication (not learning): Evidence from casemarking in artificial languages.* doi: 10.31234/osf.io/5nwhq

Smith, N. A. (2020). Contextual word representations: putting words into computers. *Communications of the ACM*, *63*(6), 66–74. doi: 10.1145/3347145

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. doi: 10.1016/j.cognition.2013.02.013

Soukka, M. (2000). *A descriptive grammar of Noon: a Cangin language of Senegal*. LINCOM Europa.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, *25*(5/6), 701–721.

Stolcke, A. (2002). SRILM-an extensible language modeling toolkit. In *7th International Conference on Spoken Language Processing (ICSLP 2002)* (pp. 901–904). doi: 10.21437/ICSLP.2002-303

Stump, G. (2014). Polyfunctionality and inflectional economy. *Linguistic Issues in Language Technology*, *11*. doi: 10.33011/lilt.v11i.1365

Sukhbaatar, S., Szlam, A., Weston, J., & Fergus, R. (2015). End-To-End Memory Networks. In *Advances in neural information processing systems* (Vol. 28).

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* (No. 54). Cambridge: Cambridge University Press.

Tal, S., Smith, K., Culbertson, J., Grossman, E., & Arnon, I. (2022). The Impact of Information Structure on the Emergence of Differential Object Marking: An Experimental Study. *Cognitive Science*, *46*(e13119). doi: 10.1111/cogs.13119

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 3, pp. 36–149).

Tenney, I., Xia, P., Chen, B., Wang, A., Poliak, A., McCoy, R. T., ... Pavlick, E. (2019). What do you learn from context? Probing for sentence structure in contextualized word representations. In *7th International Conference on Learning Representations.*

Tonhauser, J. (2012). Diagnosing (not-)at-issue content. In *Semantics of Under-represented Languages of the Americas (SULA)* (Vol. 6).

Traugott, E. C. (2003). Constructions in grammaticalizations. In *The handbook of historical linguistics* (pp. 624–647).

Ullman, T. (2023). *Large Language Models Fail on Trivial Alterations to Theory-of-Mind Tasks.* arXiv.

Vallduvi, E. (1993). *The Informational Component* (Unpublished doctoral dissertation). University of Pennsylvania.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is All you Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems.* Long Beach, CA, USA.

Wang, L., & Schumacher, P. B. (2013). New is not always costly: evidence from online processing of topic and contrast in Japanese. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00363

Washburn, M. B., Kaiser, E., & Zubizarreta, M. L. (2011). Focus Facilitation and Non-Associative Sets. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 94–102). Los Angeles, California. doi: 10.13140/2.1.3185.8561

Washburn, M. B., Kaiser, E., & Zubizarreta, M. L. (2019). The English It-Cleft: No Need to Get Exhausted. In K. von Heusinger, V. O. Gaspar, & M. Zimmermann (Eds.), *Questions in Discourse* (Vol. Semantics, pp. 198–236). Leiden, The Netherlands: BRILL. doi: 10.1163/9789004378308

Wilcox, E., & Spector, B. (2019). The Role of Prior Beliefs in The Rational Speech Act Model of Pragmatics: Exhaustivity as a Case Study. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 41, pp. 3099–3015).

Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. P. (2020). On the Predictive Power of Neural Language Models for Human Real-Time Comprehension Behavior. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* (pp. 1707–1713).

Wilcox, E. G., Vani, P., & Levy, R. P. (2021). A Targeted Assessment of Incremental Processing in Neural Language Models and Humans. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 939–952). Online: Association for Computational Linguistics.

Williams, A., Nangia, N., & Bowman, S. (2018). A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 1112–1122). New Orleans, Louisiana: Association for Computational Linguistics.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.* arXiv. Retrieved from `http://arxiv.org/abs/1609.08144` (arXiv:1609.08144 [cs])

Xue, J., & Onea, E. (2011). Correlation between Presupposition Projection and At-issueness: An Empirical Study. In *Proceedings of the ESSLLI 2011 workshop on projective meaning* (pp. 171–184). Ljubljana, Slovenia.

Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, *121*(11), 1651–

1670. doi: 10.1016/j.lingua.2011.06.002

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard studies in classical philology*, *40*, 1–95.