DOI : 10.18855/lisoko.2022.47.4.002

Communicative Efficiency and Preferred Information Structure: Evidence from Differential Subject Marking in Korean*

Hanjung Lee
(Sungkyunkwan University)

Lee, Hanjung. 2022. **Communicative Efficiency and Preferred Information Structure: Evidence from Differential Subject Marking in Korean.** *Korean Journal of Linguistics,* 47-4, 667-703. In previous theoretical analyses, caseless and case-marked subjects in Korean have been shown to contrast systematically as to their information structure status. The purpose of this paper is to explore the preferred information structure patterns associated with Korean Differential Subject Marking (DSM) on the basis of a close examination of informal conversational interaction. Evidence from analyses of conversation data demonstrates that the substantial majority of caseless subjects are restricted to thetic clauses expressing direct perceptions of an event in the here and now, whereas the majority of attested case-marked subjects are not restricted similarly and are associated with information structure patterns that involve less predictable information in situational or discourse context. Extending H. Lee's (2021) efficiency-based analysis of variable subject marking, this paper argues that the preferred information structure patterns associated with Korean DSM can be accounted for in terms of an efficient use of case marking motivated by communicative efficiency — an optimal balance between production ease and communicative success. **(Sungkyunkwan University)**

Key words: caseless subject, communicative efficiency, contextual predictability, differential subject marking, preferred information structure

1. Introduction

This paper examines the preferred information structure patterns that correlate with differential case marking (DCM) in Korean. This phenomenon describes a situation in which some subjects or objects are marked with case markers, but not others, depending on semantic and pragmatic features of

* I am grateful to three anonymous reviewers for their useful comments and suggestions. Any errors and misinterpretation are due to my own shortcomings.

the argument. In this study, we are concerned with a form of Differential Subject Marking (DSM) in Korean, namely, the alternation between the presence or absence of a case marker on the subject. The central phenomenon of interest is illustrated by the sentences in (1).

(1) a. Pesu-ka o-n-ta.
bus-Nom come-Pres-Decl
(i) (What is going on?/Why am I hearing strange noises?')
'There is the/a bus coming.'
(ii) (Is the taxi coming?) 'No, the bus is coming.'
b. Pesu o-n-ta.
bus come-Pres-Decl
(The speaker is looking at coming of the bus at the bus stop and talks to the hearer over the phone.)
'Here comes the bus.'
(adapted from Kwon & Zribi-Hertz (2008: 287-288))

According to Kwon & Zribi-Hertz (2008), both options are acceptable here, with different interpretations. In (1a), the *-ka*-subject clause is ambiguous between the neutral thetic reading glossed in (1a-i) and the argument-focus reading glossed in (1a-ii) which triggers an exhaustive-listing implicature. In both the *-ka*-subject clause in (1a) and the caseless-subject clause in (1b), the coming of the bus is assumed to be new information for the addressee. But the two clauses differ with respect to the availability of a direct perception interpretation. As noted by Kwon & Zribi-Hertz (2008), the caseless-subject clause in (1b) is understood as describing a situation that the speaker is observing directly at the moment of utterance. The nominative-subject clause in (1a) is on the other hand understood as talking about a situation that is not necessarily based on the speaker's direct observation.

Such contrasts have often been described in terms of information structural differences between the *-ka*-subject clauses and the caseless-subject clauses (Ahn & Cho 2006a, 2006b; Kwon & Zribi-Hertz 2008; E.H. Lee 2019, among others). For example, Kwon & Zribi-Hertz (2008) propose that caseless subjects and objects may be construed neither as active topics nor foci. More recently, E.H. Lee (2019) offers a different account of DSM, arguing that caseless subject NPs are the most salient discourse topics, i.e., entities that the discourse is about.

To the best of our knowledge, there are no studies that have tested the predictions of the two contrasting information structural accounts of Korean DSM on the basis of a close examination of naturally occurring data. The goals of this paper are to test the empirical validity of these two theoretical accounts against informal conversation data and to propose a novel explanation for the preferred information structure associated with Korean DSM in terms of communicative efficiency.

The remainder of this paper is structured as follows. Section 2 provides a review of two competing information structural accounts of DSM in Korean. Before presenting an analysis of conversation data, I first discuss in Section 3 notions of discourse topic relevant to the empirical analysis of the preferred information structure associated with Korean DSM. In Section 4, I present new evidence from conversation data demonstrating that the substantial majority of caseless subjects are restricted to thetic clauses expressing direct perceptions of an event in the here and now, whereas case-marked subjects are not restricted similarly and are associated with information structure patterns which involve less predictable information in situational or discourse context. In Section 5, I extend H. Lee's (2021) usage-based analysis of variable subject marking, arguing that the association of caseless-subject clauses and *-ka*-subject clauses with partly overlapping information structure can be accounted for in terms of an efficient use of case marking motivated by communicative efficiency — an optimal balance between production ease and communicative success. Section 6 concludes the paper by discussing limitations and implications of the present study.

2. Previous Accounts of Information Status and DSM

This section examines two contrasting accounts of information structure properties correlating with DSM in Korean proposed by Kwon & Zribi-Hertz (2008) and E.H. Lee (2019).[1]

[1] Ahn & Cho (2006a, 2006b) have proposed a different form of syntactic information structure-based account of Korean DSM, which takes caseless NPs to be either incorporated inside of VP or left-dislocated outside of VP. In this paper, I focus on comparison of accounts proposed by Kwon & Zribi-Hertz (2008) and E.H. Lee (2019). See H. Lee (2016) and E.H. Lee

Kwon & Zribi-Hertz (2008) have proposed an account of DCM that captures the interpretive contrasts between case-marked arguments and caseless arguments in terms of f(ocus)-structure. They attempt to derive a range of distributional and interpretive differences between case-marked subjects and caseless subjects from a single information structural parameter—f-structure visibility. Kwon & Zribi-Hertz's (2008) proposal is summarized in (2):

(2) DCM and f-structure in Korean (Kwon & Zribi-Hertz 2008: 279)
   a. NPs that support functional markers indicating structural positions in syntax are visible in f-structure.
   b. NPs that fail to support such markers are not visible in f-structure, unless some other type of marking guarantees their visibility as f-structure constituents.

The term f-structure, as used by Erteschik-Shir's (1997, 2007), identifies a level of grammatical representation where the output of syntax is annotated for information packaging. Under her theory, f-structure constituents are either topics or foci. Hence, what (2) means is that when overt subjects or objects fail to support a functional marker in morphology, they cannot be identified as topics or foci at any level of f-structure.

Adopting Erteschik-Shir's (1997, 2007) framework to represent f-structure, Kwon & Zribi-Hertz (2008) argue that like caseless objects, caseless subjects abide by (2b), in that they fail to be visible in f-structure and consequently undergo f-structure incorporation. Under this analysis, the interpretive contrasts between caseless-subject clauses and *-(n)un/-ka*-marked subject clauses follow from different f-structure properties. In (3) and (4) below, I illustrate f-structure patterns of DSM by the examples in (1). The f-structure representations Kwon & Zribi- Hertz (2008) propose for the two readings of (1a) are given in (3).[2]

(2019) for critical reviews of theoretical and empirical problems of Ahn & Cho's account.

[2] Following Kwon & Zribi-Hertz's (2008) notation to represent f-structure, category labels placed outside closing brackets indicate f-structure constituents, while labels placed inside opening brackets identify s-structure. When a sentence involves two or more levels of f-structure, I use digits to help the reader associate each focus with the appropriate topic, e.g.: [.....]$_{TOP1}$ [[.....]$_{TOP2}$ [.....]$_{FOC2}$]$_{FOC}$

(3) Pesu-ka o-n-ta.
bus-Nom come-Pres-Decl
'There is the/a bus coming.' or 'It is the/a bus that is coming.'
a. (What is going on?/Why am I hearing strange noises?)
[∅]$_{S.TOP1}$ [[the bus]$_{TOP2}$ [is coming]$_{FOC2}$]$_{FOC1}$
b. (Is the taxi coming? No.)
[<the bus>$_{FOC}$]$_{TOP}$ is coming

(3a) is a complex f-structure involving two levels: the matrix level (level 1) and the embedded level (level 2). This f-structure pattern typically responds to such questions as *What's up?* or *What is going on?.* The reply is event-reporting or presentational; hence it is spatiotemporally anchored and hosts a 'stage-level' predicate. Kwon & Zribi-Hertz (2008) assume that the thetic subject is included within the new information—the matrix focus (FOC1). Lambrecht (1994) analyzes such sentences as containing no topic. Erteschik-Shir, by contrast, assumes that such clauses contain a stage topic denoting a spatiotemporal discourse referent—the time and place to which the reported event or situation is anchored. Kwon & Zribi-Hertz (2008) further assume that a thetic (event-reporting or presentational) clause such as (1a) has an embedded f-structure correlating with the predication relation between the subject and predicate.

Kwon & Zribi-Hertz (2008) propose to analyze caseless-subject clauses as having no internal f-structure, as illustrated in the proposed f-structure in (4) for the example in (1b). Thus under this analysis, the *-ka*-subject thetic clause in (1a) and the caseless-subject clause in (1b) differ with respect to the visibility of the syntactic subject in f-structure.

(4) Pesu o-n-ta.
bus come-Pres-Decl
'(Watching coming of the bus) Here comes bus.'
[∅]$_{S.TOP}$ [the bus comes]$_{FOC}$

Kwon & Zribi-Hertz (2008) contend that caseless subjects, like caseless objects, can be construed as neither active topics nor as foci, and always occur in tense-deficient clauses construed as thetic and anchored to speech time. They account for the correlation between the simple thetic interpretation and the

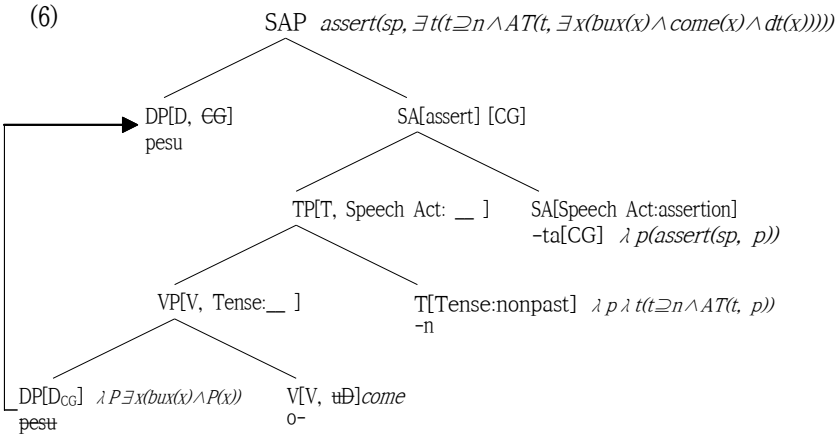
tense deficiency of caseless-subject clauses in terms of f-structure invisibility.[3] Since they are invisible in f-structure, caseless subjects must be included within the matrix focus of their clause, and hence must partake in a thetic interpretation. Unlike nominative-marked thetic subjects, however, caseless subjects are invisible in f-structure, and hence do not stand as embedded topics; it follows that caseless subjects must be incorporated into the matrix focus of their thetic f-structure. Kwon & Zribi-Hertz (2008) further claim that caseless-subject clauses may be specified for aspect, but are left unspecified for tense. Since they are construed as thetic clauses and must consequently be temporally anchored, caseless-subject clauses involve pragmatic anchoring to speech time, as in (1a), whereas topical subject and nominative-marked thetic subjects fail to be similarly restricted.

Adopting Kwon & Zribi-Hertz's (2008) f-structure analysis of DCM, E.H. Lee (2019) offers an account of the association between f-structure and syntactic structure which assumes a one-to-one mapping between information structure and clause structure (Diesing 1992; Kratzer 1995). She posits the same f-structures for *-(n)un*-marked and *-ka*-marked subjects as those proposed by Kwon & Zribi-Hertz (2008). However, in contrast to Kwon & Zribi-Hertz (2008), E.H. Lee (2019) argues that caseless NPs are the most salient discourse topics—i.e., entities that the discourse is about—and proposes the f-structure in (5) and the structure and meaning in (6) for example (1b) (E.H. Lee 2019: 175-6).

(5) Pesu o-n-ta.
bus come-Pres-Decl
'(Watching coming of the bus) Here comes the bus.'
$[\varnothing]_{S.TOP}$ $[[\text{the bus}]_{D.TOP}$ comes$]_{FOC}$

[3] Tense-deficient clauses are used here to refer to clauses that are left unspecified for tense, although they may be specified for aspect. Such tense-deficient clauses include sentences which host no tense or aspect marker and clauses which host markers signalling aspect but not tense.

(6)



On this account, the caseless subject is base-generated as a complement of the verb, but it raises to the specifier position of Speech Act Phrase (SAP) to gain a discourse topic (dt) status and get its [Commmon Ground (CG)] feature checked. This feature becomes deleted under sisterhood with the SAP head, which carries the illocutionary force feature (question, command, etc.) percolating up to SAP.[4] "assert" is taken as a two-place predicate which takes the speakers sp as the first argument and the content of the speech act as the second argument; this is basically an instruction to add the content to CG.

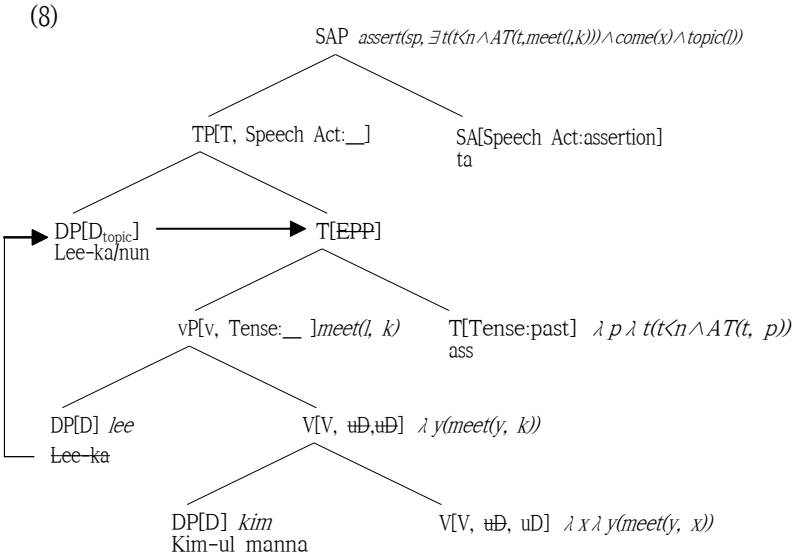
Assuming that *-ka* and *-(n)un* are discourse markers, E.H. Lee (2019) argues that *-ka-* and *-(n)un*-marked NPs are sentence topics which occupy the specifier position of TP after raising and proposing the structure and meaning in (8) for the example in (7) (E.H. Lee 2019: 189).[5]

(7) Lee-ka/-nun Kim-ul manna-ass-ta.
Lee-Nom/Top Kim-Acc meet-Pst-Decl
'Lee met Kim.'

[4] SAP is a functional projection that hosts speech act (or illocutionary force) markers such as indicative, interrogative, and imperative, while adding the semantic and pragmatic force of updating the Common Ground (CG) (Stalnaker 1978).

[5] See J. Lee (2019) for detailed discussion of conditions under which *-ka*-marked NPs mark sentence topics. See also Oh (2009) and H. Lee (2015b).

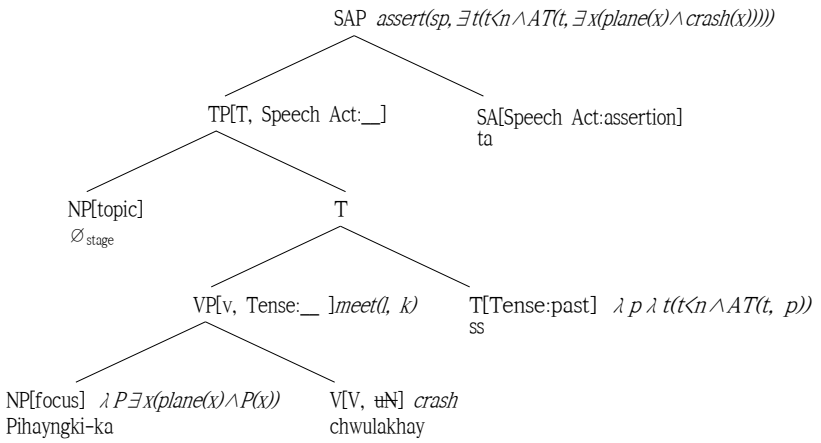
(8)



E.H. Lee (2019) posits a different syntactic structure for an event reporting reading associated with a *-ka*-marked subject clause as exemplified in (9). She assumes in this case that a *-ka*-marked subject occurs in the VP-internal position, that this subject carries the focus feature [new], and that the specifier position of TP is filled with a stage topic, as shown in (10) (E.H. Lee 2019: 191-192).

(9) (What happened?)
Pihayngki-ka/-*nun chwulakhay-ss-ta.
airplane-Nom/Top crash-Pst-Decl
'A plane crashed.'

(10)



The key proposals of E.H. Lee's (2019) analysis discussed above are summarized in (11). E.H. Lee (2019) assumes that discourse/speech context is linguistically encoded in Korean using various sentential speech act particles in the matrix clause, which occupy the head of SAP containing the feature [CG], as shown in (6). A consequence of this assumption is that caseless DPs cannot occur in past-tense sentences or in embedded clauses because they lack the [CG] feature, which leaves the [CG] feature on caseless DPs unchecked. Further, E.H. Lee's analysis (2019) predicts a strict one-to-one relationship between f-structure and syntactic structure.

(11) E.H. Lee's (2019) analysis of DSM

Subject form	Discourse function/ f-structure	Syntactic position
Caseless	Discourse topic/ $[\varnothing]_{S.TOP}[[X]_{D.TOP}\ YP]_{FOC}$	Spec, SAP
-ka-marked	Sentence topic/ $[\langle X\rangle_{FOC}]_{TOP}\ YP$	Spec, TP
	Thetic subject/ $[\varnothing]_{S.TOP1}\ [[X]_{TOP2}\ [YP]_{FOC2}]_{FOC1}$	VP-internal
-(u)un-marked	Sentence topic/ $[X]_{TOP}\ [YP]_{FOC}$ $[[\langle X\rangle_{FOC2}]_{TOP2}]_{TOP1}[YP]_{FOC1}$	Spec, TP

However, closer scrutiny of naturally occurring data suggests that the Korean data are problematic for E.H. Lee's proposal. Let us first take the restricted distribution of caseless subject DPs in tenseless clauses. H. Lee (2021) has examined the association of caseless subjects with tense deficiency by comparing the distribution of a total of 1250 tokens of caseless-subject clauses in deficient tense and past tense conditions in the data collected from sixty-seven hours of video-taped conversation between four pairs of native speakers of Korean. Her analysis shows that a majority of attested caseless subjects (76.72%) were found in the deficient tense condition, whereas 23.28% were found in the past tense condition. The finding that 23.28% of caseless subjects occurred in past tense sentences suggests that it would be premature to offer any categorical generalization about subject form and tense. Moreover, despite the reported ungrammaticality of the omission of *-ka* in the embedded subject position (E.H. Lee 2019: 173), I find from an examination of the conversation data that caseless subjects occur naturally in embedded clauses:

(12) Na-n [Yengmi o-n ke] mol-ass-e.
I-Top Youngmi come-Rel fact not.know-Pst-Decl(Inf)
'I didn't know that Youngmi came.'

(13) Na-n [kyay ka-nun tey] an ka.
I-Top he/she go-Rel place Neg go
'I won't go to where he/she goes.'

Another serious problem with categorical approaches to DSM such as Kwon & Zribi-Hertz (2008) and E.H. Lee (2019) is that the distribution of caseless subjects shows a considerable overlap with that of *-ka*-marked and *-(n)un*-marked subjects. Consider the examples of non-thetic clauses in (14) and (15) wherein the caseless-subject is in focus. The caseless-subject form is generally not required in such non-thetic clauses: ten native speakers I have consulted find both the caseless-subject form and the *-ka*-marked form acceptable in (14)-(16).

(14) A: Way kulehkey nolla? Nwuka iss-e?
why so be.surprised someone be-Int(Inf)
'How come you're so surprised? Is there someone inside?'
B: Cinhi iss-e.
Jinhi be-Decl(Inf)

(Upon witnessing Jinhi's presence in the room)
'Jinhi is in the room.'

(15) A: Nwuka wa-ss-nunci machwu-e po-a.
Who.Nom come-Pst-Comp guess-Comp try-Decl(Inf)
'Guess who has come.'
B: Swumi wa-ss-ney!
Soomi come-Pst-Excl
(Upon looking at the shoes in the entrance) 'Soomi has come!'

(16) A: I kakey-ey-nun etten sonnim-i manhi wa?
this shop-Loc-Top which customer-Nom a.lot come
'What customers do usually come to this shop?'
B: Haksayng-tul manhi o-n-tay.
students a.lot come-Pres-Evid
'(I heard that) a lot of students come (to this shop).'

The examples in (17)-(20) below exemplify non-thetic clauses with a topic subject. As in clauses with a focus subject, the caseless-subject form is not obligatory in such non-thetic clauses: all native speakers that I have consulted agree that both the caseless-subject form and the *-ka*-marked form are acceptable in B's utterances in (18) and (19), and that both the caseless-subject form and the *-(n)un*-marked form are acceptable in B's utterances in (17) and (20).

(17) A: Yengmi-nun encey tasi sewul o-n-tay?
Youngmi-Top when again Seoul come-Pres-Evid
'When is Youngmi coming to Seoul again?'

B: Kyay icey an o-n-tay.
he/she now Neg come-Pres-Evid
'(I heard that) she is not coming from now on.'

(18) A: Pang-ey ka-se Yengmi com nao-lako hay.
room-Loc go-and Youngmi please come.out-Comp tell
'Please go to the room and tell Youngmi to come here.'
B: Pang-ey Yengmi iss-ki-nun-hantey…
Room-Loc Youngmi be-Nomi-Cont-but…

(Looking at Youngmi's presence in the room)
'Youngmi is in the room but…'

(19) A: Ne yocum Minswu po-ass-ni?
you these days Minsoo see-Pst-Int
'Have you seen Minsoo these days?'
B: Ani, Minswu manhi aphu-ass-tay.
no Minsoo a.lot be.sick-Pst-Evid
'No. (I heard that) Minsoo was/has been very sick.'

(20) A: Nwukwu chac-a? Minswu chac-a?
who.Nom look.for-Int(Inf) Minsoo look.for-Int(Inf)
'Who are you looking for? Are you looking for Minsoo?'
B: Ung, Minswu eti ka-ss-eyo?
yes Minsoo where go-Pst-Int(Pol)
(Upon witnessing Minsoo's absence in the room)
'Yes. Where did he go?'

In summary, taking a closer look at the conversation data shows that the variability in the formal realization of subjects in non-thetic sentences as well as in past-tense sentences and embedded clauses conflicts with the categorical approaches such as Kwon & Zribi-Hertz (2008) and E.H. Lee (2019) which assume a strict one-to-one relationship among f-structure constituency (discourse function), subject form, and syntactic position.

3. Notions of Discourse Topic

As discussed in the previous section, E.H. Lee (2019) argues that caseless subjects are the most salient discourse topic and that different DP forms in Korean encode different types of topics: namely, discourse topic, sentence topic, and active topic. However, she does not clearly define what constitutes a discourse topic in the first place. Moreover, it is not at all clear why caseless subjects should be considered a discourse topic and how their status as a discourse topic can be verified empirically. To test the validity of the claims made in the previous accounts of the information status of caseless subjects, a brief

overview of notions of discourse topic is in order here.

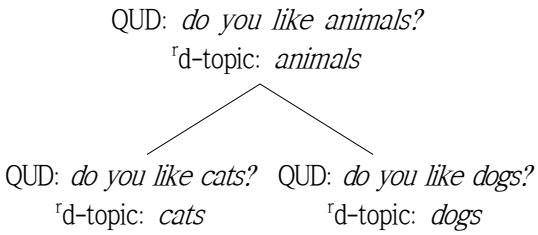
While sentence topics can be seen as a referential unit, the term discourse topic is typically applied in the literature to a propositional unit or a question which is under discussion (Roberts 1996; Ginzburg & Sag 2000). Bott (2007) argues that a referential notion of discourse topic is needed to capture the organization of information at the sentence level and at the discourse level in a tighter manner. Following Bott (2007), I assume that we need both kinds of discourse units, which can both be called discourse topics: referential and propositional ones. To avoid the ambiguous term 'discourse topic', I will follow Ginzburg & Sag (2000) and Bott (2007) in using the term question under discussion (QUD) for propositional units and only using discourse topic for referential ones, while abbreviating them as '⌜d-topics'.

According to Bott (2007), ⌜d-topics have some properties that QUDs do not have. An important difference between them is that ⌜d-topics, unlike QUDs, are referentially typed and properly bound: higher-level ⌜d-topics stand in a binding relation, namely, in a special part-of-relation to lower-level ⌜d-topics. The referent lower-level represented by ⌜d-topics must be either equal to or part of the one higher-level represented by ⌜d-topics. The example in (21), which was taken from Bott (2007:139), illustrates how ⌜d-topic binding works.

(21) Q: Do you like animals?
A1: Cats are quite CUTE.
A2: (But) dogs are HORRIBLE.

Here there is a top-level ⌜d-topic animals associated with a top-level QUD 'do you like animals?', which is broken down into two inferred sub-QUDs 'what about cats?' and 'what about dogs?'. These sub-QUDs contain the lower-level ⌜d-topics cats and dogs, which are the topics of sentences A1 and A2, respectively. Note that these lower-level ⌜d-topics are bound by the top-level one because the kinds 'cat' and 'dog' are part of the kind 'animal' (higher ⌜d-topic ≥ lower ⌜d-topic). The partial discourse tree for (21) is represented by (22).

(22)



In the following section, caseless-subject thetic clauses and non-thetic clauses with an active topic will be analyzed in terms of two questions: (i) Do caseless subjects in these clauses serve as a binder of the lower-level discourse topics or the sentence topics of the discourse segment? and (ii) Are they the prominent part of the higher-level QUD of the discourse segment? If caseless subjects are indeed the most salient discourse topic, as claimed by E.H. Lee (2019), then we can expect that they will exhibit properties of a higher-level ʳd-topic, as stated in questions (i) and (ii) above. According to which of these properties the subject possesses, clauses can be classified into four types, as presented in the table in (23).

(23) Classification of clauses according to ʳd-topic properties

Clause type	
Type 1: (i) only	The referent of a caseless subject binds lower ʳd-topics but is not contained in the higher QUD of the discourse segment
Type 2: (ii) only	The referent of a caseless subject is part of the higher QUD of the discourse segment but does not bind lower ʳd-topics
Type 3: (i) and (ii)	The referent of a caseless subject is part of the higher QUD of the discourse segment and binds lower ʳd-topics
Type 4: neither	The referent of a caseless subject is neither contained in the higher QUD of the discourse segment nor binds lower ʳd-topics

4. Empirical Study

To test the empirical validity of the two competing accounts of DSM discussed in Section 2, I returned to the conversation data collected by H. Lee (2021) and analyzed caseless subject-clauses and *-ka*-subject clauses in terms of information structure status. H. Lee's (2021) data come from sixty-seven hours of video-taped conversation between four pairs of native speakers of Korean, who were born and raised in Korea, and enrolled in the university at the time of the recording. The paired participants agreed to have their conversation video-taped in their apartments for periods ranging from four to ten days in July 2020. The participants — four females and four males — were between the ages of 22 and 28, and were mutual friends.

The procedures for identifying clausal units outlined in H. Lee (2021) produced 1250 tokens of caseless-subject and 1335 tokens of *-ka*-marked subject clauses (*-(n)un*-marked subject clauses were not included in the analysis). Of these clauses, 1087 tokens of caseless-subject clauses and 1155 tokens of *-ka*-marked subject clauses were analyzable in the sense that they had subjects corresponding to one of categories of information status of interest in the present work.

For the purposes of this study, 'information status' refers to the distinction between three different categories of information structure:[6] i) Thetic subject, ii) Focus, and iii) Topic. Here, thetic subject refers to the subject included in the entire clause that is new information (presentational or event-reporting). Focus is an answer to a *wh*-question, or information that instantiates a variable contained in the already activated open proposition.

Topic refers to information whose referent has been under discussion throughout the conversation. This type of topic is called an active topic (Kwon & Zribi-Hertz 2008). It denotes a discourse referent instantiating an entity, which is first activated in the addressee's short-term memory before something is predicated of it.

The conditions of the data analysis also reflected the claim of H. Lee's (2021) analysis by comparing the distribution of caseless subjects and *-ka*-marked subjects in two broad clause types: i) clauses with an agent that is directly identifiable by the speaker and/or the hearer in the here and now and ii)

---

[6] Consistent with lexicalist frameworks of grammar, I regard information structure as a module of grammar that encodes discourse-contextual information of sentential elements and interacts with other structures of grammar (semantic structure, argument structure, syntactic structure and prosodic structure) as an independent module.

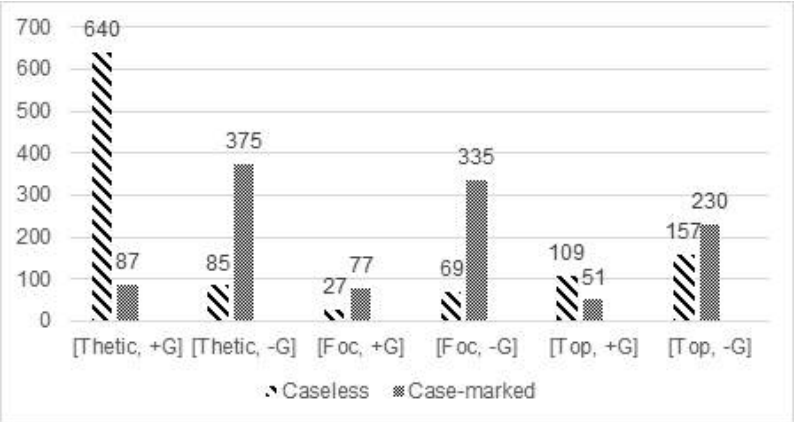
clauses with an agent that is not identifiable on the basis of information available to the speaker or the hearer from the here and now. Following H. Lee (2021), I will refer to the former clause type as [+Grounded] clauses and the latter type [–Grounded] clauses.

In the remainder of this subsection. I first present the token counts by information status and clause type. These two factors resulted in the six conditions specified in the table in (24). Examining the relative frequency of caseless-subject and *-ka*-subject clauses according to the six conditions, we find the results summarized in (25).

(24) Conditions of the analysis

Thetic		Focus		Topic	
[+Grounded]	[–Grounded]	[+Grounded]	[–Grounded]	[+Grounded]	[–Grounded]

(25) Relative frequency of caseless-subject and *-ka*-subject clauses according to clause type and information status



As shown in the figure in (25), caseless subjects were most frequently produced in the [Thetic] condition (725/1087 = 66.70%) and least frequently used in the [Focus] condition (96/1087 = 8.83%). The frequent occurrence of caseless subjects in the [Thetic] condition counters the prediction of E.H. Lee's (2019) account because on this account caseless subjects are predicted to encode topics in

non-thetic clauses. The occurrence of caseless subjects in the [Focus] (8.83%) and [Topic] (24.47%) conditions contrasts directly with the prediction of Kwon & Zribi-Hertz (2008) as well because, on their account, active topics and foci are not predicted to occur as caseless NPs.

We can also observe that the substantial majority of thetic clauses with a caseless subject were of the [+Grounded] type, whereas the relative frequency of caseless-subject clauses in the [Focus] and [Topic] conditions does not show a considerable difference according to clause type. However, in the case of *-ka*-marked subjects, the proportion of the three categories of information status was more balanced (Thetic: 462/1155 = 40.00%; Focus: 412/1155 = 35.67%; Topic: 281/1155 = 24.33%). Further, note that all of the three most frequent subtypes of *-ka*-subject clauses were of the [-Grounded] type. These results provide empirical evidence that thetic clauses were the most frequently produced subtype of the caseless-subject and *-ka*-subject clauses and that the caseless-subject thetic clauses and the *-ka*-subject thetic clauses differed with respect to grounding. The effects of information status and clause type on the choice of subject form were statistically significant when we analyzed the data using the mixed-effects logistic regression, including speaker and verb as random factors. The model showed a main effect of both information status (b = 0.03, SE = 0.02, p = .004) and clause type (b = 0.07, SE = 0.03, $p < .001$).

However, a question remains as to the exact status of caseless subjects in thetic clauses. To investigate whether caseless subject NPs can be considered the most salient discourse topic, I have further analyzed caseless-subject thetic clauses and non-thetic clauses with a caseless subject topic in terms of four categories in (23) above. In the following, I will present the results of an analysis of the proportion of these subtypes with examples.

Let us first consider the results of an analysis of the discourse topic properties of caseless-subjects in thetic clauses. In the data, we found no instances of Type 1, namely, thetic clauses whose subject is caseless and have both properties of a higher-level ⌜d-topic. Examples of the three subtypes of thetic clause appearing in the data are given in (26), (28), and (30). Consider (26), which exemplifies a Type 2 thetic clause. Here, the discourse segment is about things Speaker B plans to do with her sister. The corresponding discourse tree is partially represented by (27).

(26) Type 2

A: Ne yocum way kulehkey pappu-a?
you these.days why so be.busy-Int(Inf)
'How come you are so busy these days?'

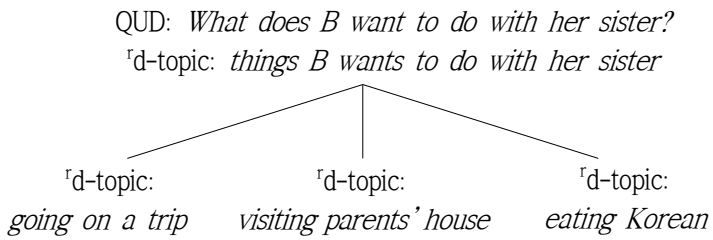
B: Nayil mikwuk-eyse wuli enni wa.
tomorrow U.S.-from our sister come
'My sister is coming from the U.S. tomorrow.'
Kulayse kathi yehayng ka-l cwunpi hay.
so together trip go-Rel preparation do
'I'm preparing for going on a trip with her.'

A: tto etten kyeyhoyk ess-e?
other what plan have-Int(Inf)
'What other plan do you have?'

B: pwumo-nim tayk-ey ka-ko, hankwuk umsik manhi
parent-Hon house-Loc go-Conj Korean food a lot
mek-ko, manhi swi-ko siph-ta.
eat-Conj a lot rest-Comp want-Decl.
'We will go to our parents' house and want to eat a lot of Korean food and rest a lot.'

A: kulem wuli-nun encey manna-l-kka?
then we-Top when meet-Fut-Int
'When are we going to meet then?'

(27) A partial discourse tree for (26)



In (26), there is a higher-level ʳd-topic *things Speaker B wants to do with her sister* associated with a QUD 'what does B want to do with her sister?'. Here, although the referent of the caseless subject in the underlined thetic clause—i.e., *my sister*—is the prominent part of the QUD, it does not fully qualify as a ʳd-topic because it fails to bind lower ʳd-topics. This is because *my sister*, a referent of a type *individuals*, cannot stand in a part-of-relation

to lower ʳd-topics of the discourse segment, i.e., *going on a trip, visiting parents' house, eating Korean food*, etc., which form a different type (activities).

An example Type 3 thetic clause is given in (28). Here, the discourse segment has *rain´* as a ʳd-topic. A partial discourse tree for (28) is given in (29).

(28) Type 3

A: eh, pi o-ney! onul pi manhi on-n-tay?
oh, my! rain come-Excl today rain a lot come-Pres-Evid
'Oh, my! It's raining!'

B: ilkiyeypo po-l-key.
weather.forecast look.at-will
'I'll look at the weather forecast.'

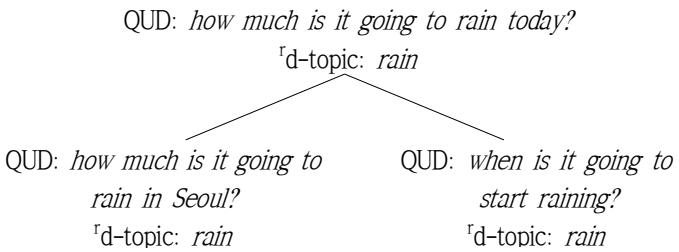
A: sewul-ey elmana on-n-tay?
Seoul-in how.much come-Pres-Evid
'How much is it going to rain in Seoul (according to the weather forecast)?'

B: payk millilithe isang.
100 ml more.than
'More than 100ml.'

A: encey manhi on-n-tay?
When a lot come-Pres-Evid
'When is it going to rain a lot (according to the weather forecast)?'

B: ocen-ey. wusan kacyeka-la.
morning-in umbrella take-Imp
'In the morning. Take your umbrella with you.'

(29) A partial discourse tree for (28)



Here there is a top-level QUD 'how much is it going to rain today?', which is broken down into two sub-QUDs 'how much is it going to rain in Seoul?'

and 'when is it going to start raining?'. Note that these QUDs share the same ⌜d-topic, i.e., *rain*. This means that, in this case, the lower ⌜d-topic is identical to the higher one. Thus, we can say that the referent of the caseless subject in the underlined thetic clause in (28), i.e., *rain* ´, qualifies as a higher ⌜d-topic by virtue of being in the prominent part in the higher QUD and binding lower ⌜d-topics under identity.

Unlike the subjects of Type 2 and Type 3 thetic clauses, the subjects of Type 4 thetic clauses show neither of the ⌜d-topic properties, as exemplified in the discourse segment (30), which has *things Speaker A and Speaker B are busy doing* as ⌜d-topics.

(30) Type 4

A: ne yocum cenhye mos po-ney. Manhi pappu-a?
you these.days never Neg see-Excl a lot be.busy-Int(Inf)
'I haven't seen you at all these days. Are you very busy?'

B: taum cwu-ey chwiep myencep iss-e.
next week-in job interview have-Decl(Inf)
colep censihoy-to cwunpi hay-ya toy.
graduation exhibition-also preparation do-Comp have.to
'I have a job interview next week. I also have to prepare for the graduation exhibition.'

A: emcheng pappu-keyss-ta.
very be.busy-Mod-Decl

B: ne-nun cal cinay?
you-Top well do
'Are you doing well?'

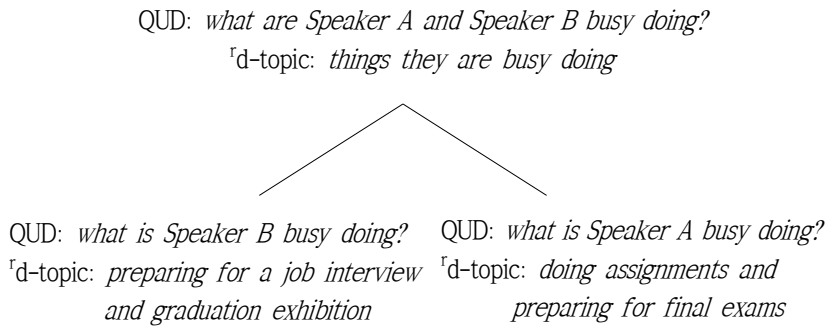
A: na-to taum cwu-ey kwacey sey kay makam-i-ya.
I-also next week-in assignment three Cl be.due-Cop-Decl(Inf)
sihem cwunpi-to hay-ya toy-nuntey…
exam preparation-also do-Comp have.to-but…
'I, too, have three assignments due next week. I also have to prepare for the exams…'

B: ya, ceki pesu o-n-ta.
hey, there bus come-Pres-Decl
taum-ey yayki ha-ca.
next.time-in talk do-Prop
'Hey, here comes the bus over there. Let's talk next time.'

A: kulay. nacwung-ey yayki hay.
OK. later-in talk do
'OK. Talk (to you) later.'

A partial discourse tree for (30) is given in (31). As shown in this tree, the referent of the caseless subject in the underlined thetic clause in (30)—i.e., *bus*—is not contained in the QUDs. Nor does it bind lower ʳd-topics of the discourse segment, i.e., *preparing for a job interview and graduation exhibition, doing assignments, preparing for final exams*, etc., which form a different type (activities). It is not mentioned in the preceding and following sentences at all, and it is far from the most salient discourse topic.

(31) A partial discourse tree for (30)



E.H. Lee (2019: 175) contends that *pesu o-n-ta* 'here comes the bus' is only acceptable when the discourse participants are waiting for the bus, and she proposes that this can be captured by assuming that the caseless subject carries the [CG] feature (see (6)). However, note that in (30), *bus* cannot be taken as part of the common ground, this conflicting with E.H. Lee (2019): as mentioned in Section 2.2, the two discourse participants do not share a spatial context, and Speaker A was not aware of the fact that Speaker B was waiting for the bus. Numerous examples of thetic clauses in the data, such as (30), present a serious problem for analyzing caseless subjects along the lines of discourse topic and common ground, suggesting that the absence of case marking on subjects does not invariably signal that they are part of the common ground and carrying the [CG] feature.

Examining the proportions of the three subtypes of attested caseless-subject

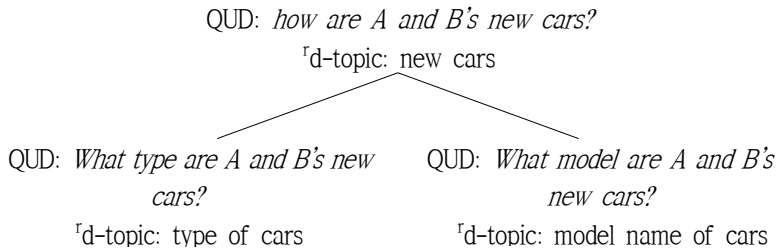
thetic clauses in the total of 1087 tokens of caseless-subject clauses, we find that Type 4 is by far the most frequently produced clause type (679/1087 = 62.47%); Type 2 and Type 3 are only 2.39% (26/1087 = 2.39%) and 1.84% (20/1087 = 1.84%), respectively. These results provide strong evidence that the majority of caseless subjects in thetic clauses do not qualify as a discourse topic and are construed as neither topics nor foci, contra E.H. Lee (2019).

Let us now examine the results of an analysis of the discourse topic properties of caseless, active topic subjects in non-thetic clauses. In the data, there were 9 instances of clauses with caseless, active topic subjects contained in a discourse segment whose top-level or higher QUD is difficult to determine. A close examination of the remaining 257 tokens of clauses with caseless, active topic subjects shows that the caseless subjects in these clauses exhibit at least one of the properties of a higher-level ⌜d-topic by virtue of being part of the higher-level QUD. In other words, all tokens of clauses with caseless, active topic subjects whose top-level or higher QUD was clearly identifiable turned out to be instances of Type 2 or Type 3 clauses. Consider (32), which exemplifies Type 2 clause. Here the discourse segment is about new cars Speaker A and Speaker B purchased. The corresponding discourse tree is partially represented by (33).

(32) Type 2
A: Ne cikum eti ka?
you now where go
'Where are you going now?'
B: Na cikum say cha kyeyyakha-le ka.
I now new car contract-purpose go
'I'm going to contract a new car now.'
A: Cengmal? Na-to say cha sa-ss-e!
really I-also new car buy-Pst-Decl(Inf)
'Really? I, too, bought a new car.'
Nay cha-nun cenkicha-ya.
my car-Top electric.car-Cop
'Mine is an electric car.'
B: Nay cha-nun cenkicha-aniya. Sinhyeng K5-ya.
my car-Top electric.car-Cop(Neg) new K5-Cop
'Mine is not an electric car. It's a new K5'
Ney cha-nun enu cenkicha moteyl-iya?
your car-Top what electric.car model-Cop(Int)

'What electric car model is yours?'
A: Sinhyeng aionik 6-ya.
new ionic 6-Cop
'It's a new Ionic 6'
B: ku cha ppalli tha po-ko siph-ta.
that car soon ride try-Comp want-Decl
'I want to try riding in it soon.'

(33) A partial discourse tree for (32)



In (33), there is a higher-level ʳd-topic *new cars* associated with a QUD 'how are A and B's new cars?'. Here, although the referent of the caseless subject in the underlined clause (Speaker B) is the prominent part of the QUD, it does not fully qualify as a ʳd-topic because it fails to bind lower ʳd-topics, i.e., *type of cars* and *model name of cars*.

An example of Type 3 clause with a caseless, active topic subject is given in (34). A partial discourse tree for (34) is given in (35).

(34) Type 3
A: Ne yocum haksayng ttay chinkwu-tul manna?
you these.days student days friend-Pl meet
'Do you meet your friends in school days these days?'

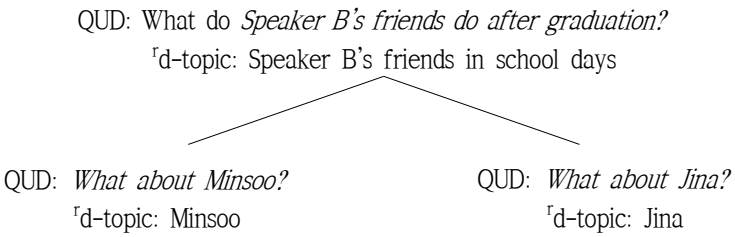
Minswu-nun colepha-konase mwe hay?
Minsoo-Top graduate-after what do
'What does Minsoo do after graduation?'
B: MInswu lophem-ey chwicik hay-ss-e.
Minsoo law.form-Loc got.a.job do-Pst-Decl
'Minsoo joined a law firm.'
A: Cengmal? ∅ cal tway-ss-ta.
really well become-Pst-Decl

'Good for him. (Lit.: (It) turned out well.)'

B: Cina-nun ettehkey tway-ss-e?
Jina-Top how become-Pst-Decl
'How about Jina? What is she doing? (Lit.: How did Jina turn out?)'

A: Kyay-nun il kumantwu-ko, hoykyeysa sihem
(s)he-Top work quit-Conj CPA exam
cwunpi hay.
preparation do
'She quit her work and is preparing for the CPA exam.'

(35) A partial discourse tree for (34)

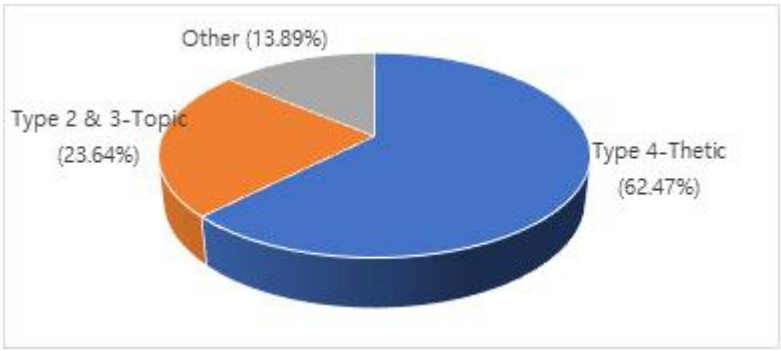


Here there is a top-level QUD 'what do Speaker B's friends do after graduation?', which is broken down into two sub-QUDs. Note that the referent of the caseless subject in the underlined clause in (34), i.e., *Minsoo ˊ*, shows both properties of a higher ʳd-topic by virtue of being in the prominent part in the higher QUD and binding lower ʳd-topics under identity.

Examining the relative frequency of Type 2 and Type 3 clauses with caseless, active topic subjects, we find that 115 instances were Type 2, and 142 instances were Type 3. Note that these clauses take up 23.64% of all instances of caseless-subject clauses (257/1087 = 23.64%).

The results of the analysis of the information status of caseless-subject clauses are summarized in (36). The figure in (36) demonstrates that thetic clauses which show neither properties of a higher ʳd-topic are by far the most frequently produced clause type; non-thetic clauses with an active topic subject are the second most frequent type:

(36) Proportion of subtypes of caseless-subject clauses



Based on the results of the analysis presented in this section, I suggest that caseless-subject clauses and *-ka*-subject clauses are associated with partly overlapping information structure which can be represented using f-structure notation as in (37):

(37) F-structure representations of the preferred information structure associated with caseless-subject clauses and *-ka*-subject clauses

Subject form	F-structure	Examples
Caseless	Most preferred: All focus $[\varnothing]_{S.TOP}[X\ YP]_{FOC/[+Grounded]}$	(1b), (4)
	Second most prefered: Active topic $[X]_{A.TOP}\ [YP]_{FOC}$	(17B)-(20B)
-ka-marked	Most preferred: All focus $[\varnothing]_{S.TOP}[X\ YP]_{FOC/[-Grounded]}$	(1a-i), (3a)
	Second most prefered: Argument focus $[\langle X\rangle_{FOC}]_{TOP}$	(1a-ii), (3b)

The table in (37) shows that the most preferred information structure associated with Korean caseless-subject clauses and *ka*-subject clauses is the simple, thetic f-structure (all-focus pattern) which differs in degree of grounding. The table further shows that the second most preferred f-structure pattern

associated with caseless-subject clauses and corresponds to the non-thetic f-structure pattern, wherein the syntactic subject stands as an independent constituent (active topic); the second most preferred f-structure pattern associated with *-ka*-subject clauses corresponds to the argument focus pattern. The strong association of *-ka*-subject clauses with low degree of grounding and the focus status (all focus and argument focus) leads to the conclusion that *-ka*-subject clauses tend to convey contextually less predictable information compared to caseless-subject clauses.

In the following section, I argue that the grammatical encoding of preferred information structure in the form of DSM can be explained as a result of an economical use of case markers motivated by communicative efficiency.

5. Communicative Efficiency, DSM and Preferred Information Structure

Linguistic encoding of the speaker's message is influenced by multiple factors. Recently, a large body of work has shown that speakers structure their utterances in ways that reduce production costs while at the same time facilitating comprehension. An optimal balance between production ease and communicative success, which is known as *communicative efficiency*, has been suggested to be a major factor shaping language structure as well as language production (Jaeger 2006, 2010; Piantadosi et al. 2011, 2012; Kurumada & Jaeger 2015; Levshina 2021; see Gibson et al. (2019) for a review). Speakers thus tend to reduce the signal of predictable components of messages in order to reduce production costs, but only to the extent that it does not jeopardize the comprehension of the message.[7]

Jaeger (2006) and Levy & Jaeger (2007) propose the principle of Uniform Information Density (UID) as a possible theoretical explanation for the effect of principles of communicative efficiency on grammatical encoding choices. In accordance with information theory (Shannon 1948), information is measured such that the more probable an item is in context, the less informative it is, and conversely the less probable it is, the more informative it is. If the rate at which information is conveyed in the speech stream is roughly constant,

[7] See Jaeger & Buz (2018) for a review. See also Ahn, Kang & Han (2002), Jaeger (2006), and Wasow, Jaeger & Orr (2011) for supporting evidence from syntactic and phonological reduction.

then more predictable words, which carry less information, should take less time to produce than less predictable words. The efficiency of this strategy for communication over a speech channel lies in the fact that it allows utterances to be shorter and easier to produce without reducing the use of less predictable words that the hearer would have the most difficulty reconstructing (Jaeger 2006; Levy & Jaeger 2007; Meister et al. 2021).

More recent research which has investigated case marking through the lens of efficiency in usage suggests that DSM can be interpreted as a result of conventionalization of language users' efficient communicative behavior (Aissen 2003; de Hoop & de Swart 2008; H. Lee 2010, 2016, 2021; Lestrade & de Hoop 2016; Levshina 2021). In their recent work on DSM in Hindi, Nepali and Manipuri, Lestrade & de Hoop (2016) convincingly demonstrate that an efficient use of case marking that is made possible by grounding plays a crucial role in motivating the tense/aspect-based differential use of ergative case and the differential use of ergative case driven by the distinction between stage- and individual-level predication. According to Dixon (1979) and DeLancey (1981), events in the present or imperfect are "A(gent)-centered" in the sense that these events have an identifiable agent: the agent function of an event participant of ongoing events in principle can be identified by the hearer on the basis of non-linguistic information available in the here and now.[8] Building on this insight, Lestrade & de Hoop (2016) propose that the agent function of an event participant of an ongoing event can be *grounded* in the here and now (Lestrade & de Hoop 2016: 403). When the activity is finished, however, the hearer can no longer see the agent at work. As such, Lestrade & de Hoop (2016) assume that the agent function of an event participant of past events cannot be grounded in the here and now.

H. Lee (2021) suggests that Lestrade & de Hoop's (2016) analysis of differential use of ergative case can straightforwardly extends to DSM in Korean if the notion grounding is broadened to incorporate the speaker's role in deriving the argument structure as well and understood as the possibility of the speaker *and/or* the hearer to determine the argument function of an event participant themselves on the basis of situational information available to them in the here and now. For direct empirical evidence for the association of the absence of case marking on subjects with grounding, she analyzed the distribution of

[8] We understand the notion of agent in a broad sense here, also applying to experiencers and entities capable of moving themselves along the lines of Dowty's (1991) theory of proto-roles.

clause types with caseless subjects distinguished by degrees of grounding. The results of the frequency analysis of subtypes of caseless-subject clauses indicate that caseless subjects most productively occur in clause types with higher degrees of grounding, that is, clause types expressing direct perception of a state of affairs by the speaker and/or the hearer in the here and now.

Building on Lestrade & de Hoop (2016), H. Lee (2021) proposes that the association of the absence of case marking on subjects with direct perception utterances follows from an efficient use of case marking that is made possible by grounding: The use of case marking on subjects becomes more redundant when there are sufficient cues to the agent of an event in immediate context, i.e., the here and now. Following a general economy principle, the speaker can omit the explicit use of case marking to minimize her effort if the here and now can be used by the speaker and/or the hearer to derive the argument structure. When identification of the agent of an event cannot be grounded in the here and now, by contrast, higher degree of explicitness in subject marking is more necessary in order to indicate to the hearer that the agent is no longer straightforwardly identifiable in immediate context.

Jaeger (2010) and Kurumada & Jaeger (2015) argue that omitting case markers for more predictable phrases and using case markers to mark less probable phrases has a processing advantage: when speakers use case markers to mark less probable phrases, they can buy more time to produce syntactic elements that are difficult to process and spread information regarding the phrase's grammatical and discourse function over a longer time, thereby leading to more uniform information density compared to leaving it unmarked. Thus, from the perspective of usage probability, the presence of case markers can be interpreted as a signal to expect the unexpected, a rational exchange of time for reduced information density or a meaningful delay. The sentence processor's preference to uniformly distribute information across linguistic signals for increased processing efficiency (by using an extra morpheme or word to mark less probable phrases) is likely to have been grammaticalized as probabilistic linguistic constraints that penalize zero marking for rare types of arguments (for supporting evidence, see H. Lee (2006, 2010, 2016)).

This view of case marking can also account for one of the main findings presented in Section 4, that the majority of attested caseless-subjects are restricted to thetic clauses expressing direct perceptions of an event in the here and now, whereas *-ka*-subject thetic clauses are not restricted similarly.

Crucially, the principle of UID predicts that contextual predictability will affect the morphosyntactic expression of information structure: the availability of strong cues such as situational information from the here and now contributes to higher redundancy of the information being conveyed, and hence renders the utterance as less information dense. Thus, when speakers report or ask about an event that can be sensed or evidenced directly in the here and now, they are more likely to omit case markers and other materials, because the immediate context provides information that reduces the information that these linguistic elements contribute to. In this way, the high contextual predictability and the low information density of direct perception utterances may explain the preferred information structure patterns associated with caseless-subject clauses, i.e., why the majority of attested caseless-subject clauses in the data are associated with a simple, thetic f-structure or the non-thetic f-structure wherein the syntactic subject stands as an active topic.

By contrast, in clauses that convey more non-redundant information, particles become more necessary to achieve more uniform information density. Thus, as predicted by the principle of UID, speakers are more likely to use case markers or particles in clauses that introduce more non-redundant information and are thus more information dense, because using case markers or particles leads to more uniform information density than omitting them. This explains why *-ka*-subject clauses are associated with f-structures which involve less predictable information such as non-grounded thetic focus and argument focus compared to caseless-subject clauses.

What remains to be explained is why both subject forms are acceptable in some examples of thetic clauses and are not in others. In type 2 and type 3 thetic clauses exemplified by (26) and (28) above, respectively, both forms are acceptable whereas in type 4 thetic clauses exemplified by (30) above, the caseless form is the better option and the case-marked form is not felicitous. This difference between subtypes of thetic clauses is summarized in (38).

(38)

Clause type	⌜d-topic properties	Grounding	Example	Subject marking
Type 2	- part of the higher QUD	[-Grounded]	(26)	√Caseless (우리 언니 와) √Case-marked (우리 언니가 와)
Type 3	- part of the higher QUD - a binder of the lower ⌜d-topic	[+Grounded]	(28)	√Caseless (비 오네) √Case-marked (비가 오네)
Type 4	neither	[+Grounded]	(30)	√Caseless (버스 온다) ??Case-marked (버스가 온다)

The strong preference for the caseless subject form in (30) may be accounted for in terms of grounding. As noted above, the use of case marking on subjects becomes more redundant when the speaker or the hearer can use situational information from the here and now to identify the argument function of an event participant. We can thus expect that following a general economy principle, the speaker is more likely to omit the explicit use of case marking to minimize her effort if the here and now can be used to derive the argument structure. This explains the acceptability contrast between the two subject forms observed in (30). By contrast, when situational cues to the argument function of an event participant are weak or unavailable, the speaker is more likely to use the explicit case marker. Thus thus explains why the case-marked form is acceptable in (26) but not in (30).

Interestingly, the acceptability contrast between the two subject forms in direct perception thetic clauses seems to disappear when the exclamative speech act sentence particle *-ney* is used, as illustrated in (28). Unlike in (30), the case-marked subject becomes felicitous in (28), in which the verb stem is followed by the exclamative speech act suffix *-ney*, thus indicating a surprise.

E.H. Lee (2019) proposes a pragmatic analysis of Korean sentential speech act particles that treats them as instructions with which to update the discourse

context by adding the semantic and pragmatic force of updating the Common Ground (CG). More specifically, she suggests that the exclamative suffix *-ney* indicates that a proposition is representing new information (E.H. Lee 2019: 125), in the process proposing that *exclaim(sp, p)* implies *new(p)* ∧ ¬*expect(sp, p)*. That is, exclamatives are treated as instructions to add the content of the proposition *p* to CG as new and contrary to expectation. Here, I will adopt this pragmatic view of the exclamative suffix *-ney*, as it has the virtue of allowing us to link the contribution of the suffix to increased information density in the following way: As the newness and unexpectedness of the content of the proposition increases, case markers or particles become increasingly necessary for more uniform information density: when speakers use them in sentences that convey more unexpected or less predictable content, they can buy more time to produce syntactic elements and spread information on the clause over a longer time, thereby leading to more uniform information density. The sentence processor's preference to uniformly distribute information across linguistic signals for increased processing efficiency (by using an extra morpheme) may explain why the *-ka*-marked subject felicitously occurs in direct perception thetic clauses such as (28) in which the exclamative speech act suffix *-ney* is used.

To summarize, extending H. Lee's (2021) efficiency-based analysis of variable DSM, I have argued that the information structure patterns associated with caseless-subject clauses *-ka*-subject clauses arise from an economical use of case marking motivated by communicative efficiency. The crucial role played by efficiency in motivating preferred information structure supports the conclusion that efficiency in communication influences not only the linguistic encoding of events and event participants but also the morphosyntactic expression of preferred information structure.

6. Conclusion

Caseless and case-marked arguments in Korean have been shown to differ systematically as to their interpretation. This paper has focused on information

status differences between caseless- and case-marked subject clauses. Evidence from conversation data demonstrates that caseless-subject clauses and *-ka*-subject clauses are associated with partly overlapping information structure: the most preferred information structure associated with caseless-subject clauses and *ka*-subject clauses is the simple, thetic f-structure (all-focus pattern) which differs in degree of grounding in the here and now. The results of the data analysis further show that the second most preferred f-structure pattern associated with caseless-subject clauses and corresponds to the non-thetic f-structure pattern, wherein the syntactic subject stands as an independent constituent (active topic); the second most preferred f-structure pattern associated with *-ka*-subject clauses corresponds to the argument focus pattern. The strong association of *-ka*-subject clauses with low degree of grounding and the focus status (all focus and argument focus) leads to the conclusion that *-ka*-subject clauses tend to convey contextually less predictable information compared to caseless-subject clauses.

Extending H. Lee's (2021) analysis of DSM, I have argued that the grammatical encoding of preferred information structure in the form of DSM can be explained as a result of an economical use of case markers motivated by communicative efficiency. Thus, this analysis shows that it is possible to develop a unifying account of the linguistic encoding of event participants and preferred information structure in the form of DSM which subsumes both kinds of encoding under the single principle of communicative efficiency. These results support efficiency-based accounts of grammatical encoding (Hawkins 2004; Haspelmath 2008; Jaeger 2010; Kurumada & Jaeger 2015; H. Lee 2010, 2016, 2021; Lestrade & de Hoop 2016; Levshina 2021) and underscore the importance of communicative efficiency in explaining and motivating patterns of language structure and usage preferences (Gibson et al. 2019; Jaeger & Buz 2018).

The present study is the first to my knowledge to investigate the preferred information structure patterns associated with Korean DSM on the basis of a close examination of informal conversational interaction, which represents the primary mode of communication. Nevertheless, this study has an important limitation in its scope in that it analyzed discourse topic properties of caseless-subject clauses only and did not compare them to properties of case-marked subject clauses. Further research is needed to investigate how case-marked subject clauses pattern differently from *-ka*-subject clauses in

naturally occurring data with respect to discourse topic properties analyzed in this study. In addition, future study will need to investigate the question of whether and how pragmatic differences between case-marked subject clauses and *-ka*-subject clauses correlate with other differences. A majority of caseless-subject clauses in the data contain verb types that can be characterized as being informationally light in context, i.e., verbs which add less information beyond what is directly perceived or inferable in immediate context. Such informaionally light verbs tend to be intransitive (e.g., verbs of directed motion and verbs of appearance, existence and coming into existence) rather than transitive. This raises questions (i) whether case-marked subject clauses pattern differently from *-ka*-subject clauses in informal conversational interaction with respect to verb type favored, and (ii) whether this difference correlates with differences in information structure status. These are important empirical questions which require a more thorough investigation in future study.

A final open question is how the f-structure representations of the preferred information structure associated with differentially-marked subject clauses are linked to their syntactic structure. The finding that caseless-subject clauses and *-ka*-subject clauses are associated with overlapping information structure indicates that the relation between f-structure and syntactic structure is much more complicated than has been assumed to date. Within a minimalist-type approach to syntax, Kwon & Zribi-Hertz (2008) suggest that morphologically marked nominals and bare nominals do not occupy the same structural positions in syntax. Assuming that functional positions in syntactic structure are those that are visible in f-structure, they propose that morphologically marked nominals occupy a functional position in syntax whereas bare (caseless) nominals occur in a basic lexical position. As discussed in Section 2, E.H. Lee (2019) proposes a different mapping between f-structure and syntactic structure in which caseless subjects and morphologically marked subjects in non-thetic clauses occupy a different functional position in syntax whereas *-ka*-subjects in thetic clauses occur in the VP-internal position.

The patterns of associations between subject bareness/marking and information status evidenced from our data do not support for such categorical approaches assuming a strict one-to-one relationship discourse function and syntactic position. The flexible associations between subject bareness/marking

and information status observed in the data highlights the importance of integrating a probabilistic dimension in a mapping between information structure and morphosyntax along the lines of probabilistic models of grammar (Bod, Hay & Jannedy 2003; Bresnan & Ford 2010; H. Lee 2016, among others). Future work will need to investigate how exactly construction of linguistic representations of differentially-marked subject clauses is shaped by probabilistic constraints relating information structure, semantic structure, and morphosyntactic structure.

References

Ahn, Hee-Don and Sungeun Cho. 2006a. "*Wh*-topics and Unpronounced Cases on *Wh*-phrases," *Studies in Generative Grammar* 16(1), 61-90.

Ahn, Hee-Don and Sungeun Cho. 2006b. "Layered Nominal Structures: Implications for Caseless Nominals," *Korean Journal of Linguistics* 31(2), 165-185.

Ahn, Hee-Don, Hyunsook Kang and Jeong-Im Han. 2002. "Towards a Functional Account of Case Marker Reduction and Coda Neutralization in Korean," *Discourse and Cognition* 9(4): 77-96.

Aissen, Judith. 2003. "Differential Object Marking: Iconicity vs. Economy," *Natural Language & Linguistic Theory* 21(3): 435-483.

Bod, Rens, Jennifer Hay and Stafanie Jannedy. (eds.), 2003. *Probabilistic Linguistics*. Cambridge: MIT Press.

Bott, Stefan. 2007. *Information Structure and Discourse Modelling*. Universitat Pompeu Fabra dissertation.

Bresnan, Joan and Marilyn Ford. 2010. "Predicting Sytax: Processing Dative Constructions in American and Australian Varieties of English," *Language* 86(1), 168-213.

Chung, Eun Seon. 2015. "Challenging a Single-Factor Analysis of Case Drop in Korean," *Language and Information* 19(1), 1-18.

DeLancey, Scott. 1981. "An Interpretation of Split Ergativity and Related Patterns," *Language* 57(3), 626-657.

Diesing, Molly. 1992. *Indefinites*. Cambridge: MIT Press.

Dixon, R. M. W. 1979. "Ergativity," *Language* 55(1), 59-138.

Dowty, David. 1991. "Thematic Proto-roles and Argument Selection," *Language* 67(3), 547-619.

Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.

Erteschik-Shir, Nomi. 2007. *The Syntax/Discourse Interface: Information Structure*. New York: Oxford University Press.

Ginzburg, Jonathan and Ivan Sag. 2000. *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. Stanford: CSLI Publications.

Haspelmath, Martin. 2008. "Frequency vs. Iconicity in Explaining Grammatical Asymmetries," *Cognitive Linguistics* 19(1), 1-33.

Hawkins, John. 2004. *Efficiency and Complexity in Grammar*. Oxford: Oxford University Press.

de Hoop, Helen and Peter de Swart (eds.). 2008. *Differential Subject Marking*. Dordrecht: Springer.

Jaeger, T. Florian. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Stanford, CA: Stanford University dissertation.

Jaeger, T. Florian. 2010. "Redundancy and Reduction: Speakers Manage Syntactic Information Density," *Cognitive Psychology* 61(1), 23-61.

Jaeger, T. Florian and Esteban Buz. 2018. "Signal Reduction and Linguistic Encoding," in Eva M. Fernández and Helen Smith Cairns, eds., The Handbook of Psycholinguistics, 38-81. Oxford: Wiley-Blackwell.

Kratzer, Angelika. 1995. "Stage-level and Individual-level Predicates," in Gregory Carlson and Francis Pelletier, eds., *The Generic Book*, 125-175. Chicago: University of Chicago Press.

Kurumada, Chigusa and Florian T. Jaeger. 2015. "Communicative Efficiency in Language Production: Optional Case-Marking in Japanese," *Journal of Memory and Language* 83, 152-178.

Kwon, Song-Nim and Anne Zribi-Hertz. 2008. "Differential Function Marking, Case and Information Structure: Evidence from Korean," *Language* 84(2), 258-299.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Lambrecht, Knud and Laura Michaelis. 1998. "Sentence Accent in Information Questions: Default and Projection," *Linguistics and Philosophy* 21(5), 477-544.

Lee, EunHee. 2019. *Korean Syntax and Semantics*. Cambridge: Cambridge University Press.

Lee, Hanjung. 2006. "Parallel Optimization in Case Systems: Evidence from Case Ellipsis in Korean," *Journal of East Asian Linguistics* 15(1), 69-96.

Lee, Hanjung. 2010. "Explaining Variation in Korean Case Ellipsis: Economy versus Iconicity," *Journal of East Asian Linguistics* 19(4), 291-318.

Lee, Hanjung. 2015a. "Case Particle Ellipsis," in Lucien Brown and Jaehoon Yoon, eds., *The Handbook of Korean Linguistics*, 196-211. Chichester: Wiley Blackwell.

Lee, Hanjung. 2015b. "Information Structure, Topic Predictability and Gradients in Korean Case Ellipsis: A Probabilistic Account," *Linguistic Research* 32(3), 749-771.

Lee, Hanjung. 2016. "Usage Probability and Subject-Object Asymmetries in Korean Case Ellipsis: Experiments with Subject Case Ellipsis," *Journal of Linguistics* 52(1), 70-110.

Lee, Hanjung. 2017. "Effects of Information Status, Subject Type and Tense on Subject Case Ellipsis in Korean: An Experimental Study," *Discourse and Cognition* 24(4), 115-141.

Lee, Hanjung. 2021. "Direct Perception, Cue Reliability and Caseless Subjects: Evidence from Conversational Korean," Korean Journal of Linguistics 46(4), 987-1013.

Lee, Jong Kun. 2019. "The Nominative Case Markers *-i/ka* and Marking Topic in Korean," *The Journal of Studies in Language* 34(4), 593-605. [이종근. 2019. "한국어 주격 표지 '-이/가'와 화제의 표시," 언어연구 34(4), 593-605.]

Lestrade, Sander and Helen de Hoop. 2016. "On Case and Tense: The Role of Grounding in Differential Subject Marking," *The Linguistic Review* 33(3), 397-410.

Levshina, Natalia. 2021. "Communicative Efficiency and Differential Case Marking: A Reverse-engineering Approach," *Linguistics Vanguard* 7(s3), 20190087.

Levy, Roger and Florian T. Jaeger. 2007. "Speakers Optimize Information Density through Syntactic Reduction," in Bernhard Schlökopf, John Platt, and Thomas Hoffman, eds., *Advances in Neural Information Processing Systems* (NIPS) 19, 849-856. Cambridge: MIT Press.

Malchukov, Andrej L. and Helen de Hoop. 2011. "Tense, Aspect, Mood Based Differential Case Marking," *Lingua* 121(1): 35-47.

Meister, Clara, Tiago Pimentel, Patrick Haller, Lena Jäger, Ryan Cotterell and Roger Levy. 2021. "Revisiting the uniform information density hypothesis," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 963-980. Association for Computational Linguistics (ACL).

Oh, Chisung. 2009. "*-Un/nun* as an Unratified Topic Marker in Korean," *Korean Journal of Linguistics* 34(3), 603-624.

Park, Myung-Kwan. 2020. "Nominative Case Unmarking in Korean," *Studies in Modern Grammar* 108, 1-17.

Roberts, Craige. 1996. "Information Structure: Towards an Integrated Formal Theory of Pragmatics," in Jae Hak Yoon and Andreas Kathol, eds., *OSU Working Papers in Linguistics 49: Papers in Semantics*, 91-136. The Ohio State University Department of Linguistics.

Shannon, Claude. 1948. "A Mathematical Theory of Communications," *Bell Systems Technical Journal* 27(4), 623-656.

Stalnaker, Robert. 1978. "Assertion," in Peter Cole, ed., *Syntax and Semantics 9: Pragmatics*, 315-332. New York: Academic Press.

Wasow, Thomas, Florian T. Jaeger and David Orr. 2011. "Lexical Variation in Relativizer Frequency," in Heike Wiese and Horst Simon, eds., *Expecting the Unexpected: Exceptions in Grammar,* 175-196. Berlin & New York: Mouton de Gruyter.

Hanjung Lee, Professor
Department of English Language and Literature, Sungkyunkwan University
25-2 Sungkyunkwan-ro, Jongro-gu, Seoul 03063, Republic of Korea
E-mail: hanjung@skku.edu

Received: October 28, 2022
Revised: December 18, 2022
Accepted: December 19, 2022